# Inventing episodic memory; a theory of dorsal and ventral hippocampus

Nicola van Rijsbergen

PhD
University of Edinburgh
2002

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Nicola van Rijsbergen

iii

# Acknowledgements

# Abstract

Despite the success of psychological and behavioural approaches in demonstrating the effects of hippocampal lesions in human amnesiac and animal subjects no single unifying theory outlining the role of the hippocampus in memory formation has been universally accepted.

In chapter one we argue that MTL amnesia is accounted for by impairment in a mechanism of conscious recall, and that we cannot easily infer the structure of memory representations in the human hippocampus from existing data. We also argue that the lesion studies involving rats, do not model amnesiacs adequately. Therefore we draw the conclusion that an understanding of the role of hippocampus in rodent memory should be developed independently of theories of amnesia.

We then examine three competing hypothesis current in the literature concerning of hippocampal information processing. The first is that the hippocampus is two circuits, one at the dorsal and one at the temporal pole; We argue strongly for a possible dissociation between dorsal and temporal hippocampus,on the basis of existing evidence and we propose a novel theory that the dissociation should be understood in terms of differences in learning speed. The second hypothesis, that the hippocampus is purely spatial, we reject on the basis of a review of evidence from recording and behavioural studies. The third hypothesis is that the hippocampus stores temporal sequences. The argument that the hippocampus may store and recall sequences, however, is considered from three different angles. We evaluate the evidence from behavioural studies and neurophysiological studies, and conclude that that existing evidence is inconclusive. We then turn to two modeling studies. First we examine a well known model of sequence learning in the hippocampus, and show that it does not meet our standard of adequate recall. The end product of the thesis is an extremely simple model that recalls sequences in one trial, that is consistent with both thee more detailed biophysical models, and known neurophysiology. It is also consistent with models that use the same architecture to recall single patterns. From this we draw the conclusion that the idea that the hippocampus learns to recall sequences is plausible. We show clearly that our simple model is more coherent, and performs better the an equivalent model.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Inventing episodic memory

The stream of thought flows on; but most of its segments fall in to the bottomless abyss of oblivion. Of some, no memory survives the instant of its passage. Of others, it is confined to a few moments, hours or days. Others again, leave vestiges which are indestructible, and by means of which they may be recalled as long as life endures. Can we explain these differences? (William James, 1890)

## 1.1 Overview

### 1.1.1 How to read this thesis

The next section provides an over view of the argument in each chapter. There are a couple of meta - instructions. At the beginning of each chapter there is a summary, which outlines the position of each chapter and its contribution to the thesis.There are two strands of argument in the thesis, one relating to Readers who are interested in the modeling chapters should ignore chapters 1-4, as they are largely independent of the modeling chapters.

## 1.2 Part I: Inventing episodic memory: Overview

Theories of event memory, where event memories are special representations in the hippocampus, are presented in chapter 1. (Marr, 1971) suggested that hippocampal

memories are one trial, simple memories, or sensory snapshots (Squire *et al.*, 2001; Frey and Morris, 1998; Norman and O'Reilly, 2001; McClelland and Goddard, 1996). In chapter 1, we argue that *episodic memory* is not this simple memory, but is instead a constructed event representation, that incorporates information learned over many trials. Furthermore, this information is learned within the hippocampus rather than borrowed from neighbouring cortex. However, we also argue that the hippocampus contains information about particular trials. We therefore propose that learning mechanisms in the hippocampus run at multiple speeds, and that information about individual and groups of trials is integrated within the hippocampus. We argue that this integration takes place along the septotemporal axis.

In chapter 2 we present further evidence that the hippocampus contains (at least) two different learning systems that learn at different speeds. We propose that the experimental work by Moser and Moser (2000) suggesting that functional differentiation on the long axis of the hippocampus can be explained by attributing different learning speeds to the dorsal and ventral regions. Moser and Moser (2000) explain their data by a functional division between spatial and non-spatial processing, supported by an anatomical division. We present evidence lesion studies in chapter 3, and anatomy in chapter four that argues against their interpretation

This argument is further supported by Chapters 3 and 4. Chapter 3 argues against the functional division of the hippocampus, by firstly arguing against the *lamellar* hypothesis needed to divide dorsal and ventral circuits. Secondly, Chapter 3 argues against the input/output connectivity described by Moser and Moser (2000). Chapter 3 also identifies two critical areas of future research. Modelling work is needed to understand the role of the distribution of inputs on the dendritic tree in CA1 pyramidal cells, and the impact of dopamine on integration of these inputs, and on plasticity at different synapses.

A critical component of integrating recall of past events with new input is the ability to rapidly recall past events. One potential mechanism for this is based on the NMDA learning mechanism, and we outline this in Chapter 3. Discussion of the evidence for this recall strategy, and two models of how it can be employed in hippocampal type architectures forms the later half of this thesis.

In Chapter 4, we examine the argument in Moser and Moser (2000) that tasks

requiring the dorsal hippocampus are best described as supporting a *spatial map* requiring special spatial input. We argue that too many non-spatial components of the hippocampal representation have been found. However the map hypothesis has led to a proposal that the hippocampus supports a predictive representation that exploits the sequential recall strategy.

The essential contributions of Part 1 are as follows:Part 1 establishes the case against two mainstream theories of the hippocampus function. Firstly, we argue against transferring a theory of declarative , or of episodic memory to explain the rat model on the following grounds.

1. Weakness of the theories of episodic and declarative memory.

2. lack of easy correspondence between recall and recognition tasks in humans and in rats.

3. Reliance of the definitions episodic and declarative memory on conscious and linguistic notions, that cannot be attributed to rats.

The case against the spatial theories of hippocampal functioning in rats is argued on the grounds that the evidence from recording cells in the hippocampus indicates that cells there correlate with a wide variety of things that are not places, and do not fit the map hypothesis.

The second contribution of part 1 of the thesis is to present evidence for a functional distinction between temporal and septal hippocampus. Evidence for this distinction is drawn from both lesion studies, in chapter 2, and from the anatomy and physiology in chapter 3. We hypothesise that this difference should be characterised in terms of the speed of learning in the two parts.

Part 1 therefore offers a novel prediction, about the effects of dorsal only, and ventral only lesions, on tasks that involve some form of alteration between trial, like the transverse patterning task.

## 1.3    Part II: Mechanisms and models of anticipatory recall

The second part of this thesis deals with how to perform exact recall of an exact sequence of items in a hippocampus-like structure.

First we examine the evidence that such a strategy is used in the hippocampus. Chapter 5 examines a case (of proactive, or predictive) representation in the prefrontal cortex. We distinguish such a representation from anticipatory use of the sequence recall strategy.

We argue that although the case that the hippocampus uses a sequence recall strategy is still unproven (we discuss this in chapter 5), a structure that in principle uses the paired recall strategy can use the sequence recall strategy.

Chapter 6 then discusses the merits of Levy (1996)'s model of sequence learning and recall over repeated trials. We present an exact replication study, and some additional experiments, through which we explore the network's behaviour at capacity. We show that the network fails to converge to a stable pattern over a long training cycle, and therefore the weights of the network do not encode the information about transitions in the patterns that Amarasingham and Levy (1997) propose. The model successfully learns correlated (but recoded) input sequences, but performs rather less well when multiple correlated sequences have to be stored separately. In this situation, learning one sequence interferes with the recall of another. The negative patterning task can be modelled as this sort of learning problem. We examine the performance of the model on a negative patterning task. We argue that despite the apparent fit between the behaviour of the model and results from Sutherland and Rudy (1989), this cannot be treated as evidence for sequence learning as a model of the negative patterning task. The argument can be summarised as follows: The representation of the task does not distinguish between hippocampal and non hippocampal dependent forms of the negative patterning task. Secondly, the model exploits assumptions about the hippocampus's contribution to behavioural errors. We present what we think is an equally good case for a an entirely, different contribution, argued from the results of the same model and experimental results in Han *et al.* (1998).

Chapter 6 concludes that

1. the model does not fit the behavioural task

2. The behaviour of the model is controlled by the threshold and the learning rule, and these perform poorly in the one trial situation that does not use a recoding strategy.

Chapter 7 follows on from chapter 6, considering the model proposed by Levy (1996) and working through an analogy to the heteroassociative network. On the basis of this analogy, we examine five different learning rules, and two different thresholding strategies that are used in pattern completion problems in partially connected heteroassociative networks. In particular, we evaluate the presynaptic covariance rule proposed by Minai (1997) on correlated input data in as sequence recall situation. This rule, according to Minai (1997) out-performs all other rules in a compression, or classification problem. However, in an error free situation, we show that the clipped learning rule still performs better, even on correlated data.

The conclusion of chapter 7 is that for single trial learning, where only the order of inputs need be recalled, a binary valued learning rule is the best choice, even if the inputs are highly correlated. We therefore argue that this kind of immediate recall of single events, if it occurs at all, will occur only in the dorsal hippocampus, using all or nothing potentiation at CA3 synapses.

The essential contributions of Part 11 are as follows. The essential contribution from chapter six is the reexamination and rejection of the model of sequence recall developed by the Levy group. This is not the only model of sequence recall available, but it the one that has been most studied (and cited by review papers claiming that sequence sequence learning is an essential hippocampal function, such as Wallenstein *et al.* (1998)). Chapter seven presents a simple model of sequence recall in one trial, that implements a presynaptic covariance rule. The rule is not novel, but has a) not been implemented in an incremental fashion in the way we describe and b) not compared against other rules or models, for this particular problem. We show that it performs as well as the binary rule, and better than the rule and threshold proposed by the model in chapter six. We suggest that a device such as this could be used for one trial recall, and perhaps training of the slower temporal hippocampus.

## 1.4 Overall approach and introduction to issues pursued in the thesis.

How is the beginning and end of an event known if not by reference to the event's meaning? McKoon *et al.* (1986)

This thesis is devoted to three issues that the above quote raises. The first chapter of the thesis argues that it is extremely difficult to assign the exact content to a memory. The hippocampus is supposed to be a memory for events, and we rely on the psychological literature for a characterisation of what these events are. In our review of the hippocampal psychological literature, we meet four different theories of what hippocampal memories are. None is satisfactory; the aim of our investigation is to extract some ideas for the kinds of computation we might expect to find in the hippocampus. One answer to McKoon and Ratcliffe's question is: Build a computational model of how the brain stores events, and then we will have a characterisation of events. Chapter 5 deals with a promising mechanism that uses the endogenous rhythms of the hippocampus to bind events. But none of the models considered in either chapter 4 or chapter 5 succeeds in characterising the computation in the hippocampus, outside the content of the representations assigned by the psychological literature. We argue that to build computational models, we need a better characterisation of the nature of learning in the hippocampus. We are particularly interested in what sort of learning is required for learning sequences.

In a way, we have too much information about the content of the representation. Correlation studies (reviewed in chapter 4) have successfully given us a correlation between neural firing patterns and outside events, and objects. Observation of correlations between stimuli and neural responses provide the data that need to be accounted for by a theory of the content of hippocampal representation -They do not amount to a theory themselves.While we are still finding new and interesting data, we are still a long way from understanding the meaning of the data.

Any theory of the function of the hippocampus also needs to account for the pattern of lost and retained memory in amnesiacs. The literature on amnesiacs has also proved to be a fertile spawning ground for divisions of human memory, for that very reason. It si still an open question whether a unified theory can account for the same sets of data.

The position held in this thesis, is that it can not. The argument for ths position is set out at the end of chapter 1.

## 1.5  The role of hippocampus in amnesia in theory

Theories of hippocampal amnesia are partly responsible for a division of human memory into implicit and explicit, episodic and semantic, procedural and declarative, phenomena most of which are identified knowledgeably by psychologists without more than a rough definition. However, for the purposes of this thesis, a rough definition is not good enough. A precise understanding of these terms of memory needs to start with an understanding of what they were originally intended to describe. Our intention is to extract psychological concepts/phenomena that may help in developing an understanding of the nature of the *neural* circuitry of the hippocampus. Figure 1.1 shows how memory phenomena are classified in three theoretical accounts of the distinction between hippocampal and non hippocampal memory.

The three distinctions arise from the following theories. The *relational memory* (RT) theory, proposed by Cohen and Eichenbaum (1993) defines hippocampal memory as declarative memory, the dual route theory, that defines hippocampal memory as *episodic* memory (E), proposed by Tulving (1983), and the *interference* theory (I), in McNaughton and Gray (2000) and Weiskrantz (1997), that regards the hippocampus as part of a behavioural inhibition system, rather than the locus of memory. The latest incarnation of this theory is extensively developed by McNaughton and Gray (2000).

One way of looking at these theories is to ask whether they propose to account for the deficits in amnesia through a loss of a particular memory content, or loss of some aspect of manipulating memory. Figure 1.2 gives an idea of how the same phenomena can be interpreted differently. A theory such as the declarative or relational theory contains both elements of strategy and representation in accounting for memory loss in amnesia. The same deficit may sometimes be explained by either loss of strategy or loss of representation (sometimes both), and as the examples below show, the two explanations are not always clearly distinguished.

| non hippocampal | | hippocampal |
|---|---|---|
| implicit<br>*no awareness* | vs | explicit<br>*conscious* |
| procedural<br>*reflex*<br>*non propositional*<br>*inflexible* | vs | declarative<br>*verbally accessible*<br>*propositional*<br>*flexible* |
| semantic<br>*facts*<br>*categories* | vs | episodic<br>*sensory experience*<br>*events* |

Figure 1.1: Proposed ways of dividing hippocampal memory phenomena from non hippocampal phenomena

| approach | content | strategy | representation form |
|---|---|---|---|
| RT/E | context | | temporal/spatial ordering |
| RT | event / episodes | | declarative |
| E/I | autoneotic experience | memory detection | |
| I | | error detection | |

Figure 1.2: The same phenomena can be captured in a theory by deficits in three different aspects of the theory. Loss of self reflective experience can be either loss of internal content of representations *or* damage to a retrieval mechanism.
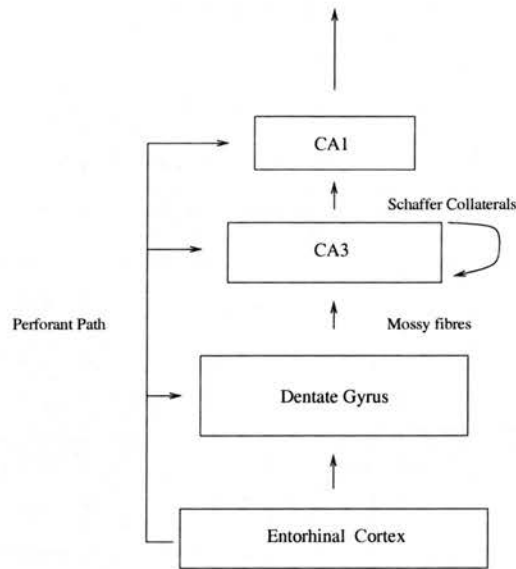
Figure 1.3: Feedforward trisynpatic circuit showing named layers of the hippocampus

## 1.6   Basic hippocampal anatomy

The anatomical circuits that support event memory are described below. A fuller description will be given in chapter 3. However, as the trisynaptic circuit is the functional circuit that is frequently employed in both models and theories, it is useful to characterise it here.

The hippocampus proper is considered to be the entorhinal cortex (EC), the dentate gyrus (DG), and layers CA3, CA2, and CA1. These layers are identified with layers of excitatory cells. Dentate gyrus has a layer of excitatory cells, called granule cells, that project upwards with mossy fibres to the pyramidal cell layer in CA3. The excitatory projection from CA3 is through the schaffer collaterals. The subiculum is variously included in the hippocampus proper and is considered to be the main output relay. Traditionally these regions were viewed as a feed forward excitatory circuit, shown in figure 1.3. Anderson in the 1970s proposed the lamellar hypothesis, which claimed that the hippocampus was made of independent bundles arranged in 'lamellae' stacked along the transverse axis. The typical hippocampal slice preparation study in vitro is a slice taken across the septo-temporal axis. The traditional trisynaptic circuit followed the lamellar hypothesis. This picture is still more or less true for the mossy fibre pro-

jection. The extent and complexity of CA3 recurrent projection in the septo-temporal directions, however, was greatly underestimated.

Within the slice preparation, the regions surrounding the layers of pyramidal cells have laminar structure. This means that the pyramidal cell, and organisation of interneurons, is layered and each layer has a separate name (see 3.3 in chapter 4). Although it has been known for some time that the pattern of organisation of both inhibitory and excitatory inputs is very precisely structured, the functional significance of this layered structure is still a matter of speculation. This will be discussed in chapter 4.

Most models of the hippocampus (Rolls, 1996; Levy, 1996) are still based on the concept of the trisynaptic circuit shown in figure 1.3, with the three synapse pathway between CA1 and entorhinal cortex together with the direct pathway from the entorhinal cortex providing the two major directions of information flow. The most striking feature of the trisynaptic pathway is the existence of the recurrent layer, in CA3. A recurrent loop is a feature associated with architectures that are pattern completion devices, i.e., content addressable memories.

## 1.7   Amnesia and the hippocampus

Amnesia is described in the Handbook of Clinical Neuropsychology (Heilman and Valenstein, 1993) as a severe memory deficit with preserved cognitive functioning, and 11 different aetiologies are included, ranging across age related degenerative disorders such as Huntington's and Parkinson's, to vascular disease, cerebral anoxia, and herpes simplex encephalitis. Chronic alcoholism giving rise to Wernike-Korsakoff syndrome also features prominently in the psychological literature on amnesia. Last but not least, the effects of surgical lesions to the medial temporal lobe (MTL) are mentioned.

It is particularly the last cause of amnesia that established the importance of the medial temporal lobe in human memory. An epileptic patient, H.M., underwent a bilateral medial temporal lobe lobotomy as treatment for severe epilepsy. H.M. suffered from upwards of 10 seizures a week, even with high dosages of anticonvulsants. Such seizures are kindled in the hippocampus, and removal of the hippocampus therefore reduced the seizures. The treatment improved the epilepsy, but left H.M. with a se-

vere *anterograde amnesia* - an inability to form new memories - and also *retrograde amnesia* - the inability to access memories from several years before the operation.

The studies of H.M. and a group of similar patients formed the backbone of the case that made the hippocampus the storage place of memory. However, study of H.M. revealed that although he lost in particular those sorts of memories we refer to as experiences - he could not remember what he had eaten just after lunch (or even that he had eaten), beginnings of conversations when he reached the end, plans formed yesterday, people met an hour ago - he still retained the ability to acquire some new knowledge. Commentary over the first sixteen or so years after the operation suggests that H.M. showed a little improvement in his ability to follow the course of events in the outside word, particularly with regard to salient events. An episode where his mother was ill in hospital left him very uneasy about her for several days, without being able to say why.

Apparent loss of memory might involve a breakdown at any point in the sequence of acquisition and retrieval of the memories. It is therefore an open question as to how much of amnesia is accounted for by deficits of acquisition, those of maintenance and consolidation of memories, and deficits due to faulty mechanisms of retrieval. A subtle distinction can also be made between faulty retrieval and problems with accessing the retrieved memory consciously. The memory may be successfully cued or retrieved, but remain *implicit*.

Despite forty years of work on amnesia, it has not yet been possible to say whether amnesia is primarily a deficit in the acquisition of memory or primarily a deficit in the retrieval of memory, although until recently acquisition theories seemed to have the upper hand. This is partly because there has been a shift towards thinking about amnesiacs as lacking the ability to create certain kinds of *memory representations*. Two theories of this kind are considered first. However, the retrieval deficit theories of Weiskrantz and colleagues (Weiskrantz, 1997; Heilman and Valenstein, 1985) have recently been revived, and will be considered here under the guise of the interference theory.

## 1.8   Theory 1: amnesia as loss of declarative memory

The relational memory theory is the latest guise of the proposal in Cohen and Squire (1980): Amnesia is loss of the ability to 'know that'.

This theory started out as a theory of amnesia, in general, though it became clear in the late 1970s (Heilman and Valenstein, 1979) that there were at least two different recognisable syndromes. Wernike-Korsakoff amnesiacs show different ability loss and retention from those with MTL amnesia. This is partly due to damage to frontal lobe functioning in Korsakoff's syndrome, associated with a greater tendency to confabulate, and loss of planning abilities. For example, Korsakoff patients have a tendency to claim faces as familiar to them, even thought they cannot produce names. Frequently they 'recognise' people with faces they have never seen. MTL amnesiacs, with their powerful awareness of their own condition, do not do this (Heilman and Valenstein, 1985).

For this reason, when amnesia is referred to in this thesis, it usually means medial temporal lobe based amnesia, or MTL amnesia. Much of the early literature on amnesia, therefore, has to be considered carefully as studies are made up of Korsakoff and MTL patients. Cohen and Eichenbaum take a different view in their 1993 book. As far as they are concerned, amnesiacs (Korsakoff and MTL) show an impairment in the same 'domain' of memory, though Korsakoff patients have additional deficits.

Cohen and Squire (1980) described two procedural tasks that can be learned by amnesiacs. Firstly a mirror writing task, and secondly the Tower of Hanoi task (described in figure 1.4). Amnesiac patients like H.M. perform normally on the Tower of Hanoi task. They learn to solve the task using the optimum number of moves (31). Comparison of learning rates between controls and amnesiacs on both tasks show the same rate of improvement. The striking difference between amnesiacs and controls is that amnesiacs *do not know they can solve the task*. They have no memory of earlier trials, and no insight into the solution when asked about their successful performance. H.M.'s inability to accurately comment on implicit tasks that he performed extremely well, was commented on in many of the early papers. Sidman (1968) also suggests that H.M.'s loss was of the ability to verbalise memory, or create memory commentary. All three current theories of amnesia existed in nascent form in the 1960s (Milner *et al.*, 1968).
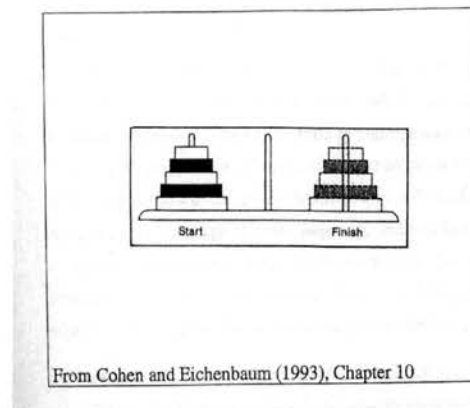
Figure 1.4: The Tower of Hanoi task involves moving the discs from the leftmost peg to the right hand peg, moving one disk at a time and never putting a bigger disk on top of a smaller one.

The Tower of Hanoi requires no memory for moves. From any midway configuration all information required to solve the task is available to perception. It requires minimal anticipation: from each midway configuration there is an obviously useful move. Some explanation of why controls and amnesiacs improve over repeated trials of this task is needed. Cohen and Squire proposed that information about the task was represented as a perceptual motor sequence, and this form of representation was not accessible to conscious reflection (or intervention). This is rather like seeing skill as a matter of requiring the right stimulus responses. This they called procedural knowledge or representation. Cohen and Eichenbaum (1993) state adamantly that

> skilled performance on the Tower of Hanoi puzzle depends no more on the
> remembering of specific puzzle configurations or moves than performance
> in tennis depends upon the explicit remembering of arm movements.

At a first glance, the procedural declarative distinction depends on the idea that some representations are inherently not articulable. However, it is important to see that there is a distinction between the nature or representational form of knowledge and the nature of the process through which it is acquired and then retrieved.

The analogy with tennis playing, intended to illustrate the distinction, highlights a problem. A procedural representation, if we believe the tennis analogy, can only be ex-

pressed in performance of the procedure. But learning to perform in tennis is not just a matter of acquiring an automatic procedure. Like playing the piano, learning requires a transition period, during which conscious reflection is needed to direct the movements. Players also learn to deautomatise movements for the purposes of improving them. So actual performance at a given moment may not depend on declarative representation access, but it can be influenced by it, and the procedure can be adjusted via conscious reflection. Even this extremely physical case does not support the idea that there is a kind of representation that is inherently inaccessible to reflection. In this kind of case, people often bring up the 'I need to see what I do before I can talk about it' experience. We are not arguing that things are never encoded as perceptual motor sequences but that it is clear, in the case of tennis and piano playing, that initial learning does proceed via conscious reflection on memory. With increasing skill, and speed, more of the movements bypass conscious reflection but at any point, if something goes wrong, these movements can be dragged back into consciousness. This transfer between explicit control and motor control and back again is probably unimpaired in amnesiacs, and the fact that they go though this process may be part of their acquired competence at the Tower of Hanoi task. It might be that it is this very process that is facilitated in both amnesiacs and controls, making the Tower of Hanoi much more like playing the violin and painting or, indeed, geometry. Learning here is an interaction between conscious and unconscious processes.

To summarise: We may characterise declarative knowledge as i) knowledge learned through conscious attention or ii)knowledge recalled through a conscious search strategy. procedural knowledge, on the other hand may be knowledge i)learned without conscious direction or awareness, or ii) knowledge recalled without awareness,e.g. a cued response. So far the evidence reviewed concerning amnesiacs shows that they are impaired only in conscious *recall*. The account presented , however, suggested that this loss of conscious recall, was due to an underlying loss of particular kinds of representations.

Now that we have established a distinction between the representation of knowledge and the process of learning and recalling it, we can see that it is not yet clear whether the procedural/declarative distinction is *intended* to be between kinds of representations, or between implicit and explicit methods of access to knowledge. It is

also not clear how much we are entitled to infer the content of the memory representations An additional ambiguity arises through the only way of assessing whether a memory/representation is declarative is whether or not the patient can verbalise or articulate the memory. We rely on the verbal utterance of the subject. If the subject can offer a verbal description of the knowledge, then we consider it declarative. However, we then need to asses the content of the knowledge expressed in the verbal utterances. Verbal utterances potentially express both explicit and implicit knowledge; vocabulary choice, pauses and inflections are automatic aspects of speech. The difficulty between drawing a boundary between the expression of explicit content and implicit content in speech is a well recognised problem (Sperber and Wilson, 1986, 1995). We therefore cannot move directly from the utterance to the content of the knowledge representation.

In Ryle (1949), the original source of the 'knowing how' and 'knowing that' distinction was intended as a distinction in methods of access only, based on a behaviourist argument as follows: if the two kinds of knowledge can be separated by modes of access then it is not necessary to infer anything further about the underlying representation. Following on from that point, if amnesiacs have lost a mode of access to memories then that is sufficient to explain their impaired performance on the task; it may not be that they have lost the memories themselves, or a particular kind of memory representation. After all, the amnesiacs can retrieve the learned skill/procedure for the Tower of Hanoi, they just can't 'know that they know' about it.

Despite the difficulties indicated above the account that Cohen and Eichenbaum propose in 1993 moves to a strongly representational interpretation of the declarative/procedural distinction. They attempt to give a more detailed account of the nature of the hippocampal contribution to declarative representation, and the properties of procedural interpretation. The contrast now draws on accounts of modular and central brain processing, as put forward by Fodor (1983). The crucial difference between declarative and procedural knowledge is now found in the relation of these kinds of knowledge to context.

Cohen and Eichenbaum describe procedural knowledge as 'context inflexible', i.e., knowledge can only be applied in the exact circumstances of acquisition and 'localised to particular input processors.' They describe this kind of knowledge as *modular*. The idea that sensory inputs are modular is largely attributable to Fodor, although Cohen

and Eichenbaum do not directly compare their account with Fodors. Fodor's account suggests that for example, a task acquired through the visual sensory domain does not prime or automatically transfer to the same task mediated through the oral domain, *unless mediated through consciousness.* Cohen and Eichenbaum do explicitly (p. 64 *ibid*) fit their description of the structure of declarative representation with some of Fodor and Pylyshn's account of the structure of the *Language of Thought'*. The parallel between these accounts is also noticeable in the language Cohen and Eichenbaum use to describe declarative memory, as we illustrate below.

Cohen and Eichenbaum describe declarative knowledge showed by normal participants in a task such as the Tower of Hanoi, as represented in a more "propositional" fashion, whereas procedural knowledge consists of a set of fixed rules or responses (they describe the move algorithm that an amnesiac might learn to apply iteratively to perform the task). The (perhaps cosmetic) use of the word "propositional" to describe the declarative representation needs further clarification, as Fodor at least associates propositions (or meanings) with rules for use in context. It is not clear how far Cohen and Eichenbaum intend to buy into Fodor's account of meaning to support their representational account. A critique of Fodor's account lies outside this discussion. They do not discuss whether this learning of rules underlying procedural knowledge involves abstraction and, if so, whether it means that procedural learning is equivalent to the kind of cognitive activity used on such a task as artificial grammar learning, which necessarily involves abstraction; or whether they really think of perceptual motor sequences as being sequences of stimulus response patterns. If procedural knowledge is a stimulus response, it can only be triggered for recall, in a replay of the exact context from which it was learned, making it fixed context.

In contrast, Cohen and Eichenbaum describe declarative knowledge as flexible between contexts, or context free. It is not immediately obvious what the context of an item of knowledge consists of, especially as they also described the declarative knowledge as context rich; the context *richness* of a particular memory could refer to a number of factors. So, let us consider the aspects of declarative memory that might render it context rich. One is content. H.M. cannot remember the particular experience of moving a piece on the Tower of Hanoi task - he lacks a memory with that particular content. Here, the context richness is the sensory or qualitative aspect of remember-

ing the move. Secondly, he does not have any meta knowledge that he is lacking this memory. So, for example, a normal person might not have a memory with that particular content, but they might remember that they did make such a move. Here the context richness is the person's commentary on their own memory. A normal person might also have context rich representations in the sense that they might remember the moves before and the moves afterwards made in the task. This a context rich memory in terms of temporal context. However, there is a problem as to whether this ordering is a property of the storage of memories, or part of their content.We can think of both as a kind of meta knowledge about memories, or as an integral part of the memory itself.

Finally, Cohen and Eichenbaum might mean context rich in the sense of restricted in application (this idea comes from language analysis, etc). To see what this means, consider the fact that the amnesiacs are evolving a strategy for solving the Tower of Hanoi. This strategy can be characterised in terms of rules. A context rich representation in this case means that the context in which the rules can be used is specified by giving the cases in which the rules applies. In Cohen and Eichenbaum's view this particular form of context richness belongs to procedural, rather than declarative knowledge. However,this is slightly counter intuitive, as the more information a memory carried about the particular experience, and its context, the less likely it would be to fit another context exactly.

Eichenbaum and Cohen draw on data from Butters in Heilman and Valenstein (1979, 1985) in their characterisation of the Tower of Hanoi task, arguing that amnesiacs can only apply their task learning skill in circumstances that it was acquired.Butters trained amnesia patients on a version of the Tower of Hanoi task in which patients were allowed to move the circles to any peg, depending on how well the solution was turning out. In these circumstances, where not every trial is a direct repeat, amnesiacs were impaired. What Cohen and Eichenbaum mention, but do not elaborate on, is the fact that most of Butter's subjects were Korsakoff amnesiacs, who frequently show frontal lobe associated planning deficits. Cohen and Eichenbaum acknowledge that this means Korsakoff patients may have performed poorly on the Tower of Hanoi task due to planning deficits.However, they don't see this as impacting on their interpretation of their characterisation of procedural knowledge.

Although amnesiacs *can* solve the Tower of Hanoi task, this is not considered evidence of the ability to use rules and extract them in a context sensitive way. This seems strange, as amnesiacs are not impaired on a task that tests the ability for rule abstraction and application directly, namely an artificial grammar learning task. It can be said that the explanation of why Cohen and Eichenbaum do not make the obvious connection between these two tasks lies in the idea that procedural knowledge is expressed as a sensory motor response to perceptual stimulus.

If one accepts the artificial grammar learning task as an analogue of the Tower of Hanoi task, it is quite clear that this task involves generalisation of rules/categorisation beyond the presented circumstance.

Artificial grammar learning involves presentation of a series of *strings*, or combinations of letters. These strings have been created by the application of grammar-like rules, that allow only certain combinations of letters and make other combinations obligatory. The rules, or *grammar*, are not given to the test subject. Instead, they are shown a series of grammatical examples. They are then tested on unseen strings and asked to guess whether they are grammatical or ungrammatical. Normal people are surprisingly good at this, although they appear to learn this task implicitly. They cannot state the rules they are using, and frequently claim to be guessing randomly.

Performance on the artificial grammar learning task is affected by the *chunk strength*; this is the frequency with which subunit strings are shown in the learning set. Kowlton and Squire (1994) distinguished the similarity of strings from the frequency effects of chunk strength. Similarity was found not to affect grammaticality judgements over and above chunk strengths.

Remarkably, amnesiac patients perform as normals on this task. (Kowlton and Squire, 1994) describes the cognitive ability measured here as part of the ability to abstract rules from instances, the same ability required for declarative memory, if declarative memory is to have propositions. Yet this is demonstrably an implicit ability that is preserved in amnesiacs.

This has proved such a problem for some commentators, such as Squire and Zola (1998), who think that semantic representations must depend on abstraction from particulars, and must therefore be impaired with the loss of the hippocampus, that they have argued that artificial grammar learning does not test powers of abstraction either.

Instead, it too is using 'procedural learning'.

Part of the problem here is the over-use of semantic memory and declarative memory to mean not just memory using semantic or propositional representations, but also memory involved in the process of creating them.

## 1.8.1    Response I: invoke neurobiology

It is also possible to take a different approach to the characterisation of the *abstraction* of representation in the hippocampus. The hippocampus is, according to Lavaneaux and Amaral (2000), sitting at the top of a neurobiological pyramid of associational cortices, starting with unimodal cortical areas, and then proceeding though the perirhinal and postrhinal cortices, entorhinal then through the circuits of the hippocampus proper (i.e. figure 1.3). Therefore, it is argued, the information reaching the hippocampus must be very abstract in nature, as processing and identification of individual stimuli happens in other cortical areas. The hippocampus is therefore in an ideal position to manufacture event descriptions: such a view is also put forward by Shastri (2001). High level models of this kind of theory give a better idea of what the hippocampal representation is supposed to look like. Cohen and Eichenbaum (1993, p63) give an example of a semantic net (Norman and Rumelhart, 1975), with nodes and relations (procedural memories that involve fine tuning and limited generalisation are represented by PDP architectures). Shastri uses a similar architecture but draws on more sophisticated notion of types. However, this is not really evidence brought to support the notion of declarative memory but rather a prediction about the content of information stored in the hippocampus, based on neuroanatomy. None of the authors is particularly concerned with how the content of the elements of the semantic structures is learned.

The issue of whether information in the hippocampus is more abstract than information in input cortices such as olfactory cortex, or perirhinal cortex, will be assessed later in the chapter.

Shastri and Cohen and Eichenbaum share a common hypothesis. Although the hippocampus is representing very complex information, the actual computational mechanism is simple association. So event description, episodic knowledge is very representationally complex. It is so in virtue of the content of its inputs; and the position of the

output on the abstraction tree. Essentially the hippocampus is acting as an associational binder.

Shastri and Cohen and Eichenbaum see this as an advantage of their theory, as it means that the computational side can be implemented very simply. The hard job of establishing and recognising complex relations between events is pushed elsewhere: of course, this is very convenient for hippocampal modelling, but perhaps does not exploit the full richness of the computation available in hippocampal architecture.

More recent papers associated with the relational theory (Eichenbaum, 1999; Eichenbaum *et al.*, 1999), discussed in the next chapter, still use the terminology of nodes and relations without offering insight into how these items that form the structure of episodic memories might be established.

## 1.8.2   Response II: additional subdivision

In the previous section, it was argued that the notion of procedural memory is too weak to capture the nature of the abilities retained by amnesiacs. Declarative memory seems too all-encompassing to be a useful description of the hippocampal memory. Simply defining hippocampal memory as 'abstracted memory' does not usefully distinguish the contribution of the hippocampus from the contribution of the surrounding areas, particularly as some forms of abstraction are clearly preserved in amnesiacs.

The neurobiological argument put forward in the previous section needs to be supported by clear evidence that representation becomes increasingly abstract as it approaches the hippocampus. That evidence will be considered when we consider correlation studies in chapter 4.

## 1.8.3   Response III: abandon the procedural/declarative distinction

A possible response to the failure to define procedural memory adequately is to take the view that the approach of phenomenologically dividing memory into independent systems is wrong. Squire and Zola (1998) argue that the hippocampus and surrounding cortical areas are interdependent. They describe semantic learning as follows:

New information is always presented as part of some event, but through

repetition or rehearsal the new information can be abstracted from its original context and represented as semantic memory.

This presentation of particulars occurs through or in the hippocampus. However, without something to contrast it with, and with damage to the hippocampus, the entire system of memory collapses.

Data that purports to show selective impairment of only one of two particular kinds of memories can still be explained, as hippocampal damage in humans is never complete, and the extent of extra-hippocampal damage hard to assess (see Squire and Zola's response to Vargha-Khadem *et al.* (1997) in the following section).

However, in other publications, such as Squire and Knowlton (1995); Reed *et al.* (1999), it is made clear that there are subtypes of implicit memory, such as those underlying category formation, that are hippocampal independent. This looks like a reintroduction of the procedural/declarative distinction, with procedural memory reworked as those processes underlying category formation.

Let us turn now to a second theory of amnesia that attempts to draw another qualitative distinction between two kinds of memory. This theory shares many of the features of the previous theory.

## 1.9  Theory 2: amnesia is loss of episodic memory

Human memories are divisible at the phenomenal level into two kinds of memories: memories of experience and memories for facts. James (1890) draws a distinction between memories that arise from, and are part of the conscious stream of thought, and memories that are incidently captured from unattended experience. In this section, we will consider the possibility that two dissociable memory systems underly these two kinds of memories. New evidence based on a study of three hippocampal amnesiacs who acquired their disorder during early childhood (ages 4, 7, 9), described in Vargha-Khadem *et al.* (1997), suggests an alternative view of the hippocampus' role in declarative memory. These three children, despite having little recall of daily events, appointments, tasks and places, managed to progress through normal school, and perform close to average on standard verbal IQ and general knowledge tests. This developmental study has been cited as strong support for a distinction in kinds of declarative

memory into semantic (the kind of memories that are tested in general knowledge and verbal IQ tests) and episodic memories (that are specifically autobiographical, event memories). This distinction originates from Tulving (1983) (updated in Tulving and Markowitsch (1998)), who holds that these two forms of memory are not necessarily different in representational demands, but are different in content. Episodic memories have a kind of content that Tulving refers to as *autonoetic*. This word is intended to cover the "feeling" of memory, sometimes a vivid sensory image, or an unspecified experience of recall.

**How does episodic memory compare to semantic memory?**

The original episodic/semantic distinction as first formulated by Tulving also puts more emphasis on the difference in representational structures. Episodes are organised by spatial and temporal co-occurrence, as opposed to a 'semantic' ordering of inference or similarity relations. Episodic knowledge is context dependent, in the sense of being context specific, and context restricted. This is therefore almost exactly the opposite of the context rich but flexible representations in Cohen and Eichenbaum's theory. Neither episodic memory nor semantic memory needs be explicit, i.e., accessible to consciousness. The episodic system exists for its own sake, because episodes are useful. However, the episodic memory is best thought of as a specific, specialised version of semantic memory that draws heavily on semantic notions of content. As a consequence, this theory has the same property as the previous one: all the hard work constructing meaning is done elsewhere.

It is an important component of this theory that both semantic and episodic memory should have direct pathways from sensory information, and that information should flow from semantic to episodic memory. But it does not rule out a pathway from episodic memory.

## 1.9.1    Response I: abandoning the episodic/semantic distinction

Squire and Zola (1998) argue that in the Vargha-Khadem study there is no way of evaluating either the extent to which there is residual episodic memory, nor what the verbal IQ type scores would have been without the episodic impairment. In other words, they

claim that the data is still consistent with a general impairment in declarative memory proportional to hippocampal damage.

## 1.9.2 Response II: undermining episodic priming

The original episodic/semantic distinction proposed by Tulving hypothesised the existence of a phenomenon called episodic priming, and considered a body of evidence meant to support the existence of this phenomenon. Episodic priming is supposed to occur in a condition where the subject is presented with unrelated paired words and then cued for recall. Tulving originally argued that episodic priming (within newly learned paired associates) was a process proceeding via the semantic and episodic representation spaces, and was therefore a slower priming effect than activation that underlay semantic priming, occurring through the semantic space only.

McKoon and Ratcliffe (1986a), McKoon *et al.* (1986) and McKoon and Ratcliffe (1986b) attacked the episodic/semantic distinction on the grounds that it is not possible to separate semantic and episodic priming effects in normals. McKoon and Ratcliffe (1986a) found no differences in speed of priming in replications of the experiments Tulving cited. Episodic priming as described by Tulving did not appear to exist.

On semantic space accounts of semantic priming (see, for example, Lowe and Mc-Donald, 2000) it would be surprising if episodic priming existed as a separate phenomenon, for the following reason. Semantic relatedness is modelled as derived from co-occurrence statistics. Priming is supposed to reflect the degree of semantic relatedness. Random co-occurence is locally biasing the co-occurrence of words. There is no reason to think this is a separate process from co-occurrence detection in any other situation.

## 1.9.3 Reinventing the episodic/semantic distinction I

Tulving (1986) then argued that the episodic/semantic memory system makes sense as a description of human experience of memory, and that it was a category mistake, to expect the underlying causal structure to reflect this distinction. Fortunately[1] since

---

[1]This extreme position is contradicted by Tulving elsewhere. He is right in the sense that episodic memory does seem like a natural kind and, as far as psychology goes, that makes it worth investigation.

then, it has been possible to relate subjects' phenomenal experience of performing semantic and episodic tasks to the activated brain regions. As is shown below, this has given (some) new support to the idea that there is a different mechanism for memories that are *experienced* as episodic.

## 1.9.4    Reinventing the episodic/semantic distinction II: autonoetic content

Essentially, autonoetic content is a new phrase for the same phenomenon that reared its head when we discussed meta-knowledge, and awareness or 'knowing that.' In section1.8 on declarative memory, we discussed the problem that awareness could be a property of the representation, or of access to the representation. In the autonoetic versions, awareness is a special property of hippocampal representations. Experimentally, the first challenge for the autonoetic theory of hippocampus is to show that awareness is a defining part of the content, and always accompanies an episodic memory. It might be that awareness triggers the creation of an episodic memory; it has an essential role in acquisition. This should be distinguished from the possibility that awareness plays an essential role in recall, or always accompanies recall.

Finally, there is a possibility that awareness is a sufficient but not a necessary condition for either acquisition or recall. The date of Maguire and Mummery (1999) suggests that awareness should be grouped alongside other kinds of salience. The alternative is that awareness is interpreted rather like a body state cue, as just another kind of context, and that it should not have any additional role outside use as a possible contextual cue. In this case, only some hippocampal memories are autonoetic.

## 1.9.5    Autonoetic content in recall

Eldridge *et al.* (2000) support the idea that hippocampal activity is phenomenologically detectable and accompanies recall. This experiment looked for the fMRI activa-

---

However, it could just be distinguished phenomenally by its content. Without evidence that the different content produced different behavioural effects, there would be no grounds for saying that it was a separate memory system, rather than just a separate word - linguistic behaviour effects might be enough to support separate meaning. Ryle is a two edged sword here.

tion correlates of the experience of remembering, in a classic list recognition task. In a word list recognition or perceptual recognition task, the subject is allowed to study one or two lists. Then, at some later point, the subject is asked to recall the contents of the list, in an explicit test. Or, the subject is presented with a series of words, or pictures and asked if the current item is in the study list. In this particular study, subjects were presented with a series of words on a screen, and then placed inside the scanner. They were presented with test words, some of which were from the training set. They were asked to judge how confident they were that the word had appeared in the test set, and then asked to rate the basis of their response as either knowing or recalling of the word. Subjects were asked to classify a memory item as recalled, if they had some internal (sensory) experience on which they based their judgment. Eldridge's data clearly showed raised activation in the hippocampus proper (i.e., DG, CA1, CA3) only for the correct "recalling" responses. The other categories of responses, false alarms and misses showed no raised activation in the hippocampus. Perhaps the most interesting aspect of this experiment was the distribution of false alarms. Only 1 percent of the recall responses were false alarms against 33 percent of the knowing responses. However, the question of whether awareness is a side product or a causal factor in recall, or created by awareness in acquisition, is still unanswered. There is also counter evidence that suggests hippocampal memories are not always accompanied by awareness. A study that looked at contextual, or pattern, priming put the emphasis back onto contextual information, by comparing explicit and implicit contextual conditioning. Chun and Phelps (1999) found amnesiacs were impaired on a visual search task, designed so that the background distractor predicted the position of a sought symbol. Control subjects were tested for awareness. Although most (6/10), claimed to notice some relationship between the backgrounds, this awareness did not affect performance. This is an example of a hippocampal dependent task, where the hippocampal mechanism is not affected by awareness.

### 1.9.6 Consciousness can (sometimes) predict recall

Experiments by Clark and Squire (1998) compared amnesiacs and controls in the acquisition of eye blink trace conditioning. Both groups acquired an eye blink response to a tone that overlapped with an air puff in a normal conditioning paradigm. When

a delay intervened between air puff and tone, amnesiacs were not able to acquire the conditioned response. However, both groups were then distracted by a silent movie that they were instructed to remember, through the 'distractions' of sounds. Amnesiacs showed no conditioning in this paradigm, while in controls only those subjects that showed awareness of a relationship between tone and air puff showed the conditioned response. These results initially seemed to suggest that accessibility to conscious awareness was an essential part of the functioning of the hippocampus (Eichenbaum, 1999). A follow up study (Manns *et al.*, 2000) on conditioning without the delay found that awareness in this case also predicted the magnitude of the conditioned response. Conditioning without the delay is not hippocampally dependent; it therefore looks as if awareness generally affects conditioning. Are there further grounds for assuming a special relationship between awareness and hippocampal memory?

Maguire and Mummery (1999) suggested that it may be that an episode with heightened awareness is more likely to become an episodic memory, on the basis of their finding of raised activation in the hippocampus on retrieval of contextual and "personally relevant" memories, rather than factual, or simply autobiographical memories.

An alternative theory of why amnesiacs do not form these kinds of memories derives from the studies that suggest that it is not the special accessibility of these memories that is lost, but the self monitoring content.

The data concerns H.M.'s inability to report on his own physical body states. He was, of course, not able to remember when he had just eaten, but neither was he able to consult his body's internal state to judge whether he was or was not hungry. He acted on implicit signals of hunger and satiety, and refused food, but never gave "feeling full" as a reason for his actions. Due to the extensive nature of H.M.'s lesions, it may be that this is a result of damage to the amygdala, rather than the hippocampus. However, Davidson and Jarrard (1993) and Clifton *et al.* (1998) showed behavioural evidence that rats were not able to monitor their internal state. This raises the issue of whether one can talk about the contribution of awareness or self awareness to animal behaviour.

It has been argued (Shanks and John, 1994) that it is practically impossible to establish the absence of a contribution of awareness to a memory task in a human subject. Another approach might be to look for another characteristic that distinguished

semantic memory from episodic memories. One such characteristic has been proposed in theories of categories, where category construction is considered a prime example of semantic memory.

In section 1.8, we said that meta knowledge or awareness could be thought of as an inherent quality of representations, or as second order representation of representation, or second order representation. In this section, we have found evidence that shows awareness is not a necessary part of recall or acquisition of episodic memories. In the next section we consider the plausibility of Tulving's claim that the hippocampus is a specialised subtype of semantic memory, and does not play the role of providing particulars for abstraction in semantic memory. In other words, the opposing view, that Squire and Zola (1998) (although see also Squire and Knowlton, 1995 and Reed *et al.*, 1999) set out in section 1.8.3, says that the hippocampus is needed for the creation of semantic categories in amnesiacs. This means that MTL amnesiacs should not show normal category formation. Note that this is different from the classification judgements required in the artificial grammar task, see discussion in Reed *et al.*, 1999). There is a vast literature on category creation. In amnesiacs it has been studied through the use of list recognition tasks, of the kind described in Eldridge *et al.* (2000). Although list recognition does not explicitly test category formation, the performance of subjects is affected by factors related to implicit category structure. So, for example, in a word recognition test, if (some) study words are semantically related and then in the test list another related word occurs, that related word is likely to generate a false judgement that it occurred in the study list. Such a word is often referred to as a lure.

## 1.10   Recognition judgements and category creation

There are a number of memory tasks that amnesiacs generally perform well, such as those involving perceptual priming, and word or sentence completion. They perform better on tasks that require them to complete words with *the first thing that comes into their head*, rather than those that require them to explicitly recall the contents of a list of words or objects. Amnesiacs are impaired on explicit novelty judgements, where they are asked simply to say whether a test item appeared in a previously presented set. However, they are unimpaired on categorisation tasks and artificial grammar learnin-

ing (see below) where they are required to make a category decision, rather than recall an item. Work on the influence of semantic knowledge and structure on recognition has suggested that these two memory systems (category and recognition) are not completely independent, contra Squire and Knowlton (1995).

The initial work on word recognition from lists found that amnesiacs were impaired in false recognition due to similarity effects, (Schacter *et al.*, 1996), but enhanced on unrelated false recognition cases. This means they did not respond to category lures normally, producing fewer false alarm responses of this type. Koutstaal *et al.* (1999) replicated these results using perceptual items, which were scored for similarity. They also included *prototypical* items for the categories in the testing set. Difficulties in interpreting evidence regarding the existence of false alarm effects created by prototype or 'gist memory' are compounded by the failure to distinguish Korsakoff from MTL amnesiacs (Schacter *et al.*, 1996) and hippocampal from non hippocampal plus wider MTL damage in Koutstaal *et al.* (1999).

However, these findings in Schacter *et al.* (1996) are disputed by Melo *et al.* (1999), who treated the Korsakoff amnesiacs and those with frontal lobe damage as a separate group from MTL amnesiacs. These results show that amnesiacs are impaired in recognition, and false recognition relative to controls. However, in all responses (both old and new judgements if one considers the proportion of responses of false recognition, relative to "old" responses - 'category lure intrusions'), amnesiacs show a greater proportion of gist based false alarms to controls. On the other hand, amnesiacs with frontal lobe damage show category lure intrusions, thus partly explaining Schacter's findings.

Overall, these results support the idea that recognition judgements are due to two factors, category/semantic memory and episodic, and that prototype activation is intact in impaired recognition performance. However, the two experiments with perceptual and verbal list items are not equivalent. Data from the word recognition tasks do not tell us anything about impairments in novel category creation. The second paper (Koutstaal *et al.*, 1999) is better characterised as examining the effects of similarity on category acquisition, whereas Melo and Schacter *et al.* (1996) were examining the effects of pre-existing semantic structure.

Melo suggests that perhaps the best explanation of increased lure intrusions is that control subjects can use recall of particular items to exclude or suppress cases during

recall. It would be extremely interesting to know if indeed amnesiacs are more or less affected by similar items in the perceptual analogue task, where they are acquiring a new category. Such a case is partially explored in the next section where a theory that considers the suppression of interfering item to be the main contribution of the hippocampus will be discussed.

## 1.11    Theory 3: interference

The apparent failure of amnesiacs to judge correctly whether they have seen an item before is accounted for by two different theories of amnesia as interference, rather than memory loss. The original version put forward by Warrington and Weiskrantz (1970) proposed that amnesia was caused by deficits in recall. We will also outline a more recent version that claims the hippocampus has an equally important role in preventing interference during acquisition.

### 1.11.1    Weiskrantz's version

Warrington and Weiskrantz (1970) suggested an interpretation of amnesiac behaviour in recognition, an account that makes amnesia largely a retrieval deficit. They had observed that on a list recognition task, amnesiacs performed at the levels of controls. However, if asked to learn a second list, the amnesiacs continued to produce the primed responses from the first list. They argued that good performance on the first list indicated that the amnesiacs were capable of acquiring the memories; what they lacked was the ability to suppress, or adjust to, competing responses. In particular, they also had no way of distinguishing between remembered items. Under Weiskrantz's latter day theory, expressed fully in 'Consciousness Lost and Found' (1997), the hippocampus is a tool for comparing and commenting on the comparison between more than one memory or, indeed, memory and current representation of the present, which is also a kind of memory. This account takes amnesia to be purely a retrieval deficit.

On this account, it is difficult to explain why conditions during acquisition should affect amnesiacs' performance in the recall task. Weiskrantz may be characterising something important about the features of tasks amnesiacs cannot perform. However, both the previous theories, in a way, offer an explanation of this; there is something

about hippocampal representations that mean they support such commentary and comparison. To be a serious alternative to the memory theories, Weiskrantz's theory needs to offer us something more about the contribution of the hippocampus to comparing memories.

## 1.11.2   A behavioural inhibitor: Gray's version

McNaughton and Gray (2000) propose that the hippocampus is not a temporary memory store, or in fact a memory store of any kind. In their theory the hippocampus is primarily an inhibition system with a twofold function; firstly to prevent inference - that will be explained below - and secondly to inhibit inappropriate behavioural responses. The latter links the hippocampus with the anxiety (fight/fight/panic) system, and indeed their book is called the 'Neuropsychology of anxiety'.

This theory is particularly interesting for its explanation of the behaviour of human amnesia "forgetting", and the discussion of *interference*. They claim not that amnesiacs have lost stored memories, but that they cannot successfully retrieve them, because they cannot suppress competing cued responses. This sits comfortably alongside the Weiskrantz theory that amnesiacs have lost insight into whether or not they are recalling something - rather like blind sight for memories.

McNaughton and Gray's theory is that the hippocampus' role in recall and in *learning* is to suppress the creation of competing alternatives. They give as an example the well known problem of symmetrical association forming. In a symmetrically connected hebbian system, discussed in section 3.7, learned association (due to correlation) of A with B automatically produces strengthening of connections between A and B, and also between B and A. The task of the hippocampus is to prevent B and A becoming co-associated. Amnesiac failure in memory tasks is not due to failure of recall, but over-abundance of recalled conflicting and interfering alternatives.

McNaughton and Gray take amnesia to be a problem with both acquisition and retrieval, as they think the same mechanism for suppression of similar or competing alternatives is needed in both.

### 1.11.3 Basic Hippocampal anatomy from the point of view of Mc-Naughton and Gray

Hippocampal regions are regarded as 'logic gates.' The functional view of hippocampal anatomy taken by McNaughton and Gray proposes that rather than looking at the feed forward trisynaptic loop as the essential computation unit, each individual region of the hippocampus performs an independent, separate computation.

The other aspect of the interference theory is reflected in the dispute about the components of recognition judgements. Weiskrantz (1970) and McNaughton and Gray (2000) both suggest that implicit and explicit recall can interfere with one another. Therefore one might expect this possibility to be reflected by mutual inhibition, or some form of neuro hormonal switching between the hippocampus and surrounding cortical areas. Figure 1.5 shows some of the ways the information flow through the hippocampal circuit can switch, according to the behavioural stimulus.

### 1.11.4 Response to the interference theory: retrograde amnesia

One strong plank of support for memory based theories of hippocampal literature is the phenomenon of temporally graded retrograde amnesia. Retrograde amnesia, as we defined in section 1.7 is loss of memories before the amnesia causing event. Amnesiacs do not seem to lose a class of memories completely, but rather a time period of memory before the operation. This is the phenomenon that has led to the postulation of a consolidation period. In this period, memories are moved from temporary storage in the hippocampus to permanent storage in the neocortex.

A considerable body of evidence (Kapur and Brooks, 1999; Jarrard, 2001; Squire *et al.*, 2001) has documented time limited retrograde amnesia in amnesiac patients. Memory theories such as McClelland and Goddard (1996) argue that retrograde amnesia only makes sense if theories are being slowly transfered out of the hippocampus.

There are also well documented objections to the consolidation theory. First of all, it takes a very long time, longer than any studied plasticity process. Secondly, the extent of retrograde amnesia varies depending on the the extent of extra hippocampal damage. Reed *et al.* (1999) describe a case of a few years of retrograde amnesia with hippocampal damage, versus a case of 25 years, in the case of damage to entorhinal

From McNaughton and Gray (2000), chapter 10
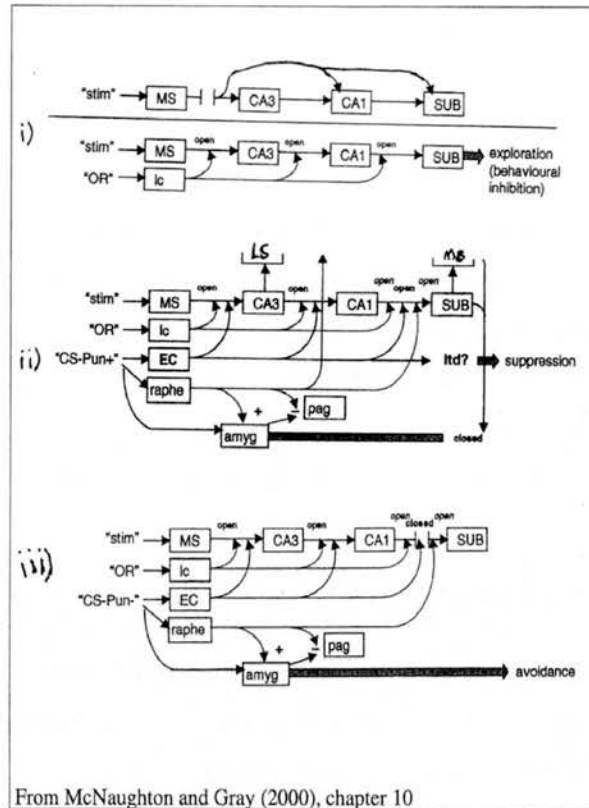
Figure 1.5: These examples show information flow though the hippocampus in different behavioural scenarios; i) Continuous exploration. ii) A situation of conflict, of goals, or between expectations, and curent input . iii) An entirely novel environment. McNaughton and Gray's idea is that parts of the circuit entorhinal cortex inputs suppress hippocampal output. They are also interested in the idea that input of some content, say from the amygdala, can selectively alter the composition of the hippocampal circuit.

cortex, plus hippocampus, plus subiculum.

Finally, it is not clear that loss of stored memories *is* a better explanation than loss of a particular form of recall. It may be that conscious recall becomes more and more difficult with the age of a memory.

## 1.12 Animal models of amnesia

Studying memory in other species (particularly rats, monkeys and chimpanzees). allows the study of the effects of precise lesions of the hippocampus and related areas. However, as animals cannot directly report the effects on memory, behavioural tasks have been devised that attempt to demonstrate amnesia like effects in the animals responses to the tasks. In these section we consider the merits of some of these tasks as models of human amnesia.

### 1.12.1 DNMS

The role of hippocampus and the contribution of perirhinal cortex has been investigated using the delayed non match to sample (DNMS) task as an analogy of antrograde amnesia (Aggleton and Brown, 1999; Wan *et al.*, 1999). DNMS involves the animal being shown an object, or an arrangement of objects. After a variable length delay, the animal is then offered a choice of novel object, and the object shown previously. The animal is required to choose the novel object. This task tests either the ability of the animal to hold an object "in mind" over a length of time, or it tests the recognition capacity of the animal; it is thus very similar to a recognition judgement task in humans. It can be solved simply by familiarity or novelty *detection*. Thus a dispute has broken out as to whether this task is or is not impaired after hippocampal lesions. The dual route model in humans would predict little or no impairment on this task, contrary to most experimental findings until a study by Murray and Mishkin (1998) that claimed that perirhinal lesions alone resulted in impairment on this task and ibotonic lesions, that spared that pathway between the perirhinal and the fornix, left performance on the task intact. This study was then contradicted by a similar study by Zola *et al.* (2000) that compared the various types of lesions, and claimed to find an impairment on the DNMS task with all of them. One difference in pre-operative training may be the root

of this discrepancy; Murray trained the monkeys pre operatively on the task over short delays. The animals were also not returned to home cages. However, Zola and Squire's monkeys are only impaired on long delays, over 40 seconds. And even on long delays, they still performed the task (albeit at around 65% success, rather than 80%. Murray's response to this discrepancy is to point out that the important thing to notice is that under some conditions the task is not impaired after hippocampal lesions, and it is impaired in *proportion* to perirhinal lesions. Most recently, Baxter and Murray (2001a,b) have published a reanalysis suggesting that hippocampal animals may perform better than controls, or partially lesioned animals. This result is not universally accepted in the literature (Clark *et al.*, 2001 looks agains at the reanalysis), but would be strong evidence for interference between recall pathways.

We have already observed that both the explicit recall and implicit memory recognition contribute to recognition judgements. It is not straightforward to compare these two components with hippocampal and perirhinal/parahippocampal regions. A much greater impairment was shown on the visual match task, that simply measured the amount of time or attention the animal devoted to looking at the novel stimulus.

## 1.12.2 Additional evidence regarding the parahippocampal regions in humans

One of the contributions of the post (homologous to parahippocampal regions in the monkey) and perirhinal cortices to recognition seems to be a novelty signal. Neurons have been found in these areas that signal both familiarity and recency (Aggleton and Brown, 1999). It is clear that a novelty signal is enough to do most tasks based on recognition. The perirhinal cortex has been found to be active by C-fos staining (Wan *et al.*, 1999), on the presentation of novel single items - alongside the temporal lobe. The postrhinal cortex and area CA1 showed increased activation for novel spatial arrangements of familiar items. This has led to the claim that it is the postrhinal cortex that is particularly concerned with space and spatial memory. This should be compared to the findings of Kirchhoff *et al.* (2000) in human subjects that, in general, a novel stimulus resulted in raised activation in a wide variety of areas concerned with memory, including the prefrontal areas. It was also found that the response to novelty in the hippocampal and parahippocampal cortices predicted success in recall from explicit

memory. It may be that DNMS, in not being hippocampal dependent, actually is an accurate model of human recognition judgements that are composed of a hippocampal and a non hippocampal component.

### 1.12.3   Spatial tasks/maze tasks

Maze tasks, such as the water maze, the radial arm maze and the T-maze, are used in reference memory tasks where the animal has to remember which arms have been baited, or the location of a hidden submerged platform, relative to landmarks or non visual cues.

Maze tasks are quintessential hippocampally dependent tasks. However, it is difficult to immediately see how they model amnesia. Again, relational or episodic theorists claim that the abstract nature of spatial relations make the memories hippocampal dependent. The other two theories account for deficits in the maze tasks, not by loss of a particular kind of information, but by loss of ability to privilege correct recall and successful trials, from current sensory data, or incorrect trials. This position predicts that rats might perform better in an error free learning situation, just as amnesiacs do.

There is a very influential account of hippocampal functioning in rats, that we call the strong cognitive map hypothesis, that basically divorces rat hippocampal functioning from amnesia. Chapter 4 will consider arguments against this theory, and discuss the nature of the impairment in maze tasks.

Looking ahead, the conclusion of chapter 4 is that the hippocampal representation supports both amnesia theories. Chapter 5 considers the 'sequence learning' interpretation of the contribution to spatial tasks, that also encompasses the configural learning tasks described below.

### 1.12.4   Configural learning

Configural learning tasks are particularly interesting as subtle variation in the task led to different hippocampal effects. They have been regarded as tasks that model the effects of a special kind of hippocampal retrieval 'contextual retrieval' (O'Reilly and Rudy, 2000; Norman and O'Reilly, 2001) and a special kind of hippocampal learning; 'relational' learning (Dusek and Eichenbaum, 1997) or 'conjunctive' (McClelland and

Goddard, 1996) or 'configural' (Sutherland and Rudy, 1989; Alvarado and Rudy, 1995, 1992; McDonald *et al.*, 1997), or 'contextual' (Good *et al.*, 1998) learning. Commentators frequently claim that these tasks are hippocampal lesion sensitive in rats because they require this special kind of learning or retrieval, and this is comparable to the kinds of learning and retrieval found in human episodic memories.

An alternative interpretation of the learning and retrieval operations required for configural learning problems has emerged from the modelling literature. Configural tasks have been modelled as sequence learning tasks (Sohal and Hasselmo, 1998; Levy, 1996; Wallenstein *et al.*, 1998). It has been argued that learning parts of an event in correct temporal order is a contribution of the hippocampus to event representation. This hypothesis potentially unifies accounts of amnesia and hippocampal deficits in rats. It has also been argued that since these configural tasks can be represented as sequence learning tasks, and learned by a hippocampal type architecture, this constitutes evidence for the hippocampal sequence learning. We shall come back to this point in chapter 6.

In the modelling section of this thesis we shall consider how well the representation of configural learning tasks as sequence tasks works. There is a problem (side stepped by these models), in that not all the configural learning tasks that can be modelled as sequence learning tasks are in fact hippocampally dependent. How much of a problem this is will become clear after we outline the tasks. The easiest task to represent as a sequence task is the transitive reasoning task. Dusek and Eichenbaum (1997) suggest that the transitive reasoning task in rats is a suitable analogue of the kinds of flexible relations that the hippocampus is supposed to encode. The transitive reasoning task is hippocampally dependent in rats. Transitive inference involves the transfer of properties between premise pairs. The subjects are trained to select one of a pair of items: $a$ is rewarded rather than $b$ (represented as $a > b$), and then three similar pairs $b > c, c > d, d > e$. They are then tested on the pairs $b > d$ and $a > e$. As $b$ and $d$ have been both rewarded and unrewarded choices with equal frequency, the correct selection of $b$ in test pair is considered an example of transitive inference between pairs. However, it is about applying a relation from one set of pairs to a new pair. It is therefore much closer to the artificial grammar learning task, in humans, which is not impaired in amnesia.

| 'feature' negative patterning | negative patterning | biconditional |
|:---:|:---:|:---:|
| XA + | A+ | AX+ |
| A- | B+ | AY- |
| X- | AB- | BY+ |
| B+ | | BX- |
| not impaired | slowed | mildly slowed |

Table 1.1: Impairments due to hippocampal lesions on configural learning tasks

Other configural learning tasks, such as negative patterning, capture interference effects. Negative patterning and 'feature' negative patterning both involve reversal learning. In negative patterning, stimulus AB+ represents the case where A is positively reinforced when combined with B. B- and A- represent the cases where A and B by themselves are not reinforced. The rat is trained on these alternative cases, and then its degree of learning is evaluated by its readiness to respond in these three cases.

Negative patterning was thought by configural association theorists (Alvarado and Rudy, 1995, 1992) to be another defining hippocampal dependent tasks in rats. However, studies with ibotonic lesions (more selective, fibre sparing) (McDonald *et al.*, 1997; Gallagher and Holland, 1992) found that in fact hippocampal rats were not completely impaired, but acquisition was slowed. A similar effect is found for the biconditional task. Both these tasks are described in figure 1.1.

The 'feature' negative patterning task, in figure 1.1 however, is not hippocampal dependent. The modelling of these tasks as sequence learning tasks, discussed in chapter 6, cannot easily distinguish between these two types of tasks.

It is possible to use the feature negative patterning task to examine a hippocampal memory dependent known as proactive interference. 'Proactive interference' occurs where learning a new item can interfere with the recall of an previously learned item. This effect only occurs in the intact animals. Hippocampally lesioned animals learn this task better than controls. It has been proposed (O'Reilly and Rudy, 2000) that the negative patterning task is a reasonable model of the negative transfer effect seen in list learning in amnesiacs described in section 1.11.1. However, the negative patterning task is not fully hippocampal dependent; it is a difficult task for an intact animal. An

alternative explanation of the acquisition impairment is failure to control responding.

It is clear that to show light on the impairment in amnesia, the precise nature of the impairment on configural learning tasks must be identified. We shall discuss the cause of proactive interference in the intact hippocampus, and the hypothesis that these tasks should be represented as temporal sequence problems in chapter 6.

## 1.13    Retrograde amnesia and animal models

Although antrograde amnesia can easily be modelled as the failure to acquire new memories, it is far harder to test animals for retrograde amnesia.

Despite technical difficulties, two recent papers, (Sutherland *et al.*, 2001; Riedel *et al.*, 1999) have considered disrupting storage of spatial memories retroactively. Although Riedel *et al.* (1999) have found a time limited period in which disruption can take place, they have not found clear evidence of a temporal gradient. Contrary evidence is presented in Ramos (1998). Winocur *et al.* (2001) claim to have found retrograde amnesia in a socially transmitted food preferences task.

However, these papers all find a role in short term consolidation of memories (less than a month). This does not compare very well to the human time scale of years. Although Riedel *et al.* (1999) clearly establish a time limited role for the hippocampus, they do not rule out the possibility that the role is more like that which McNaughton and Gray (2000) propose.

### 1.13.1    Conclusion

Overall, it does not seem that there is a definitive rat model of human amnesia. However, this does not mean we can learn nothing about amnesia through studying rodent lesion studies. It does mean that we should be suspicious of broad brush hippocampal theory that claims to describe both rodent and human. One of the themes of this thesis is that the rodent hippocampus does not perform one unified function. In the final chapter we will consider whether this is borne out by studies of intact human hippocampus. The next section discusses the possible developmental contribution of the hippocampus, that turns out to depend heavily on other human specialisations, such as language.

# 1.14 Learning and consolidation

## 1.14.1 Episodes are constructed not just represented

Goswami (1998) shows that children remember events using an event schema. They use the pattern of causal relations in an event to *construct* a memory of it. Goswami's data demonstrates that children make more errors on non causally related event recall, and also tend to displace details that are irrelevant or not causally contributory. The observation was made that in free or prompted recall, the child's discourse in reply to a question like "tell me about going shopping", is very causally structured and goal orientated. The episodic memory perhaps exploits this structure of propositions to assist in the recall of a difficult memory task.

The phenomenon of so called infantile amnesia bears on the development of episodic memory. Adults do not remember events below ages three or four. There are two proposed developmental explanations for this. The first is that young children's memories are coded perceptually, and with sensory motion coding that becomes unreadable by the adult after they switch to linguistic type encoding of events. The second explanation is that the brain areas such as the neocortex and hippocampal areas simply aren't fully functional, as they undergo extensive development. One aspect of explicit memory that is specifically episodic clearly shows developmental change in the children's ability to distinguish particular events in recall. Goswami describes a study that compared four year olds to seven year olds. Both groups were taken to the laboratory five times over two weeks, and shown puppet animals playing games. These games were the same on every visit, and occured in the same order. On one visit, the animals produced a novel game. The children were asked to describe what happened during their visits. The younger children tended to report that the novel event had happened on each of their visits, and included it in their general 'schema' of events. The older children made a distinction between that particular deviant occurrence and the general pattern of the event.

This study suggests that the ability to recall distinct episodes without interference is acquired late in development, alongside the development of the pre-frontal cortex and the hippocampus. As it happens, explicit memory follows a similar developmental path, while implicit memory does not improve with age. This kind of memory is tested

on, for example, perceptual matching tasks and fragment completion type tasks, where children are primed with a list of words, and then asked to complete presented letter pairs, e.g., CH **** with Cherry. In the implicit version of the task, the participants were simply asked to complete the letters with the first word that came to mind. In the explicit task, they were asked to complete the word as one presented in the list. (Of course, this is the same task that amnesiacs perform so well in the implicit version, and struggle with in the explicit version).

A study by Naito (1990) found that adults and children performed equivalently on the implicit completion task, but adults had far superior explicit recall of the words originally presented. This study has been replicated with picture fragments, and implicit memory found to be fully functional in children as young as three. Here the implicit task was to say what the fragments "looked like" as quickly as possible. The explicit task was to try and use the picture fragments to recall the list of previously presented picture. Performance on the explicit task correlated with age, reaching adult performance at about age 12.

This evidence strongly suggests the picture of human episodic memory as a specialised form of semantic memory, as described in Tulving's theory. Episodes are remembered as structured events. The interesting question is how is this event structure learned? Is it derived from neocortex, or does the hippocampus itself develop generalised knowledge about events, and event structure? The contrast already touched on in section 1.8.3, between the requirements of learning about generalised structure and learning about individual events has been well expressed in McClelland and Goddard (1996):

> ... The brain makes use of complementary learning systems... One of these the neocortical learning system, is specialised for the gradual extraction of structure from ensembles of events and experiences, leading to the acquisition of connection weights among neurons that support generalisation

This describes inductive learning, located firmly in the neocortex.

> The function of the second system, the hippocampal memory system, is to store new memories and retrieve them while they remain in storage, so they can be played back to the neocortical system for interleaved learning with learning with other memories and experience.

This picture of two complementary systems, related by a process described as learning, abstraction and consolidation, is still a useful one and much exploited (O'Reilly and Rudy, 2000).

We argue that it is no longer possible to neatly attribute the two varieties of learning to these two locations. We argue this from three points: (i) episodic memory depends on the existence of generalised knowledge acquired over multiple trials; (ii) the hippocampus shows evidence of representing information abstracted over many trials - see in particular chapter 4; (iii) it also shows evidence that a slower learning system exists within the hippocampus - see chapter 2.

One of the arguments put forward in this thesis is that consolidation should be thought of as a purely biological process of stabilising memories, in the medium or even the short term. It has no obvious implications for the kind of information being stored. Abstraction, or generalisation, is a description of the kind of information learned that says nothing about the stability or longevity of the information. The separation of consolidation from generalisation is an example of a switch from a content based approach to the hippocampus, to a strategy based approach. In general, the memory strategy used by the hippocampus for long term memory storage tells us nothing about the content of that memory.

## 1.14.2 Summary

We have examined ways of characterising the kind of representations stored in the hippocampus. Chapter one reviews standard attempts to characterise memory representations. Intuitively we have a notion of episodic memory, but attempts to refine it within psychological theories are not convincing. We show this in the first part of Chapter 1, by examining firstly the definition of declarative memory, defined relative to procedural memory, then episodic as defined relative to semantic memory. We argue that there is equally good evidence that memory loss caused by the damage to hippocampus, is due to interference between recall mediated by the hippocampus, and recall mediated elsewhere. Apparent memory loss may therefore be due to loss of a recall strategy, rather than loss of actual memories.

However, this evidence does not apply to the best known model of MTL amnesia, the rat with a lesioned hippocampus. We discuss various memory tasks for rats that

have been claimed to be equivalent to tasks taxing the human hippocampus, and draw the conclusion the match is not particularly good. We also suggested that episodic memory is closely bound up with language and with consciousness, and therefore is likely to be quite different in the rat. The majority of the remaining chapters are on hippocampal memories in the rat.

Finally we have established the idea of ' an episode' as containing information gained over many similar instances, and also some information from one particular instance. In chapter two we propose that information is learned at different speeds in the ventral and dorsal hippocampus respectively, and that dorsal hippocamus learns information about particular single instances.

# Chapter 2

# The contributions of dorsal and ventral hippocampus to memory.

## 2.1 Summary

This chapter presents a theory that dorsal and ventral hippocampus play different roles in the learning and recall. The best known interpretation of the failure of hippocampal rats on some learning tasks, has tended to identify those tasks as 'spatial'. We examine a claim Moser and Moser (2000) that these spatial tasks depend particularly on learning and recall in the dorsal hippocampus, and reject in favour of the idea that the dorsal hippocampus plays a particular role in recall from single instances. The dorsal hippocampus has been shown to learn more quickly than the ventral hippocampus; lesioning the dorsal hippocampus removes an effect of interference from the recall of successful behaviour on a previous trial. We contrast this with the idea that the ventral hippocampus, as it has been shown to learn more slowly, may learn more general attributes across sets of similar trails. We therefore propose an analysis of behavioural tasks, that suggests that some tasks might be better performed by animals with dorsal lesions.

By the end of chapter two , we have established the following critical points.

- The differences in LTP and LTD between ventral and Dorsal hippocampus, support the idea that the learning speed is slower in the ventral hippocampus.

- Behavioural data from lesion studies, selectively lesioning the poles of the hip-

pocampus, offers support for the idea that the ventral hippocampus learns more slowly.

- We therefore hypothesise that the Dorsal hippocampus is specialised to recall 'single trial' events. We suggest that the memory output from the dorsal hippocampus biases the animals behaviour towards repeating its (successful) behaviour on the previous similar trial.

- We therefore predict that animals with only temporal hippocampus will perform better than animals with only dorsal hippocampus on transverse patterning tasks, and perhaps even better than controls.

## 2.2   Introduction

In most accounts of the human hippocampus, the aim has been to provide a unified account of hippocampal functioning. However, a new attractive hypothesis has appeared in the rat literature that suggests this may not be a feasible objective. The idea is that the hippocampus is dissociable into two functional circuits along the dorsal-ventral axis. Experimental support for this hypothesis was provided in 1993 by Moser *et al.* (1993), and then reinvestigated in a number of studies(Hoz, 2000; Bannerman *et al.*, 1999; Hock and Bunsey, 1998). Moser and Moser (1998a) proposed that the dorsal hippocampus was primarily concerned with spatial processing and the temporal, less accessible, pole with some unspecified function. Candidates for that function range across most of the hippocampal dependent behaviours that do not fit comfortably into memory based accounts, such as behavioural disinhibition, impairment to conditioned freezing, etc. These behaviours are those that in particular interference accounts of hippocampal functioning (described as theory 3 in chapter 1) adopted as primary. Obviously, Moser and Moser's proposal (1998a) potentially provides a way of reconciling these two accounts. This thesis proposes a different theory that is consistent with Moser's results, but which draws on some more recent experimental results.

The three dimensional organisation of the hippocampus
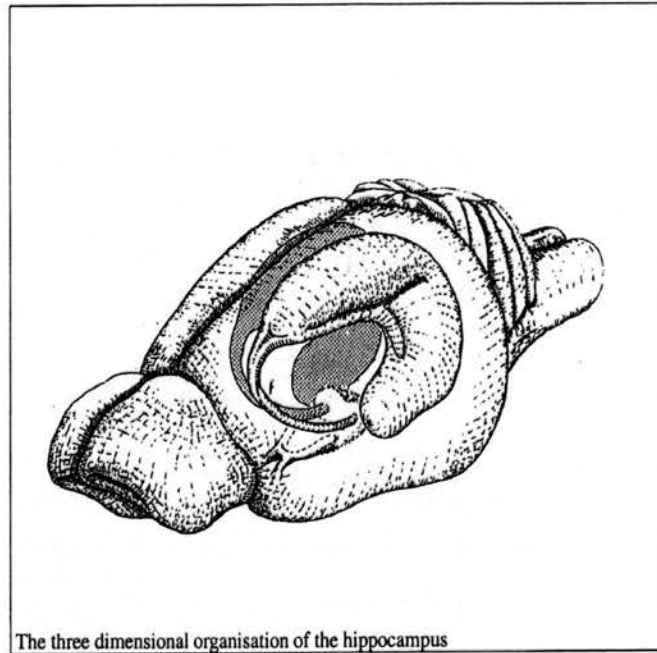
Figure 2.1: This figure shows the position and orientation of the hippocampus. The fornix is marked at the dorsal tip. The drawing shows the output fibres that travel through the fornix extending backwards, on the hippocampal surface, to ventral area.(Amaral and de Witter, 1995)

## 2.3   Dorsal and ventral hippocampus

Dorsal and ventral hippocampus are often also referred to as the septal and temporal poles of the hippocampus, particularly in anatomy papers. In the monkey, dorsal hippocampus is often referred to as posterior hippocampus, while ventral hippocampus is known as anterior hippocampus. In this chapter dorsal and ventral hippocampus will be referred to, whereas in chapter 3.10 the anatomical convention will be followed.

## 2.4    A theory of content differentiation

### 2.4.1    Moser's original proposal

Lesions of the dorsal hippocampus affect both acquisition (Moser *et al.*, 1993; Moser and Moser, 1998a; Moser *et al.*, 1995; Hock and Bunsey, 1998) and recall (Moser and Moser, 1998b) of a quintessential spatial memory task, such as the reference memory task in the Morris water maze. Moser divided the hippocampus into sections along the dorsal-ventral axis. The results in Moser and Moser (1998a) show that animals with 70% of hippocampus left intact (30% lesioned from the ventral pole) performed as intact animals. The closeness in location of the fornix to dorsal hippocampus could suggest that the fornix/fimbria input/output system might be damaged by dorsal lesions. However, these recent lesion studies are ibotonic fibre sparing, in an effort to leave as much of the input and output connections intact as possible. Moser and Moser (1998a) argue quite convincingly that animals with ventral end spared show improvement on the spatial task, as the lesion moves past the mid-line, so it is not just the dorsal tip/beginning of the fornix that is required. Also, fornix lesioned animals show severe impairments of conditioned freezing, and which is associated with ventral lesions (see below) rather than dorsal lesions. Crucially, animals with progressively less dorsal hippocampus showed a decline in performance proportional to the amount of dorsal hippocampus removed.

It was not clear whether the effect of dorsal lesions only produces a deficit in recall, or in acquisition of spatial memory. Moser and Moser (1998b) therefore conducted a reversible lesion study to try to clarify this. It was found that a small drug-induced lesion of the dorsal end of the hippocampus was sufficient to disrupt recall in pre-operatively trained animals. A control group of animals tested 48 hours after injection of the drug in the same protocol showed no impairment, confirming that the lesion was genuinely reversible. This is not quite the same as showing that the effects of the lesion are always reversible; animals were killed in the test case to examine the extent of the lesion. It is therefore not known whether the stored memory could have been retrieved successfully, as recalling a memory under damaged conditions might affect the actual content or stability of the memory. Recent results on reconsolidation after recall in the amygdala in Nader *et al.* (2000) suggest the memory in that region becomes labile

during recall even if it has previously been consolidated.

The most interesting finding was that animals trained and tested during the period of the lesion showed no impairment on either learning or recall. Moser *et al.* (1995) claims that this shows that the spatial memory is a relational structure that uses the entire hippocampus available, and that successful recall depends on the structure of the hippocampus remaining the same between learning and recall. There is an implicit suggestion, therefore, that the ventral hippocampus does not form part of this relational structure to the same degree. Moser *et al.* (1995) did not investigate whether, after the lesion had disappeared, recall became impaired for the animals trained during the lesion, as their theory predicts.

Although Moser *et al.* (1995) argues that the data suggest that the representation in the hippocampus is disrupted, the data could equally well support the idea that the dorsal hippocampus has a particular role in recall. Moser and Moser (1998a) claim further support for the division of the hippocampus, by claiming that the inputs to the hippocampus via the perforant path were divided in content. Visual (i.e. spatial[1]) information is channelled to the rostro-medial zones only, while only olfactory information reaches all three. The anatomical picture supporting this claim is as follows. Entorhinal cortex can be divided into three zones; the caudo-lateral zone projects to the dorsal half of the dentate gyrus, the intermediate zone to the adjacent quarter, while only the rostro-medial zone projects to the ventral hippocampus. Consequently, they claim that the temporal pole of hippocampus does not receive as much spatial information as the ventral two thirds. For this separation of inputs to provide a basis for two separate circuits, the information needs to be kept segregated through the hippocampus. Moser assumes "....a largely unidirectional transverse loop through the dentate gyrus, CA3, CA1 and subiculum" repeated along the longitudinal axis. Consistent with the lammellar hypothesis, Moser also draws attention to the greater number of mossy cells, and of bioaminergic receptors at the ventral pole but does not speculate on their possible functional significance. For the moment, we will leave the anatomical aspect of Moser's theory on one side, returning to it in the next chapter (in particular section 3.9.1) and concentrate on the behavioural aspect.

---

[1]spatial information is not necessarily visual, as there is also self motion information, path integration, timing intervals between other cues

## 2.4.2    The putative role of ventral lesions

Support for a difference in ventral and dorsal hippocampal functioning based on different content would be greatly increased if a double dissociation between tasks associated with lesions of ventral and dorsal areas could be found. Having identified dorsal lesions with spatial tasks, attempts have been made to identify a ventral dependent task. Some place cells[2] (Granger *et al.*, 1996) have been found in the ventral hippocampus, but those experimenters who support a strong distinction in the functioning of the dorsal and ventral parts claim that as the division (if there is one) occurs closer to the ventral end, the more likely it is that these place cells may not be 'real' place cells (Moser and Moser, 1998a). It has been found that place cells in the ventral hippocampus were found to have larger fields and apparently less spatial discrimination (Jung *et al.*, 1994). Several recent studies (Richmond *et al.*, 1999; Bannerman *et al.*, 1999) claim that conditioned freezing is dependent on ventral hippocampus. Freezing behaviour is an automatic response to the expectation of an electric shock. When footshock is paired with a light or a tone stimulus, the rat quickly becomes conditioned to expect the footshock, and produces the freezing response. Freezing is therefore regarded as providing a potential measure of the rat's ability to remember the association.

Conditioned freezing is not, however, an unambiguous task. Firstly, it is not a task, in the sense that navigating a maze is a task, as it is the animals automatic or implicit response to the environment that is measured, rather than the animals ability to *use* previously learning information. Secondly, a distinction should be drawn between freezing that is conditioned on context, where the context in which the conditioning occurs is considered to be the box, and freezing conditioned on a stimulus, such as a tone that occurs while the rat is n a stable context. Rats can also learn to associate the background context with footshock. Philips and LeDoux (1992, 1994) found lesions of the amygdala disrupted both conditioned responses, while lesions to hippocampus disrupted only conditioning to context.

Bannerman *et al.* (1999) do find an impairment in freezing response after ventral lesions, but find it on *both* these tasks. There are two different possible explanations. Failure to show conditioned freezing could be an impairment of the freezing motor

---

[2]Place cells fire when an animal fires in a particular location in a particular environment; see chapter

4

response, rather than a failure of memory (Maren and Fanselow, 1997). Secondly, the ventral end does receive input from the amygdala and thalamus, and outputs back to those regions (see discussion of the limbic system in chapter 3, section 3.5). So a break of contextual information flow to the amygdala would account for the task impairment, without pointing to particular computational deficit within the hippocampus. Another point worth bearing in mind is that the impairment in freezing response has been found not to correlate well with other measures of fear conditioning, such as the startle response and behaviour (Desmedt *et al.*, 1998), again pointing to an impairment of freezing response rather than in memory. In summary, we conclude that at present it is not possible to identify the role of the ventral hippocampus in the conditioned freezing response.

It is therefore proposed here that the reason why it seems so difficult to identify a ventrally dependent task is because of the bias towards thinking about tasks in terms of content of information needed, rather than the computational demands of the task.

The failure to identify a unique role of the ventral hippocampus becomes more damaging in the light of work reviewed in chapter 4. It is clear that spatial versus non-spatial content division partly relies on being able to identify spatial-only cell firing, and its role in behaviour. Chapter 4 will argue that the pure spatial theory is no longer supported by evidence from the place cell paradigm.

However, a replication of Moser's experiment suggests that a difference between dorsal and ventral hippocampus is not a difference in information content but in computational mechanism, as is now explained.

## 2.5   A theory of computational differentiation

Moser's original results have also been replicated in Hoz (2000) and Hoz and Morris (1999). Hoz (2000) showed that using a slightly different training protocol removed the differences between the ventral and dorsally lesioned animals. The essential differences are these: the rats in Hoz's experiments were trained for two additional days, with fewer trials on each day; Moser's rats were trained for four days, with six trials per day. Hoz trained two groups, one as Moser, for six trials over four days, the second group for four trials per day, over eight days. At the end of the Moser training group,

there was a significant split between the animals lesioned in ventral and dorsal parts. At the end of the eight days, the groups were indistinguishable. Both were unimpaired relative to the control group of completely lesioned animals.

These startling results suggest strongly that there are significant computational differences on the dorsal ventral axis, perhaps to do with the either the speed of learning, or the speed of consolidation. Hoz's interpretation of her own data is ambivalent between the possibilities that the ventral hippocampus learns more slowly and that it consolidates more slowly. Under the latter interpretation, information is being slowly transferred to the neocortex; once there, the task is no longer hippocampally dependent, so the ventrally lesioned rats are using their neocortical representations to solve the task. Hoz does not develop this interpretation. However, an alternative interpretation suggests that the difference might be found in the learning mechanisms in the hippocampus.

Another replication of Moser's study by Richmond *et al.* (1999) found dorsal lesioned rats able to learn the water maze task in six days, with four trials per day. The identical protocol in Bannerman *et al.* (1999) on the same task found an impairment. These studies differed in the kind of pre-training the rats were give. Richmond *et al.* (1999) pre-trained on a fear conditioning task, whereas Bannerman *et al.* (1999) pretrained in a T-maze. Hoz suggests that this discrepancy may be due to the ventral hippocampus being engaged in the fear conditioning task. However, there is also a possibility that performance was affected by the number and selection of start positions. Bannerman *et al.* (1999) used eight positions, chosen randomly, whereas Richmond *et al.* (1999) used four, one for each daily trial, in random order. Dorsal lesioned rats still performed worse than ventral lesioned rats, even though they performed better than controls.

## 2.5.1  Physiological evidence supporting differences in learning

Let us suppose that the Hoz replication is explained by faster learning speed in the dorsal hippocampus. Are there any physiological differences between ventral and dorsal hippocampus that might help to account for differences in the speed of learning? First of all let us outline the nature of the biological changes that are thought to underly learning in the hippocampus.

## 2.6   Long term potentiation as the basis of learning

Long term potentiation (LTP) fits the prescription that Donald Hebb (1949) proposed as the basis of biological substrate of learning. The idea of hebbian learning is simply that the synaptic strength between two neurons changes in proportion to the degree of correlation in their firing. Long term potentiation is an increase in synaptic strength that occurs when a presynaptic neuron fires up to 50 ms before the post synaptic neuron firing. It was discovered first by Bliss and Lomo (1973) and has been described in detail at hippocampal pyramidal cell synapses. Depression also occurs at these synapses if the post synaptic neuron fires before the presynaptic neuron.

The best understood type of LTP is mediated by the NMDA receptor. The post synaptic NMDA receptor is blocked by Mg2+ at resting potential, but on depolarisation the Mg2+ block is relieved, permitting an influx of Ca2+. This activates at least three protein kinases that produce the change in synaptic strength. These may involve the insertion of new AMPA receptors into the postsynaptic cell or the phosphorylation of AMPA receptors. NMDA receptors provide a triggering signal for LTP, though they do not participate in normal synaptic transmission. Other triggering signals are known to use similar, sometimes overlapping effects (Cowan *et al.*, 2001).The same mechanisms are activated during long term depression (LTD).

There are several protocols for inducing LTP and LTD in the hippocampus: tetanic LTP (high frequency stimulation of the neural afferents), paired pulse facilitation, (input pulses paired with post synaptic depolarisation), theta burst stimulation (a series of high frequency bursts at the theta frequency). LTD is harder to induce and stabilise, but can usually be induced with low frequency stimulation. LTP or LTD can appear within half an hour, and last (depending on protocol and type) for over 24 hours.

Short term potentiation or synaptic facilitation can appear within a spike train.

Recent studies of LTP, LTD and short term depression (STD) in region CA3 in vivo and slice[3] have found significant differences between the ventral and dorsal poles in the schaffer collateral synapses. This data is presented here, in this chapter, rather than in the anatomy chapter, because the papers reviewed were clearly looking for effects that fitted with Moser's hypothesis, which is referred to directly here. Papatheodoropoulos and Kostopoulos (2000a) find that tetanic (100 Hz stimulus) LTP, which produces

---

[3]see chapter 3, figure 4 for a description of regions of the hippocampus, and laminar organisation

a reliable effect of 30% increase in EPSP slope in the dorsal hippocampus, produces only a 5% increase in the ventral pole. This effect was blocked by AP5, which blocks NMDA receptors. They therefore suggest that it may be caused by the lower density of NMDA receptors that has been found in the ventral pole. Papatheodoropoulos and Kostopoulos (2000b) found a deficit in short term potentiation across the full range of inter pulse intervals. A second set of results by Izaki *et al.* (2000) found preferential induction of LTD in ventral hippocampus by a low frequency pulse bursts. This proto-col, used in slice and in vivo, also produced LTD in dorsal hippocampus, but the LTD was not stable. The same protocol used in the ventral hippocampus produced LTD that remained stable for an hour and a half after induction. These results have been replicated recently by Maruki *et al.* (2001b).

Potentially, differences in LTP and LTD provide evidence that the substrate of learning is different in the two different poles, and the difference is of the right kind to explain the effect of increasing the number of trials, or the number of days on learning.

However, it does not explain why performance in the dorsal hippocampus should correlate with the number of trials on the reference memory task, while performance in the ventral hippocampus should correlate with the number of days. One possibility is that the smaller effects of NMDA LTP mean that effective learning requires some other mediating process that takes a long time.

This data suggests two ways in which learning may be different. Firstly, learning in the ventral hippocampus may still be NMDA dependent, but require more repetition of firing patterns to induce comparable levels of potentiation. Secondly, learning in the ventral hippocampus may be mediated in the hippocampus via a mechanism that is complementary, or even independent of the NMDA receptor.

## 2.6.1 Alternative substrates for learning in the ventral hippocam-pus

One kind of NMDA mediated LTP (late LTP) has turned out to be blocked by D1 re-ceptor antagonists. It is also known that dopamine receptors are strongly differentially distributed to favour the ventral pole. It might be that ventral hippocampus shows a stronger effect of late LTP - this form of LTP cannot be induced through the protocol used by Papthegopopulous. Huang and Kandel (1995) found an enhancement of late

LTP by the activation of D1/D5 receptors. They also found a non-NMDA component that was enhanced by dopamine, that Kumar (1999) attributes to $\alpha$- amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) mediation. As described in the next chapter, both AMPA receptors and dopamine are distributed to favour the ventral pole.

This late LTP although quickly induced has much longer lifetime of duration. There are thus two possible ways this learning mechanisms might account for slower learning in the ventral hippocampus.

The first possibility is that the ventral end plays a greater role in learning when acting in response to a reinforced signal. Dopamine responses lower LTP threshold or increase the potential response. The dopamine can act retrospectively, to increase or even reverse from LTD after a tetanus is applied or it can act to prime the synapses, and increase the probability and the magnitude of the LTP response. Dopamine triggers potentiation that acts very like LTP but is independent of NMDA receptors, known as dopamine evoked potentiation (DEP). However, although it is easy to see that dopamine has a potentiating effect at synapses, its overall effect on the integration of inputs and the firing threshold of the neuron is not so clear. This matter will be discussed at length in chapter 3, section 3.8. The second postulates an interference effect, akin to that described over a longer time scale. LTP expression and consolidation needs a longer interval between trials.

The crucial difference might not turn out to be the number of trials per day, but rather the intertrial interval. Tetanic LTP cannot be induced in the same pathway under a certain interval. The same might be true of late LTP, on a longer time scale. This issue has not yet been investigated.

### 2.6.2 The functional significance of slow learning

One suggestion has been tentatively made by Jung *et al.* (1994), who claimed to find fewer and more diffuse place cells in the dorsal hippocampus. They suggest that the dorsal units encode more general information. This is consistent with the idea that dorsal cells require more accumulated firing events. The relational memory theory discussed in chapter 1 in fact predicts that some cells are able to accumulate information over multiple trials. Wallenstein *et al.* (1998) and Eichenbaum (1999) have identified the need for two different types of representations. The two do not have different

content in the Moser sense but have different levels of abstraction.

Eichenbaum (1999) describes two different kinds of representations in the hippocampus memory space: (i) sequentially organised cell firing that represents the 'automatic recording of unattended experience'. These cell firings represent the passage of events within a particular episode; (ii) however, some cells form nodes, defined as common elements across these episodes, which may be either common places, tasks, or behaviours. These nodes are intended to capture the structure across episodes.

How are these two different kinds of nodes constructed? Eichenbaum and colleagues put forward an argument that network models such as the models of Levy, Wallenstein and Hasselmo automatically create these two kinds of nodes, when solving the problem of how to link discontinuous events through time. They require cells performing different kinds of learning. Oddly enough, they also require the different kinds of learning proposed for the dorsal and ventral poles of the hippocampus.

It is therefore proposed here that we should be able to separate learning on two time scales. Learning within the dorsal hippocampus is genuinely one trial, recall in the dorsal hippocampus is therefore biased towards recalling that single trial. Learning in ventral hippocampus may be biased towards learning information derived from multiple trials. The next section describes an effect that this theory predicts will be limited to the dorsal hippocampus, as it appears to be derived from the previous trial memory.

Learning across trials may potentially exploit information flow along the dorsal-ventral axis of the hippocampus, thus departing from the trisynaptic, forward flow that has been presented so far.

## 2.7  Proactive Interference

Nobody has looked directly yet at whether rats can distinguish the memory of the previous trial, from a more general memory of the task and environment. However, the effect of the inter-trial interval has been investigated, in both learning and recall. Potential support for the possibility that the dorsal hippocampus recalls the memory of a previous trial, rather than an accumulated memory is provided by the phenomenon of proactive interference. This phenomenon is described in Han *et al.* (1998), where they looked at the effect of intertrial spacing in both learning and recall. They were looking

$$
\boxed{\begin{array}{c} X \to A+ \\ A- \\ X \to B - \\ B+ \end{array}}
$$

Table 2.1: Reward contingencies for the negative patterning task.

at a serial negative patterning task, represented in figure 2.1. The animal learns to associate X followed by A with reward, and A alone with no reward. X and B together are not rewarded, while B alone is rewarded.

This task was once thought to be hippocampal, as configural learning theorists claimed that reversal tasks of this kind were generally dependent on a hippocampal mechanism. However, Gallagher and Holland (1992) showed otherwise, and in 1998 found that hippocampal animals performed better than control animals on this task, in certain circumstances. Han *et al.* (1998) compared animals trained at 1 min intervals, with animals trained at 8 min intervals. Both lesioned and control animals performed equivalently during training on the long interval but, at the short interval, lesioned animals performed better.

They also found an effect of recall. Intact animals were vulnerable to proactive interference. Animals trained on long intervals and tested on short intervals performed worse than when long intervals were allowed for testing. They were also worse than lesioned animals trained on long intervals and tested at the same short intervals.

This effect was found to be due to the nature of the previous trial: animals performed worse when the negative (unreinforced) trial followed a reinforced trial containing the same stimulus item. So it is as if the memory for the previous trial 'lingers'. Han *et al.* (1998) claim this is an effect of the hippocampus based recall of the previous trial. The preceding theory would suggest that the effect should be limited to the dorsal hippocampus.

Their results also suggest a way of dividing up tasks that might correctly predict which tasks would show better performance with intact dorsal hippocampus versus intact ventral hippocampus. If dorsal hippocampus has a bias towards recall of the previous trial then any task in which the effects across trials are additive should be prefer-

entially learned by rats with dorsal hippocampus spared. Dorsal hippocampus spared animals should also show gradual improvement. However, ventral spared rats should perform as well as, or better, than dorsal spared, or intact rats on tasks (like reversal tasks) where performance requires inconsistent responses. However, they should be relatively impaired on additive tasks, like the water maze. They should also show different pattern of improvement, as in Hoz's replication; they should improve suddenly, as the slow learning system moves over threshold.

It is necessary to distinguish the interference effect found by Han *et al.* (1998) from the kind of negative transfer interference usually associated with this task. Effects of negative transfer, as we said in the previous chapter, are effects in learning. The negative patterning does produce such an effect, but it is not specifically hippocampally dependent. Proactive interference is closer to accelerated forgetting, due to the recall of the opposite reward contingency. It is worth pointing out that the rats do not know they have passed into a phase that is testing previously acquired responses - a behavioural strategy that always updates its response by the most recent piece of information about reward consistency might be a sensible approach.

However, the consolidation hypothesis suggesst that it is not trial spacing but the long/overnight rest period that enabled ventral hippocampus to learn or consolidate elsewhere. The only way to discriminate between these hypotheses is through further experimental work. It is possible to model a task like Gallagher's, and show that a similar effect of accelerated forgetting is associated with having a large rate of potentiation/depression. This is discussed in chapter 5.

## 2.8   Recent work on tasks that show dorsal and ventral differentiation

In the time since Moser's study was published, a number of studies of the effects of dorsal, and ventral lesions on hippocampal tasks have been made.

### 2.8.1 Conditioned place preference

Fertinbeau and McDonald (2001) investigated the effects of amygdala, fornix, complete, dorsal and ventral hippocampus lesions on conditioned place preference. This task measures the tendency of the rat to stay out of half of a conditioning box in which it has previously received an electric shock. Dorsal, complete and amygdala lesions were found to impair learning on this task, while fornix and ventral hippocampus lesions enhance learning. These results support the idea of a dorsal/ventral disassociation, but do not support the idea that ventral hippocampus is more closely associated with the amygdala. This is an example of a task that is consistent across trials, in which the memory of the previous trial gives as much information as memory of all the trials. Therefore, ventral lesion might result in enhanced learning through biasing hippocampal output towards recall of the previous trial.

### 2.8.2 Delayed alteration

Delayed alteration is another task requiring memory of performance on the previous trial. Indeed, Maruki *et al.* (2001a) found that dorsal inactivation resulted in impaired performance, at long delays, while ventral inactivation produced no impairment. However, Maruki *et al.* (2001a) remarks that ventral inactivation produces a tendency to persevere in a particular response. Whether this observation supports the hypothesis under discussion here is a matter of interpretation. Perseverance errors are associated with pre-frontal deficits, another circuit that is preferentially engaged with ventral hippocampus. However, such a response is also consistent with the rat possessing accumulated information that 'rewards are (sometimes) available on the left arm', without being able to recall that it had just previously visited the left arm.

## 2.9 Summary

The weight of experimental evidence that there is some sort of dissociation between ventral and dorsal hippocampal function in rats is now quite impressive. However, the emerging data no longer fits the spatial/non-spatial explanation proposed by Moser. The hypothesis proposed here is a distinction between two memory functions within

the hippocampus, where one is to recall the immediate past occasion and the other is to record general properties of the environment and task contingencies. A bias in the dorsal hippocampus towards recall of the most recent trial is one possible explanation of the dorsal/ventral data that has not yet been tested explicitly.

The next chapter follows on from this by considering the anatomical support for dissociation between ventral and dorsal hippocampal circuits. It also puts the observations on NMDA dependent LTP in this context. Chapter four considers the hypothesis that the hippocampus is a 'spatial' cognitive map. Our ability to identify a clear task set that the ventral hippocampus performs, appears to compare unfavorably to the clear identification that 'spatial' tasks are performed by the dorsal hippocampus. However, chapter four, when we get there, will argue that the computation role of the dorsal part is also underspecified.

# Chapter 3

# Anatomy

## 3.1 Summary

In the previous chapter we discussed the theory that dorsal and ventral hippocampus processed information with differing content; Spatial information through the dorsal hippocampus. We argued against this possibility, and suggested a difference in the recall and learning strategies of dorsal and ventral hippocampus. In this chapter we present a systematic review of the and internal circuitry of the hippocampus, and describe the connections to and from hippocampus. We use this review to argue against the idea that dorsal and ventral hippocampus receive different inputs that are kept segregated through hippocampal processing. The chapter also describes some features of the micro-circuitry of the hippocampus( function unknown), that differ between ventral and dorsal areas. We also argue that these micro-circuits should be taken into account by hippocampal modellers, as the assumptions used in neural network models about hippocampal connectivity are at best wrong, and occasionally misleading.

By the end of chapter 3 we have established that theories that rely on separation of different types of input remaining segregated within the hippocampus are not consistent with anatomy.

## 3.2    Three ways of looking at hippocampal circuits

There are three views of information processing in the hippocampus that can be fitted to hippocampal anatomy. We have already mentioned the trisynaptic circuit, where the hippocampus is viewed as a feed forward excitatory system. This circuit forms the basis of a number of models from Rolls (1996) through to Guazzelli *et al.* (2001). Another view, outlined in chapter 1, treats each layer of the hippocampus as a separately gated computation unit. Finally, a third view (outlined below) regards the hippocampus as divisible into two parts along the septo-temporal axis. In this chapter, the anatomical and physiological support for these three views will be evaluated. In particular, it will be argued that the lamellar/trisynaptic circuit picture of information flow through the hippocampus is insufficiently precise.

## 3.3    Functional differentiation along the septo-temporal axis: the neurobiological basis

In the previous chapter, the proposal due to Moser was presented in section 2.4.1. He argues that hippocampus can be divided into two parts: a dorsal circuit that deals with spatial memory and a ventral part that controls either memory for reward or behavioural inhibition. Part of the supporting evidence for this proposal was the observation that the input coming through the direct cortical pathway that terminated in dorsal hippocampus is predominantly spatial. However, in chapter 2 we put forward a counter proposal suggesting there are computational differences between the septal and temporal ends related to differences in speed and substrate of learning. This chapter supplies the essential anatomy and neurophysiology needed to make sense of this proposal.

Figure 3.1, adapted from Amaral and de Witter (1995), shows the position of the hippocampus and the relative arrangement of septal and temporal poles. These poles correspond to the areas lesioned in ventral and dorsal hippocampal lesions. Also shown is the orientation of the hippocampal slice and some of the positions of the subcortical and cortical areas that project to the hippocampus. Figure 3.3 shows the full range and variety of afferents the hippocampal region receives, without showing the details of the connectivity map. What follows is not a comprehensive review of all projections

Adapted from Amaral and de Witter (1995)
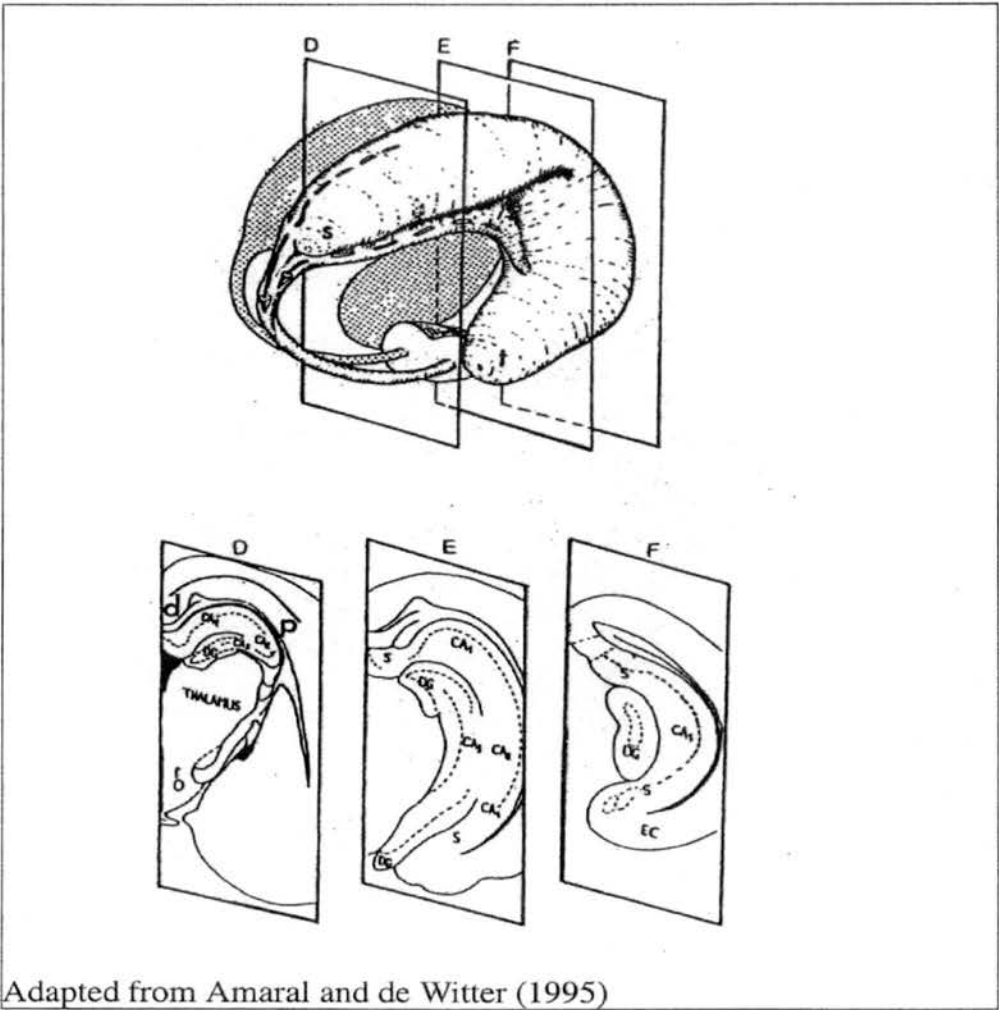
Figure 3.1: The location of septal (dorsal) and temporal (ventral) poles seen in horizontal section.

| subcortical inputs | subcortical outputs |
|---|---|
| Amygdala | Amygdala |
| Claustrum | Olfactory regions |
| Supramamillary nucleus | Mamillary nuclei |
| Lateral hypothalamus | Hypothalamus |
| Anterior thalamus | Anterior thalamus |
| Midline thalamus | Midline thalamus |
| Ventral tegmental area (VTA) | Nucleus accumbens |
| Raphe nuclei | |
| Locus coerulus | |

Adapted from Amaral and de Witter (1995)
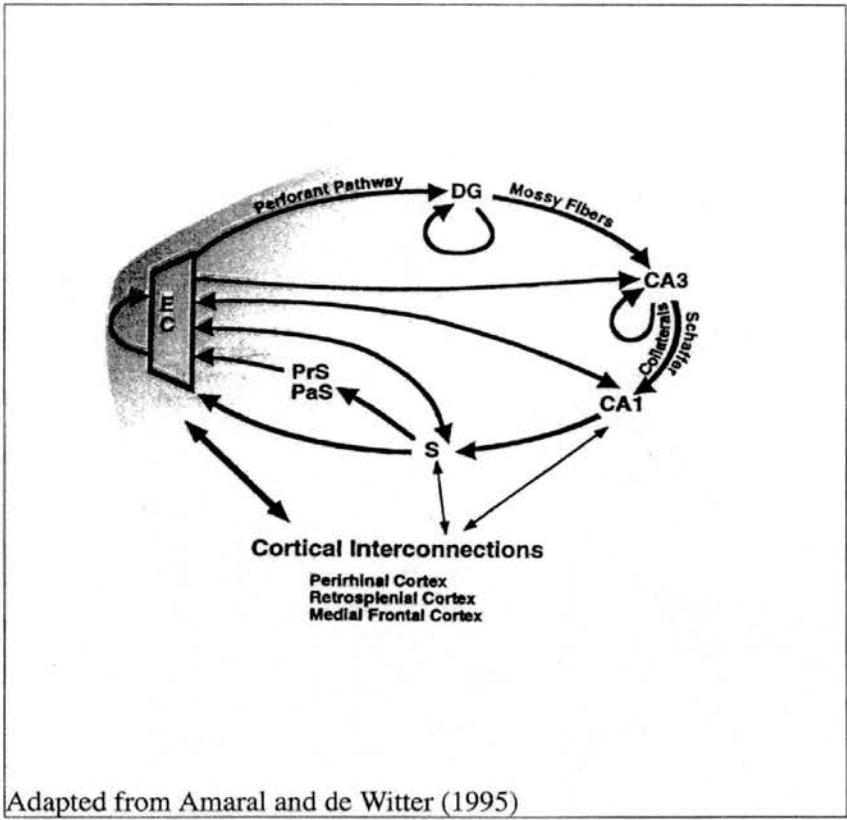
Figure 3.2: Hippocampal circuit connections.

| DG | Dentate gyrus |
|-----|------------------|
| EC | Entorhinal cortex |
| PrS | Presubiculum |
| PaS | Parasubiculum |

to and from the hippocampus, but merely enough to show that functional differences are possible.

## 3.4    The lamellar hypothesis

The lamellar hypothesis was proposed by Anderson *et al.* (1971) and Anderson and de Witter (1989) and is still heavily relied on in the modelling literature. In Moser's hypothesis it is necessary to keep the content on the dorsal circuit separated from the ventral circuit. Evidence for organisation in lamellae is strongest for the mossy fibre projection to CA3, as described in Tamamaki and Nojyo (1991).

### 3.4.1    Dentate gyrus

The dentate gyrus is divided into two excitatory layers. The granule cells form the polymorphic layer and send a mossy fibre projection, at the same septo-temporal level, to the molecular layer, that contains mossy cells. The molecular and polymorphic layers (also known as the hilus) are shown on figure 3.3. The mossy fibre connection is point to point, organised in fibre bundles, and travels onward to the CA3 pyramidal cells. The only cortical input to the dentate gyrus is via the perforant path projection from entorhinal cortex. Subcortical, partly cholinergic input from the septum targets the polymorphic cell layer, as does a substantial noradrenergic and seratonergic input from the brain stem. There is a very small dopaminergic input from the ventral tegmental area (VTA). A narrow band of the molecular layer also receives input from the hypothalamus.

According to the trisynaptic model, 'the mossy fibres form the excitatory input, therefore their functional role is clear.' This point of view is expressed explicitly in Lisman (1999), who talks of the necessity of one granule cell firing one CA3 pyramidal CA3.

There are some differences between dorsal and ventral ends. Most conspicuously, the density of granule cell packing varies from septal to temporal pole. At the septal pole, the ratio of granule cells to pyramidal cells in CA3 is 12:1, while at the temporal pole the ratio is 2:3. Perhaps consequently, the ratio between excitatory and inhibitory cells is also affected by septo-temporal position in the DG/CA3 circuits. The ratio
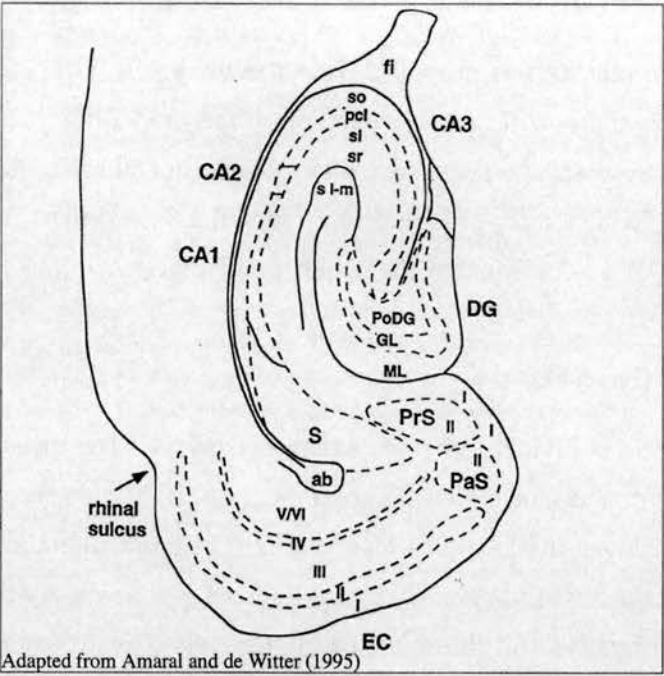
Figure 3.3: Cell layers of the hippocampal formation.

| DG | Dentate Gyrus | | CA1-3 |
|---|---|---|---|
| Po | Polymorphic layer | so | stratum oriens |
| Ml | Molecular layer | pcl | pyramidal cell layer |
| Gl | Granule layer | sl | stratum lucidium |
|  |  | sr | stratum radiatum |
|  |  | sl-m | stratum lacunosum-moleculare |
| S | subiculum | | |

of inhibitory basket cells to granule cells increases from 1:220 to 1:40 (Amaral and de Witter, 1995). Whether the change in ratio affects the amount of inhibition or the distribution is as yet unknown. It is known that the mossy fibre projection is largely organised in a lamellar fashion; the same septo-temporal layer of DG connects to the same layer of CA3. However, the feed forward inhibitory connections are predominantly to the neighbouring lamellae. There are also feedback inhibitory and possibly excitatory connections from the mossy cells. Mossy cells described in Acsady *et al.* (1998) seem to inhibit granule cells in the same septo-temporal layer, but also send a widely divergent back projection that may be either excitatory or inhibitory, thus forming disynaptic recurrence. Li *et al.* (1994) found that the back projection (about which there is still some dispute) predominantly originated from temporal CA3 cells.

### 3.4.2   Can one mossy fibre fire a pyramidal cell?

The numbers given in Henze *et al.* (2000) estimate that one granule cell contacts as many as 14 pyramidal cells, and each CA3 pyramidal cell receives input from as many as 50 cells, via giant mossy fibre boutons. Each bouton may have up 30 active zones; assuming they operate simultaneously, this potentially provides massively strong excitatory drive. However, pyramidal cells are not the only cells that receive giant boutons. The mossy cells of the hilus (see figure 3.3), also excitatory, receive more convergent excitatory input. Mossy cells have a larger spread of dendritic tree and appear to "see" more mossy fibre cells. However, they contact gabanergic interneurons within the DG, so their overall effect is probably inhibitory. But most striking of all, the greatest number of synaptic contacts made by the mossy fibres are actually with inhibitory interneurons, up to 150 per fibre. These boutons forming on inhibitory interneurons are not as large, but have relatively larger active zones.

The implication of these details is that the overall effect of a mossy fibre input to a pyramidal cell is not known. For example, it is not known whether the inhibitory cells fired by the mossy fibre inhibit the same pyramidal cell contacted by the mossy fibre, or the neighbours of that cell, or what the time course of the disynaptic inhibition is relative to the firing of the pyramidal cell.

The overall picture of sparse firing in CA3 caused by a point to point connection from dentate gyrus seems roughly correct. However, the existence of divergent recur-

Adapted from Ishizuka *et al.* (1990)
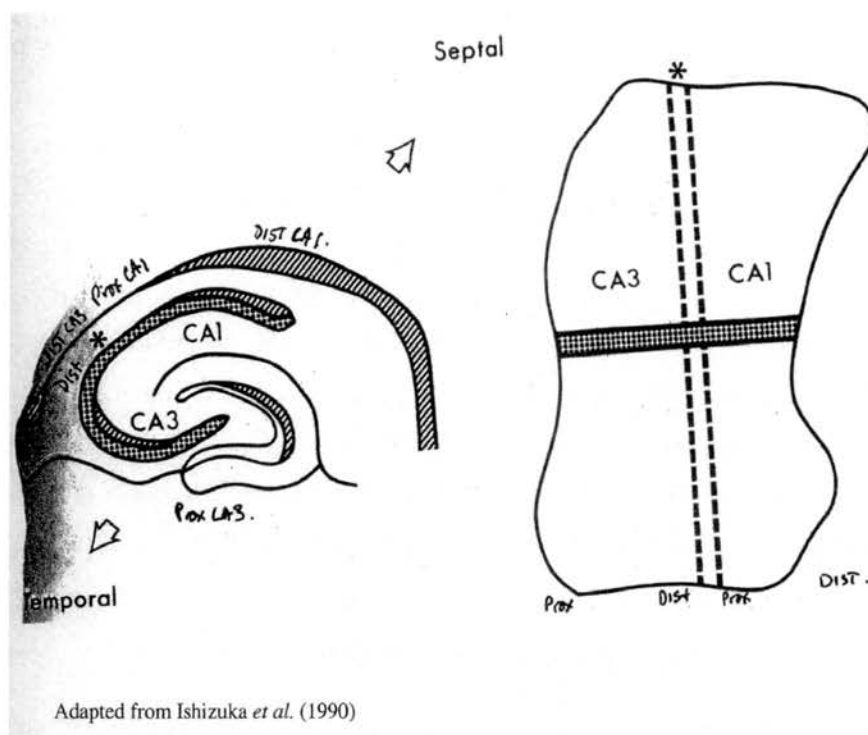
Figure 3.4: Proximal and distal CA3 and CA1 positioned on the transverse axis.

rence within dentate gyrus between septo-temporal layers makes it unlikely that the lamellar structure of the projection is sufficient to preserve the septo-temporal division of input information that Moser's anatomical hypothesis requires.

### 3.4.3 The structure of CA3

The excitatory cells of CA3 are the pyramidal cells. On the trisynaptic view, CA3 is sparsely, and randomly recurrent, sending topographic projections to CA1. The projection from CA3 to CA1 is composed of the schaffer collaterals, whereas the internal recurrent collaterals are sometimes referred to as the associational fibres. First, data from Ishizuka *et al.* (1995) and Ishizuka *et al.* (1990) suggest that CA3 is not a uniform circuit along the transverse slice. In general, the position of cells relative to the transverse axis (shown in figure 3.4) is described by their proximity to the preceding region; proximal CA3 is next to the dentate gyrus, proximal CA1 is next to CA2.

Proximal CA3 connects mostly to itself, while mid CA3 connects sparsely across

the whole transverse region. Cells in distal CA3, which has the thickest projection to CA1, have much longer axon collaterals and are connected across the transverse axis. However, distal CA3 does not connect with itself, in some ways not being a recurrent circuit at all. These cells have their dendrites in the stratum lacunosum-moelculare (SLM) rather than in stratum radiatum. Proximal cells show distinctive termination on the radial axis, confining most of their dendrites to the deep layers of the stratum radiatum and stratum oriens.

However, each circuit on the transverse axis projects along the septo-temporal axis. Radial termination and direction of CA3 connections also vary systematically with the cell position on the transverse axis. Distal cells project in a temporal direction and at temporal levels labelling is found deeper in stratum radiatum and oriens. At more septal levels connections move out towards superficial radiatum. Proximal cells tend to project septally, but show the same pattern of radial termination. The unresolved issue is whether the positioning of the recurrent input on the dendritic tree has any significance for information processing. The distant distal dendrites in SLM are further away from the soma and also co-terminate with perforant path input. The issue will be discussed further in the section 3.6.5 on perforant path input below.

A remarkable feature of almost all studies of intrahippocampal connections in the 1990s has been the reporting of patches of unexpectedly high density of arborisation and, by implication, synaptic contact of the recurrent collaterals forming the CA3 - CA3 connection. Ignoring the detail of the *radial* distribution of CA3 input, the picture of sparse random recurrence in CA3 connections needs to be replaced by the following picture. By extensive axon reconstruction and counting boutons, Li *et al.* (1994) discovered that actually the CA3 recurrent projection had a peak area of very high probability of synaptic contact at a different septo-temporal level. In a 1.5mm slice containing 20,000 pyramidal cells, the axon showed perhaps 14,000 boutons. We do not, of course, know that all the boutons, or potential active zones inside boutons, *are* active. But these numbers mean in practice that a pyramidal cell may contact almost all the cells within its area of maximum contact.

Data from Ishizuka *et al.* (1990) is shown in figure 3.5 that indicates the neighbourhood structure of these connection patches. What is not known is the probability of reciprocal connection between these patches. However, it is possible to observe that
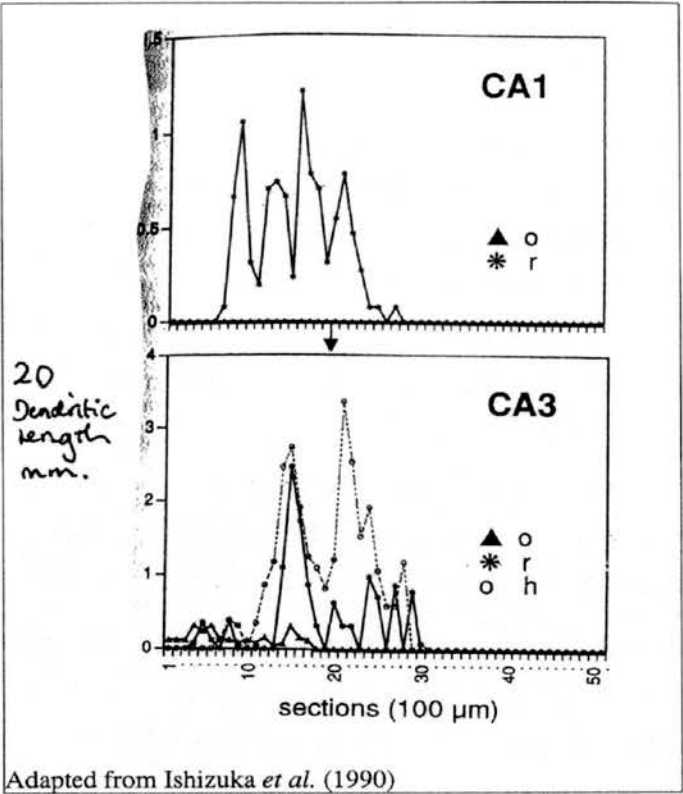
Figure 3.5: This graph shows the distribution of dendritic tree in septo-temporal levels, for the CA3 cell positioned at the black arrow. Note that peak axon distribution is not at the same septo-temporal level as the cell, but in nearby levels. H is the hilar region/proximal CA3, i.e., contains the back projection to the dentate gyrus.

for *temporally* situated neurons there is a tendency to project septally, and for *septally* situated neurons, a tendency to project towards the temporal pole. This has lead Amaral and de Witter (1995) to remark that information flow along the longitudinal axis is at least as important as output to CA1. It has also been reported (Hoz, 2000) that axon propagation in temporal pyramidal cells is slower than in septal cells.

It is clear that the ratio of excitatory and inhibitory input to the pyramidal cells may be different at the septal and temporal poles. To identify the functional consequences, a theory about the relation of the proportion of inhibition to task related activity is required. This type of theory has been provided in models such as Levy (1996) and Hasselmo *et al.* (1996). However, it is not clear that the models put forward accurately reflect the inhibition structure and, although they discuss the consequences of varying the inhibition/excitation ratio, they do not relate it to longitudinal differences in CA3.

### 3.4.4   Projections to CA1

The extensive studies of intra hippocampal projections (Ishizuka *et al.*, 1990, 1995) show that schaffer collateral projections to CA1 were affected not only by the location of the projecting cell on the temporal axis, but also by its position on the transverse axis. Ishizuka's study is very detailed, based on labelling and then tracing the axons of only a small number of cells. Figure 3.6 illustrates the pattern of the projection from CA3 to CA1, which show the same gradients as the intrinsic CA3 connections, but in the opposite direction. Again, the interesting issue here concerns the changing position of inputs on the radial axis, shown in 3.7 e.g., movement from stratum oriens in the septal levels, to stratum radiatum at temporal levels. In general, heavier labelling (hence higher density of termination) is found within stratum oriens in the septal pole of CA1. There are, however, variations due to the transverse structure of the projection.

### 3.4.5   Projections from CA1 to the subiculum

CA1 projects topographically to the subiculum along the transverse axis. But the projection diverges along the longitudinal axis. Cells at any given septo-temporal level project along a third of the extent of the septo-temporal axis.

Adapted from Ishizuka *et al.* (1990)
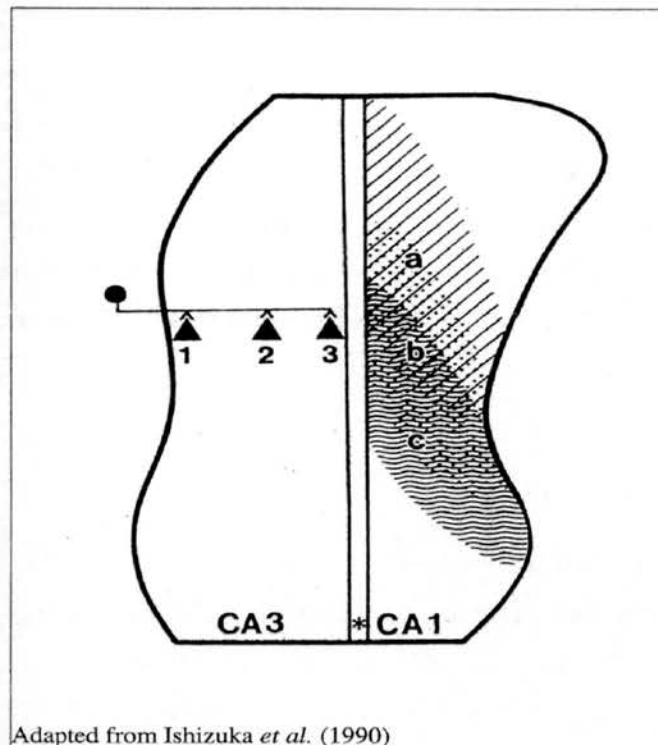
Figure 3.6: This figure is adapted from Ishizuka *et al.* (1990). Based on the varying distribution of inputs, the diagram needs to be decoded as follows:

Cell 1 in *Proximal* CA3 terminates as *Oblique lines*

Cell 2 in *Mid* CA3 terminates as *dots*

Cell 3 in *Distal* CA3 terminates as *wavy lines*

Cell 1 projects more towards septal levels. Cell 3 projects to more temporal levels, where a is positioned towards the septal pole.
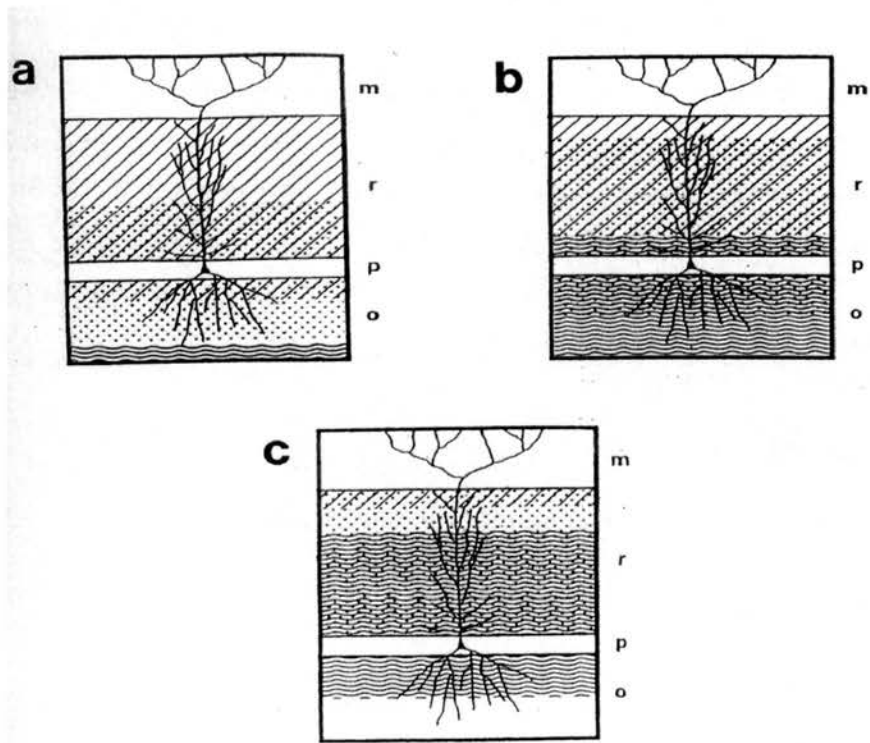
Adapted from Ishizuka *et al.* (1990)

Figure 3.7: Picture (a) shows the radial distribution at a more septal level of cells 1,2,3. Notice how the radial inputs more or less preserve the transverse topography, but move deeper in to the stratum oriens in cell 3 at the septal pole. Picture (c) shows the radial distribution at a more temporal level.

### 3.4.6    The subiculum

The subiculum is still regarded as the major output region, projecting back to entorhinal, perirhinal, retrosplinal and medial prefrontal cortices. It is engaged in the extrahippocampal circuits described below as the prefrontal and limbic circuits. In general the projections from the subiculum preserve topographic organisation along the longitudinal axis (Amaral and de Witter, 1995).

The back projection to the entorhinal cortex is topographically organised along the longitudinal axis. Temporal subiculum projects to more medial entorhinal cortex, while septal subiculum projects laterally. The forward projection in the perforant path (described below) shows the same topography. The projection to the amygdala originates from temporal levels only, while the projection to perirhinal and retrosplinal originates only from the septal two-thirds.

### 3.4.7    Summary

Three candidate elements of the intrinsic circuit vary along the septo-temporal axis:

- The ratio of granule cells to pyramidal cells

- The number and possible spread of basket cells

- The radial distribution of recurrent and projecting connections

The extent of CA3 recurrent and forward connectivity along the longitudinal axis means that the old lamellar picture needs revision. However, if CA3 could be switched out of the circuit, the adjacent circuits via the amygdala and the perirhinal to entorhinal cortex to perforant path could be kept independent.

## 3.5    Hippocampus and extrinsic circuits

The hippocampus is involved in at least three extrinsic circuits that show different projection to the septo-temporal axis, namely the visual, limbic and prefrontal areas. The "limbic" system was originally thought to be concerned with olfaction. It is a horseshoe shaped rim of cortex surrounding the junction between diencephalon and cerebral

hemispheres, joined by the olfactory system at the base of the brain. In humans, it is usually considered to comprise cingulate, the parahippocampal gyrus, septum, amygdala, hypothalamus and thalamus, olfactory system and the hippocampus. The system is thought to govern both autonomic behaviours such as thermoregulation, hunger and thirst signals and also the relationships between automatic responses, emotional responses, memories and voluntary actions.

Moser's hypothesis primarily speaks of the preferential interaction between amygdala and temporal hippocampus and the apparent streaming of the input via the perforant path from the entorhinal cortex. A weaker form of the content streaming hypothesis is proposed by McNaughton and Gray (2000), relating to the so called dorsal and ventral streams of "what" and "where" information in the visual system to dorsal and ventral in the hippocampus. This is a huge hypothesis considered by Mishkin, about the organisation of the entire visual system, and falls outside the scope of this chapter.

## 3.6 The limbic system

### 3.6.1 The septum

Jakab (1995) in his review of the anatomy of the septum, describes the structure as being involved in "a variety of physiological behavioural processes .... e.g., learning and memory, emotions, fear, aggression and stress as well as autonomic regulation" thereby encompassing almost the entire spectrum of behaviour relevant to the hippocampus. The septum projects to the hippocampus in a topographical fashion, medial septum to septal hippocampus, and lateral septum to temporal. These projections contain cholinergic fibres (excitatory) and gabanergic (inhibitory) input, concerned with the generation of the theta rhythm, as discussed in the chapter 5, section 3 on the theta rhythm. The septal part of the projection, originating from the diagonal band, is significantly more excitatory, composed of 50-70% cholinergic fibres. The temporal half is more inhibitory, being 30-50% cholinergic (Amaral and de Witter, 1995; Freund and Buzsaki, 1996).

The effects of lesioning the septum are very similar to those of lesioning the hippocampus. Lesioning the septum shows impairments on water maze radial arm maze,

and conditioned freezing. Table 1 in Appendix C shows a table reproduced from Mc-Naughton and Gray (2000) that gives an overview of tasks and lesioned areas. The close resemblance of septal and hippocampal lesions, however, suggests that processing in the two regions is interdependent.

### 3.6.2   The amygdala

The amygdala is part of two circuits that project to the hippocampus. The direct projection is reciprocally connected to the temporal third of CA1 hippocampus only. There is also a second circuit via entorhinal cortex and the perforant path. But this also only projects to the temporal part of the hippocampus.

It has been suggested that dopamine may selectively gate the activation of amygdala and the hippocampus and modulate their interactions with the prefrontal cortex (Thierry *et al.*, 2000). Conditioning and startle and avoidance responses are depressed after amygdala lesions but unchanged after septo-hippocampal lesions. This suggests that complete hippocampal lesions do not disrupt learning in the amygdala. Amygdala lesions do not disrupt water maze performance, but do affect alteration tasks.

### 3.6.3   The hypothalamus and thalamus

Finally the hypothalamus also projects to the hippocampus, in a mutually inhibitory circuit.

The thalamus provides excitatory drive to CA1 that appears to be independent and complementary to the intra-hippocampal excitatory pathways. Projection to CA1 from the thalamus is found in the SLM of CA1 pyramidal cells (mid to ventral regions) and Bertram and Zhang (1999) found the fastest response latency to the excitatory thalamic input in that area. Previously, it was thought that the thalamus provided only a neuromodulatory affect to the hippocampus. Thalamic input appears to target the GabaB interneurons along side pyramidal distal dendrites.

### 3.6.4   Prefrontal cortex

Prefrontal cortex lesions are associated with deficits in goal directed planning and working memory. The prefrontal cortex is innervated by temporal CA1, which is also

the area that receives a dopamine projection from the VTA, via the pathway referred to as mesocortical-dopamine pathway. Both systems project to layers V-V1 in close proximity (Gurden *et al.*, 1999). Work by Thierry *et al.* (2000) has suggested that hippocampal projections to prefrontal cortex are in apposition to the dopaminergic projection. The possibility exists then that the dopamine system selectively gates the interaction of prefrontal and hippocampus. Stimulation of the VTA, thought to release dopamine in the pre-frontal cortex, appears to enhance long term potentiation of the hippocampal projection synapses at some stimulation frequencies, while at other frequencies reducing the post synaptic response. Dopamine acting via the D1/D5 receptors is thought to enhance the NMDA receptor component of the response in prefrontal cortex, the apparent opposite from its effect in CA1 (Seaman and Durstwitz, 2001).

### 3.6.5 Entorhinal cortex and the perforant path

The majority of cortical input travels through the perforant path from the entorhinal cortex. As we have already said, the connections are roughly topographical from entorhinal through to DG and then CA3. The cortical projection from peri and postrhinal (or parahippocampal) areas is separated into projections from layer two and layer three of the entorhinal cortex.

In a recent paper, Witter (2000) identified two different information processing pathways through the hippocampus rostromedial areas, which do not divide neatly, cutting across what are referred to as the lateral and medial parts of the entorhinal cortex. The layer II lateral entorhinal cortex (LEC) projects to the dentate gyrus and the CA3, terminating in the outer layer, which is the stratum lacunosum moleculare (SLM). This is shown in figure 3.9, while the projection from the medial entorhinal cortex (MEC) terminates on the inner layer on the same cells in the stratum radiatum. This means the inputs stratify on the radial axis of CA3, according to their origin. Proximal CA3 cells do not have dendrites in the SLM, leading to speculation in Ishizuka *et al.* (1990) that they do not receive input from the LEC.

In CA1 and the subiculum, layer II from the LEC and MEC stratifies along the transverse access. This is shown in figure 3.8.

The temporal and septal parts of the hippocampus receive different projections. Figure 3.8 illustrates the projection from the caudolateral zones of LEC and MEC

Adapted from Amaral and de Witter (1995)
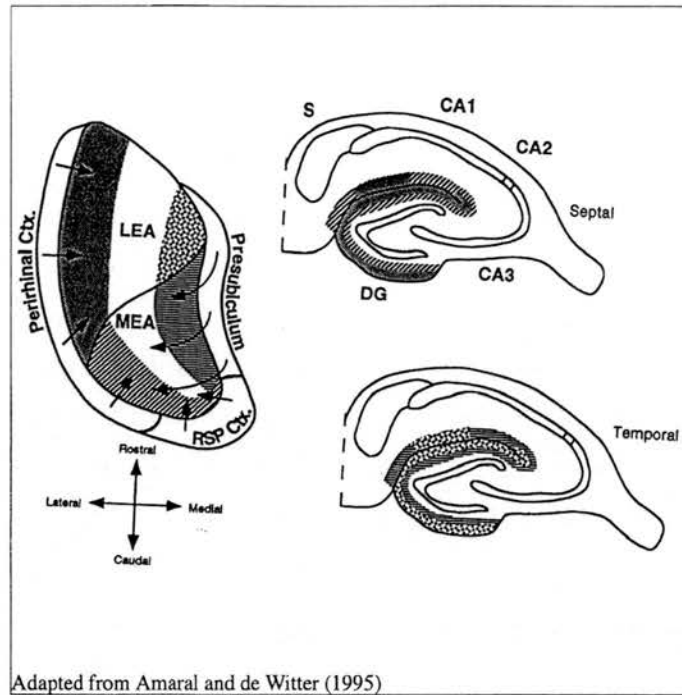
Figure 3.8: This figure shows the stratification of perforant path across the radial axis, and the origin of the projections at the septal and temporal poles.

projecting to the septal pole. This half of the entorhinal cortex receives input from the perirhinal and postrhinal cortex. This is the division in the pathway that has received particular attention from Moser and Moser (2000, 1998a), as the postrhinal cortex has been linked with recognition of spatial arrangements and is a major contributor of spatial information, as well as novelty and familiarity signals for individual items (see chapter 1, section 1.12.1). However, it is worth mentioning that spatial tasks such as the water maze do not depend on the postrhinal cortex. The rostromedial zone receives input from presubiculum and retrosplinal cortices and projects to the temporal pole of the hippocampus.

More interesting from the point of view of this thesis is the distribution of inputs on the radial axis.

The significance of this is that if we are thinking about the information processing path of the information coming through the perirhinal and postrhinal cortices, we can
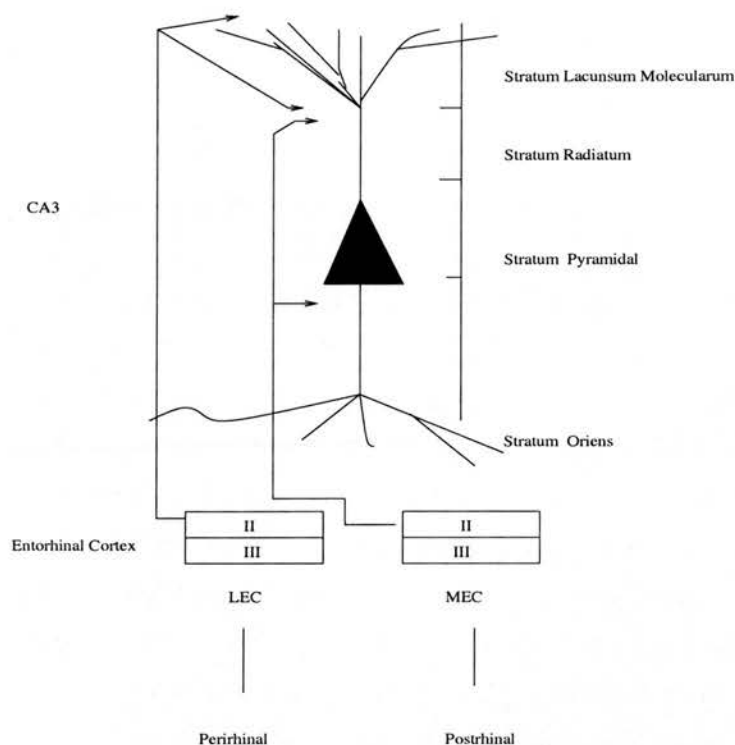
Figure 3.9: Pyramidal cell illustrating radial positioning of temporammonic inputs to the outer layers

see that it is being integrated in a different fashion in different areas. In DG and area CA3 it is stratified to different positions within the *same* pyramidal cell, whereas in the subiculum and CA1 it is distributed to *different* cells.

We can ask the question: what difference does stratification of inputs along the dendritic tree make? Can we think of a way of exploiting the stratification of inputs? These questions will be discussed in the following section, which looks at how synaptic plasticity and neuromodulation differ in their effects on different strata of the dendritic tree.

### 3.6.6 Summary

The septum, amygdala and prefrontal cortex are all associated with neuromodulatory inputs to the hippocampus, the exact functions of which are unknown. Lesioning the septum appears to disrupt all, or most, of the functions associated with the hippocam-

pus, while lesioning amygdala, thalamus, hypothalamus or prefrontal areas produces dissociated effects.

## 3.7    Hippocampal synaptic potentiation and learning

Models of synaptic plasticity at the network level are usually based on some form of hebbian learning. In chapters 6 and 7, a number of simple rules for storing association in a hebbian type way will be discussed. These rules usually exploit some simple way of combining inputs, and a threshold that the combined sum must exceed to enable the neuron to fire. In principle, this is what neurons do. The situation in a real neuron, however, is that a great deal of variation in exactly how inputs are put together is possible within the cell. The factors that control this variability (e.g the relative timing of inputs or dopamine influx from the prefrontal cortex), give us information about *which* signals (signals occurring at the same time) are important to a cell, and *when* (e.g., the rat receives a reward) those signals are important.

As we said in chapter 2, long term potentiation is thought to be the biological correlation of learning. The conditions under which synaptic plasticity occurs in the hippocampus vary from pathway to pathway.

The protocol for inducing associative LTP is either tetanic, high frequency bursts or paired pulse: depolarisation of target neuron, paired with afferent stimulation. The resultant LTP can be dissociated into an early phase that lasts less than three hours, induced by weak tetanus (one burst of 100Hz lasting 1s) and a late phase, dependent on protein synthesis, that lasts for much longer, induced by a strong or theta phase tetanus (three 100Hz bursts) at 8Hz. Chapter 5 describes the theta rhythm and part of the reason why that form of stimulation is optimal (Larson *et al.*, 1986). LTD can be induced with low frequency stimulation, or a paired pulse protocol where the post synaptic neuron is depolarised prior to afferent stimulation - as one might expect from Bi and Poo (1998).

The best documented, and most understood, form of LTP (both early and late) is NMDA receptor mediated, as we have already outlined. The conditions that control the direction of change at the NMDA synapse are actually the order of occurrence of synaptic transmission, and postsynaptic depolarisation. This temporal asymmetry was

*Figure 7.* Critical window for the induction of synaptic potentiation and depression. The percentage change in the EPSC amplitude at 20–30 min after the repetitive correlated spiking (60 pulses at 1 Hz) was plotted against the spike timing. Spike timing was defined by the time interval (Δt) between the onset of the EPSP and the peak of the postsynaptic action potential during each cycle of repetitive stimulation, as illustrated by the traces above. For this analysis, we included only synapses with initial EPSC amplitude of <500 pA, and all EPSPs were subthreshold for data associated with negatively correlated spiking. Calibration: 50 mV, 10 msec.

Adapted from Bi and Poo (1998)

Figure 3.10: Bi and Poo (1998)'s synaptic modifications in CA3. The change in EPSC amplitude is shown to depend on the order of the pre and post synaptic spike. Positively correlated spiking means the post synaptic spike occured up to $50ms^{-1}$ after the presynaptic spike.

first discovered by Levy and Steward (1983a); Levy *et al.* (1983); Levy and Steward (1983b); Wilson *et al.* (1981) in dentate gyrus synapses and found to be a general property of NMDA LTP (Wickliffe, 1991). Levy's results show *LTP* occurring if the presynaptic neuron fires up to 50ms *before* the postsynaptic neuron and *LTD* occurring if the postsynaptic neuron fires in the 10ms-1 before the presynaptic neuron fires. A more recent description (see figure 3.10) of this phenomenon is shown in the diagram taken from Bi and Poo (1998). This data was, crucially, taken from CA3-CA3 synapses in culture. The dependence of the amplitude of potentiation on the initial size of the synapse is shown in figure 3.12. Bi and Poo (1998) also describe NMDA dependent potentiation in supra-threshold synapses that showed no corresponding LTD. The trigger is thought to be depolarisation in the postsynaptic cell, causing the voltage dependent

*Figure 4.* Effect of repetitive stimulation with negatively correlated postsynaptic spiking on suprathreshold connections. *A*, Results from an experiment similar to that described in Figure 3*A*, except that the synaptic activation was capable of initiating spiking of the postsynaptic neuron. The spike initiated by current pulse injection peaked at ~10 msec before the onset of each EPSP during repetitive stimulation. Calibration: 100 pA, 10 msec for EPSCs; 40 mV, 10 msec for the EPSP. *B*, Summary of all experiments similar to that described in *A*. Data points represent mean ± SEM (*n* = 3). The mean percentage change in synaptic strength after induction was 31.9 ± 9.3% (±SEM). Significant potentiation was observed (*p* < 0.05, *t* test).

Adapted from Bi and Poo (1998)

Figure 3.11: The behaviour of supra threshold synapses described in the same culture in response to positively and negatively correlated spiking
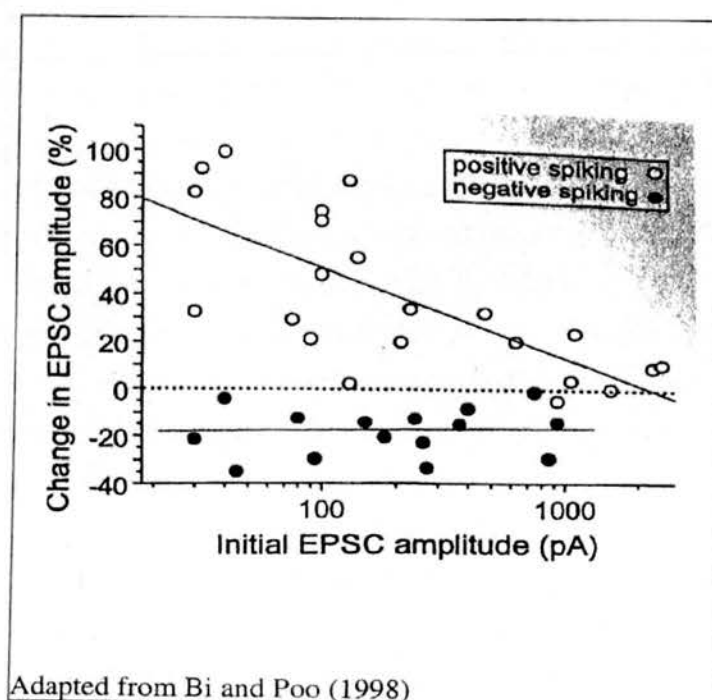
Figure 3.12: The amplitude of potentiation depends on the initial size of the synaptic current, indicating saturation occurs.

Mg++ block of the NMDA receptor to disappear. This allows calcium entry to the post-synaptic cell, initiating the biochemical changes underlying LTP. These changes may result in the phosphorylation of AMPA receptors, or even the recruitment of stored or silent AMPA receptors to the cell membrane, thus increasing synaptic strength. Conversely, NMDA mediated LTD is thought to be mediated by dephosphorylation, or the removal of AMPA receptors from the membrane.

Depolarisation of the postsynaptic cell is caused by an action potential propagating backwards in the dendritic tree from a spike at the soma. Therefore, control of spiking at the soma by inhibition or neuromodulators and control of the active channels in axon and dendrites that impede or facilitate backpropagation are crucial factors in synaptic plasticity, sometimes called metaplasticity.

NMDA dependent receptor LTP has been found at schaffer collateral synapses in CA1 and CA3, and also in the dentate gyrus, and the perforant path synapse.

In CA1 and CA3 pyramidals, postsynaptic firing is required to induce LTP in the schaffer collateral synapses, in particular bursting in CA1 (Pike *et al.*, 1999). The same protocols applied to perforant path synapses do not induce LTP so readily. Low frequency input produced NMDA dependent LTD very readily, but the usual LTP protocols, although producing a recovery from LTD, did not potentiate above baseline, except in the presence of bicuculline (i.e., Gaba suppression).

NMDA dependent LTP has also been found to propagate either to neighbouring synapses, or synapses on the same presynaptic neuron, in a very un-hebbian type way (Maccaferri and McBain, 1996; Engert and Bonhoffer, 1997). This may mean that the colocation of synapses in the dendritic tree may prove to link them. However, this has not been adequately demonstrated in vivo, and it is also not clear whether this can act to link inhibitory and excitatory inputs.

### 3.7.1    NMDA independent LTP

NMDA mediated LTP is not the only form of synaptic potentiation found in the hippocampus. Non-NMDA potentiation has been found in mossy fibre synapses, probably caused by a presynaptic rise in intra-cellular calcium, triggered by tetanus, but non associative, in that it does not require postsynaptic activity.

Alongside NMDA dependent LTP, the basal dendrites and stratum radiatum of CA1

and CA3 also show a form of plasticity that is not NMDA dependent, but it usually occluded by NMDA responses (Cavus and Teyler, 1998).

The next section describes some of the major neuromodulators that are known to have meta-plastic effects within the hippocampal system.

## 3.8 Neuromodulation

There is a trend for the ventral/temporal pole to be favoured in the distribution of neuromodulator (Martens *et al.*, 1998) receptors. Of particular importance are muscarinic acetylcholine receptors, and dopamine receptors. However, as we have seen, the temporal pole lacks NMDA receptors.

### 3.8.1 Dopamine

The mesocortical dopamine pathway from the VTA projects to stratum oriens of CA1, at temporal layers, and also to the subiculum (Verney *et al.*, 1985).

Dopamine has been shown to produce a potentiating effect on synaptic transmission similar to early LTP by Yang (2000) and Otmakhova and Lisman (1996). Otmakhova and Lisman (1998) claim that D1/D5 agonists enhance early long term potentiation and reduces depotentiation. However, Kumar and Faber (1999) claim that dopamine evoked potentiation (DEP) can be occluded by early LTP, if it is applied afterwards. They conclude that it is a separate phenomenon that partially overlaps mechanisms in LTP. They also showed that the AMPA component of the current which is modulated by dopamine is voltage dependent, and maximally activates at a more hyperpolarized voltage than the NMDA component. Their experiments were not conducted within the CA1/hippocampus system, but in goldfish mauthner neurons of the reticularsplinal cortex. These neurons also show persistent LTP mediated both by NMDA, and non NMDA mechanisms. However, the hypothesis that this is also true for CA1 DEP/LTP interactions is also worth considering, due to the following evidence.

D1 receptors have been found to have an inhibitory effect on the CA1 perforant path input (to SLM), while leaving the schaffer collateral input unaffected (Otmakhova and Lisman, 1999a). In particular, it seems to be the NMDA component of synaptic transmission that is affected. These authors showed that CA1 perforant path synapses are

composed of NMDA and AMPA receptors, but proportionally more NMDA receptors. The excitatory postsynaptic potential (EPSP) was decomposable into two components. Application of dopamine reduced the EPSP at the schaffer collateral slightly, while suppressing the perforant path response almost completely. Paired pulse facilitation at the schaffer collateral synapses was left intact. This suppression effect could be ameliorated by D1/D2 receptor blockade. However, the enhancement of LTP found by Yang (2000) in CA1 was found for currents mediated by both AMPA and NMDA, but the NMDA activation was not needed to trigger this potentiation, whereas a rise in intra-cellular Ca2+ was required.

It has also been reported that dopamine (D1, D5) receptor agonists can enhance late phase LTP, while D1 receptor antagonists block late LTP.

There is evidence that the effects of dopamine are region specific; Swanson-Park *et al.* (1999) also looked at the D1/ D5 receptor blockade in both CA1 and DG. Consolidation of LTP in DG was not affected by the receptor blockade, while the persistence of LTP in CA1 was reduced (presumably through interference with late LTP). Note that exactly the reverse of this effect was found with blockade of Beta-adranergic receptors.

Overall, it is clear that dopamine can act as a learning trigger, (particularly, perhaps in the temporal hippocampus which lacks NMDA receptors) and also as a switch between inputs (schaffer collateral and perforant path), particularly in the septal hippocampus.

### 3.8.2    Acetylcholine

Acetylcholine is particularly associated with theta rhythm, as cholinergic input comes via the projection from the septum and is responsible for pacing hippocampal theta. Acetylcholine can have both an inhibitory affect and an excitatory one. Acting via muscarinic receptors, acetylcholine suppresses the postsynaptic response to schaffer collateral intrinsic input, while enhancing the response to external input. It also has been proposed as a switch between a recall and a learning mode in the hippocampus. Models of this will be considered in the next chapter.

### 3.8.3   Noradrenaline

Noradrenaline is primarily linked to mossy fibre input; the highest concentration is found in the dentate gyrus, and then areas of CA3 where mossy fibres projects to stratum lucidium, SLM, and the hilus. The noradrenaline system therefore hypothetically acts on CA3 cells that have dendrites in the appropriate regions; i.e., distal/septal (see below). Noradrenaline and seratonin, however, act primarily by extra cellular diffusion (Vizi and Kiss, 1998), and therefore it is hard to conclude anything about the distribution of action from the distribution of release sites. Only 15 (noradrenaline) to 20 (seratonin)% of release sites were synaptic. Hoz claims that there is greater distribution towards the temporal end. Noradrenaline is taken up by the beta-adrenergic receptors, and blockade of these has been found to affect the consolidation - or long term persistence - of LTP (Swanson-Park *et al.*, 1999), particularly in the dentate gyrus.

Noradrenaline also may suppress perforant path input to CA1 in the same manner as dopamine (Otmakhova and Lisman, 1999b)

### 3.8.4   Common mechanisms and effects

Activation of beta-adrenergic, muscarinic and dopamine receptors (depending on density of D1/D5 receptors) enhances the activation of protein kinase C and A. This had the effect of increasing the amplitude of back-propagating action potentials in the distal dendrites in CA1. Generally back-propagating action potentials attenuate due to transient K+ channels. However, they provide a signal that increases local calcium concentration, the key ingredient in almost all forms of NMDA and non-NMDA mediated synaptic plasticity.

This has implications for the integration of perforant path and schaffer collateral input in CA1, as these neuromodulators also enhance EPSPs travelling from the apical dendrites.

Dvorak-Carbonne and Schuman (1999b) and Dvorak-Carbonne and Schuman (1999a) have found that the entorhinal cortex inputs (Layer III) to the SLM are able to selectively suppress spiking in response to the schaffer collateral inputs, via a gabanergic interneuron. It is known that the basal dendrites and stratum radiatum where the schaffer collaterals terminate in CA1 provide a stronger signal at the soma (reviewed and

studied in Yeckel and Berger, 1998). Boosting the back propagating signal into the apical dendrites potentially provides a way of enhancing the signal from the perforant path, by enhancing LTP at these remote synapses. This appears to contradict the finding that dopamine suppresses perforant path input. Lisman and Otmakhova (2001) take the view that perforant path input is suppressed during learning to prevent current sensory (via the perforant path) input from interfering with information buffered in CA3 and CA1. Their argument partly depends on the assumption that dopamine only enhances learning at schaffer collateral synapses through acting on NMDA receptors. They therefore see dopamine as a learning from input/recall switch. As we saw, dopamine can in the right conditions trigger learning independently. Dopamine may therefore act as a switch between learning systems.

It is proposed here that the schaffer collateral are signalling retrieved memories from CA3 and dentate gyrus. The stratified inputs are therefore also balancing two types of memory signals. For input from the entorhinal cortex is not just sensory input as such, but more like information about individual item and complex item recognition.

However, we propose here that CA1 can selectively learn from and output information to the schaffer collateral pathway or the perforant pathway, or combine the information from both.

The lack of dendrites in the SLM in both proximal and temporal parts of CA1 suggests that these regions may not have the ability to selectively regulate perforant and schaffer collateral inputs, and are therefore combining recalled information and novel information in an incremental way.

### 3.8.5   Perforant path input

How does information coming from the distant distal dendrites in the SLM and those in the stratum radiatum summate? This question has been addressed in the CA3 region, where Urban and Barrionuevo (1998) found that inputs summed sub-linearly if the mossy fibres were situated before the perforant path, and linearly otherwise. This effect was thought to be due to transient K+ channels (Hoffman et al., 1997). Similar results have been found in CA1, where schaffer collateral and perforant path inputs also sum sublinearly if the schaffer collateral input arrives between 2ms before the perforant path input, and up to 30ms after it. This study by Hoffman and Johnston (1999b) found

that the sublinear effect was due to GabaA and GabaB activation, probably mediated by feedforward and feedback inhibitory signals from both perforant and pyramidal afferents. In the absence of GabaA and GabaB, summation was supra-linear in the period of coincident stimulation. Although the normal situation is that the pyramidal cell sums linearly, or sublinearly, in situations of disinhibition, (such as perhaps gamma oscillations, see the next chapter) the cells may be able to take advantage of supra linear summation

On the other hand, stimulation at the SLM perforant path afferents results in a spike blocking effect of the response to schaffer collateral input (Dvorak-Carbonne and Schuman, 1999b). The same effect has been observed in CA3. This effect is mediated by GabaB interneurons, targeted by perforant path and thalamic input.

### 3.8.6 Summary

It is known that back propagating action potentials, important signals for NMDA receptors, attenuate as they move away from the soma into the distal dendrites. In CA1, at least, that attenuation can be removed by down regulation of the K+ channels, through activation of the enzymes protein kinase C and A (PKC, PKA), regulated by the neuro-modulator systems (dopamine, acetylcholine, noradrenaline). We also know that these very same neuromodulator systems can suppress transmission through the perforant path synapses.

The overall picture here seems to be that inputs are stratified along the radial axis, so that they can selectively suppress one another.

As proximal CA3 cells do not have dendrites in the SLM, it is possible that they do not receive any perforant path input. This suggests there is a functional difference in the way these cells (and perhaps cells in the temporal pole) regulate the balance between current sensory input, and recognition signals coming through the perforant path and internal recall within the hippocampus.

It is still, however, unclear what the functional difference between input positioned in SLM, and that in stratum radiatum or stratum oriens, actually amounts to in the case of a particular cell. This issue needs to be explored by accurate modelling work.

It is clear that in order to understand how information is integrated in principal cells from different sources, we cannot neglect both the radial structure and the interaction

with neuromodulators in these cells.

## 3.9     What the anatomy tells us

### 3.9.1    Consequence for Moser's hypothesis

We can already see that anatomical grounds for proposing two pathways through septal
and temporal hippocampus carrying different information sets are weak. The extensive
interconnectivity within CA3 and the spread through CA1 suggest the hippocampus
integrates more or less all the information it receives. There is perhaps a special circuit
relating amygdala, temporal hippocampus and prefrontal cortex. But that circuit is
more likely to be concerned with controlling the output of the hippocampus. This
negative conclusion helps to support the idea that the difference in learning between
septal and temporal hippocampus described in chapter 2 is due to differences in the
substrate of learning.

### 3.9.2    Mechanisms for interleaving learning and recall

In this chapter we have developed the idea that an individual pyramidal cell may se-
lectively integrate its inputs, via neuromodulators that affect individual pathways. The
same mechanisms have been proposed to separate learning and cell firing, by prevent-
ing signals passing along dendrites and axon. There is no doubt that these mechanisms
are plausible.

### 3.9.3    Supra-linear and sublinear summation

We have seen that the same cell can modulate whether its inputs sum sublinearly, or
supra-linearly. The mechanism postulated for this is a mixture of feedforward and feed
back signals. In chapter 6, we will look at a very simple implementation of this idea.

### 3.9.4    NMDA LTP

The data we showed from Bi and Poo's paper is frequently cited. But one of the feature
of it is the presence of supra threshold synapses that do not depress. In chapter 7, we

will make the case that such synapses, modelled as all-or-nothing binary synapses, are the most suitable for the single trial, immediate recall that we propose is found in septal hippocampus.

## 3.10 Conclusion

In this chapter we argue that on the basis of the known neuroantomy and connections to and within the hippocampal region, it is very unlikely that spatial content is confined to the dorsal hippocampus.

# Chapter 4

# Spatial learning

## 4.1 Summary

In chapter four we consider the 'Spatial Hypothesis', that the hippocampus is preeminently concerned with learning spatial layouts, and relationships. The importance of this chapter for the thesis is that overall, we are arguing against the idea that the hippocampus deals particularly with information having a certain kind of content. We therefore argue strongly against the spatial hypothesis- and critically review the evidence supporting it.

## 4.2 Cognitive maps

Discovery of precise place correlates (O'Keefe and Dostrovsky, 1971) of cells firing in the hippocampal subfields CA1 and CA3 led to the idea that the hippocampus builds a map-like representation of the rat's environment. Prior to the discovery and description of place cells, it was already known that the hippocampus played an important role in the rat's ability to navigate in a maze environment as reviewed in Best and White (1999). The cognitive map theory, propounded at length in 1978 by O'Keefe and Nadel, suggested that the place cell representation might potentially form a map of the rat's environment. They also made the case that a map-like representation was the most desirable mental representation for navigation. A map-like representation can be understood as a perspective free, task free and perhaps even Euclidean representation

of space. Task free and perspective free, means the map can be used in any task, and
from every view point. If the rat does not know in advance what problems the environ-
ment is likely to throw forward then, arguably, its representation of the environment
should not contain many preconceptions either. This compares nicely with the idea of
context free and context restricted representations, discussed in chapter 1. The more
precisely a representation is configured to fit a particular environment, the fewer en-
vironments it will be useful for. The idea is that spatial layout is constant across the
tasks in the same environment, and space has properties that are constants. Muller
*et al.* (1996) discusses the property of connectedness. Every two points in space are
connected by a third and are properties transitive: if A is accessible from B, and B
is accessible from C, then A is accessible from C. This is prior knowledge about the
structure of space that potentially might be innate, rather than learned.

O'Keefe and Nadel did not necessarily assume that the map-like representation
was purely spatial, but the spatial component was their particular contribution to un-
derstanding the hippocampus.

Current hippocampal theory still draws on the idea that place cells form some sort
of a map. The idea of explicit spatial maps are found in a strong and weak version
in the modelling literature. The strong version is the preconfigured attractor map that
assumes maps are preconstructed in the hippocampus. A weaker version assumes that
some spatial elements that control place cell firing are preconfigured. The weak version
claims less about the overall structure of the representation, but seeks to model the
content of the place cell representation.

## 4.3    What is an attractor map?

Neural networks can sometimes be analysed as dynamic systems. The most famous
case is that of the Hopfield (1982) network, for which an energy function can be de-
fined. The network updates the firing of each neuron between 1 and -1, untill eventually
settling into a low energy state, known as an 'attractor', as the network does not move
out of this state without a significant perturbation. Learning adjusts the attractors by
altering $w_{ij}$ between units i and j, to coincide with the pattern to be learned. When
the network is presented with a noisy version of the learned pattern, recall can occur

as the network settles back into the attractor. The Hopfield model is a discrete system, whereas most of the models of the hippocampus are continuous.

## 4.4 Map-like characteristics of the hippocampal representation

This section describes the characteristics of CA3 and CA1 representation that have been heavily emphasised by experimentalists examining cell firing within the frame work of a spatial task. This means that the experiment has been designed so that cell firing patterns can be related to location and cue, and location information is also useful in solving the task - the rat, for example, has to remember where something is.

The most basic task is a simple exploratory, or pellet finding task. Here, the rat is simply placed in the area and encouraged to explore it. Early studies looking at the control of place cell fields by cues such as entry point, darkness, coloured cards used a simple cylindrical arena, with a curtain placed around the edge to limit the cues available to the rat. Experimenters recorded the current position of the rat, and the firing rats of selected cells. Figure 4.1 shows the typical firing field of a place cell relative to the rat's location, and a controlling cue card. Peak firing is shown as the darkest areas.

When the cue card is moved relative to the rat's entry point, the place fields move with the cue (Muller and Kubie, 1987). However, when the rat is allowed to enter from a different point, the fields remain stable. If the cue is moved in front of the rat, however, the field remains stable, indicating that is (perhaps) tied more strongly to location than to cue (Rotenberg and Muller, 1997). How to describe this finding depends on interpretation as Sharp (1999b) regards this as the place field developing a new relationship to the cue, and therefore a kind of remapping. The place fields are tied to other cues besides visual. If the rat is allowed to explore the environment, then placed back in darkness, the same place fields reappear, indicating that they depend on olfactory and path integration input (Quirk et al., 1990). When the arena is expanded, or contracted, place fields change in size, remaining in the same relative positions. Place fields were also found to be remembered; a rat taken out and replaced in the arena for short periods over six months retains the same pattern of place field firing
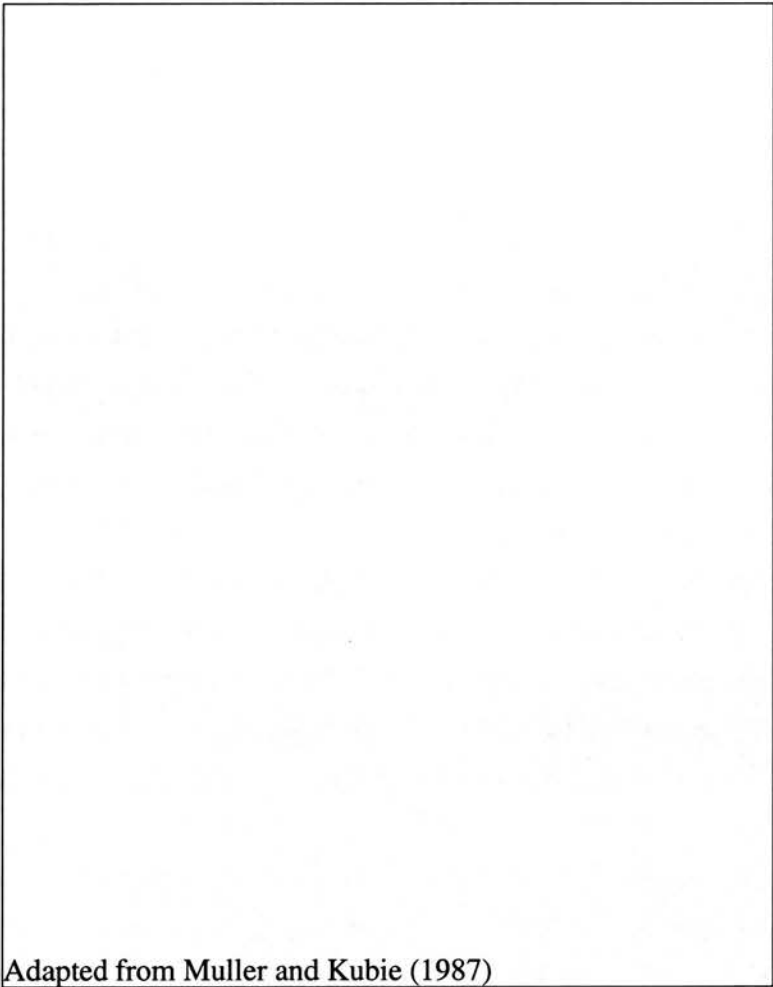
Adapted from Muller and Kubie (1987)

Figure 4.1: Place cell firing.

(Thompson and Best, 1990). Place cells are also affected by direction, but only where the environment is constrained, so that the rat does not cross the place field from every direction. For example, the rat is forced to run down a narrow track.

In summary, place cell firing appears to be controlled by distal (far) cues, environmental boundaries, multi modal sensory input, and the recent history of the rat (path integration). None of this is inconsistent with the idea of a task independent map.

These studies are essentially descriptive, monitoring place cell activity rather than explicitly testing the precision of the representation. The next step required a demonstration that place cell representation is used in more demanding tasks.

Place cells have also been recorded in memory dependent tasks. One such task is the water maze reference memory task. The Morris water maze is a circular enclosure filled with milky water. It contains a submerged platform that the rat cannot see. Typically, a rat is allowed to find the platform, or is placed on it, and then is repeatedly released in to the water to navigate to the platform. This task depends on the hippocampus functioning (Morris, 1981), as hippocampal lesions result in the rats searching the wrong part of the pool. Performance on the task is usually measured by the amount of time the rat spends in each quadrant of the pool. A second measure sometimes used is the number of annular crossings (crossing through the middle of the pool). Escape latency is another possible measure.

However, rats with hippocampal lesions become relatively hyperactive and swim faster than intact rats. The existence of place cells during the water maze task has been verified. Hollup *et al.* (2001) adapted the water maze and compared it to a dry task using the same distal cue set. Interestingly, they saw some of the place fields transferred from the wet to dry version.

However, when the task was transferred to navigation to goals or, in particular, variable food sources, indications found in Markus *et al.* (1995) and Gothard *et al.* (1996) began to suggest that the place cell representation might not be entirely task independent. Both papers reported the clustering of place cells around goal areas and variability in place cell firing due to a change in the demands of the task. Gothard *et al.* (1996) attempted to explain this phenomenon using the idea of multiple maps controlled by different reference points. This theory has been used in the development of a number of models of the hippocampal representation. The next section will discuss

these models briefly, before considering evidence that is incompatible with the idea of multiple maps.

## 4.5 Strong spatial maps - preconfigur ed attractor networks

The idea that the place cells were a coherent ensemble reached a peak in Muller *et al.* (1996). Muller proposed that synaptic delays or the 'resistance' of a pathway between two place cells should be set to represent, and be proportional to, the distance between those two points. The cognitive graph can then find the shortest distance between any two points, using the analogy of current flow. The graph has the ability to represent both barriers by breaks of connectivity and scaling with size. A similar model is proposed by Samsonovich and McNaughton (1997) that treats the CA3 region as a continuous attractor map. They make the prediction that no two cells with overlapping place fields after environmental distortion can be active simultaneously if their place fields did not originally overlap. Both these models have in common the mapping of place cell reference onto a *pre-existing* attractor structure.

The models assume no learning, the correspondence between a set of cells, and a chart random assignment of cells to places. It was assumed that random assignment of cells to places was needed to explain the lack of relationship between a cell's position in one environment, and another, and various theoretical arguments Rolls (1996) about the need to orthogonalise similar environments for accurate retrieval have been advanced to support that assumption. Models such as the hopfield network described in section 4.3 and heteroassociative network, described in chapter 7 perform best on orthogonal input data.

These models assume that the hippocampus does not contain one universal attractor structure, or chart, to which an environment is mapped, but a number of different charts, between which the hippocampus can rapidly switch. The assumption was originally required to explain how the place cell structure can be set up almost instantly, and then changed quickly when the rat is moved from one environment to another - known as remapping. These three assumptions have been all challenged experimentally, as we will discuss in sections 4.5.1, 4.5.2 and 4.6; The challenge to the last assumption has

implications for modelling studies, as the possibility that input patterns are not uncorrelated must be considered. In chapter six and seven, we therefore consider the effect of correlated data on simple hippocampal models.

### 4.5.1 Non topographic organisation of cells

Cells in CA1 and CA3 are unlike object recognition cells in perirhinal cortex, or cells organised in columnar fashion in MT. Similar stimuli are not represented in a similar location, either within an environment, or across environments. Part of the cognitive map theory is that the identity of individual objects is not represented at all; rather, only the layout or relative organisation.

A crucial test of this idea is to examine the relationship between the representations of two visually identical environments. Skaggs and McNaughton (1998) examined the relationship between two visually identical regions, and found strong correlation between CA1 place cell activity in these two regions. These cells, the authors suggest, cannot be assigned randomly. However, some cells remapped or, more accurately, responded to different environmental features. Skaggs and McNaughton argue that their results suggest that some cells respond to visual combinations, while others are affected by a randomising effect. However, the regions were joined by a corridor, and the rat was allowed to pass from one region to another. So, arguably, both rooms were part of the same environment, and this experiment does not show anything about relations between different assigned reference frames, but only relationships within a reference frame.

In a second experiment reported in Tanila (1999), the rat was taken from one region to another, visually identical but differently located. In this case, the two visually identical regions are clearly signalled as separate. The rat presumably is able to tell these regions apart using non visual cues. Tanila reports no overlap for CA3 cells firing in the second environment and a small degree of overlap for a CA1 ensemble recording. This means that cells that fired in the first region either did not fire, or fired in a completely different region of the second chamber. This appears to support the idea of random assignment, contra Skaggs and McNaughton. However, on returning the rat to the original environment, the cells responded with an overlapping representation. This means that some of the cells that fired in both the initial response to the first

chamber and during the initial response to the second chamber, fired on the subsequent presentation of the first chamber. Tanila does not report whether this representation remained overlapping, on further presentations.

It is unclear how to interpret these results. One possibility is that the rat is recalling the two separate regions, in response to the visually identical cues. This might indicate the presence of interference in recall. However, it also suggests that the hippocampus may be capable of simultaneously recalling more than one memory representation. The multiple maps hypothesis put forward by Reddish (1996) and Wood *et al.* (2000) claim that effects like the one described in Tanila are actually dual memories, rather than interference. The question remains open, as to whether this is the explanation for the correlations Skaggs and McNaughton observed.

Although these results do not directly contradict the idea of independent pre-configured charts, they do suggest that information is available across environments about common properties.

## 4.5.2   Place cell development and replacement

A link was thought to be established between LTP and spatial learning in water maze type tasks, as rats treated with AP5 (NMDA antagonist) showed impairments on the task. Bannerman *et al.* (1995), and independently Saucier and Cain (1995), showed that the impairment on the task can be eliminated by pretraining on a similar type of task. So rats treated with AP5 pretrained on a different water maze in another room were able to learn normally on a new maze, even though the two water mazes had no stimulus overlap. Bannerman *et al.* (1995) found that only specifically spatial pretraining removed the learning impairment. So it may be possible to transfer the spatial aspects of a task from one trial to another, but NMDA learning is still required to learn them in the first place. These results also suggest that the hippocampus has a very important role in identifying and learning the constraints of a task. Bannerman *et al.* (1995)'s data alone suggests that NMDA receptors are critical for learning, but not either retrieval, or maintenance and adaptation of learned material.

Additional evidence suggests that the medium term stability of place fields does depend on NMDA LTP. Acute blockade of NMDA resulted in newly established place fields disappearing. The fact that NMDA receptors are responsible for stability rather

than initial learning does not return us to the idea that the map structure must be innately preconfigured. Place fields have been found to be very stable over repeated short trials in the same arena (Thompson and Best, 1990). A recent study by Ludvig (1999) tracked the behaviour of place cells over prolonged exposure to the same environment. He found a number of place cells terminated their firing response between 6 and 12 hours after the rat was placed in the environment, while other cells that had been firing at background level for the first few hours developed place fields during the same period. He also describes one cell that terminated its firing and then showed a new place field response that remained stable until the end of the 24 hour period. Ludvig's data do not give an impression of how much the ensemble changed over the 24 hour period. But his results are more consistent with a learned representation that evolves with use.

### 4.5.3  Partial remapping

Remapping is the phenomenon of sudden global reassignments of cell firing correlations, when part of the environment is changed. In so far as this has been observed, it has been argued that it is evidence of some global representation. The phenomenon of partial remapping has also been observed, suggesting that the place cell representation is not a single coherent entity. Shapiro *et al.* (1997) looked at the effect on the location of place field of rotating and scrambling cues. They found that the majority of cells rotated with distal cues (47%) and 16% followed distal cues. Repeated double rotation trials pushed the number of cells that formed new representations (i.e., totally re-mapped), from 10% to 60%. Shapiro *et al.* (1997) proposes a model whereby stimuli compete to control place fields. The system has a bias towards distal cues, but that bias is overridden by instability. Overall, they view the place fields as responding to cue conjunctions, rather than encoding spatial relationships. The effect on performance error of remapping due to cue manipulations has been examined in the alternation task in a radial arm maze. Lenck-Santini *et al.* (2001) found that alternation errors correlated with the degree to which place cell fields had been disrupted by cue changes.

Guazzelli *et al.* (2001) have a model that can dynamically remap the place cell representation, using a mechanism designed to model the way the retinal map handles visual saccades. The model uses an environmental anchor and path integration to remap relative to the environmental anchor.

The unresolved issue is still the exact contribution the hippocampus makes to the rat's performance on these tasks. Even Guazzali's model still relies on the 'forced attractor' method of route finding. So to use the map, the crude method of activating the desired location and allowing the next layer to compute a path is still required. Map models rely on the output layers, (meaning the subiculum) to control them in some unspecified way. It is possible that if there is a 'map' it is actually located in the subiculum, as will be discussed in the section 4.8.

It is possible to design a map-type model that copies the behaviour of place cells. However, there are two further problems. Firstly, such a model does not really address the question of how this map is used by surrounding cortex. Secondly, the presence of task related firing suggests that the map theory may be on the wrong track, or perhaps only able to account for some of the hippocampus functioning. This last point gives rise to the weak spatial approach.

## 4.6  Weak spatial maps

Some of the strongest evidence for the place cell representation acting as a coherent body is provided by remapping and rescaling phenomena. The evidence for this, outlined below, has given rise to the weak spatial map hypothesis, where a subset of hippocampal cells are designated place cells, that are used to capture the spatial properties of an environment.

### 4.6.1  Rescaling

Such a view is taken by O'Keefe and Burgess (1996), who also manipulated place fields by altering the shape of the rat environment. Place fields were found to scale as the environment grew larger, and stretch as the environment changed shape from square to rectangle. They do not indicate what proportion of cells show this property. A much earlier paper (Muller and Kubie, 1987) in addition to observing rotation of place fields in response to rotating cue also observed scaling. In this case, 52% of cells showed completely different behaviour. These cells showing rescaling behaviour have formed the basis of a number of models.

## 4.6.2  Place cell models

Place cells can be modelled directly, in terms of their representational content. This approach is exemplified by O'Keefe and Burgess (1996). Place cell fields, as we have already described, are controlled by relationships between particular cues. O'Keefe and Burgess (1996) identify the cues controlling a cell's behaviour, and attempt to model the geometric determinants of place fields, using boundary vector cells. Their aim is to model the way the fields operate and distort in response to the shift in the environment. They assume that each cell is controlled independently by a subset of boundary vector cells. The model assumes no interaction between the cells. Each cell has a threshold linear response to the summed inputs. Each cell has its representational content determined independently.

This more qualified version of the spatial map theory assumes only that some of the cells are used to capture objective spatial relations. A later model from the same lab (Hartley *et al.*, 2000) now claims to capture the behaviour of those true place cells. They also assume that cell boundary vectors are predetermined. One consequence of their view is that the location of place cells should be predictable across geometrically different environments. They claim to have found that these cell locations are predictable but, so far, these results are published only in abstract form. Their claim is in direct contradiction to results published in Wood *et al.* (2000), who claim that cells that are obviously place cells in one arena can be non place cells (have odour, or task correlates) in another. Whether these specific boundary cells exist in the hippocampus, or in the subiculum (see below), a model of their behaviour is not an adequate model of overall hippocampus functioning, as it considers only a small fraction of cell behaviour.

## 4.7  Other correlates of 'place' cells

Other features of the environment, such as odours, have fields associated with them Wood *et al.* (1999). These are not significantly different from place fields, but simply are controlled by a different modality. However, a change in experimental approach has revealed a more exciting feature of the cell representation.

Most of the evidence we have considered so far has been from experiments where

the rat has been allowed to explore freely, or given a simple memory task, that involves correctly selecting the location of a food reward. The behaviour of place cells in more demanding memory tasks has also been considered. In the T-maze, the Y-maze and the radial arm maze, the rat can also be asked to perform a simple reference memory task. These mazes can also be used for alternation tasks. Here, the rat is required to alternate between the two ends of the T-maze to receive a reward. Again, successful navigation depends on the positioning of cues outside the maze, but also on the rats memory of its previous trip through the maze and on its memory of the task constraints.

There is a growing body of evidence from these types of tasks that the same cells that appear to represent place can represent other invariants related to a particular task. Wood et al. (2000) showed that the 'place cells' at a central location for an alternation task were selective between left and right turn trials. Their data do not indicate whether these cells predicted the expected direction of turn, or recorded the previous turn.

The DNMS task has also revealed a tendency in hippocampal cells to represent abstract properties of the task. Rather than representing individual match, non match cases, hippocampal cells fired for all match, or all non match cases, as described in Hampson et al. (2000); Suzuki and Eichenbaum (2000). Kobayashi et al. (1997) found task dependent correlates of hippocampal cells for different tasks in the same spatial arena, as did Markus et al. (1995). Task dependent remapping supports the idea that the rats may have multiple representations of the same environment. This is an extremely difficult property to model - it raises the question of what exactly the learned representation in the hippocampus contains.

In section 4.5.2 on LTP and learning, we saw that pretraining on a water maze led to normal performance in NMDA blocked animals. This might also be interpreted as indicating that the important contribution of the hippocampus is to supply important information that has been generalised about the nature of the task. Further evidence for this has come from consideration of the representations in the areas surrounding the hippocampus. The simple model put forward by Burgess, Recce and O'Keefe in 1994 treated the entorhinal/perirhinal areas as supplying sensory information, and the subiculum as supplying simple goal representation relative to the map. If the hippocampus is not a map, perhaps a better understanding of what the hippocampus does supply can be achieved by comparing its representation with those the region receives,

and those it outputs. As it turns out, the map idea may be better implemented in the subiculum.

## 4.8   Place inputs and outputs of the hippocampus proper

Subiculum cells show place correlates very similar to hippocampal cells, but spread over a wider area. However, under remapping conditions for hippocampal cells, subicular place fields remain stable. Sharp (1999a) compared hippocampal and subicular place fields in an experiment that involved placing the rat in a box, and then shrinking or enlarging the box. The majority (89%) of subicular cells followed the change in dimensions of the box while only 16% of hippocampal cells did so. These maneouvres were performed while the rat remained in the environment. This contrasts with the results of the experiment by Burgess and O'Keefe (1996). Their model (see below) assumes that subiculum place fields are formed by the superposition of hippocampal fields. However, the data found by Sharp suggest that subicular fields are relatively independent and complementary to the hippocampal fields, and fields in both regions may encode sensory information directly from the perforant path, as suggested by Witter (2000). Sharp's hypothesis is that the subicular fields do something more like hippocampal place fields were originally supposed to do: it is suggested that the subiculum has a universal map, that computes distance and direction information by combining path integration, and composite sensory cues.

Different representations have also been found in the entorhinal cortex, perirhinal, postrhinal or parahippocampal region. In particular, entorhinal cortex has been found to respond to conjunctions of features in context; perirhinal cells have been found to respond to spatial arrangements and objects in place. The response of the perirhinal cortex was found to reflect the demands of the task, but in a different fashion from hippocampal cells. Suzuki and Eichenbaum (2000) compared the behaviour of entorhinal cells and hippocampal cells on match to sample and non-match to sample tasks. The animals were presented with two variants of the task: a case where only the match was repeated, and a case where the "filler objects" were also repeated between presentation of the sample and the match. Changes in cell responses were item specific, indicating recognition of a particular item. In the first condition, some cells in the perirhinal cor-

tex showed an increase in firing to a match condition, while others firing suppressed firing. This second condition was found to raise the proportion of cells that showed suppressing responses to a match condition. These neural patternings lasted for the duration of a trial, forming a short term memory. Suzuki and Eichenbaum do not speculate on how a suppressed response could carry information. Speculation elsewhere has suggested that object recognition systems in perirhinal, hippocampus and perhaps amygdala are in competition (Murray and Mishkin, 1998), as we have already discussed in chapter 1.

In contrast, under the same pair of conditions, cells in the hippocampus responded to match/non-match conditions in a way that was abstracted from the particular item on trial; cells responded to either all match conditions, or all non-match conditions. Suzuki and Eichenbaum suggest that the hippocampus contributes information about the demands of the test, that feeds back and modulates the response of the perirhinal cortex. They suggest that the representation in the hippocampus is as much a memory for the constraints of a task as it is of an environment.

The subiculum, however, does not store information *about* trial type (Hampson *et al.*, 2000). Subicular cells store information *across* trials that persists after the hippocampal cells ceased to fire.

The complex representations found in the cortical and subcortical areas surrounding the hippocampus clearly play an important role in spatial and working memory based tasks. The representation of place is not a property that the hippocampus alone possess. If we want to isolate the contribution of the hippocampus, perhaps we should turn instead to accounts of the computation performed in hippocampus, regardless of the content of the inputs.

## 4.9   Consequences for Moser's hypothesis

We discussed in chapter 2 that Moser proposed that the hippocampus was divided into a spatial and non spatial computational circuit. The evidence presented in this chapter suggests that there is now evidence that the hippocampus does not store pure spatial information. This is dorsal hippocampus as the majority of the cell recordings discussed here have been recordings from the dorsal hippocampus. If there are two computa-

tional circuits, they need to be characterised in a different manner. It was proposed in the previous chapter that this difference was in terms of the speed of learning, and perhaps in the degree of generalisation of particular episodes.

In the next chapter, we discuss the relationship of generalisation to prediction, and the need for prediction to incorporate information about the relationships of events over time.

# Chapter 5

# Predictions, Anticipation and Theta Rhythm

## 5.1  Summary

The purpose of chapter five is to consider the hypothesis that 'sequence learning' is a primary function of the hippocampus. This hypothesis claims that the hippocampus is specialised for learning the order of events. Evidence supporting this device is of two kinds; deriving from the presence of endogenous rhythms in the hippocampus, that can be used to store the order of events and evidence deriving from the interpretation of hippocampal dependent tasks as tests of memory for event order. The conclusion of this chapter is that evidence of both kinds is inconclusive. Nevertheless, the issue of whether a structure like the hippocampus can store memories and recall them in the order stored is usefully addressed in modelling studies.

## 5.2  Introduction

This chapter is about models of the hippocampus built around the idea that the hippocampus is concerned with prediction. The key point about prediction is that it involves anticipating one event following another. Most of the models considered in this chapter have related this to the idea of sequence processing. In this view the mechanics of prediction involve learning the order in which events (inputs) occur, and then being

able to recall by replaying the order in which they occured. This enables anticipation. It is not quite prediction, as prediction involves making a jump from a representation about a past event to a future event. Anticipation may be only a conditioned response, like the rat freezing in a box in which it has previously had an electric shock. To claim that the rat is predicting an electric shock is a very strong claim. The cognitive map hypothesis, as we saw, involved the claim that the rat has a mental representation. However, models of that representation tied it very closely to current sensory input. For it to be a full blown *mental representation*, that representation needs to be disconnected from current sensory input. It needs to be capable of standing in for the environment when the environment is not there. In philosophical terminology this property is the property of *aboutness*. These remarks follow the tradition of philosophy that defines meaning in terms of the conditions of correct application, or use. In order for a representation to be a mental representation, it must be capable of being evaluated as a misrepresentation. An important part of being able to establish the content of a representation is an examination of conditions under which that representation will be revised, or corrected.

In the previous chapter, we discussed whether or not the case of the overlapping firing patterns from two visually similar environments could be understood as a genuinely ambiguous representation. An ambiguous representation here means that the rat is representing uncertainty about which of the two environments it is in - it represents the ambiguity in the environment. For this, the place cell firing would have to have the potential to represent two environments separately, but simultaneously. For the hippocampal system to have the possibility of representing two alternative possibilities, the system would have to somehow represent that the rat can only be in one environment at a time, and that these possibilities are exclusive possibilities. Only if that were the case could this be viewed as a case of prediction.

The alternative, less cognitive (but perhaps more plausible) possibility is that the rat is confused, and then representation recalled corresponds not to both places, but neither - rather some composite place that has the characteristics of both.

# 5.3 Formal prediction

A population code representation can encode information that includes a probability distribution across alternatives. A formal model is described in Zemel *et al.* (1998). But as they rightly point out, the difficulty is not encoding this kind of information in a population encoding but in decoding it. And, further, finding a method of decoding that the rat brain might be using.

Whether or not the rat brain can perform some operation that corresponds to a formal prediction model will only become an interesting question if the place cell firing can be shown to be genuinely tied to the future, (anticipatory) *and* genuinely representational, as that is the prerequisite for the encoding of events as alternatives.

Under some theories of representational content, the correlation of the environment with the occurrence of the mental representation is enough to assign content to it. In this case, it is the occurrence of the neural firing. Neural firing that correlates with anticipated events has been observed in monkey prefrontal cortex, but not yet in the hippocampus. The paradigm in which this has been observed, reported in Rainer *et al.* (1999), looked at a paired pattern associative memory task, and a delay matched to sample task. The animal was taught to associate pairs of patterned objects. Rainer *et al.* (1999) distinguish neural responses controlled by current sensory items, from neural responses based on expected sensory items. The animal was shown the first of the learned pair and then a second object after a 1000ms delay. If it matched the correct target object the animal pressed a lever to obtain a fruit juice reward. Neural firing recorded in the lateral prefrontal cortex anticipated the appearance of the second object and, more crucially, errors were predictable by the similarities between target objects. Where two target items were similar, they interfered with one another by creating false alarm responses. However, when the two items were used as sensory triggers (i.e., the first items in the learned pair) they did not interfere with one another.

The interesting halfway house is the idea that the rat can recall a past sequence of events and use it to anticipate a following sequence. The neural firing patterns corresponding to a recalled sequence need not correspond to a predicted future event, and therefore will not correlate with one. The rest of this chapter is concerned with the data that has led to a theory of anticipatory recall, and a critique of the existing models. Many of these models draw on the existence of theta and gamma oscillation within the

hippocampus, as vehicles for indexing or ordering neural firing. Therefore the order of this chapter is as follows. Firstly, a section will be devoted to the physiological aspects of theta with a little about sharp wave and gamma waves. Then the evidence that place cells anticipate the location of the rat will be discussed. Three models of the theta rhythm based on the evidence will then be evaluated

## 5.4    Rhythms in the hippocampus

This section describes some of the in vivo firing characteristics of hippocampal cells. Place cells are the pyramidal cells. They are characteristically complex spike cells that produce a train of spikes in response to the place correlate. However, pyramidal cells do not produce uniform responses to an artificially applied stimulus. Vinogradova (1995) describes both tonic and phasic firing and, particularly in CA3, rapid habituation to a repeatedly applied stimulus. Pyramidal cells are, of course, not the only type of cell present in the CA3 and CA1 areas. The majority of cells are interneurons, most of which are thought to be inhibitory, described extensively in Freund and Buzsaki (1996). It has been found that some interneurons show place specific firing.

The hippocampus shows three forms of background oscillatory activity. Background is taken to mean a mode of firing that is not associated with a particular environmental stimulus. These forms of background population activity, theta, gamma and sharp wave, are associated with certain phases of behaviour. Individual pyramidal cells show random low frequency firing. However, there are cells entrained to the theta rhythm (8-12Hz) that is paced by cells in the medial septum (Vinogradova, 1995; Freund and Buzsaki, 1996). The rhythmic input from the septum is, as described in the previous chapter, both cholinergic and gabanergic. Acetylcholine acts both pre and post synaptically via two receptor subtypes. Nicotinic and muscarinic receptors are found presynaptically and modulate glutamate, producing an inhibitory effect. Muscarinic receptors are also found postsynaptically, where the effect is on K+ current and depolarises the postsynaptic neuron, producing an excitatory effect. Gaba (generally) has an inhibitory (i.e., hyperpolarising) action, but the inputs from the septum may act on inhibitory interneurons, thus producing an overall disinhibitory effect. The models discussed in section 5.9 have attempted to capture some of the effects and mechanisms

of theta, but there is not a definitive model.

In the developing rat, there are extra-synaptic Gaba receptors that have a depolarising action. There are speculations that these receptors are still functional in the adult rat, and may be activated by 'overspill' of Gaba (Traub *et al.*, 1999).

Place cells interact with the theta rhythm, and can sometimes become entrained to it. Other cells, which may be interneurons but may also be pyramidal cells, turn off during theta (Vinogradova, 1995). It is the place cell/theta interaction that has proved of particular interest to modellers.

The second form of background firing is high frequency irregular activity known as sharp wave. Regular oscillations known as gamma oscillations occur at 40-120Hz. Sharp wave and theta are mutually exclusive, and occur in sequence in both sleeping and resting animal. However, gamma oscillation is found during theta, and can be detected in the dendrites of pyramidal cells that are also modulated by theta. This subthreshold fast oscillation also occurs during sharp wave. Both sharp wave and theta have been found to control the induction of LTP and LTD.

## 5.5   Prediction without theta

It does not help to view all forms of learning as forms of prediction, as we saw in chapter 1. What we seek here is evidence that the rat is mentally ahead of its current situation or location.

### 5.5.1   Coherence as a measure of anticipation

Usually the current location of the rat when the cell fires is used to assign the content of the place cell. Muller and Kubie (1989) proposed a variation of this paradigm, where place cell firing was used to predict the location of the rat. They devised a measure of the coherence of place cell firing which, they proposed, measured the point at which place cells made their best prediction, independently of the correlation between location and firing.

If this prediction is then compared to the actual position of the rat when the place cell showed its most coherent firing, the predicted location of the rat is slightly ahead (120 ms) of the actual location of the rat. Muller and Kubie found this positive shift in

about two thirds of cells recorded from both CA1 and CA3. This positive displacement is small enough that it might be accounted for by the positioning of the tracking light between the animal's ears, rather than its nose. It is also worth noticing that the rat's computed position was based on partitioning the arena into a 64 by 64 grid, of about $3cm^2$, and the study did not take not account of running speed. Muller and Kubie claim their effect is significant, but their analysis does not take into account the rather coarse sampling of the data. However, their findings did lead them to speculate on the possibility that the place cell ensemble represents 'what might be', in addition to 'what is'.

Blum and Abbott (1996) claimed that a temporal asymmetry was an immediate consequence of asymmetry in the NMDA dependent LTP. They modelled the Muller and Kubie data on this basis. LTP was modelled as a function of cell firing rates, and time interval between pre and post synaptic firing, that occurred only when presynaptic activity precedes postsynaptic activity. However, they also used a kind of retroactive reward system, where synapses were only potentiated after the platform was reached, and the amount of retroactive potentiation decayed over four seconds.

## 5.5.2 Transitive reasoning as a form of prediction.

Two papers (Dusek and Eichenbaum, 1997; Bunsey and Eichenbaum, 1996) on transitive inference in rats can be viewed as examples of predictive inference. These were described as the flexible use of representations. As we discussed in chapter 1, transitive inference involves the transfer of properties between premise pairs. The subjects are trained to select one of a pair of items: $a$ is rewarded rather than $b$ (represented as $a > b$), and then three similar pairs $b > c, c > d, d > e$. They are then tested on the pairs $b > d$ and $a > e$. As $b$ and $d$ have been both rewarded and unrewarded choices with equal frequency, the correct selection of $b$ in test pair $b > d$ is considered an example of transitive inference between pairs. This task was compared with Piaget's theory of human cognitive development, where relational transfer is usually tested with the examples of relations like 'longer than', and does not develop until age 6 or 7. However, there are some important differences. The rats were tested using odour discriminations. Therefore, the pairing between items was essentially arbitrary. The rats are learning to associate arbitrary co-occurrences, just as the kind of path

learning shown in Blum and Abbott (1996) resulted in transfer of association between pairs of places.

The key thing that the rats need to learn is to associate the environments or episodes in which the odours are presented. This distinguishes the demands of this task from the exploratory task. Eichenbaum and colleagues found that rats learned to transfer between odour pairs in a single trial, performing at above 80% success on the $b > d$ discrimination. However, hippocampal, fornix, or perirhinal and entorhinal lesions abolished this transfer, without affecting performance or learning rate on the individual pairs, or performance on the $a > e$ pairing that can be learned as a reward stimulus pairing. This was claimed as evidence that the hippocampus mediates a kind of inference. It certainly seems to show the hippocampus has a role beyond simple recall of a previous environment.

This task has been modelled as a form of sequence learning and recall, where it is simply a matter of additive associativity. It looks like a case of genuine anticipation. It may depend on the background context to create the chain of associations; this has yet to be determined experimentally.

Representing this task as a sequence is fairly intuitive. The ordered pairs are presented A followed by B, followed by 'rewarded B', as described in Smith and Levy (2000).

## 5.6 The role of theta

It was proposed by O'Keefe and Recce (1993) that the theta rhythm provided a mechanism of prediction, by moving the firing of place fields in advance of the rat's current location. The observed phenomenon is now called the theta phase procession.

### 5.6.1 Phase precession

Theta, as we have said, is observed as rhythmic oscillations in the firing of cell populations in the hippocampus, at a rate between 8 and 12Hz. The entire hippocampus is not synchronised at this rhythm, although theta in the dentate gyrus is often considered to be reference theta (it is the dentate rhythm that shows up in the EEG). Rather, the peak of theta firing moves through the areas of the hippocampus, the response in each
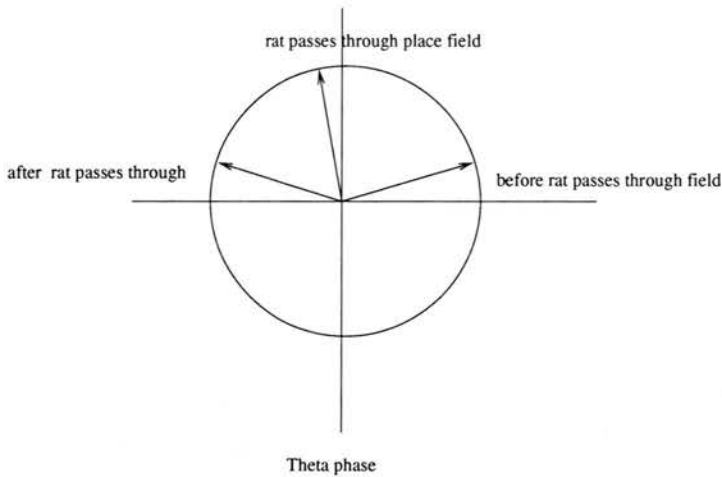
Figure 5.1: Theta phase precession. This figure shows the change in timing of a place cell firing relative to the *local* phase of theta oscillation. The cell fires in advance of the local peak before the rat passes through the field and after the local peak when the rat is leaving the place field.

area being at some phase relative to the dentate theta. A particular cell fires at a point in the theta phase, relative to the local peak of theta firing, and also at a point in phase relative to the dentate theta. O'Keefe and Recce (1993) found that the timing of the firing of an individual place cell moved relative to the local peak theta firing, as the rat moved through the location correlated with the place field. In particular, when the rat ran along tracks, the cell firing advanced in phase. The picture presented, illustrated in figure 5.1, is that when the rat is approaching a place field the cell fires at the end of the theta cycle, and when the rat is leaving the field the cell fires at the beginning of the cycle.

The interpretation proposed by O'Keefe and Recce (1993), Burgess *et al.* (1994), Skaggs *et al.* (1996) and Wallenstein and Hasselmo (1997) is that the cycle recalls a series of locations in the order that the rat expects to traverse them. The cycle begins with the locations the rat has just moved through, and predicts the future location.

It is important to recognise that this is only an interpretation. The firing of an individual cell relative to the theta cycle will be related to the strength of excitatory inputs, and the time constant of the cell. If the cell is oscillating faster than the theta rhythm, it will move in relative phase. A number of modelling studies (Bose and

Recce, 2001; Burgess *et al.*, 1994) showed how to set up the theta phase procession using a simple inhibitory circuit, and also how to set up phase locking.

Vertes and Kocsis (1997) found that oriens alveus interneurons fired ahead of the pyramidal cells as did pyramidal layer interneurons, thus fitting the model described in Bose and Recce (2001). This is sufficient to achieve phase procession.

The theta rhythm is generated partly by the external modulation of theta driving input from the septum. It is also controlled by the basket cell/interneuron network, but not the pyramidal cell population within each hippocampus region, as theta is sustained even under conditions where pyramidal cell firing was blocked (Traub *et al.*, 1999).

### 5.6.2   Behavioural correlates of theta

The theta rhythm occurs in the waking rat, during sniffing, exploring and grooming. It has been suggested that theta is particularly associated with movement and also with states of raised attention. However, it has also been argued that theta activity suppresses hippocampal transmission (output), and therefore theta plays an important gating role (Vinogradova, 1995).

Theta also occurs during REM sleep. Again, it has been proposed that this is during the replay of memories during sleep that may be required for memory consolidation, evidence for which was provided in Wilson and McNaughton (1994), discussed below. However, it has also been proposed that consolidation occurs during sharp wave activity, which occurs *after* exploratory, theta associated behaviours, and during slow wave sleep (Buzsaki, 1989).

### 5.6.3   Theta and plasticity

Plasticity during theta appears to be critically dependent on NMDA receptors (Lynch and Larson, 1989; Larson and Lynch, 1989; Larson *et al.*, 1986; Wickliffe, 1991). Wyble *et al.* (2000) found modulation of EPSPs in CA1 during theta rhythm. They stimulated either CA1 directly, through stratum radiatum, or contra-laterally, via the firing of CA3 cells, and they recorded from CA1 stratum radiatum. Overall, theta phase depressed the size of EPSP's from both kinds of stimulation by 18%. The slope of EPSP's varied with theta phase, so that peak slope was $18 - 90$ deg after the positive

peak of dentate theta, and after the negative peak of the local CA1 theta (as recording electrodes were positioned to measure roughly anti-phase theta). This means that maximum EPSP slope is during the rise of local theta rhythm. The minimum EPSP slope occured about 90 deg later. Lynch and Larson (1989); Wickliffe (1991) originally identified the cause of the timing of point of maximum EPSP as a refractory response to the peak feedforward inhibition triggered by the granule cells firing at the dentate theta peak. For more details about the connections between dentate gyrus and the pyramidal cell layers, see chapter 3, section 3.4.1.

These results are consistent with the findings that LTP is preferentially induced on the positive phase of local theta. They are also consistent with the idea that there is a short window of simultaneous firing that might be modelled as synchronous updating in a network model.

Stimulation in the negative phase of local theta results in LTD (in dentate gyrus) or LTP reversal in the pyramidal cell layers (Holscher *et al.*, 1997).


## 5.6.4   Evidence for sequence replay during sleep.

As already mentioned, theta also occurs during REM sleep. In humans, deprivation of REM sleep affects both short term memory and cognitive functioning. Work in McNaughton's lab found evidence of a correlation between the sequence of cell firing patterns during waking theta, while the rat was running on a track, to sleeping theta in the sleep session following exploratory activity (Skaggs *et al.*, 1996; Wilson and McNaughton, 1994). In later experiments (Kudrimoti *et al.*, 1999; Poe *et al.*, 2000), a phase reversal of the firing patterns of these cells was observed. So the cells that had fired in positive theta now fired in the theta trough, under the conditions that produce depotentiation of theta induced LTP described in Holscher *et al.* (1997) and Xu *et al.* (1998). Kudrimoti *et al.* (1999) and Poe *et al.* (2000) claim this as evidence that theta plays a role firstly in storing sequences, and then in the necessary forgetting that is required to prevent the associative memory in the hippocampus from overflowing.

In addition to the idea that theta phase indexes memories, thereby allowing meta properties (such as order) to be represented, Buzsaki (1989); Sohal and Hasselmo (1998) and Borisyuk *et al.* (1999) have proposed that theta modulated gamma is a binding mechanism for simultaneous events. The idea of synchrony as a binding sig-

nal was proposed early on by von der Malsburg, and has been used in models of all levels. This is in line with Engel *et al.* (1991)'s proposals for the function of the 40Hz oscillation in the cat visual system visual system. The idea that the phase of cell firing relative to an underlying oscillation can carry information has also been proposed for other systems. A phase coding system using oscillations has also been proposed for the olfactory system (Hopfield, 1995), and also for working memory. The idea that the order of cell firings represents the order of events is perhaps the simplest application of a phase coding for memory.

## 5.7 Evidence for the role of gamma oscillations

The sequence learning models of Hasselmo and Lisman (Hasselmo *et al.*, 1996; Sohal and Hasselmo, 1998; Lisman, 1999; Lisman and Otmakhova, 2001) have taken up the idea that gamma, within a theta envelope, may regulate the packaging of information transmission. However, its functional significance is still uncertain.

Most of the study of gamma frequency oscillation in vitro in hippocampus has come about through recognition of the role of gamma in generating epileptic type synchronised population bursts (Penttonnen *et al.*, 1998). A satisfactory link between gamma and memory has not yet been established. Traub *et al.* (1999) attempt to link gamma frequency with changes in plasticity in vitro, in the following manner. In order to induce gamma oscillations, a one or two point tetanic stimulus is applied, close to threshold. However, a twice threshold stimulus produces first gamma and then a frequency shift to the beta range (10-30Hz) with gamma superimposed, that Traub *et al.* (1998) speculate is caused by changes in AMPA receptors.

For the moment, the lack of evidence linking memory and plasticity with gamma in vivo, and lack of information about the role of gamma in the hippocampus means that its inclusion in models of sequence learning remains speculative. Indeed, there is some negative evidence: LTP is not preferentially induced at gamma type frequencies in CA1 and CA3, although high frequency stimulation in dentate gyrus has been found to be sufficient to promote (NMDA independent) LTP in the mossy fibre pathway to CA3. More evidence is available that suggests the creation of transient, fast synchronising patches of synchrony in the entorhinal, dentate and pyramidal cell layers, via

gap junctions.

## 5.8   Sharp wave and plasticity

It has been suggested by Cavus and Teyler (1998) that repeated 200Hz tetanus is a
reasonable analogue of sharp wave stimulation. It induces non-NMDA LTP in area
CA1, in particular in stratum radiatum, and also in the stratum oriens and stratum
radiatum of CA3. It probably achieves this by elevating intracellular calcium. This
form of stimulation also induced NMDA LTP, whereas theta type tetanus did not appear
to induce non-NMDA LTP, suggesting that there might be oscillation specific forms of
plasticity.

## 5.9   Models of sequence recall

There are three network models that attempt to use sequence learning to model hip-
pocampal dependent tasks. Both Lisman and Hasselmo combine sequence learning
and neuromodulation to store and retrieve sequences within a theta cycle. Both models
are influenced by a conceptual model developed by Levy's lab (Minai and Levy, 1993;
Levy and Wu, 1996; Levy, 1996; Wu *et al.*, 1996; Amarasingham and Levy, 1997;
Smith and Levy, 2000), which explores the way hippocampal architectures might store
and retrieve sequences.

### 5.9.1   Sequence learning and recall using theta modulation

Lisman's model used the idea of theta enveloped gamma to cycle through a sequence
of items. While similar to the models of Hasselmo and of Levy, using a different
mechanism it also utilises the perforant path to provide excitatory context. Items of
recall are primed by perforant path inputs to be close to threshold, and then detonated
by mossy fibres. As it seems likely that the perforant path and mossy fibre inputs CA3
are antagonistic, rather than reinforcing, there is good reason to reject this model as
it does not add anything to the ideas already discussed. Lisman does speculate on the
possibility of reverberatory activity between CA3 and DG, allowing for cycles of recall.
It is required to take place within a gamma cycle (i.e., very quickly). Although, as we

saw, CA3 seems to use a rebound effect from inhibition from dentate gyrus to optimise plasticity, this is at the rate of the theta oscillation.

Hasselmo *et al.* (1996) have proposed a model whereby the level of cholinergic input gates learning by controlling output (pyramidal cell) firing and enhancing LTP by modulating the GabaB current during theta. Heteroassociative memory occurs between the patterns in CA1, by the perforant path, and the patterns of activity in CA3 induced by mossy fibre firing. Retrieval was by presentation of CA3 activity. We have already noted that the schaffer collaterals and the perforant path synapse are located in different parts of the dendritic tree. Potentially this means they can be modulated independently. Hasselmo used cholinergic input to achieve exactly this. Acetylcholine was used to suppress transmission through the stratum radiatum, with lower suppression in the perforant path's region of termination. Successful retrieval results in higher activity in the CA1 region, which in turn suppresses the septal input. CA3 is assumed to be an autoassociator.

An initial criticism of this model was that cholinergic input diffuses through the extra cellular space; it may not be specific or rapid enough to mediate fast switching between recall and learning. Sohal and Hasselmo (1998)'s version of the sequence learning device uses GabaB as a mechanism for switching between recall, and learning. The model uses the gamma rhythm within the theta rhythm to cycle through possible sequences, and slow acting inhibition acts as an annealing parameter to settle on the final sequence. Particular attention is given to this model here, as it is very close to the sequence learning model discussed in the next chapter, but at a lower level. In Sohal and Hasselmo's model, CA3 is assumed to be a sequence memory, performing heteroassociative recall through the recurrent connections, and the focus has switched from the modulation due to cholinergic input, to GabaB dynamics. GabaB receptors operate on a slower time course than GabaA. The model hypotheses that the GabaB receptors are on an interneuron that receives only septal (theta) input. GabaA receptors, however, are modulated by both theta and feedback excitation from pyramidal cells. This input is disinhibitory and, as GabaB modulation decreases, GabaA modulation increases, due to the feedback activity inhibitory cells that receive excitatory input from pyramidal cells.

The difficulty with this model is partly that it incorporates some quite low level

details, without incorporating all of them.The argument in the paper is that this population of GabaB theta interneurons is required for sequence disambiguation. However, in principle (see next chapter) sequence disambiguation of overlapping sequences does not require annealing. Levy and Hasselmo both use *context* units - units that do not represent items in the sequence, but link items. The construction of these units means that sequences that have some ambiguous section are *recoded* to be less ambiguous, as they will activate different context units.

On the other hand, the level of Gaba inhibition will be affected by the gabanergic and cholinergic inputs that carry the theta rhythm. However, other commentators have stressed the fact that it is not known whether these have a primary effect on the pyramidal cells, or the interneuron population. The overall effect of GabaB due to the septum could be either inhibitory, or disinhibitory.

A very similar inhibitory system, with a theta modulated disinhibition, is analysed in Bose and Recce (2001), to explain phase precession of pyramidal cells. The difficulty with models at this level is simply that that different types of interneurons and neurotransmitters provide too much freedom for model building.

### 5.9.2    Modelling at the conceptual level

In the next chapter, we shall consider a model of sequence learning in the hippocampus that attempts to divorce the implementation from the biophysical mechanisms. Modeling the hippocampus as sequence learning device, has been approached from a top down , and also a bottom up perspective. Firstly, from the top down, one can construct the simplest computational model consistent with hippocampal anatomy, and show that it performs the task adequately. The disadvantage of this approach is that the resulting model is rather under-determined. If the consequences of this were that the model was extremely simple computationally, then this might not matter. The model considered the next chapter is of this kind. In fact, as we will see, the model has replaced biophysical parameters (that may or may not be correct) with computational parameters that are necessary to make the model work, but do not necessarily have any grounding in reality. In chapter seven we will reduce the parameter space by considering an even simpler version of the model. Secondly, starting bottom up one can model the hippocampus in full biophysical detail, including the theta rhythm, and show that it

performs the task. However, this approach also leaves the model under-determined, as there are inevitably bound to be gaps in the biophysical description. Models like those described in this chapter(e.g Hasselmo *et al.* (1996)) are more bottom up than top down, but the choice of the details incorporated has been guided by top down considerations.

## 5.10   Conclusion

We need to keep three issues separate.

1. It needs to be shown conclusively that the oscillatory pattern in theta and gamma are used for selective learning and recall. The evidence we have so far presented merely suggests this. Abolishing theta by lesioning the septum damages normal learning and recall. However, this will damage the overall behaviour of inhibition in the hippocampus. It might be possible to do more delicate pharmacological manipulations, e.g., slow the rhythm, or partially abolish it.

2. There is the issue of whether theta provides synchrony, or whether phase shifting relative to the oscillation is critical information bearing signal. The role of gamma oscillations in creating synchrony is an exciting area of research, but it does not fit comfortably yet into the rather simple network models.

3. There is an issue of interpretation. Is the hippocampal representational code genuinely predictive? We gave an example of a genuinely anticipatory code in the pre-frontal cortex. The evidence from, for example, Wood *et al.* (2000) discussed in the previous chapter is still ambiguous as to whether it is controlled by the rat's anticipated action or its previous action. We suggest that it is more plausible that the hippocampus provides what we call anticipatory recall of the previous, individual trial. This information may be used by the prefrontal cortex for planning, and other parts of the hippocampus (the hypothesised slow learning system, perhaps in the temporal hippocampus). This proposal needs to be entirely divorced from speculation about the functioning of the theta rhythm. As the next two chapters set out, sequence learning is possible in a hippocampus-like structure without imposed oscillations.

# Chapter 6

# Sequence learning in the hippocampus: Part I

Memory should allow the present situation to act as a retrieval cue, and when presented with a situation that is similar to some previously encountered situation, it should retrieve the consequences...*Pentti Kanerva 1984*

## 6.1   summary

The contribution of the modeling study in chapter six is twofold- First, we examine an existing model of sequence recall in a network model based on hippocampus type architecture. Firstly we replicate the results of a number of papers by W.B. Levy's group, but give as different analysis of the internal behaviour of the model. Secondly, we attack the basis on which the learning rule and threshold setting mechanisms were selected. The choice of threshold mechanism means that the model does not converge to a stable activity state.This in turn means that the weight values in the network as not as predicated by the analysis proposed by Levy 's group. Therefore, part of the motivation for selecting a learning rule that computes the conditional probability relationships between pairs of input and output units is lost. The problem of learning and recalling highly correlated sequences of data is better addressed by using the tools that have been developed to allow recall of correlated pattern pairs in heteroassociative networks. These mechanisms will be examined and tested in chapter seven. A Pseudo

code appendix contains details of the exact procedures used for training and testing the networks described in this chapter.

## 6.2   Episodic memory

Marr's 'Theory of Simple Memory' (1971), described in his paper 'A theory of Archicortex', prescribes how the hippocampus should be understood as a simple memory. On this view,

- Simple memory is a fast, perhaps even *one trial* method of storing associations.

- It is for events that only occur once

- No prior knowledge is represented in the memory. In practice this means that the memory is *connected randomly*

- It does not generalise across similar events, or categorise them. That is one reason why the memory is *content addressable*

- Storage is temporary

Many of these features are familiar from the description of episodic memory in chapter 1. This is not surprising, as Marr's motivation was to fit the computational nature of his simple memory to the requirements of an *episodic memory*. Marr's theory is a paradigm example of a top-down approach to modelling the hippocampus (see Willshaw and Buckingham (1990) for a useful explication of his theory). The subject of this chapter is Levy (1996)'s model for the recall of sequences, which should be viewed in the same spirit.

We are considering how memories can be used to inform and support action choices. There is a top-down distinction we can draw between two different strategies for making use of past experience.

**Anticipatory recall**   Cued by the current situation, this strategy recalls the last time a similar situation was presented and the immediate consequence of that situation. Hidden within this is a judgement of similarity between the current and past situations, as this determines whether or not the cue will be successful.

**Predictive memory** Essentially the difference here is that the similarity judgement is
represented explicitly. How similar the current situation is to a past event weights
the representation of how similar the consequence is predicted to be. Secondly,
the prediction of the consequence is based on considering all the events of a
similar type.

For simple memories, such as the associative network (Willshaw *et al.*, 1969), the
(Hopfield, 1982) network, and the networks studied here, recall involves a clean up
operation. The network will either recall a pattern or fail. It will recall a pattern that is
more similar to the stored pattern than the cue was to the pattern part. In fact the more
dissimilar the cue is from the pattern to be retrieved for recall still to occur, the 'better'
the memory. However, if you are trying to predict a future consequence, you should
not be more certain about the consequence than you are about the observation. So
already we can see that consequence and sequential recall are opposed to one another.

In chapter 3 we discussed the temporal asymmetry of NMDA dependent LTP, and
in chapter 5 we discussed the use of this property to recall sequences. In this chapter,
we will examine the implementation of this idea, as explored by W.B. Levy's publi-
cations (Amarasingham and Levy, 1997; Levy, 1996; Levy and Wu, 1996; Minai and
Levy, 1993; Smith and Levy, 2000; Wu *et al.*, 1996, 1998). This network modelling
approach attempts to relate higher level tasks to general properties of the hippocampal
system. It does not, therefore, seek to model explicit neurobiology, or neurobiologi-
cal phenomena, such as the theta rhythm. However, it is important that the model be
consistent with known neurobiology, as the functional/computational properties need
to be preserved in a more realistic model, or real system.

In the previous chapter, we concluded that the theta rhythm provided a window
of synchronised activity, during which learning occurred. This model takes a step
back and invites us to consider the properties of more or less synchronous update
schemes, where the only two important properties are the order of neural firing and
the distribution of synaptic strengths.
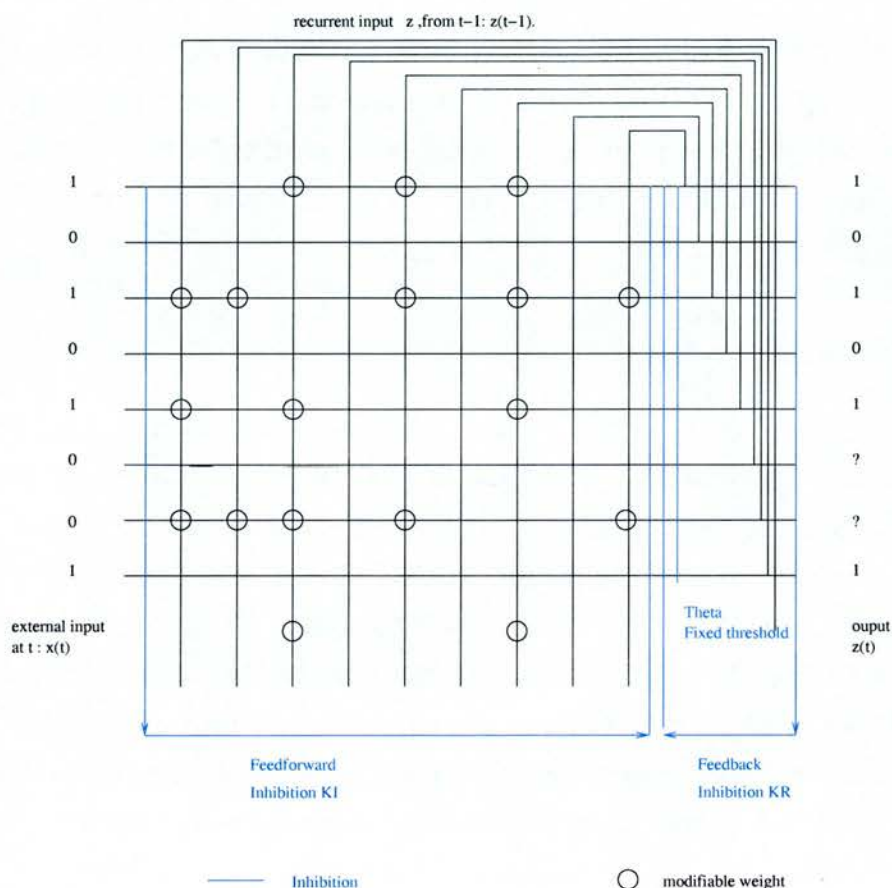
Figure 6.1: This is the recurrent network architecture used for the simulations in this chapter.

## 6.3    A sequence learning model of the hippocampal CA3

First of all, a pattern of neural firing in the hippocampus is represented as a binary string, for both input and output. To enable the model to learn transitions between a sequence of binary strings, it is assumed that the interval between string presentation is equal to the conduction time through the recurrent loop, and this time interval is also equal to the delay for feed forward inhibition, and feed back inhibition (conduction plus synaptic delays). As interneuron synapses appear to be faster than pyramidal cells, this is not unreasonable.

## 6.3.1 Architecture

The architecture is loosely based on CA3 and the Marr network (described below). Pyramidal cells in a single layer are connected recurrently to one another. A pyramidal cell integrates input from two sources: 'teaching' input cells, that directly fire the cell, representing the required output, and weaker input from the previous output. Whether the previous time step's input can fire the cell is controlled by an inhibitory neuron that provides inhibition proportional to both the teaching input and the output.

Increased activity in the inputs results in both higher excitation and higher inhibition. This is a realistic assumption.

Figure 6.1 shows the structure of the network. The dimensions of the network are as follows. The input layer, $X$, has 1000 neurons, as has the output layer, $Z$. As described in Levy (1996); Levy and Wu (1996); Amarasingham and Levy (1997) the connection matrix is randomly constructed with a connection probability 10%.

Note that there is no explicit role here for the perforant path input. However, in chapter 3, we did discuss the possibility that perforant path input might simply act to raise the firing threshold of the neuron in response to the schaffer collateral input.

## 6.3.2 Symbol definitions

| Variables | definition |
|---|---|
| $W_{ij}$ | weight between neurons $i,j$ |
| $t$ | time step at current input |
| $X_i^t$ | external input to neuron $i$ |
| $Y_i$ | output before step function |
| $Z_i^t$ | value of output |
| Free parameters | definition |
| $e$ | learning factor |
| $K_I$ | inhibitory factor |
| $K_I X_i^t$ | inhibition due to feed forward excitation |
| $K_R$ | inhibitory factor |
| $K_R Z_j^{t-1}$ | inhibition due to recurrent excitation |
| $\theta$ | threshold of step function |

| O | 1 | Pre( t-1) | 0 |
|---|---|-----------|---|
| 1<br>Post<br>(t) | + | | - |
| 0 | 0 | | 0 |

Figure 6.2: Contingency table shown for the postsynaptic rule

### 6.3.3   CA3 and Levy model

The following points are assumed in the architecture of the model described in the previous chapters.

1. One granule cell is capable of firing one pyramidal cell

2. No LTP in mossy fibre synapses.

3. One to one correspondence between pyramidal and granule cells

4. Mossy fibre input very much greater than recurrent input.

All these assumptions are false, except possibly the last one, but arguably they are reasonable simplifications.

### 6.3.4   Learning rule

The learning rule that is used in the sequence model described in this chapter is known as the postsynaptic rule. If $Z_i^t$ denotes the activity of cell $i$ at time $t$, the change in weight $W_{ij}$ between presynaptic cell $i$ and post synaptic cell $j$ is given by:

$$\Delta W_{ij} = e Z_j^t (Z_i^{t-1} - W_{ij}^{t-1}) \tag{6.1}$$

For binary neurons, learning (weight change in either direction) can take place in response to four possibilities. The postsynaptic rule exploits only two of these; learning occurs only when the output cell is active. Figure 6.2 shows this in table form, for ease of comparison with the variants of this rule described in the next chapter. The postsynaptic rule is a simple possible rule interpreting the temporal asymmetry. Secondly, it has the property that individual weights saturate at 1, thus realistically preventing

the weights from growing indefinitely (as in the case of the Hebb rule). The amount of potentiation depends on the size of the synapse. This has the consequence that the amount of potentiation due to an input pattern depends on the patterns previously presented to that synapse and the current state of the synapse. This is an interpretation of the Bi and Poo (1998) data that we discussed in chapter 3, section 3.7.

The postsynaptic rule belongs to a family of simple rules customarily used in similar architectures, where modification occurs at different places in the contingency table shown in fig 6.2. The advantage in using this one is only its *relative* claim to realism. There is no claim that this is the *optimal* learning rule for this problem.

### 6.3.5 Interpretation of the parameters

One reason given for the particular form of the learning rule is concerned with its convergence properties.

The activity of the *postsynaptic* neuron, *Post*, is given by $Z_i^t$ in the model.

The activity of the *presynaptic* neuron, *Pre*, is given by $Z_j^{t-1}$ in the model.

For $\Delta W_{ij} \to 0$,

$$W_{ij} \to \frac{P(Post = 1, Pre = 1)}{P(Post = 1)} \tag{6.2}$$

This is just the conditional probability $P(Pre|Post)$.

For anticipatory learning, one might think that a rule that set the weights to $P(Post|Pre)$ would be a more natural choice, as it is easy to see how such weights could be combined to select the most probable postsynaptic neuronal firing from presynaptic inputs. In the next chapter we discuss such a rule.

Choosing a learning rule that reflects the conditional dependencies between firing events is inconsistent with the original idea of simple memory. Marr selected binary synapses for the hippocampus; he did so on the grounds that real valued synapses, reflecting probabilities of occurrence were inappropriate. Why? The simple memory was designed to be a one shot learning system, automatically storing and recalling events that occured at most once. The notion that you could extract meaningful information about the frequency of co-occurrence of singular events simply had no purchase.

One further point is that it is not clear what the function of LTD is. It may not be to represent information but rather to erase previously learned information. Chapter 5

outlined some evidence that LTD occurs during sleep, triggered in the negative phase of theta, that may depotentiate LTP from the positive phase of theta. As we saw in Chapter 3, the NMDA dependent LTD depends on back propagating spikes. It may be that back propagating spikes are blocked (see discussion in section 3.8) and LTD does not often occur. This would be consistent with the numerous observations that LTD is very difficult to stabilise *in vivo* in schaffer collateral synapses.

In Levy's model, the sequence is presented over a large number of cycles. This does not seem to be a realistic training paradigm, for two reasons. Firstly, it requires too many iterations. Animals learn much faster, and they learn in the presence of noise. Minai and Levy (1993) investigated learning in the presence of noise; we shall not consider it here. The use of multiple iterations is common to many neural network models. Also, recall in chapter 1 that the standard of theory of learning via the hippocampus assumes that the *neocortex* requires multiple iterations from the hippocampus. Here we suggest that the fast learning part of the hippocampus might train the slow learning part, by using the septo-temporal connections.

This particular network can be trained on sequence disambiguation cycles in a single trial. Even so, in our experiments, we are interested in whether we gain any usable information about the structure of the sequence over the long training time. Therefore the question of whether the learning rule converges as expected and assumed will be examined.

Even if the weights converged to the expected values, there is still the question of whether the network uses the information stored in the weights effectively. We shall return to this question in chapter 7.

### 6.3.6   Threshold setting

Marr (1971) proposed a dual threshold where the firing condition of a neuron depended on (i) the absolute value of the weighted summed inputs (the *dendritic sum*) exceeding threshold and (ii) the number of modified active synapses. This takes into account the fact that a neuron that fires more often will have more modified synapses. Obviously, a special case that has been much analysed, as discussed in the next chapter, is the binary valued (*clipped*) network where weights are either 1 or zero, and the dendritic sum is equal to the number of modified active synapses.

Here, the network implements a similar dual-control setting strategy, as defined in equation 6.3. Firstly, the quantity $Y_i^t$ is calculated, which depends on the global activity. This transformed sum must exceed the threshold $\theta$ for the neuron to fire. The behaviour of the transform in response to variable levels of activity is shown for a range of parameters in figure 6.3.

This threshold transforms the sum of active weights for each neuron to lie between 0 and 1.

$$Y_i^t = \frac{\overbrace{\sum_j W_{ij}^{t-1} Z_j^{t-1}}^{input\ activity}}{\sum_j W_{ij}^{t-1} Z_j^{t-1} + \underbrace{K_I \sum_j X_j^t}_{feed\ forward\ inhibition} + \underbrace{K_R \sum_j Z_j^{t-1}}_{recurrent\ inhibition}} \tag{6.3}$$

$Z_i^t = 1$ if $Y_i^t > \theta$,
$Z_i^t = 0$ otherwise.

Intuitively, whether a neuron fires as part of the next pattern depends on how many of the inputs it has successfully captured from the previous pattern. However, it *also* depends on how many other neurons it "expects" to fire. This expectation is calculated in the two inhibitory terms that count the number of neurons that fired last time step, $K_R \sum_j Z_j^{t-1}$, and how many externally fired neurons there are at the current step, $K_I \sum_j X_j^t$. $K_R$ and $K_I$ are free parameters that dictate the slope of the transform applied to the dendritic sum. The dependence of the $K_R$ term on global activity means that the neurons do not fire independently. Figure 6.3 shows the threshold dependence on activity. At low activities the neuron $i$ transforms its input in a supra linear fashion, while at high activities it transforms its input in a sublinear fashion.

The effect on $Y_i^t$ of the values of $K_I$ and $K_R$, for a given activity level, is shown below in figure 6.4.

### 6.3.7 Choice of Parameters

Amarsingham and Levy (1998) chose parameters to produce the desired range of activity in the network but they did not explain how these parameters were determined.
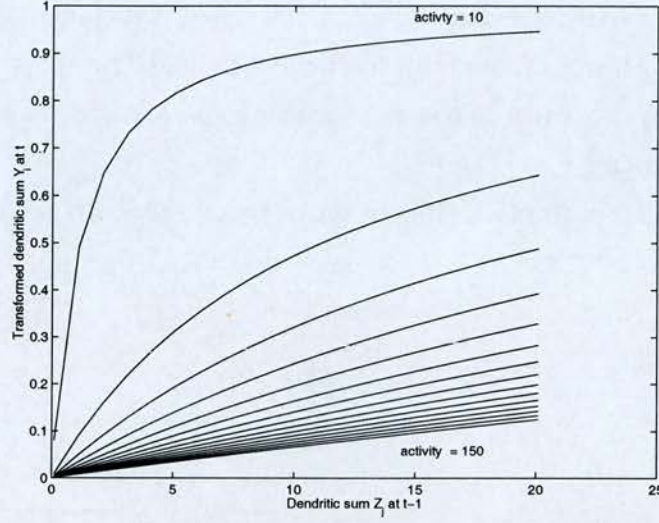
Figure 6.3: The shape of the transformed sum $Y_i^t$ as a function of its dendritic sum, $Z_j$, for different levels of input activity (10 : 10 : 150), with $K_I = 0.016$, $K_R = 0.016$. Each line relates to a particular level of activity.



Figure 6.4: The effect of changing parameters $K_I$, $K_R$, for fixed activities. As $X_j^t$ is of fixed size (in simulation) and much smaller than $Z_j^{t-1}$, it is possible to treat it as a fixed parameter, and scale $K_R$ accordingly.
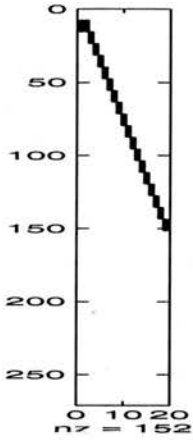
Figure 6.5: Sequence data: the input sequence. Only the first 260 of the 1000 neurons are shown.
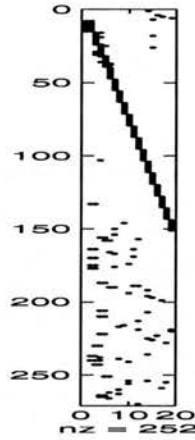


Figure 6.6: Sequence data: the recoded sequence

They give the value of $\theta$ as 0.8, and values for $K_I$ and $K_R$ as varying between 0.011 and 0.18. We found values for $K_R$ throughout this range that gave the same variability 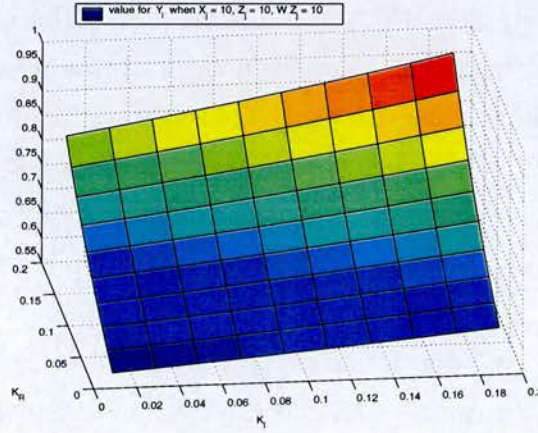in activity, with $K_I$ constant at 0.016, in the middle of this range. The effect of varying both parameters is shown in Figure 6.4. $K_I$ has a very much smaller effect on the behaviour of the network as $K_I X_i^t \ll K_R Z_i^{(t-1)}$

It might be possible to think of a more principled way of choosing parameters. For example, to take $K_R$ as a measure of the background probability of firing within a connected patch. However, Levy and Amarsingham have not adopted this approach. In the appendix is a pseudocode description *Proceedure 0* of how parameter searches were conducted.

## 6.4   Learning a single sequence

This set of simulations set out to replicate the results presented in Amarsingham and Levy (1998).

## 6.4.1   Sequence recall

There is a sequence of $L$ input patterns, each input pattern being a set of $M$ active neurons, made up of consecutively numbered neurons. Each successive member of the sequence is generated by turning off the $Ch$ lowest numbered members of the given pattern and turning on the $Ch$ next highest numbered neurons that are not already active. Thus adjacent patterns in the sequence have $M - Ch$ active neurons in common, referred to as the *overlap*. See figure 6.5 for an example of such a sequence of patterns. Amarasingham and Levy (1997) and Levy (1996) are interested in the tolerance in the recall of a sequence to the degree of overlap between patterns in the sequence.

The network is trained on this sequence, the criterion for convergence being how the final output of the network trained on the complete sequence compares to the output on the previous iteration of training. The final output from the last training sequence, composed of the original sequence and the activity of the context units, is called the *recoded sequence* by Amarasingham and Levy (1997); Smith and Levy (2000); Wu *et al.* (1996). During the testing procedure, this is then taken to be the stored sequence. Figures 6.5 and 6.6 show examples of the original sequence and recoded sequence.

None of the published papers Amarasingham and Levy (1997); Levy (1996); Levy and Wu (1996); Smith and Levy (2000); Wu *et al.* (1996) gives a complete characterisation of the sensitivity of network behaviour to different parameter values. For the purposes of comparison, the results of Amarasingham and Levy (1997) were replicated, and compared to the capacity estimates shown there.

### Simulations

There were $N = 1000$ neurons and $M = 8$ active neurons per pattern. Parameter searches were conducted across the range $0.008 < K_R < 0.20$ for sequence length $L = 36$, for overlaps of $0, 2, 5, 6$. The parameters that produced the best performance were then tested over sequence lengths of up to 200. All simulations were written in Matlab 5.0 and run on Compaq 600MHz 2-processor workstations. A typical single run, involving storing and retrieving a single sequence, took 1 hour compute time. Results from parameters searches are shown in the appendix, in table 1.

**Storage.**  Each pattern in a sequence had a total of $M$ active neurons, the next pattern in the sequence having $Ch$ neurons turned off and $Ch$ turned on. This was done

in two different ways in the experiments reported here: *shifted* sequences had neurons $1, 2, 3, ..., M$ active in pattern $1$, $Ch, Ch+1, Ch+2, ..., M, M+1, M+2, ..., M+Ch$ active in pattern 2 and so on; in *randomly overlapping* sequences, pattern 1 had $M$ randomly chosen neurons out of $N$ made active, pattern 2 had $Ch$ of the active neurons in pattern 1, chosen at random, switched off and a further $Ch$, chosen at random, switched on to replace them; pattern 3 was generated from pattern 2 in a similar way, and so on.

New sequences were created for each training run. Each network was created and initialised for each individual training run with stated parameters.

Algorithm A in the appendix is a pseudo code description of the training and testing procedures for this network.

**Testing.** Recall was tested by presenting the first transition in the sequence (i.e. the first and second patterns of the recoded sequence). The network was then allowed to retrieve the successive members of the sequence for the number of time steps of the learned input sequence.

The recoded sequence consists of between 20 and 80 neurons per time step in a typical run. Amarasingham and Levy (1997) define the *activity* as the average number firing at a time. The original sequence therefore can account for a maximum of 50% of the recoded sequence.

Amarasingham and Levy (1997) define successful recall as follows

> the network produces a minimum of 75% ordered recall of the recoded
> sequence during testing

This is taken to mean that 75% of the responses to the last presentation of the input sequence are identical in the test condition. There is an ambiguity here, as to whether this is 75% of the individual patterns or 75% percent of the neural responses in each pattern; it is also not clear whether false positives during recall are included. Other papers use 70% (Levy and Wu, 1996) to 85% (Wu *et al.*, 1996) of the sequence.

Amarasingham and Levy (1997) further impose the requirement that successful recall performance of this kind must be produced in 4 out of 5 randomly constructed networks. The table of their results shown in Figure 6.7 gives the maximum length of sequence that could be reliably recalled under the conditions stated, for a given amount

of pattern overlap between one pattern and the next in the original sequence. They call this maximum length of reliable recall the *capacity* of the network.

Rather than classifying the results into successful and unsuccessful cases, the results presented here show the quantitative relationship of performance to sequence length - we are just as interested in what happens when recall fails as when it succeeds. To capture the closeness of the input sequence and the recalled sequence, we decided to use the normalised cosine and average it across all patterns in the sequence. The figure of 75% overlap between input and recalled sequences cited by Amarasingham and Levy (1997) corresponds to a cosine score of between 0.7 and 0.8.

To calculate the Performance involving vectors **A** and **B** of equal length, we calculate the normalised scalar product

$$Performance = \mathbf{A}.\mathbf{B}/|\mathbf{A}||\mathbf{B}|.$$

Presented below are results from networks trained on the original *shifted* sequences used by Amarasingham and Levy (1997) and also on a set of *randomly overlapping* sequences constructed to have the same amount of pairwise overlap, but where the active neurons are selected randomly, as explained above.

## 6.4.2  Results

Figure 6.7 shows results from Amarasingham and Levy (1997). As their analysis is based on the asssumption that *each neuron is made active and then becomes inactive only once* in the recoded sequence, they included in the table the number of neurons that violate that assumption. In addition, they included the number of neurons that do not fire ever in the course of the recoded sequence.

## 6.4.3  Experiment 1 - shifted sequences

We replicated some of Amarasingham and Levy (1997) Amarsingham's results exactly, for shifted sequences with giving overlaps of $0, 2, 5, 6$. in batches of three. As already explained, parameter searches had to be conducted for each value of overlap.

| Overlap | Capacity | Activity | Unused | Multiple |
|---------|----------|----------|--------|----------|
| 0 | 20 | 0.051 | 144 | 155 |
| 1 | 24 | 0.048 | 153 | 105 |
| 2 | 33 | 0.049 | 122 | 121 |
| 3 | 47 | 0.064 | 64 | 45 |
| 4 | 57 | 0.061 | 59 | 15 |
| 5 | 90 | 0.055 | 45 | 15 |
| 6 | 110 | 0.050 | 63 | 6 |
| 7 | 160 | 0.054 | 113 | 15 |
| variable | 22 | 0.138 | 77 | 15 |

Figure 6.7: Results reproduced from Amarasingham and Levy (1997). Capacity is the maximum length of sequence for which there was good recall, according to the given criterion.

### 6.4.4 Experiment 2 - randomly overlapping sequences

We ran an additional set of overlapped sequences, at the same set of overlaps: 0, 2, 5, 6. The purpose of Experiments 1 and 2 was to compare our empirical estimates of the capacity of the network with Amarasingham and Levy (1997) and Wu *et al.* (1996).

### 6.4.5 Results

Figure 6.8 shows successful performance on the two types of sequences (where an average cosine score of 0.75 is taken as indicating successful performance) compared to the results obtained by Amarasingham and Levy (1997). Our results showed great variability and are averages over three different rounds. The discrepancy between Amarsingham's results (figure 6.7) and the results reported here may be caused by the fact that in our case parameters were optimised for performance and additionally, training was run for 500 rather than 300 cycles. We achieved better results with the random overlap case.

Amarasingham and Levy (1997) argue that the network performs better on overlapping sequences than on sequences with zero overlap if the total number of neurons in-

| Overlap | Capacity Levy | activity | Capacity shifted(Exp. 1) | activity | Capacity overlapped (Exp.2) | activity |
|---------|---------------|----------|--------------------------|----------|------------------------------|----------|
| 0 | 20 | 0.051 | 45 | 0.036 | 45 | 0.038 |
| 2 | 33 | 0.049 | 50 | 0.040 | 65 | 0.041 |
| 5 | 90 | 0.055 | 100 | 0.080 | 100 | 0.081 |
| 6 | 110 | 0.050 | 120 | 0.096 | 130 | 0.097 |

Figure 6.8: Capacity: results from experiments 1 and 2

volved in the retrieved patterns is taken into account. For example, for zero overlap, the capacity is 20, meaning that $20 \times 8 = 160$ neurons are involved in successful retrieval; for an overlap of 4, the capacity is 57, involving $4 + 57 \times 4 = 232$ neurons. However, for our revised results it is a matter of interpretation. For example, if the number of "free units" is taken into account, the network performs best on non-overlapping sequences. A sequence of zero overlap with 45 patterns, requires $45 \times 8 = 360$, whereas a sequence with overlap 6, and 130 patterns requires $6 + 2 \times 130 = 266$ neurons.

## 6.4.6   Quantitative investigation of performance

Figure 6.9 shows how the performance varies as the network exceeds capacity, for different values of the parameter $K_R$ which yield stable behaviour. The network performs better and at higher capacity with the lowest stable value of $K_R = 0.012$. The outlying points shown in red for $K_R = 0.011$ apparently have best performance but the patterns in the recoded sequence do not form a distinguishable sequence. These results illustrate the sensitivity of the performance measure to the closeness of the learned pattern to the original sequence and the importance of the network showing the correct dynamics. Tables of the results of training on sequences with different degrees of overlap are included in the appendix, in table 2 together with graphs that show information about the comparable structure of the two types of training sequences, in figures 2, 3, and 4. Note that, since the more recurrent inhibition there is, the lower the level of activity, activity is inversely related to $K_R$.
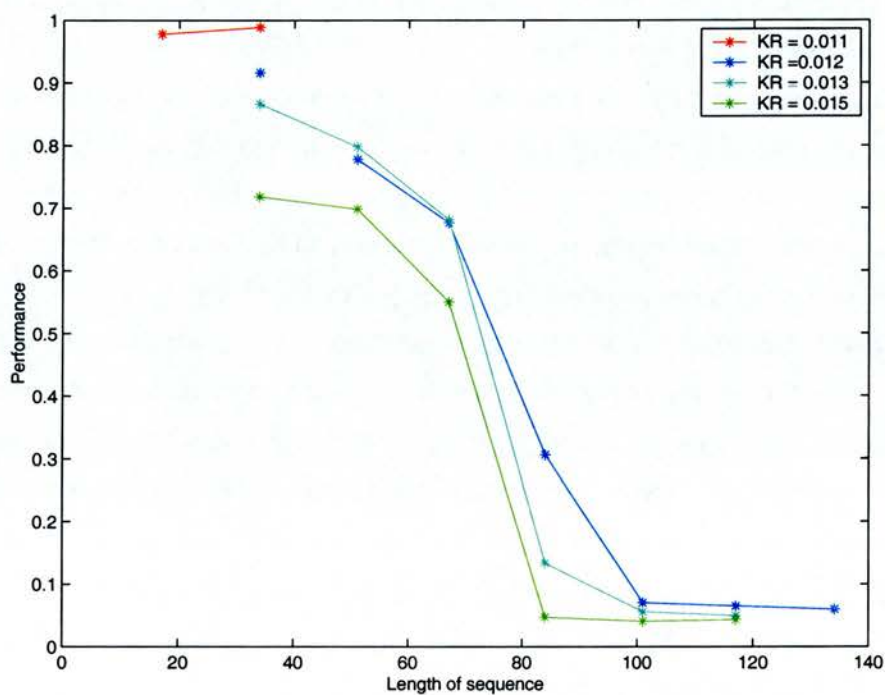
Figure 6.9: The figure shows the performance on sample sequences for an overlap of 2, over values of $K_R$ yielding stable behaviour. Outlying points in red show the cosine performance of the degenerate case for $K_R = 0.011$.

## 6.5    Performance and Capacity

The expression given in Wu *et al.* (1996); Amarasingham and Levy (1997) for estimating sequence capacity, $C$, is given in terms of $E$, the expected number of time steps on which a neuron is active and $a$, the fraction of the $N$ neurons active per time step.

$$C = \frac{E}{a} \tag{6.4}$$

Wu *et al.* (1996) state this relationship but do not derive it. On their assumption that during recall of the entire sequence, each neuron fires in a single burst of activity for $E$ successive time steps, the total number of occasions on which neurons are active during retrieval can be calculated in two ways: the total number of neurons active per time step ($aN$) multiplied by the number of time steps, $C$, giving $aNC$; or the number of times for which each neuron is active, $E$, multiplied by the number of neurons, $N$, giving $EN$. Setting these two quantities equal yields equation 6.4.

Their assumption is not justified; their own results show, for example, that some neurons are never active (figure 6.7). Figure 6.10 shows the actual capacity of the network compared with estimates using equation 6.4, for various different amounts of pattern overlap. The crosses show the actual length of the input sequence, and the circles are the expected maximum capacity from equation 6.4. This gives a spread of capacity estimates. The empirically determined maximum is in fact (from figure 6.8), roughly the lowest estimate.

### The effect of the level of activity

The data in figure 6.9 suggests that the network performs better at higher activities (lower value of $K_R$). Although this isn't guaranteed, it suggests that capacity and activity are not inversely related. We therefore examined the effect of sequence length on activity, as shown in figure 6.11, and then activity on performance, in figure 6.12. This confirms the trend for higher activity to result in better performance. The explanation is twofold. Firstly, activity is a *dependent variable*, and one of the things it depends on is the length of the input sequence. The more patterns that are stored, the more constrained is the recoded firing pattern. Secondly, the higher the activity of the recoded sequence, the more the network behaviour is controlled by its internal dynamics, and the fewer of the neurons in the original sequence have to be retained to give the same performance score.
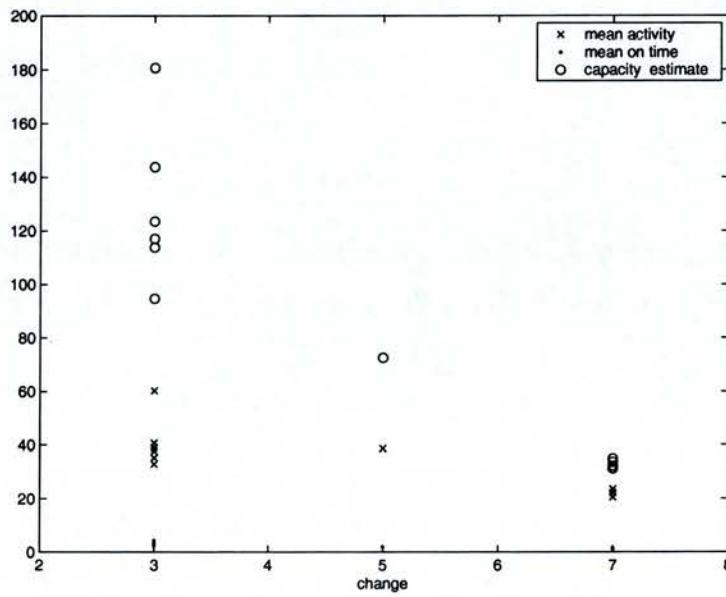
Figure 6.10: Estimation of capacity derived from sequences of length 40. For a particular values of the parameter $Ch$, the actual capacity (cross) is compared with the estimate (circle) from equation 6.4, using a mean figure for $E$. 6 runs were carried out for $Ch = 3$, 1 for $Ch = 5$ and 4 for $Ch = 7$.
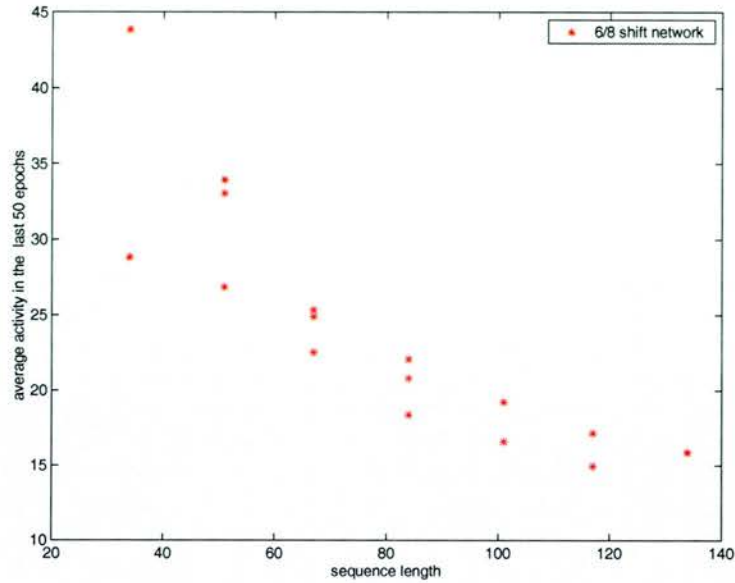
Figure 6.11: This graph shows how activity varies with sequence length, for networks run for 3 different values of $K_R$, with an overlap of 2, within the bounds of stable performance.

Equation 6.4 describes a simple inverse relationship between activity and capacity. One might expect that performance at recall might show the same relationship, as performance ought to be correlated with length. And in fact it is. But the relationship between capacity and length is not the relationship described in 6.4. Compare figure 6.12 with figure 6.9. There is clearly a nonlinear transition between recall and unsuccessful recall at the lower end of the activity scale.

A similar transition can be seen in the region of high activity. However, the sharp red line in 6.9 shows that this transition is not detectable in the measurement of performance. In each case, these transitions reflect changes in the underlying oscillations in the network.

## 6.6   Oscillations and performance

Amarasingham and Levy (1997) observe that the network oscillates, meaning that the activity level fluctuates widely, from time step to time step. They assume that "the
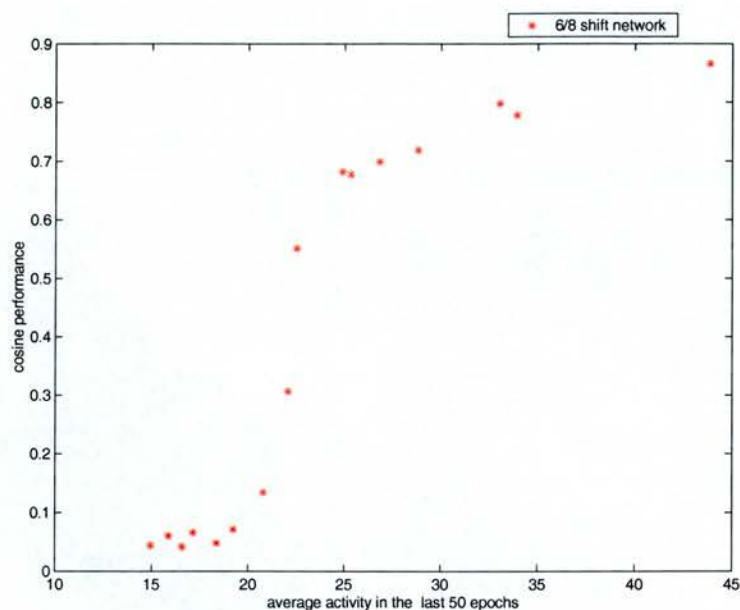
Figure 6.12: This shows the data used for figure 6.11, plotted to show the effect of activity on performance.

network's coding converges [stochastically] if the activity converges [stochastically]". The importance of convergence is that only once this has happened can one expect the learning rule to produce convergent weights. Amarasingham and Levy (1997) assume convergent weights.

Figure 6.13 shows an example of successful recall, accompanied by non-convergence of activity. Close inspection of the results of training over a long period reveals that the criterion of 75% for successful performance is reached by the 300th cycle, by which time the output is stable (green crosses, left hand diagram). Yet by the 300th cycle, the activity is still oscillating wildly.

In Figure 6.13 we compare three ways of evaluating performance on the task. The graph shows the similarity of the recoded sequence to the original sequence, alongside the similarity between the recoded sequence and the recalled recoded sequence. Each sequence was recalled twice, with the feed forward component threshold during recall set either to zero to the level during learning. Systematic comparison of the threshold showed the at the best recall generally occured with the threshold set at the value used during learning. An additional example is included in the Appendix, figure 6.
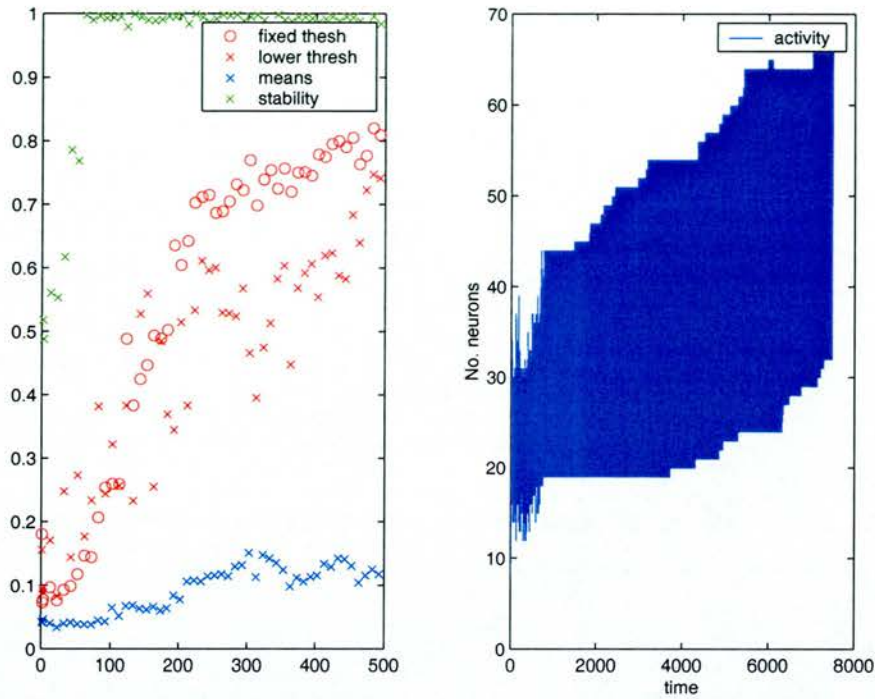
Figure 6.13: Network A: Sequences of length 40 and overlap 2 were stored over 500 cycles. Graphs show the evolution of performance (LHS) and activity (RHS) over the training epochs. On the LHS; Green crosses: average similarity between each output and the corresponding output on the previous cycle; red circles and dots indicate: performance on recall compared to the recoded sequence, for two different threshold schemes; *lower threshold*: Feed forward inhibition disappears during recall as $Z_i = zero$. *fixed threshold*: Feedforward inhibition treated as constant, set during recall to the value during learning. Blue crosses: similarity between the test output sequence and the original input sequence. On the right, neural activity over the $40 \times 500$ timesteps of the 500 cycles of training is shown.

The conclusion is that successful performance on this task does not require convergence of activity. Performance as calculated is summed over all patterns, thereby ignoring the possible variation between individual patterns. The next section shows that the underlying oscillations can be qualitatively characterised into four types.

## 6.7   Oscillations

There are four ways in which the number of active neurons changes over time. Here we classify network oscillations qualitatively. In networks of type A, figure 6.13 both the mean number of neurons and the oscillatory fluctuations from the mean drift over time; in type B the oscillations are unstable; in type C, the oscillations are stable but in recall the performance does not meet Amarasingham and Levy's criterion; in type D the oscillations are stable and do meet the criterion. Degenerate non-oscillatory behaviour is also seen where the same proportion of units (which may be 100%) are active each time step. These behaviours are not illustrated.

We are interested in how stability in learning and testing conditions, as measured by the performance criteria, compares with network stability. Ideally we want stable conditions, which are also convergent conditions, to correlate with good performance. In figures 6.14, 6.15, 6.16, (network B), oscillatory behaviour is illustrated that is unstable yet correlated with good performance. This network would have been classified as capable of successful recall at the 330th epoch. However, network C (figures 6.17, 6.18), run with sequence length $L = 101$ but otherwise the same parameters, produced a stable oscillation, but did not perform above criterion on recall. Finally, network D ($L = 111$; figures 6.19, 6.20) was both relatively stable and on average above criterion. However, at the 300th epoch it would have been classified as unsuccessful. The divergence in behaviour of networks B and D is not predictable from the parameter values.

These three examples serve to illustrate that the criterion employed by Amarasingham and Levy (1997) aren't nearly strict enough (even if the principle of assessing the recoded sequence rather than the original sequence is accepted). The large oscillations in the stable network suggest a reason why convergence may not occur. Although the sums over all weight change within an epoch may (in fact does) converge to zero, the

modulus of the sum may be quite large.

We conclude that the stability of the network is not predictable from input parameters When learning is permitted, this is a response to the particular connection matrix and the demands of the input sequence, both of which vary randomly. Smith and Levy (2000) assume normally distributed dendritic sums and considers the network as a dynamic system, and they show if the oscillation remains bounded between one standard deviation of the expected mean activity, the network converges to a stable oscillation. Although they show the relationship between mean activity and inhibitory parameters in the case without learning, they do not show a relationship between the bounds of the activity and parameters. Nor do they show that the relationship between mean activity and parameters holds for the case of learning.

One implication of these results is that the learning rule cannot 'learn' the conditional probabilities expected, if the network does not stabilise, as those probabilities do not stabilise. The results shown in the next section confirm this.

## 6.8   Distribution of weight values

A realistic weight distribution in a pyramidal, hippocampal neuron should be either flat or approximately normal. Amarasingham and Levy (1997) and Smith and Levy (2000) claim that the learning rule produces an approximately flat weight distribution. However, the weight distribution found empirically looks bipolar, as shown in figure 6.23.

In the situation when the network apparently converges, does the learning rule then converge to produce the expected weight values? We calculated the expected weight values given the final recoded sequence for two cases when performance had reached criterion. Figure 6.21 shows the histograms of the distributions of the expected weight values, for each particular weight value. If each weight value equalled its expected value, then all points should lie on the diagonal. 6.13 shows the weight distributions at 800 epochs.

Figure 6.22 shows the typical distribution at the end of a 500 training cycle. To confirm convergence of weight values, the network in 6.21 was trained for a further 1000 epochs, in 100-epoch slices. Weight convergence did not improve.
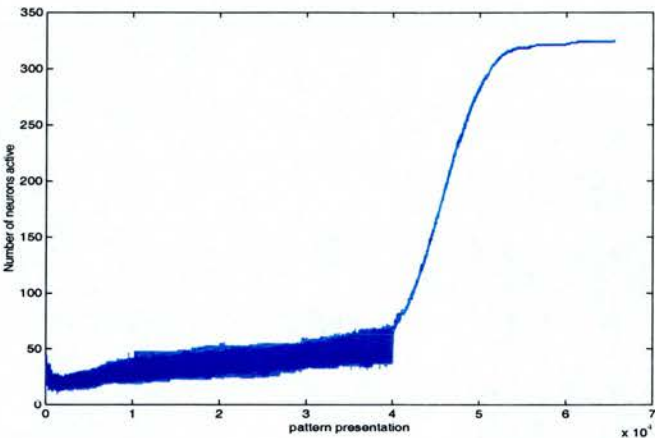
Figure 6.14: Network B: Activity shown over the entire training time. Parameters: $K_R = 0.015$, $L = 131$, overlap $= 6$, $40 \times 131 = 65500$ time steps. A sharp transition between a relatively stable oscillation, and a stationary state is seen. This network would have just failed to have been classified as successful by Amarsingham and Levy, reaching their criterion at the 330th epoch; see figure 6.16.
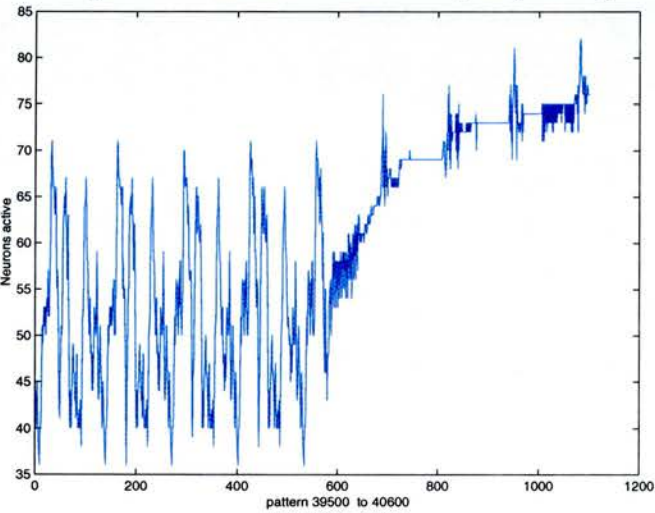


Figure 6.15: Enlargement of figure 6.14 for pattern presentations 38500 to 40600.
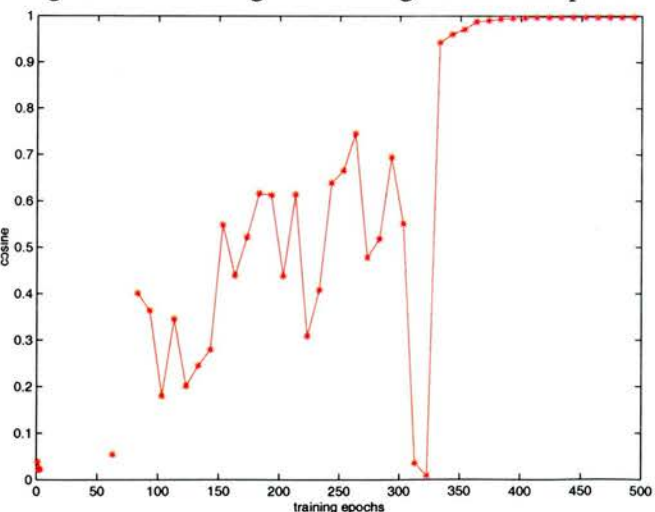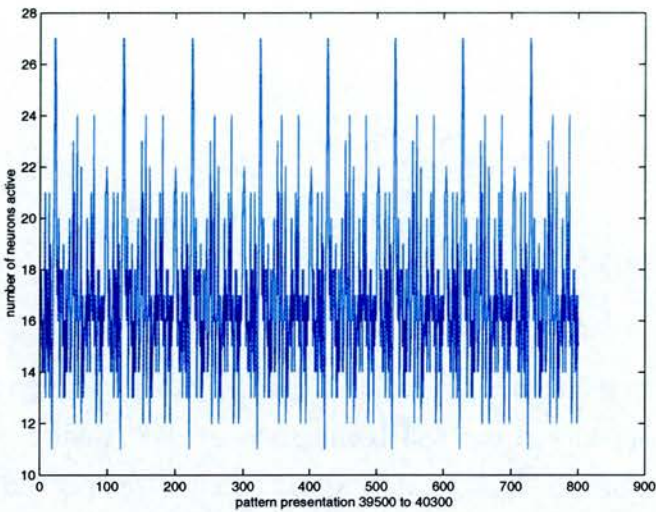
Figure 6.17: Network C: This figure is an enlargement from a stably oscillating network Parameters are $K_R = 0.015$, length = 101, the same parameters as above, at the same time step that network A showed the transition to a stationary state.
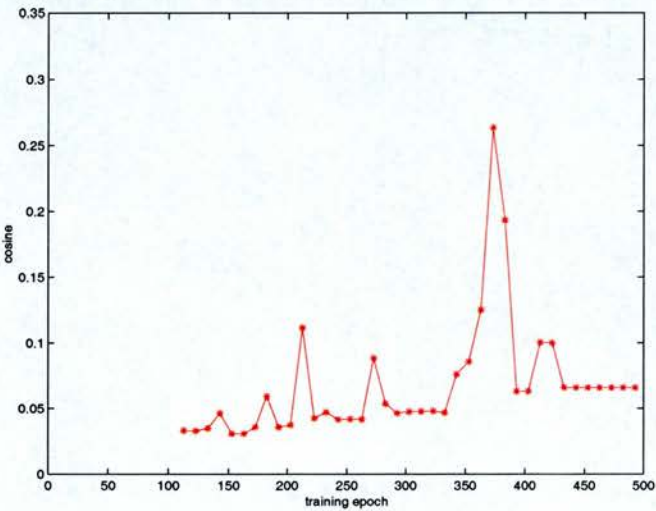


Figure 6.18: Network C, while stable, would have been classified as unsuccessful by Amarsingham's performance criteria.
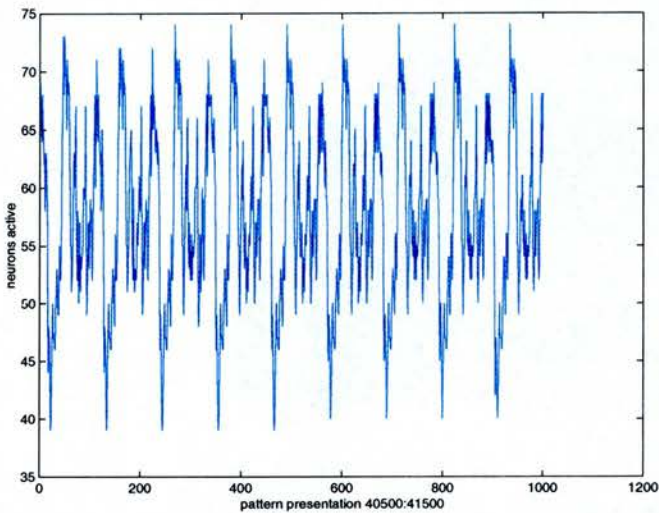
Figure 6.19: Network D shows a stable oscillation.Parameters are $K_R = 0.015$, length = 111, overlap = 6
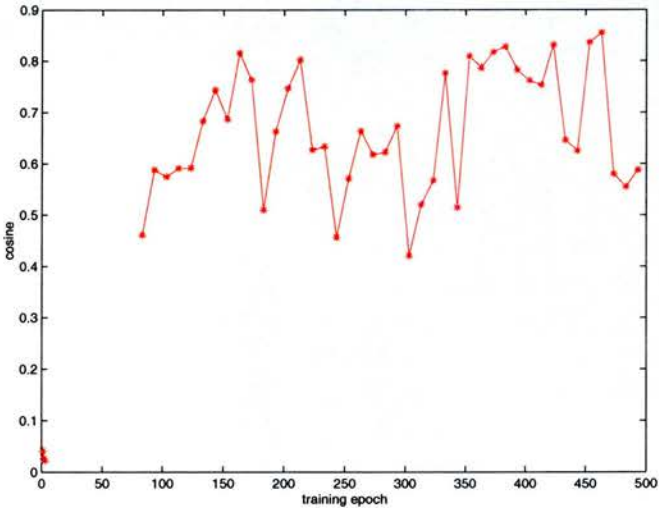


Figure 6.20: While Network D shows a stable oscillation, and satisfies the recall crite-rion, it does not do so for every oscillation. In fact at the 300th oscillation, it is below performance criterion, and would therefore have been classified as unsuccessful. Pa-rameters are $K_R = 0.015$, length = 11, overlap = 6
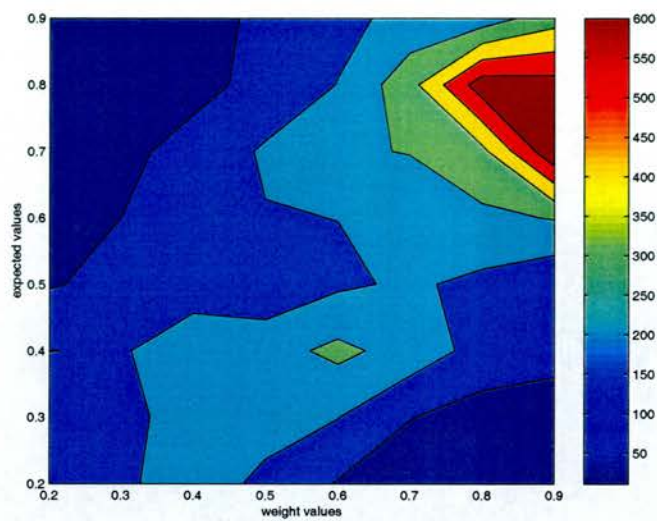
Figure 6.21: Expected weight values versus actual weight values I : Contour plot of the frequency histograms of the expected values of weights of a particular actual value. In this case, convergence of weights and expected values is approximately correct.
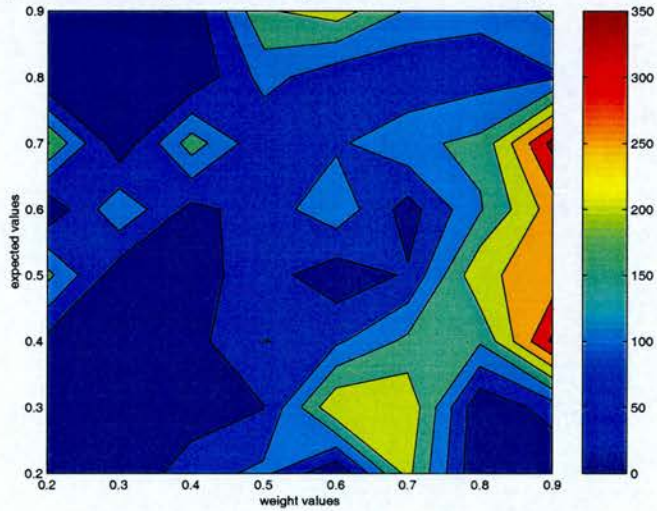


Figure 6.22: Expected weight values versus actual weight values II: Contour plot of the frequency histogram of expected weight values, of weights of a particular actual value, for a network that is apparently in stable oscillation. In this case, weight convergence is very poor.

Reproduced from Amarasingham and Levy (1997)
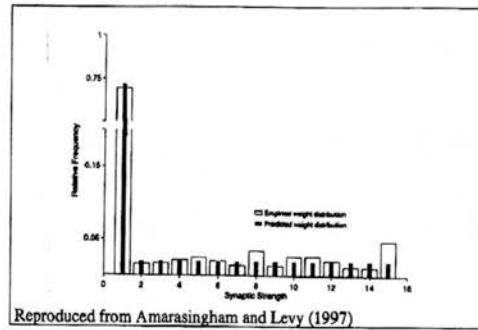
Figure 6.23: Weight distribution that is modelled as flat, rather than as bi-polar by Amarasingham and Levy (1997)

The conclusion is that good performance on the task does not require convergence of either weights or activity.

## 6.9 Conclusions

Amarasingham and Levy (1997)'s analysis is based on the assumptions that the weights had converged and the weight distribution is flat. It turns out that in the cases they considered the weights had not converged. We found that when there was convergence the weight distribution was not flat.

The implications of the replication study above are as follows.

1. Convergence is not necessary for the sequence recall task.

2. This is fortunate, as this study shows that convergence, which has been claimed, does not take place.

3. Lack of convergence means that the desired conditional distribution is not represented in the weight matrix.

We noted at the beginning of the chapter that Levy's model was an attempt to marry idea of the sequence recall with that of prediction. We also noted that this attempted to combine two inherently conflicting goals. However, if the network is not required to learn probabilities, the lack of convergence may not be very important, particularly

if we view the hippocampus as a temporary memory store, which is thus necessarily unstable. Part of the instability is caused by the wayward activity dynamics, caused by activity based thresholding. The next chapter will consider a solution to this problem.

The remainder of this chapter will consider two other aspects of the model:

1. whether introducing a more structured weight matrix has any effect on the model;

2. whether sequence disambiguation really is a good model of transverse patterning.

## 6.10   Changing the weight matrix

Ishizuka *et al.* (1995)'s data (see chapter 3, section 3.4.4) indicates that the CA3 region of hippocampus is constructed more like a densely woven rug incorporating several layers than the random 'conjunction' box of the associative net-type architecture. We have noted that temporally situated neurons tend to project septally and septally situated neurons project temporally.

It is an assumption of the model that the network connection matrix is asymmetric. In some circumstances, symmetric and asymmetrically connected attractor networks can show very different dynamic properties. If the zones of maximum probability of connection are roughly the same in both directions, as it appears from Li *et al.* (1994), then the probability of symmetric connection becomes high, even if the connections are sparsity of connections remains the same. The increased likelihood of reciprocal connection lead us to experiment with the tolerance of this kind of network for symmetrical connection.

The crucial dynamic behaviour shown by the network is that of aperiodic oscillation. As we saw in the previous section, periodic oscillation, where the network oscillates stably through a series of patterns, does not necessarily converge to an oscillation through the patterns in the learned sequence. Increasing the proportion of symmetrical connections allows the periodic behaviour to increasingly dominate the networks behaviour. The dynamic behaviour of periodic and aperiodic oscillation is described in Minai and Levy (1993) for fixed weights, controlled by the parameter settings that control the relative strengths of the excitatory and inhibitory connection strengths. The

learned sequence is imposed on these underlying dynamics, most successfully when behaviour is aperiodic. Minai and Levy (1993) identify the ratio of excitatory and inhibitory inputs as critical.

### 6.10.1    Experiment 3: symmetrical connection

For networks trained on an overlapping sequence of 25 patterns, we ran a parameter search for the asymmetric weight matrix. We then ran a series of trials using stable parameters taken from the middle of the stable range, for different proportions of symmetrical weights. Construction and training of networks was performed according to Algorithm A in the Appendix, with modifications described in the section in the Appendix *Experiment 3*. Figure 6.24 indicates that the proportion of symmetrical connections that disrupts this behaviour is much higher than the 10% of the connections being symmetrical which represents the amount of assymetry in a network with 10% connectivity if the connectivions are distributed at random.

The issue of the extent of symmetrical connection in CA3 is one that must resolved empirically.

This experiment does not show conclusively that this type of aperiodic oscillatory behaviour cannot occur in symmetrical networks. It does however suggest that the proportion of symmetrical connection is a critical parameter in dictating the behaviour of any particular network. Li *et al.* (1994)'s observations also suggest another intriguing possibility. This is that some of the observed boutons are silent. Here the idea is that the network is sparsely connected in the sense that only a few connections are active at any one time, but each neuron has much larger set of potential connections that it can activate; ie, these symmetrical connections do exist but are silent.

### 6.10.2    Conclusions

This section has considered oscillatory behaviours in this neural network in their own right. However, it is extremely unlikely that CA3 ever shows these kinds of spontaneous oscillatory behaviours. In the previous chapter, we discussed the nature and biological control of oscillations in CA3; none of them is generated by the spontaneous firing of pyramidal cells. Models of these oscillations, including this one, have
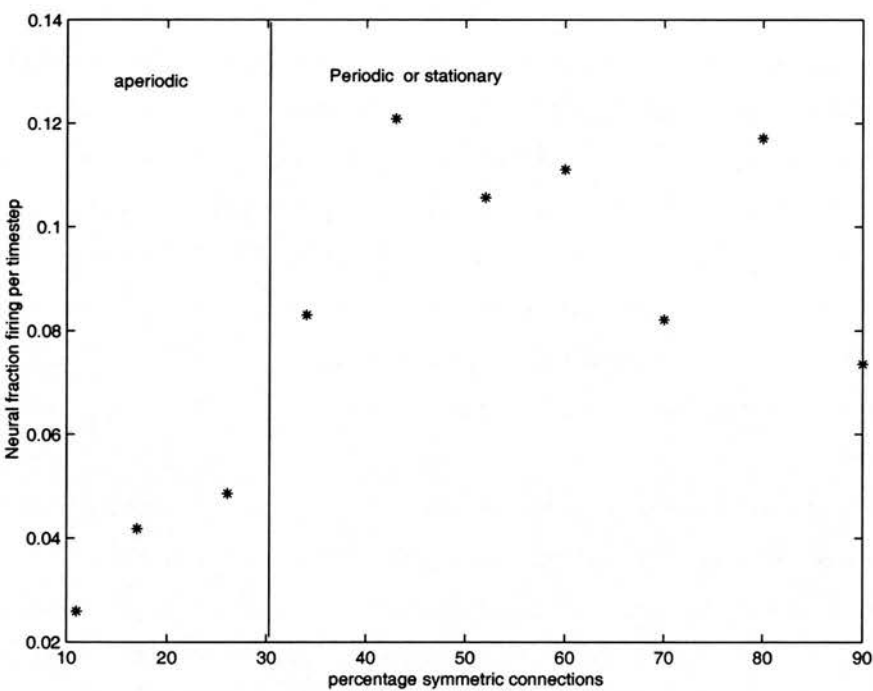
Figure 6.24: The boundary between aperiodic and periodic behaviour for networks trained at constant parameter, connection density: 10%, percentage of symmetric connections as shown

all treated sequence recall as a property to be imposed on the underlying oscillations. The next chapter will consider sequence storage and recall without oscillations. It turns out that storing sequences is easy, whereas stabilising oscillations is much harder.

## 6.11   Learning more than one sequence

The problem of learning and recalling more than one sequence has been approached using the sequence disambiguation problem as a model for certain kinds of configural learning tasks. In particular the negative patterning task in Wu *et al.* (1996, 1998) and the transitive reasoning task in Smith and Levy (2000); Levy (1996).

The possibility of modelling these configural tasks as sequence learning tasks has been treated as indirect support for sequence learning models of the hippocampus. We argue here that the sequence learning version of the negative patterning task does not discriminate between the hippocampal, and nonhippocampal versions.

### 6.11.1   Multiple sequences

An experiment was performed to tested the performance of the network on disconnected sequences to see how well a sequence survives when trained on a new episode. A sequence of randomly selected patterns was divided into five separate input events, and then each input sequence of ten 8 patterns trained and recalled separately. The experiment was performed at five different values of $K_R$.The accuracy of recall from the network is shown in figure 6.25. The most recently stored sequence, shows the best recall and the sequence stored first was recalled least well. This suggests that the degradation of recall with further learning is a general property of the LTD, rather than an effect of low capacity. A recency effect is also reported in Greene *et al.* (2002). Greene *et al.* (2002) also claim to find a primacy effect at high levels of activity, but in an overall context of poor retention.

These results suggest that the storage of separated sequences reduces the capacity of the network. Greene *et al.* (2002) overload their network . They argue that this is a realistic model of memory in list learning, as list recall and recognition tasks aim to overload working and medium term memory. Although our results are consistent with Greene *et al.* (2002), in our data at most activities sequences 'aged'. Were this an
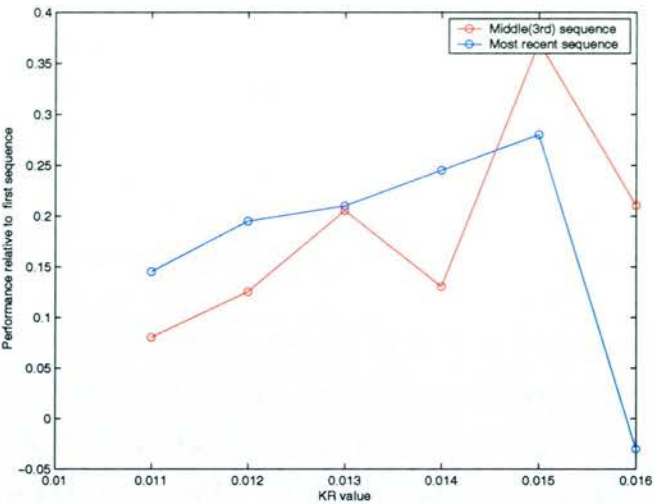
Figure 6.25: Recall of the first sequence was poor at all values of $K_R$, remaining within 0.4-0.5. The graph shows the relative performance of the middle and most recently stored sequences. All sequences improved slightly as $K_R$ increased, until the final data point where all became suddenly worse, apart from the middle sequence.

accurate model of episode storage, it would be psychologically possible to judge the recency of an event by the clarity of the memory! Recency effects in list learning are usually limited to the last few, or last items.

The next section discusses how the sequence disambiguation problem has been used to model configural learning tasks. We will also consider whether there serial position effects in an underloaded situation. To do this, we invesitigate the effects of string and recalling sequence in just one trial

We first discussed the configural learning problems in chapter 1, section 1.12.4, as animal models of 'relational memory'. Wu *et al.* (1998) present a version of the sequence disambiguation problem constructed to resemble the negative patterning task, as described in Alvarado and Rudy (1995).

| A+ vs B - |
| B+ vs C- |
| A- vs C + |

This is essentially a sequence disambiguation problem with six sequences, shown in Figure 6.26

Clearly, this is not only a model of configural learning, but an implicit theory about

the role of sequence recall in action selection. Both rewarded and unrewarded action selections contribute equally to the established memories. Errors during learning are entirely due to interference between the recall of learned sequences with overlapping elements. It seems unlikely that in the experimental case these memory traces in the hippocampus are the only memory traces contributing to the hippocampus. Wu *et al.* (1998)'s method of learning the problem is to *stutter*, or repeat each input pattern three times. This doctoring of the input is necessary as they claim that otherwise the network cannot form " context units" and therefore cannot learn the problem. We interpret this rather ad hoc solution as a solution to the problem of controlling activity dynamics in the network. In section 6.5 we mention that activity is dependent on length( or , more accurately, the number of modified synapses in the network, pattern set pair- see next chapter for explanation). If the sequences are too short, the network doesn't function very well. Intuitively, this is because learning a small number of very sparse patterns is not sufficiant to move the network into a region of stable activity.

However, there is a difficulty in interpreting this task. Wu *et al.* (1998) emphasise the importance of this being a study of context dependent learning. They then equate context dependent learning with learning to discriminate on the basis of recent past. In Wu *et al.* (1998)'s problem, their representation is more akin to the presentation of slightly similar inputs for discrimination. Stimuli A and B form a complex stimulus, as do B and C, A and C. Discrimination based on complex stimuli formed by conjunctions of stimuli is not necessarily hippocampally dependent. In chapters 1 and 2 we discussed the 'feature' negative patterning task, where the rat has to use a context feature to discriminate the cases where A and B are positively reinforced. Wu *et al.* (1998) treat the reinforced and unreinforced training instances as equivalent.

```
X -> A+
   A-
X->B -
   B+
```

This version of the task is not hippocampally dependent.

There is also an important distinction to be made between an ordinary conditioning task, and conditioning task plus context. Philips and LeDoux (1994) found that dorsal hippocampal lesions affected conditioned freezing (a measure of the rat's expectation of footshock) to context, defined as the box in which conditioning took place, only
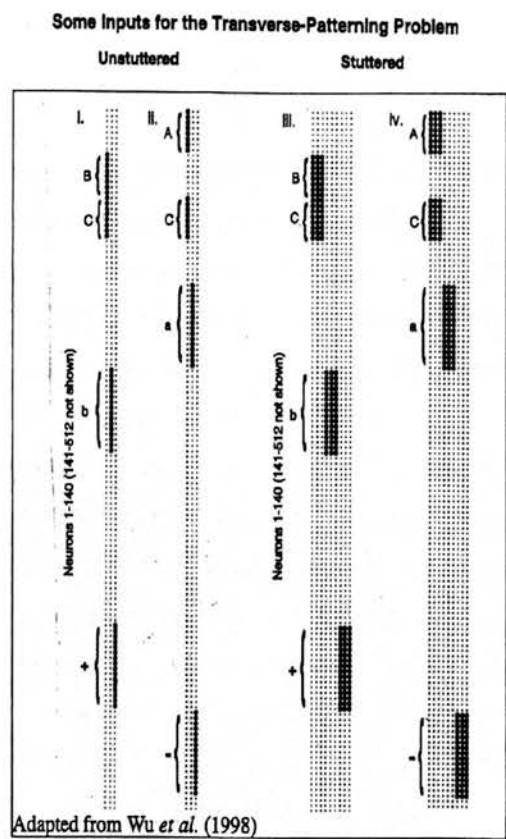
Figure 6.26: Sequence representation of the transverse pattern task

when it was the background to a conditioning CS and US pairing within the task. When the background was there in the *foreground* of the task (i.e., the direct stimulus) paired with a single element conditioned stimulus, this conditioned freezing was unaffected by hippocampal lesions. Straightforward stimulus response, or fear conditioning, is not hippocampally dependent. This experiment reinforces the point that one cannot necessarily treat any stimulus that provides means of discrimination as context. It also cannot be assumed that interfering contexts necessarily have a detrimental effect on task performance. In recall tasks studied by Gaffan and Parker (1996), interfering contexts in fact removed them from consideration as relevant stimuli, thereby improving the performance of intact animals on a hippocampally dependent task (Gaffan and Parker, 1996).

However, in chapter 2 we saw that there was an effect on recall of the proactive interference that was trial interval dependent. The negative patterning task showed effects of interference from the previous trial, that we characterised as proactive interference. This effect is hippocampally dependent. In chapter 2 we argued that this was best explained as an example of recall from the previous trial acting to bias recall in the next trial. In chapter 1 we discussed the importance of distinguishing effects of recall from acquisition in these sorts of tasks, but observed that in practice it is impossible to prevent new acquisition affecting a so-called test trail. In this section we follow Levy's reinterpretation of the negative patterning task, and show that what appeared to be an effect of recall, in Han *et al.* (1998), can be modelled as an effect of learning, provided we assume that the rat continues to learn during the recall phase.

## 6.11.2   Disambiguation

Experiment 4.1 tested the network's ability, in one trial learning, to correctly disambiguate two sequences with an common middle section. Two sequences which shared three patterns in a sequence of eight were trained and performance was compared with that in a control task involving a pair of independently constructed sequences. The networks were tested, in 20 batches of 25, on these two versions of the task. Firstly, a parameter search was run to check the parameters for performance on a continuous sequence of length 18 (see also table included in Appendix A). Then the network was trained on the first sequence and then tested on the recall of it. After training the net-

work on the second sequence, it was retested on the first sequence. Assuming that the requirement for success on the task is disambiguation, the number of patterns in the part of the sequence following the common section that failed to be retrieved was recorded (figure 6.27). Failure could occur in two ways: failure to recall any pattern, or recall of the incorrect pattern.

The underlying failure to learn the sequence was taken into account for both first and second trials, and was not different in the two cases. The results in Figure 6.27 were generated by subtracting the underlying failures of learning from the failures of recall. The bars show the distribution of number of failures of recall across the 20 sets. Failure is defined as failure to recall one or more patterns of the sequence.

This task was designed to ressemble the task described by Han *et al.* (1998) but without representing reward contingencies specifically. The comparison depends on the observation that it is the learning of the one half of the task that is impairing recall of the other half.

This system thus possess the same qualitative behaviour as shown in the Han *et al.* (1998) experiment, thus supporting the idea that learning of the previous test trial is interfering, or biasing recall of the second trial. However, this effect is partly accounted for by the relatively low capacity of the system. The other contributing factor is probably the fact that since the amount of LTD is proportional to the size of the synapses, LTD caused by the overlap will have a relatively large impact on the synaptic weights.

The same behaviour is seen in decay across recall of sequences in the context unit network. This is therefore a general property of the recall in the network with LTD that does not converge.

### 6.11.3   Conclusions

The conclusion to be drawn from these experiments is two fold. Firstly, it is possible to create proactive interference in models, However, this may be an artifact of this particular learning rule, As we have seen, the Bi and Poo (1998) data suggested that LTP was a proportionally much larger effect than LTD.

This does highlight, however, the need to keep the recall of previously learned information apart from new learning. We know from recent studies on re-consolidation in the amygdala that such information may not be kept separate. However, the hy-
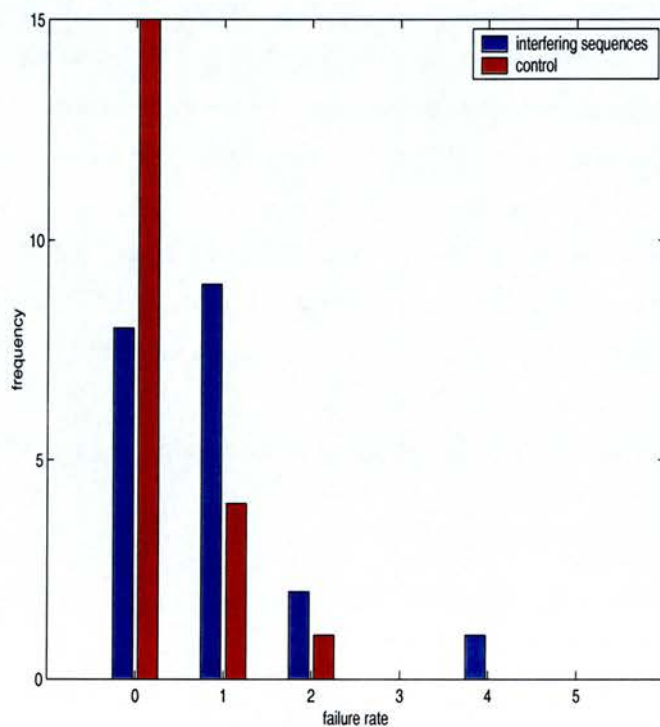
Figure 6.27: Failure rates on two overlapping versus two random sequences. This figure shows a frequency histogram of failure rates, as defined in the text, for the first stored sequence of the pair when retrieved on the second test trial compared to the first trial. Two cases are shown: interfering sequences with an overlapping section (blue); control case where the sequences were constructed randomly. There was zero failure rate for the second sequence. Mean of initial weights: $0.03$, $K_I : 0.014$, $K_R : 0.014$, $\theta : 0.8$, $e = 1$

pothesis developed in chapter 2 is that this may be the special contribution of the hippocampus, especially the dorsal hippocampus.

## 6.12   Summary of chapter

This chapter shows that sequence disambiguation may be performed in a sequence learning network without any special reliance on physiological features such as theta rhythm modulation. This tends to invalidate the arguments made in some of the more biological models, e.g., Sohal and Hasselmo (1998), that something like the theta rhythm, or acetylcholine modulation is necessary.

However, we also saw that the learning rule studied here does not produce the desired convergence in this architecture, probably due to unlearning through LTD, and through unstable dynamics. Note that in our simulations, the parameters used constrained the activity dynamics.

In the next chapter we shall consider what the minimum requirements are for a model of sequence learning.

# Chapter 7

# Sequence learning in the hippocampus: Part II

## 7.1 Introduction

The model described in the previous chapter, although it is presented as a simple conceptual model, is not the simplest one. For the purposes of comparison, this chapter describes a simpler model, the Associative Net, a heteroassociative matrix memory (Willshaw *et al.*, 1969; Willshaw, 1971) with clipped weights. Although this is normally used for a cued recall task it can be employed as a sequence recall device when trained on a special set of a patterns where each recalled pattern acts as the cue for the next pattern, a preliminary investigation of which is given in Willshaw (1971).

This simple device works surprisingly well. The second part of this chapter investigates how performance is affected by the use of real valued learning rules. Performance with the learning rule and threshold scheme proposed in the Levy model is evaluated against the simpler architecture.

The model is tested on its ability to learn and recall sequences over one trial, as one trial learning is closer to the real situation in the hippocampus than using an algorithm that takes many iterations.

We are interested in hippocampus-like models, and therefore we are interested in the learning of sequences in partially connected networks. It is impossible for the sparse networks considered in the previous chapter to learn and recall small size pat-

163

terns in a one trial situation, as the probability of neurons hitting modified synapses is too low. Evidence was also presented in chapter 3.10 that sparsity levels in CA3 may have been over exaggerated. That chapter discussed the existence of patches of very dense connectivity, between 50 and 70%. We suggest that a patch of CA3 (dorsal, perhaps particularly proximal) is particularly suitable for sequence recall, so that sequence order information can be passed to other parts of CA3. Proximal CA3 has the local recurrence consistent with these memory models

The problem of recalling patterns from fully connected associative network type devices has been well studied for binary and linear Hebb rules (Willshaw and Dayan, 1990; Dayan and Willshaw, 1991).

## 7.2   Matrix memory

The matrix memory shown in figure 7.1 stores the outer product of two vectors with binary valued components on a grid of on/off switches. Each weight takes the value 1 when the corresponding components in the two vectors are 1, thus storing the correlation between the vectors. This matrix of on and off switches stores a binarised superposition of outer products. To achieve recall, for each column of the matrix, representing one component of the required output, the scalar product of the input vector ('cue') is formed, with the weight column vector and then the columns are summed. The sum is divided by the number of 1's in the input. The network outputs 1 when this fraction is greater or equal to 1 and outputs 0 otherwise.

The simple fully connected associative network of 1000 by 1000 weights, storing pairs of patterns and using one pattern as a cue, can store up to about 6955 patterns. An approximate capacity calculation is given by Willshaw $et\ al.$ (1969) where:

$R$ is the number of pattern pairs

$M_a$ the number of active units out of $N_a$ per input pattern

$M_b$ the number of active units out of $N_b$ per output pattern

$\rho$ is the probability of of synaptic modification

$$\rho = 1 - \left(1 - \frac{M_a}{N_a}\frac{M_b}{N_b}\right)^R$$

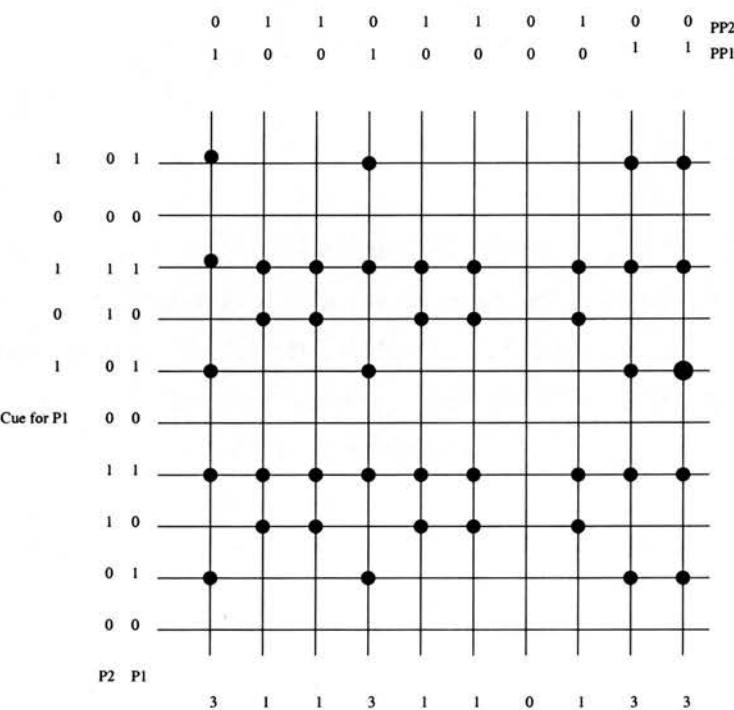The number of pattern pairs that corresponds to a particular loading is

Figure 7.1: Pattern pairs [P1; PP1] and [P2;PP2] are stored in the network. A partial cue (part of P2) is presented, and the output is divided by the number of elements in the cue, resulting in the successful recall of PP2.

$R = -\frac{N_a N_b}{M_a M_b} ln(1 - \rho)$

The error free limit on the value $\rho$ is defined as

$N_b \rho^{M_a} = 1$

Hence for networks of dimensions $N_a = N_b = 1000$,

Setting $M_a = 10$ (close to optimal, calculated on information theory grounds)

$R = \frac{-1000^2}{10^2} ln(1 - .5012)$

$= 6955$

Setting $M_a = 30$

$R = \frac{-1000^2}{30^2} ln(1 - .7943)$

$= 1757$

These are only approximate figures, as they do not take into account the fact that the expected probability of synaptic modification, $\rho$, has a distribution that depends on the actual number of times a unit is active in a particular pattern set, not the expected unit usage (Buckingham and Willshaw, 1992). This dependence is given as

$\rho_i = 1 - (1 - M_a/N_a)_i^r$ where output unit $i$ is active on $r$ occasions.

To turn this into a sequence recall device, the output has to be returned to the input via a set of one to one connections (Willshaw, 1971). The condition for this to work is that the recurrent feedback is silent during learning.

## 7.3    The sequence recall strategy

The sequence recall strategy is a good recall strategy in circumstances where the network is required to do not just pattern completion, but also prediction, ahead of the changing external cues. However, it is a fragile strategy as the introduction of an error may cause the remainder of the sequence to be lost. Figure 7.2 demonstrates the fragility of this strategy, and that sequential recall works only when the network [1] is underloaded.

One can think of this as impacting on the evaluation of the information efficiency of

---

[1]Further details of the experimental procedures used for training and testing networks in this chapter are described in the Appendix in the section *Experiments in Chapter 7*
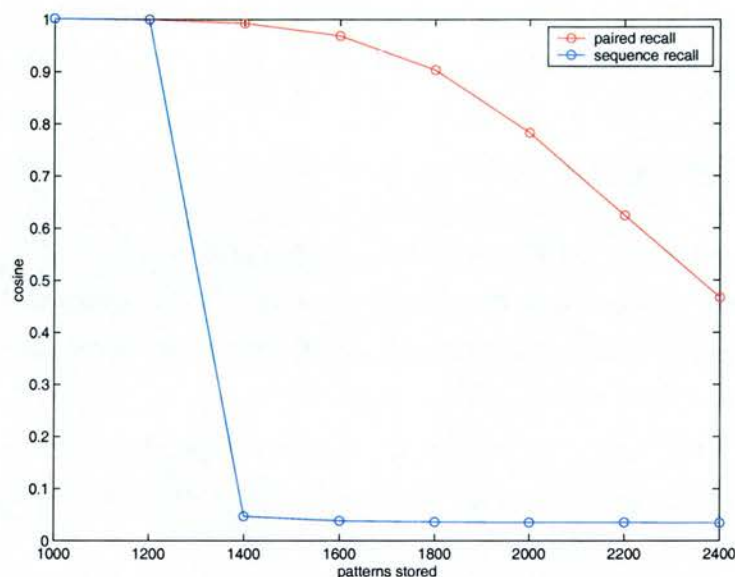
Figure 7.2: The fragility of the recall strategy. Paired pattern recall is compared with sequential recall of the same pattern set. $M_a = 30$, patterns are selected randomly

such a device. We know that the capacity of the basic (linear associative modification, and binary valued) associative network decreases exponentially with noise in the cue. Therefore for some level of external noise, sequential recall is going to be a better strategy. However, this relies on the lack of noise during learning. Such a case pushes us back into the issue of network dynamics. In the network of the previous chapter, some random variation, or noise, was created by instabilities during learning. We argued in the previous chapter this was partly created by the global threshold that varied with the activity of the previous time step, and therefore lagged the actual activity in the network. Winners take all thresholding, used in this chapter, eliminates noise during learning.

Many of the problems faced by the sequence recall devices are already familiar from other pattern completion devices. A pattern completion device such as the heteroassociative network requires a weight matrix and a firing threshold. The issue of how to set this threshold for each unit is already a well described problem (Buckingham, 1992; Buckingham and Willshaw, 1993; Dayan and Willshaw, 1991; Willshaw and Dayan, 1990; Graham and Willshaw, 1995, 1997) to which a number of solutions are available.

The aim of this work is to compare the biologically inspired thresholding strategy of the previous chapter with strategies that have been developed for theoretical reasons.

## 7.3.1   Winners-take-all (WTA)

The simplest thresholding strategy for a network of this type simply lowers the threshold until the required number of neurons turns on in the output. This is frequently assumed to be a reasonable approximation to activity-based thresholding, of the kind deployed in the hippocampal model in the previous chapter.

A network of the same dimensions as the network discussed in the previous chapter ($M_a = 8, N_a = N_b = 1000$) was trained with winners-take-all thresholding. The weights were trained incrementally with three different learning rules:

**Clipped** The Willshaw rule mentioned in the introduction Willshaw *et al.* (1969), which is a binary hebb rule sometimes referred to as clipped learning;

**Unclipped** A linear real valued hebbian rule, where the size of the weight is proportional to the number of correlations stored,
$$\delta W_{ij} = Pre * Post;$$

**Presynaptic** A rule similar to the post synaptic rule used in Amarasingham and Levy (1997), except that modification takes place only when the presynaptic neuron is active (Willshaw *et al.*, 1996),
$$\delta W_{ij} = Pre * (Post - W_{ij})$$

As each weight used in the clipped rule contains only one bit of information, using a real valued rule means that potentially more information can be stored in the weights. However, the question is how successful this sequence learning device will be in utilising this information.

Figure 7.3 shows the results of storing up to 6000 patterns with binary-valued components using the WTA threshold and these three learning rules. The relatively poor results with unclipped and presynaptic rules are due to WTA thresholding being unable to exploit the information stored in the weights to discriminate learned patterns from unlearned patterns. We can demonstrate this by measuring the potential degree
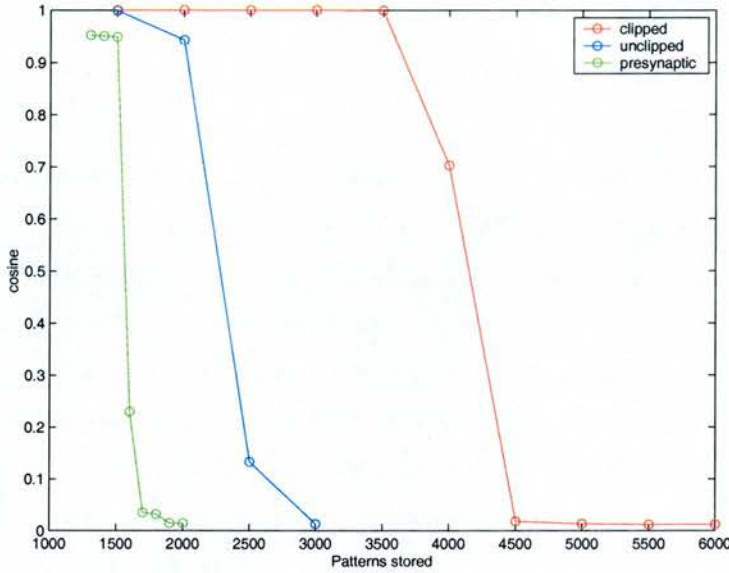
Figure 7.3: Performance of the three types of hebbian learning with WTA thresholding, using a sequence recall strategy. $M_a = 10$

of discrimination in terms of the signal to noise ratio (here called $\zeta$), which sidesteps difficulties with setting thresholds.

Signal to noise is calculated from the two distributions of dendritic sums (the dendritic sum is the scalar product of the input vector with the weight vector calculated for each unit of each output). One distribution relates to the cases when the output unit should be active ('high'), one to the case when it should be inactive ('low'). Signal to noise is equal to the squared distance between the means of the two distributions divided by the average variance.

Signal to noise, using standard symbols, is given by

$$\zeta = \frac{(\mu_h - \mu_l)^2}{1/2(\sigma_h^2 + \sigma_l^2)} \tag{7.1}$$

This analysis suffers from the fact that in the clipped rule the two distributions are markedly different. The high distribution has zero variance, leading to a low average variance and therefore a high signal to noise. The distributions for the presynaptic rule are different from these and it is therefore difficult to say which rule performs better. A modified version of average variance uses a term that approaches the low distribution variance as the variance of the high distribution approaches zero, as given

in Buckingham (1992), which we use here.

$$\zeta* = \frac{(\mu_h - \mu_l)^2}{\sigma_h^2 + \sigma_l^2 - \sigma_h\sigma_l} \tag{7.2}$$

Two different ways of measuring signal to noise are discussed in the associative network literature - see especially Dayan and Willshaw (1991); Minai (1997). One way is to calculate dendritic sums for each unit individually (Dayan and Willshaw, 1991). This measure is appropriate where each unit fires independently. The second way is to measure the discriminability of the units for a given pattern, which reflects the firing condition in a winner takes all situation. Neither method gives the expected probability of error, which depends on the shape of the distribution and on the threshold.

Signal to noise ratios for 1500 stored patterns with $N_{a,b} = 1000$, $M_{a,b} = 10$ are.

| Rule | $\zeta$ | $\zeta*$ | error | limit | performance |
|---|---|---|---|---|---|
| unclipped | $66.7(calculated)$ | $45.5 \pm 0.5$ | 0 | 1880 | 0.85 |
| clipped | | $91.5 \pm 0.5$ | 0 | 4000 | 0.70 |
| presynaptic | $46.5 \pm 0.5$ | $23.3 \pm 0.5$ | $194 \pm 23$ | 1600 | 0.22 |

For the unclipped rule, the expected value of the signal to noise, $\zeta*$, is shown only, and is calculated from Dayan and Willshaw (1991). For the other two rules, the limit (point where the recall strategy collapses) and the performance at that limit are also shown in figure 7.3. The value of $\zeta*$ for the presynaptic rule, while less than for the clipped rule, is sufficiently high to suggest it should perform better than it in fact does.

## 7.3.2   Sources of error

The difficulty with a winners-take-all thresholding strategy is that it relies on the high units having numerically greater dendritic sums than the low units. For any pattern set, the distribution of dendritic sums will, as already mentioned, reflect the individual distributions of unit usages. Although we can say that an output unit is expected to fire $R(M_a/N_a)$ times, we cannot know in advance how often a unit will actually fire. A low unit (a unit that should be inactive for a particular recall situation) that fires often in the storage of all the patterns presented will have a high dendritic sum, which may be similar to that of a high unit (a unit that should be active in a given situation) that fire infrequently. This is particularly the case for the unclipped rule, as weights can

grow arbitrarily large. This may therefore account for the relatively poor performance of the unclipped rule. The LTD component in the presynaptic rule might be thought to overcome this problem as it contains both saturation and LTD. However, LTD reduces the dendritic sums of *both* low and high units. This will act to increase the variance of both high and low distributions.

## 7.4    Comparison with Levy thresholds and learning rule

We are now in a position to compare these standard learning rules in our sequential heteroassociative network with the threshold strategy and the learning rule that were proposed in the previous chapter. The following results were obtained with $M_a = 8$ $K_I = 0.1$. $K_R = 0.12$ were adjusted to keep activity stable at $M_b = 8$ for at least one iteration of the sequence. $\theta$ was still fixed at 0.8.

## 7.5    Learning in one trial

A parameter search was conducted over different values of connectivity to try and get the best performance for a particular length of sequence. The results are summarised below with best performance found with the learning constant at 0.1.

The limit, the largest number of patterns that can be stored after which sequential recall is no longer possible, is given together with the performance at that limit. Results of the simulations with pre and post synaptic rules are given in tables7.1 and 7.3.

Inspection of these tables indicates that this postsynaptic system performs worse than the associative network. The presynaptic rule applied to the heteroassociative network with WTA produced a capacity limit of 1600 patterns in the fully connected case, whereas both presynaptic and postsynaptic rules combined with the activity based threshold used by Levy and colleagues store a maximum of 560 and 507 patterns respectively.

We now examine how the inhibition and the form of the learning rule influences signal to noise properties and hence performance.

| Connectivity | performance | limit |
|:---:|:---:|:---:|
| 100 | .955 ± .1 | 560 ± 40 |
| 50 | .812 ± .24 | 132 ± 23 |
| 10 | .74 | 18 |

Table 7.1: Post synaptic rule

| % connection | performance | limit |
|:---:|:---:|:---:|
| 100 | .997 | 507 ± 50 |
| 50 | .77 | 68 ± 13 |

Table 7.2: Presynaptic rule

| % connectivity | performance | limit |
|:---:|:---:|:---:|
| 100 | .92 | 900 ± 50 |
| 50 | .75 | 15 ±5 |
| 10 | .86 | 0 ± 5 |

Table 7.3: Saturating Hebbian rule

## 7.6   Pre and Post synaptic rules

Figure 7.4 shows a sample plot of the high and low distributions in the absence of the inhibition function. Figure 7.5 below shows the effect of adding the inhibition. Clearly the dendritic sums are better separated after the inhibition has transformed the sum. This suggests that the poor performance of the one trial postsynaptic network is not due to inhibition but to the learning rule. Since the effect of LTD in the learning rule in the associative network is to move the two distributions closer together, one would predict that recall can be improved in this one trial situation, by removing the contribution of the LTD, and approximating the clipped hebbian rule as

$$\delta W_{ij} = Pre * Post(1 - W_{ij})$$

The characteristics of this rule are shown in the table 7.3:

Performance using this rule is still well below that using WTA. In this case the activity based thresholding strategy is not improving performance. Comparison of figures
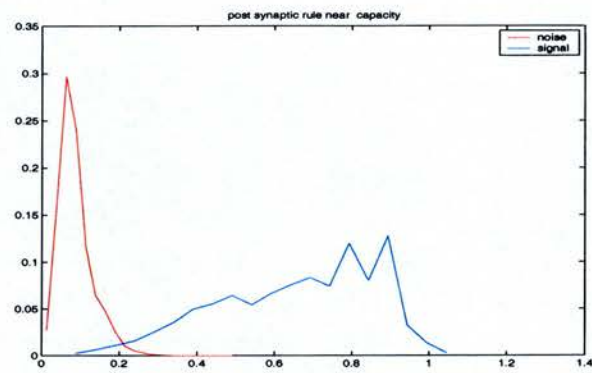
Figure 7.4: After 400 patterns the graph shows the dendritic sum distributions before inhibition trained using the postsynaptic rule.
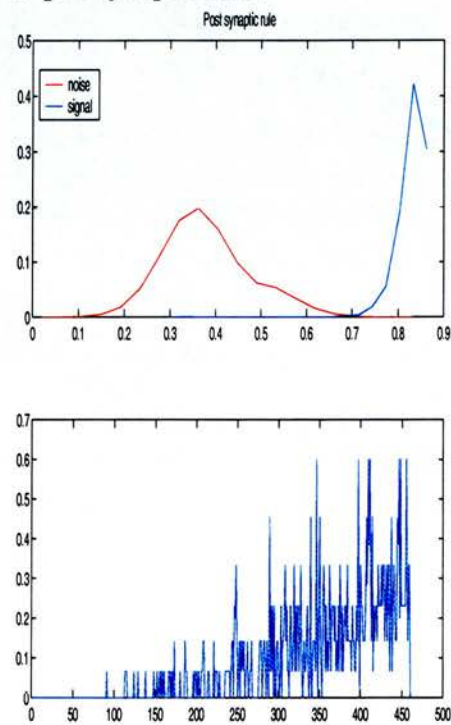


Figure 7.5: b) After inhibition. Also shows the distribution of errors through the sequence.
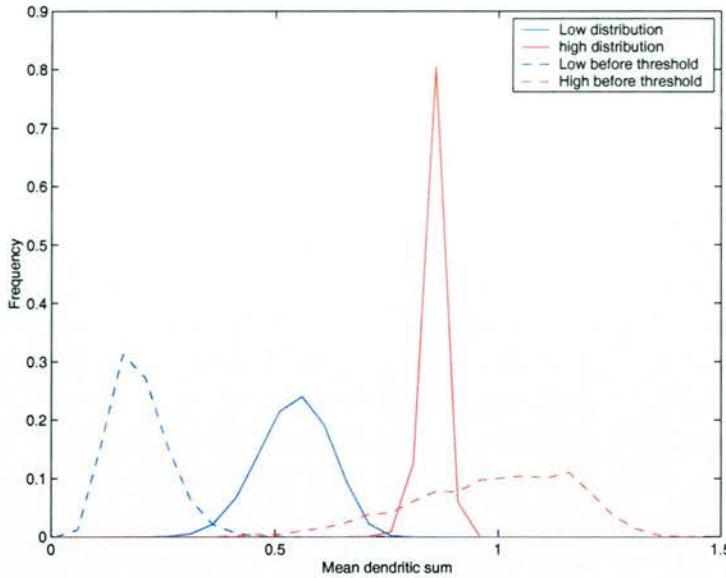
Figure 7.6: Presynaptic rule: After 400 patterns, the graph shows the distribution of mean dendritic sums before inhibition, $\zeta = 30.5, \zeta* = 34.2$ and the distribution of after inhibition $\zeta = 22, \zeta* = 25.6$.

7.7a and b shows that the inclusion of inhibition is actually moving the distributions closer together. The signal to noise goes from 90 to 15.

The explanation for the poor performance of the thresholding is embarrassingly obvious. Consider figure 7.8. As the low dendritic sum falls in the region of the graph that is steepest, effectively lowering the threshold. While this is a desirable property where the creation of context units relies on 'low' units firing spuriously, it is clearly undesirable under these conditions of constant activity.

## 7.6.1   Additional sources of error

The occasional appearance of spurious units during learning is an unavoidable consequence of the activity based thresholding strategy. The ideal threshold strategy/potentiation would permit a context unit to form only when necessary. For this to work, a minimum of two iterations may be necessary: learning, followed by recall, followed by compensation for the disappearance of activity. This would also assist recall in overcoming

Figure 7.7: Saturating Hebbian rule: After 400 patterns,the graph shows the distribution of mean dendritic sums before inhibition, $\zeta = 90$; b shows the distribution of dendritic sums after inhibition $\zeta = 15$.
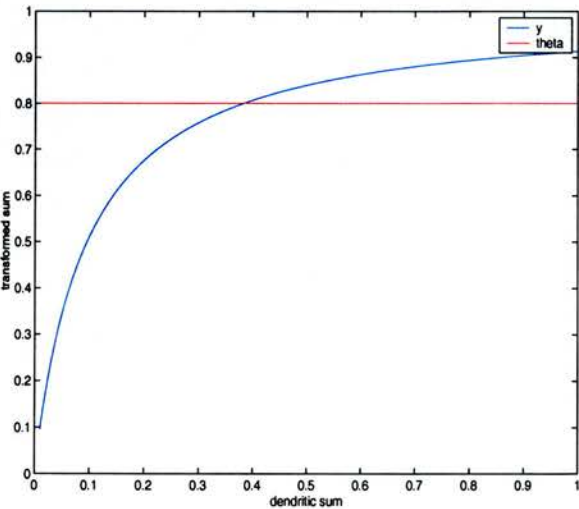


Figure 7.8: Transformed dendritic sum at sample parameters $K_I = 0.016$, $K_R = 0.016$

failure on one pattern, as in the current architecture failure on one pattern can cause a failure on all succeeding patterns, Recall on a lower threshold allows 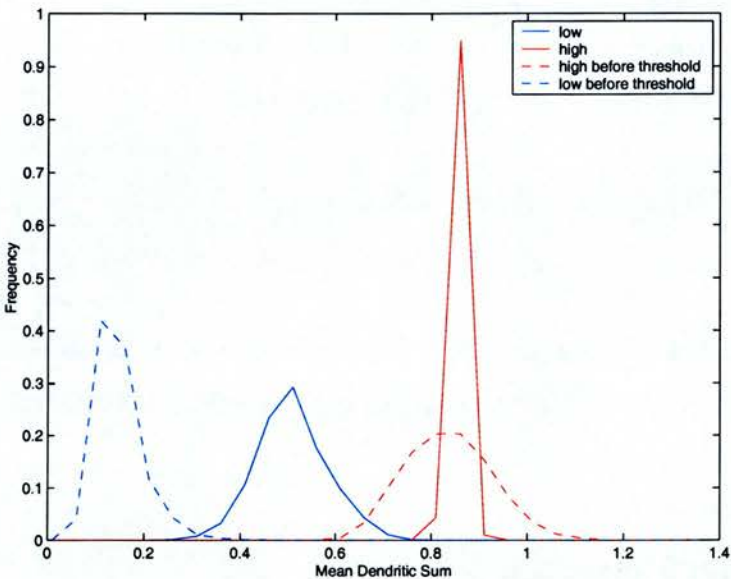the network to produce the remains of the sequence in a noisy fashion. This is a possible use for the 'annealing' properties of acetylcholine, proposed in Sohal and Hasselmo (1998) but exactly reversed.

## 7.7   Conclusion

So far, we conclude that the postsynaptic rule threshold combination is not particularly good at storing the structure of a long sequence of patterns, and enabling the sequence recall strategy.

## 7.8   Improving performance

A strategy for improving recall in the single iteration can be taken from the Associative Net.

It is easy to improve recall in the WTA situation for the linear hebbian rule. Graham and Willshaw (1995) proposed eliminating the dependence of the dendritic sum (d) on activity by normalising by unit activity (a). WTA is applied to the normalised sum

$$d' = \frac{d}{a}$$

This is only appropriate for partially connected networks, as in a fully connected network there are no variations in input activity. In the sparse situation, dividing by the number of active input lines should help distinguish the case of a few large weights from a large number of small weights. Alternatively, the sum can be transformed exploiting unit usage or transformed sum

$$d* = 1 - (1 - \frac{d}{a})^{\frac{1}{r}}$$

This second, more powerful strategy is not necessarily appropriate for a rule such as the presynaptic rule. It is based on the fact that for the unclipped rule, we know that the relationship between the mean of the low distribution, and the unit usage holds, as given in equation 7.2. This relationship does not hold for the presynaptic, and post

synaptic rules, where synaptic modifications are both positive and negative, and the dendritic mean low sum is not simply proportional to unit usage.

A related way of improving the performance of the presynaptic rule is to exploit information about the unit usage. Minai (1997) suggests a presynaptic covariance rule, where each weight is adjusted by the probability of the output unit firing. In his study, where weights were set directly, each weight is set to $P(post|pre) - P(post)$. Using a WTA thresholding strategy, this rule outperformed the presynaptic and clipped rules on correlated data.

This rule is very close to the activity based 'theta rule' strategy (Willshaw *et al.*, 1996) which proposes a modifiable threshold calculated by

$$\delta theta_j = \alpha(Pre * theta_{max} - theta)$$

to be combined with the presynaptic sum rule on a similar problem.

The theta rule and the presynaptic covariance rules are almost identical, the threshold being treated as an additional weight with input permanently one. It depends on the $theta_{max}$ parameter. In Minai's formulation, each individual weight has a covariance (inhibitory) weight set alongside it. The presence of these inhibitory weights means that the learning rule is robust to correlations in the data. For a fully connected network theta and the sum of the covariance weights will approach one another. However, for a sparsely connected network, the sum of the covariance term actually utilised for a given unit will depend on the *number* of active *existing* connections, not just the unit usage.

These strategies for sparse networks and correlated data are complementary. Minai did not test the extent to which the incremental learning rule performs. He loaded the probability values described straight into the network under test. In the case of the correlation matrix from the clipped rule, this will have made no difference. However, in the case of the presynaptic, and presynaptic covariance rules, we saw in the last chapter that these rules do not converge quickly. It is likely that a number of iterations of the learning rule will produce better convergence, and hence better results. The mismatch between the incremental rule and the calculated probabilities is demonstrated in figure 7.9. This partly accounts for the discrepancy between the results reported in Minai (1997), that show better performance for the presynaptic rules. A second, perhaps crucial, difference is that the Willshaw rule, and the presynaptic rules were being
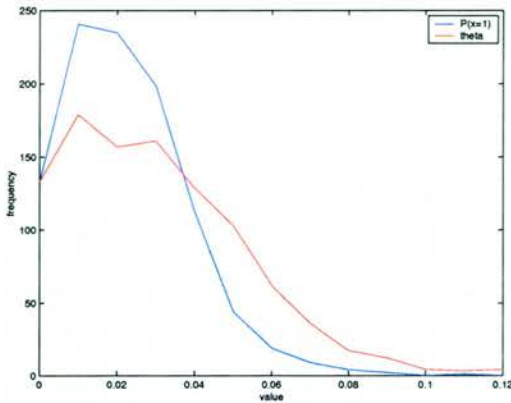
Figure 7.9: Frequency distribution of values of theta versus probability of output unit firing for a given pattern set

compared well outside the level of optimal of sparsity for the Willshaw rule. They are not therefore being compared in an error free regime. The sequence recall strategy simply does not work under error-full conditions.

## 7.9   Sparse networks

It is apparent that learning sequences with a sparse matrix is going to be significantly harder, as this introduces the possibility of errors cause by absent weights. However, this is the case we are particularly interested in, as the hippocampus is partially, if not sparsely connected. Here we briefly consider network of 50% connectivity. This is not particularly more or less realistic than any other figure, but does give a reasonable indication of how well our threshold strategies are performing under incompletely connected conditions.

### 7.9.1   Comparing thresholds

The addition of the covariance threshold makes a dramatic difference, under sparse conditions with correlated data. Data below is for $M_a = 30$, for the overlapping case, overlap $= \frac{9}{M_a}$. Data in the table compares signal to noise and errors at 75 patterns, with the presynaptic rule with and without the covariance term.

Figures 7.10,7.11 show the effect on performance of the presynaptic covariance
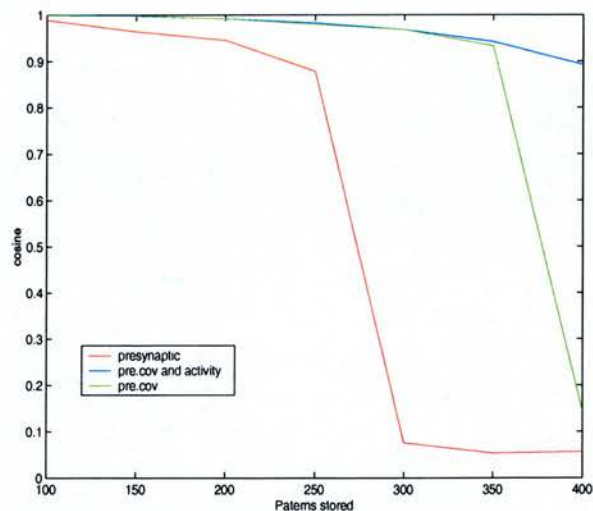
Figure 7.10: Performance in sparse networks of the presynaptic, presynaptic covariance, and clipped rules $M_a = 30$ random patterns
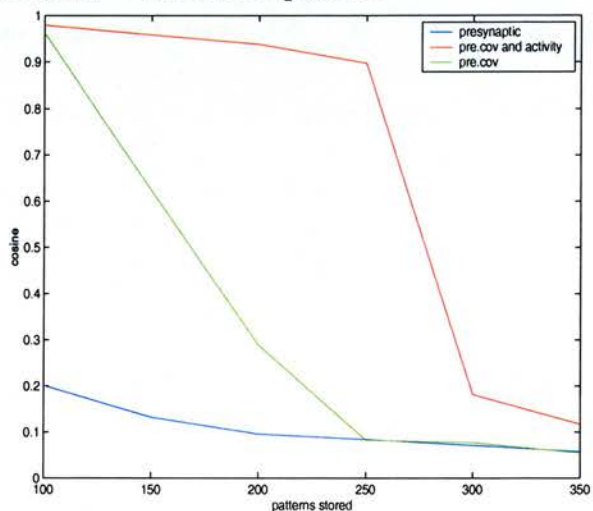


Figure 7.11: Performance in sparse networks of presynaptic, presynaptic covariance and presynaptic covariance and activity transform. using overlapping sequence data. The combined strategy of using the covariance theta term and an activity transform produces the best results.
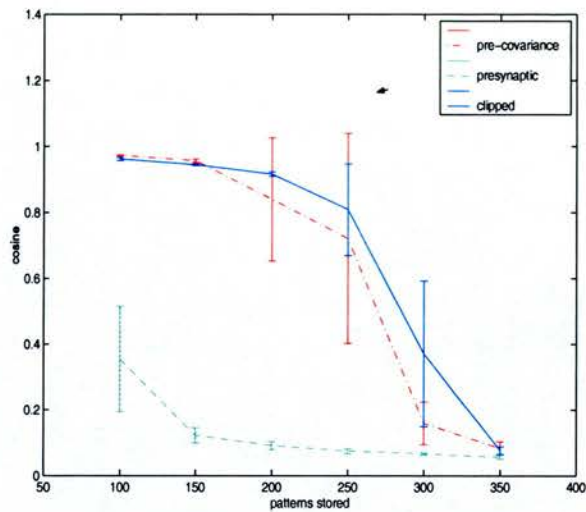
Figure 7.12: The best of the presynaptic rules, using covariance and activity perfor-mance almost as well as the clipped rule. Performance compared over 5 runs, overlap of 21

| rule | S/N | cosine | errors |
|---|---|---|---|
| Pre | $126 \pm 5$ | .99 | $28 \pm 16$ |
| Pre.cov | $142 \pm 3$ | .99 | $3.2 \pm .8$ |

Table 7.4: Signal to noise ratio at 75 patterns, over five runs.

rule and also the addition of a divisive threshold, based on the normalised WTA strategy. Figure 7.12 shows that the clipped rule and the presynaptic-covariance rule are now performing almost equivalently.

We now have a threshold system for recall that works much better than the scheme described in the Levy model. So the next stage is to adapt this thresholding system so that it can set its own activity level. We know in theory that we can set a firing, non firing condition threshold at somewhere between high and low distributions for each unit. The Levy model needs three threshold parameters to be hand tuned. $K_R$ and $K_I$ are contributing to a divisive threshold, while $\theta$ is a subtractive threshold.The presynaptic covariance rule plus activity needs only theta max to be prespecified for the subtractive threshold, and replaces the activation function 6.3 by normalisation by unit activity.

Considering the signal to noise ratio in each of these cases, it is clear that an even better thresholding strategy should push performance much higher. The model in the previous chapter proposed a thresholding strategy using a transform based on average activity, and relied on setting a subtractive threshold by hand. We have already seen that setting a subtractive threshold is unnecessary, given that the covariance term can be used instead. Considerations from the real valued associative network suggest that a good divisive strategy will be based on individual unit activity, rather than average activity.

It is clear that for binary patterns, no advantage is conferred by using the conditional rules on this particular problem. There is a good argument, originating from Marr (1971) put forward in the previous chapter, that for this type of learning, one trial, or snapshot, the conditional probabilities purportedly represented in the weight matrix have no real interpretation, even if they are trained over multiple iterations to accurately reflect the input set.

There is the further question of realism. The presynaptic rule/post synaptic rules seem to be closer to the behaviour of NMDA LTP. However, some studies of LTP in the hippocampal region suggest that real valued weights may not be an accurate reflection of synaptic behaviour.

## 7.10    Return to the hippocampus

The post synaptic rule is perhaps the most plausible from the point of view of abstracting from the temporally asymmetric mechanism of LTP in the region of CA3, but how does it compare with more standard learning rules? The saturating hebbian rule roughly corresponds to one interpretation of Bi and Poo's (1998) finding for supra threshold synapses that did not show LTD.

These authors show that LTD (unlike in the rules considered here) is (not) synaptic amplitude dependent. This may actually be better under the non convergent conditions.

It has been claimed that some forms of LTP in CA3 are effectively binary. Petersen (1998), claims that the NMDA component of the schaffer collateral synapses is in fact all or nothing, and that apparent graded potentiation, was due to the presence of differing thresholds for potentiation. They discuss the fact that the associative network provided one example of an information storage device that improves with binary weights. It is therefore interesting to observe that the learning of sequences by a similar architecture is also improved, especially, as it is these very same NMDA receptors that provide the inspiration for the graded temporal learning that the post synaptic rule is intended to approximate.

## 7.11    Conclusion

One conclusion that can be drawn from this work is that under these one, or few trial conditions, LTD is actually detrimental to learning. In chapter 3, the hypothesis that the hippocampus supported different speeds of learning. The slow learning system, it was suggested, combined small amplitude LTP, with greater LTD. The studies in this chapter suggest this is not a suitable regime for one trial learning. However graded synapses can be made to perform one tral learning, and recall with adjustment of the theta threshold, as we show in figre 7.12. This implies a requirement for learning in inhibitory interneurons to track the activity of the presynaptic neuron. The proximal area of CA3 that shows local recurrence would be particularly suitable for sequence strategy recall.

## 7.12   Summary

In this chapter, we have developed the insight that the sequence learning device is a heteroassociative network trained on a special conditions, and with a special pattern set. We have observed that real valued learning rules, in particular rules like the presynaptic rule, do not work particularly well in this architecture. However, the problem of recall in heteroassociative type architectures has been studied well for the real valued rule. We also saw that the presynaptic rule performs better when modified into the presynaptic covariance rule. Best results were achieved by the joint application of these solutions. However, results with the clipped rule are still better.

# Chapter 8

# Conclusion

## 8.1   Part I: Inventing episodic memory: conclusions

Theories of event memory that we presented in chapter 1 suggested that hippocampal memories are one trial, simple memories, (Marr, 1971), or sensory snapshots (Squire *et al.*, 2001; Frey and Morris, 1998; Norman and O'Reilly, 2001; McClelland and Goddard, 1996). In chapter 1, we argue that *episodic memory* is not this simple memory, but is instead a constructed event representation, that incorporates information learned over many trials. Furthermore, this information is learned within the hippocampus rather than borrowed from neighbouring cortex. However, we also argue that the hippocampus contains information about particular trials. We therefore propose that learning mechanisms in the hippocampus run at multiple speeds, and that information about individual and groups of trials is integrated within the hippocampus. We argue that this integration takes place along the septotemporal axis.

Some slight evidence has appeared that suggests (Small *et al.*, 2001; Strang *et al.*, 1999) that it may be possible to track effects of recall and acquisition on activation patterns on the long axis of the hippocampus. The study Small *et al.* (2001) as it stands shows experience dependent change in response to repeated trials of the same paired stimulus. Their main conclusion is that the activation of a paired stimulus of name and face does not decompose into name pattern and face pattern. However, the spatial organisation of patterns on the long axis emerged only on the third block of testing. The significance of the movement of activation patterns is still unknown, suggesting that the final representation is gradually constructed, and is not recalled from the same part of the hippocampus that is initially activated.

In chapter 2 we present further evidence that the hippocampus contains (at least) two different learning systems that learn at different speeds. We propose that the experimental work by Moser and Moser (2000) suggesting that functional differentiation on the long axis of the hippocampus can be explained by attributing different learning speeds to the dorsal and ventral regions. Moser and Moser (2000) explain their data by a functional division between spatial and non-spatial processing, supported by an anatomical division.

This argument is further supported by Chapters 3 and 4. Chapter 3 argues against the functional division of the hippocampus, by firstly arguing against the *lamellar* hypothesis needed to divide dorsal and ventral circuits. Secondly, Chapter 3 argues

against the input/output connectivity described by Moser and Moser (2000). Chapter 3 also identifies two critical area of future research. Our proposal is that the input stratification on the dendritic tree of CA3 and CA1 pyramidal cells means that these inputs can be selectively integrated and excluded. In particular we identify a role in this for dopamine, a neurohormone associated with control of prefrontal lobe activity, a region that is critical for planning, and therefore using hippocampal representations. We say a little more in section 8.3 below on future work.

A critical component of integrating recall of past events with new input is the ability to rapidly recall past events. One potential mechanism for this is based on the NMDA learning mechanism outlined in Chapter 3. We call this the sequence recall strategy: recall of ordered sequence of events learned over a single trial. Discussion of the evidence for this recall strategy, and two models of how it can be employed in hippocampal type architectures forms the later half of this thesis.

In Chapter 4, we examine the argument in Moser and Moser (2000) that tasks requiring the dorsal hippocampus are best described as supporting a *spatial map* requiring special spatial input. We argue that too many non-spatial components of the hippocampal representation have been found. However the map hypothesis has led to a proposal that the hippocampus supports a predictive representation that exploits the sequential recall strategy.

The conclusion of Part 1 of the thesis is that although there is reasonable evidence for a functional distinction between temporal and septal hippocampus, the best way to characterise this difference is in terms of the speed of learning in the two parts.

## 8.2 Part II: Mechanisms and models of anticipatory recall

The second part of this thesis deals with how to perform exact recall of an exact sequence of items in a hippocampus-like structure.

First we examine the evidence that such a strategy is used in the hippocampus. Chapter 5 examines a case (of proactive, or predictive) representation in the prefrontal cortex. We distinguish such a representation from anticipatory use of the sequence recall strategy.

We argue that although the case that the hippocampus uses a sequence recall strategy is still unproven (we discuss this in chapter 5), a structure that in principle uses the paired recall strategy can use the sequence recall strategy.

Chapter 6 then discusses the merits of Levy (1996)'s model of sequence learning and recall over repeated trials. We present an exact replication study, and some additional experiments, through which we explore the network's behaviour at capacity. We show that the network fails to converge to a stable pattern over a long training cycle, and therefore the weights of the network do not encode the information about transitions in the patterns that Amarasingham and Levy (1997) propose. The model successfully learns correlated (but recoded) input sequences, but performs rather less well when multiple correlated sequences have to be stored separately. In this situation, learning one sequence interferes with the recall of another. The negative patterning task can be modelled as this sort of learning problem. We examine the performance of the model on a negative patterning task. We argue that despite the apparent fit between the behaviour of the model and results from Sutherland and Rudy (1989), this cannot be treated as evidence for sequence learning as a model of the negative patterning task. The argument can be summarised as follows: The representation of the task does not distinguish between hippocampal and non hippocampal dependent forms of the negative patterning task. Secondly, the models exploits assumptions about the hippocampus's contribution to behavioural errors. We present what we think is an equally good case for a an entirely, different contribution, argued from the results of the same model and experimental results in Han *et al.* (1998).

Chapter 6 concludes that

1. the model does not fit the behavioural task

2. The behaviour of the model is controlled by the threshold and the learning rule, and these perform poorly in the one trial situation that does not use a recoding strategy.

Chapter 7 follows on from chapter 6, considering the model proposed by Levy (1996) and working through an analogy to the heteroassociative network. On the basis of this analogy, we examine five different learning rules, and two different thresholding strategies that are used in pattern completion problems in partially connected heteroassociative networks. In particular, we evaluate the presynaptic covariance rule proposed

by Minai (1997) on correlated input data in as sequence recall situation. This rule, according to Minai (1997) out-performs all other rules in a compression, or classification problem. However, in an error free situation, we show that the clipped learning rule still performs better, even on correlated data.

The conclusion of chapter 7 is that for single trial learning, where only the order of inputs need be recalled, a binary valued learning rule is the best choice, even if the inputs are highly correlated. We therefore argue that this kind of immediate recall of single events, if it occurs at all, will occur only in the dorsal hippocampus, using all or nothing potentiation at CA3 synapses.

## 8.3   Future work

The model described in chapter 7 could be embedded in a larger model that incorporated a slower system of learning trained by the fast, one trial system.

In chapter 1, we argued that episodic memory has a special relationship to language, and memory construction. We also argued that it is extremely difficult to distinguish affects of acquisition and recall in human subjects. One reason for this is that recall can play a role in acquisition. For example, when learning a list of words, or lines for a play, the best learning method is repeated testing. Practising recall helps with acquisition. However, as illustrated in Chapter 1 proactive interference can mean that recall may interfere with acquisition.

In chapter 2 we argued that learning and recall of general properties of tasks and environment and immediate recall of a particular trial require different types of learning. Fast learning is needed to capture a single trial; information from multiple trials is stored more slowly in the temporal hippocampus. Information flow along long axis of the hippocampus can be used to integrate information from different stages of learning. This lead us to a new division of hippocampal tasks. Tasks where learning is additive are those where a behavioural strategy based on what the rat did last time works well are likely to be learned faster by the dorsal hippocampus. Tasks where information about the previous trial is required are likely to be dorsal hippocampus dependent. We are arguing here that the dorsal hippocampus provides something that the ventral hippocampus lacks. In some situations, such as the interference tasks like

negative patterning where the recall of previous trial consequently biases towards the action rewarded on the previous trial, performance might improve with dorsal lesions. This prediction might be tested by comparing pattern of errors of intact rats with the errors of dorsal, ventral, and complete hippocampal lesioned rats in the early stages of learning on an interference type task.

Chapter 3 set out some of the potential mechanisms of changing the integration of information from different sources, that is stratified through the dendritic trees of hippocampal pyramidal cells. It is an open question as to whether we should think of these cells as switching between learning and recall (Hasselmo *et al.*, 1996; Paulsen and Moser, 1998; Lisman, 1999) or as switching between inputs (Dvorak-Carbonne and Schuman, 1999b; Hoffman and Johnston, 1999a). It is these mechanisms that will hold the key to understanding how information from multiple trials of a behavioural task can be integrated.

An obvious development of ideas in these three chapters, would be to examine whether dopamine could be used to switch between type of learning within a single cell. Such a model needs to start by accounting for the effects due to the absence of dendrites in SL-M in proximal cells. If mechanisms modulated through the SL-M are crucial for switching between cell computations, then considering realistic models with and without those dendrites should offer different predictions for the cells' behaviours in response to inputs. A model developed by Lisman and Otmakhova (2001) has modelled some of the behaviour shown in Otmakhova and Lisman (1999a, 1998) in the context of his sequence learning model. His model does not take into account additional data available on the interaction between perforant path and schaffer collateral inputs without dopamine.

There is a more basic issue that emerges from the outline of the change in radial distribution of inputs along the transverse and septo-temporal axis. We claim that this may be computationally significant. This claim needs to be further investigated.Hoffman and Johnston (1999a) provide a basic model of dendritic integration that could be used to compare cells from the dorsal and ventral poles.

# Pseudocode

# Pseudo Code

Part one of Appendix A describes training and testing for the networks in chapter six.

# Part 1

## Algorithm A

First specify the distance between input pattern pairs (e.g 6 bits). specify number of pattern pairs (i.e L/2). Decide on number of desired trials of sequences (t) - usually three or five. Construct t sequences of type " shifted" and t sequences of type "overlapped".

## Algorithm A

1. Construct an n by n weight matrix initialized to random values between 0.01, and 0.3,with a mean value of approximately 0.1.

2. Create sparse matrix by deleting (say 90%) of connections chosen randomly. Store a copy of the matrix.

3. Select parameters *theta, e, KI, KR* according to Procedure 0.

4. Present the first two pattern pairs to the network. The first pattern is clamped to the output neurons. The second pattern is clamped to the input neurons.

5. Apply the learning rule for all w. Record the value of the sum of the weight changes. Calculate the dendritic sums of all output units using equation 6.3 and allow all units with a sum greater than *theta* to fire. each input unit is also capable of firing one output unit during training. In practice, superimpose the current input pattern on the output units. Thus the output unit fires if the recurrent input from the previous time step puts it over threshold, or if it receives direct input. This combined pattern forms the new output pattern.

6. Apply the next input pattern to the input units.

7. Repeat steps 5-7 for all patterns.

8. At the last pattern in the sequence include a null step with no learning, and clear the output.

9. Repeat steps 4-9 for as many trials as desired; until either the sum of dw has reached zero or according to the criteria specified in procedures 3 and 4.

10. Repeat steps 1-10 for each sequence of type "shifted" and type "overlapped".

## Procedure 0: Procedure for selecting parameters.

There are five parameters. Parameter *theta*, and parameter *e* the learning rate, were treated as fixed as given in Amarsingham(1998), theta =0.8, e = .01. The mean of the initial weights was also a parameter, and its value was selected to give approximately the right number of neurons firing per time step, with a random input and no learning. KI and KR were treated as free, and the procedures for setting their values were as follows.

1. Set KI to a values between .001 and .1, in intervals of .002, then for each value of KI, chose a set of values for KR at .001, to .2 in intervals of .02. The first network will have parameters KI = .001, KR = .001, the second KI = .001, KR .003 etc.

2. A run consists of a network for each parameter pair, with an input sequence of length 20, trained for 500 trials. The activity of the network, defined as the number of neurons firing per time step, was monitored at each time step. If the activity increased above 50 percent, or dropped to zero, the parameter pair was discarded.

3. The performance for each network of each parameter pair is measured according to procedure two. Chose the five best performing parameter pairs

4. Repeat training with Algorithm A for five best pairs.

5. For sequence of different lengths,, but same correlation repeat with same five parameter pairs. If performance appears to be better on the outlying parameter

pairs, repeat with with fixed KI, but adjusting KR to the next five pairs, until performance stops improving.

Plots of the parameter space indicated that activity and performance was very much less sensitive to the value of KI than to the value of KR, as expected from equation 6.3.

## Procedure 1: Evaluating performance.

1. On the last iteration of algorithm A for a given network, the output of the network to the entire sequence, presented pattern by pattern, is saved. This is the *recoded* sequence.

2. The first pattern pair, is then presented to the network, step six from algorithm A is then performed.

3. Then, no additional patterns are presented, and the network output is calculated only from the dendritic sums, based on the previous steps output.

4. Repeat 3) until a string of outputs has been created as long as the original input sequence.

5. The output pattern is a sequence of the same length as the *recoded* output pattern from step 1. Compare the output from step 4 and step 1 by taking the normalised dot product of pattern 1 from step 4, and pattern 1 from step 1.

6. Repeat for all patterns. Discard the first pair of patterns.

7. Repeat testing procedure for each network with the same parameters, and then average cosine scores.

8. Where the capacity of the network is to be determined for a particular set of parameters, this is defined as the pattern length, that produces an average cosine score, as calculated above, greater than 0.75.

## Procedure 2

A measure of stability of the network is considered to be the similarity of the output sequence produced in step 1 of procedure 2, to the output sequence produced on a step

similar to step one, but applied earlier, or later, in the training procedure. A cosine score, computed as step 5 in procedure 2 quantified this similarity.

## Procedure in Experiment 3: Described in chapter 6.

The assumption made in Levy and Amarsingham is that parameters KI and KR control the level of activity in the network. This assumption is not entirely true, as we discuss in the chapter.

We used Algorithm A, with the following modifications. In step 2 in stead of deleting connections randomly, we deleted most of the connections, and then selectively pruned the matrix, so that no symmetric connections existed. We then deleted connections, until the number of existing weights 10 percent. For networks where a matrix of a certain percentage of symmetric connections was required, we selected a random weight, created a symmetric connection, and then deleted a randomly chosen weight.

The clarification of the kind of oscillations produced by the network was done by hand. A stationary network is a network where any input pattern of the sequence produces the same output. A periodic oscillation is where the network moves through the output pattern in response to a randomly chosen input pattern from the sequence, and this cycle is shorter than the length of the sequence the network is trained on.

# Part II: Chapter seven

## Chapter seven: Algorithm B

Algorithm B describes steps used to construct the one trial network.

1. Create a weight matrix, where the initial mean value of w is extremely small $\gg 0.01$ for the cases of the pre, post, pre-covariance, and saturating hebbian rules. For the cases of the clipped, and unclipped rules, the weight matrix should initially be zeros.

2. Clamp the first pattern pair, pattern 1 to output, and pattern 2 to input, and calculate and calculate dw according to the learning rule, and update the weights.

3. Repeat 2) for all pattern pairs.

4. Recall was tested after the addition of each 10 pattern pairs, parameters for threshold during recall were set after parameter searches, described below. Recall was tested as described in procedure 1.

## Parameter search 2: For a one iteration net with levy threshold.

Parameter searches were performed as described in procedure 0, with the following modifications. As training took a single iteration, and recall was also for a single iteration, the parameters were not set to to keep activity stable of many iterations. Instead they were set to mimic a Winners Take All threshold, and permit exactly the same number of neurons to fire at a time as were firing in the input pattern. The region of parameter space in which activity remained at the desired level was very large (see Appendix, figure 7). This was established by a very large (but rapid) search through values of KR and KI.00.1 to .2, in interval of .001. The exact values during learning made very little difference, provided they were within this region. To determine which parameters produced the best possible performance during recall, the lowest possible values of KR, and KI was run and 20 additional pairs values were tested, on the trained network. Fortuitously, the lowest possible value KR proved to be generally the best values for recall. For recall, the value of KI is irrelevant. We experimented with adding a constant as described in smith(2000), and found little difference.

## Chapter 7: paired recall

Figure 7.2 describes an experiment comparing pattern paired recall, with sequential recall. Each Sequence is made up of a string pair of input and output patterns. Paired recall is assessed by looking at the response to each input pattern presented separately, and comparing it to the clamped output pattern using cosine.

## Setting thresholds

Setting theta max: For the interpretation of the presynaptic covariance rule, the value of theta max is critical. For a given set of presynaptically trained weights, a range of

theta max values were tried (0.1- 2, in intervals of 0.1) and the value that produced the fewest errors selected. Where the strategy of training inhibitory weights is used in the partially connected network,the problem is to set the learning constant of the weights, but as the rule for setting the weights is linear, it could be scaled post hoc to the value that yielded the best performance.

## Figures 7.10, 7.11

Graphs show recall from the five same network and pattern pair sets. The weights are set with the presynaptic rule, and recall is then compared using three different strategies: i) Dendritic sum normalised by unit activiy, ii) Weight subtraction using trained inhibitory weights (Pre-cov), and iii) A combination strategy (Precov-act). Where we compare the clipped rule to the presynaptic rule, the same five connection matrices and pattern sets are used. This ensures that we really are comparing the performance of a rule on a particular data set, as well, as the rules' performance on average.
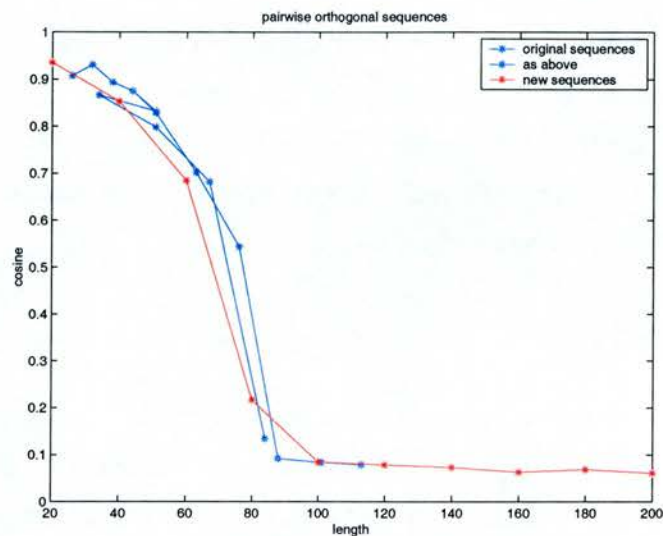
Figure 1: Sequences which are pairwise orthogonal (shifted), compared to randomly selected sequences, $overlap = 0$

# Part III: Additional data from chapter 6

Pairwise orthogonal sequence were constructed by both sequence generating programs, show in figure 1 The results of the parameters searches are shown in table 1 The initial weight distribution with mean .30, std .17. set the initial activity of the networks described in chapter 6, section 6.4.

Equivalent searches were also conducted for the shifted sequences, but the data is not included here. Performance is compared in table 1 at two or three values of $K_R$. The best performing parameters at length 20 were not guaranteed to give the best results at longer lengths, as can be seen in table 2. Bold indicates the performance at the best performing value.

| $0.009 < K_R < .02,\ overlap = 0,\ Ch = 8$ | | | $0.009 < K_R < .02,\ overlap = 2,\ Ch = 6$ | | |
|---|---|---|---|---|---|
| $K_R$ | state | Cosine | $K_R$ | state | Cosine |
| .009 | stationary | NA | .009 | stationary | NA |
| .011 | osc | .82 | .010 | stationary | NA |
| .010 | osc | .84 | .011 | osc | .89 |
| .011 | osc | .82 | .012 | osc | .87 |
| .012 | osc | .77 | .013 | osc | .76 |
| .013 | osc | .74 | .014 | osc | .65 |
| .014 | osc | .71 | .014 | osc | .65 |
| .015 | osc | .65 | .015 | osc | .46 |
| .016 | stationary | NA | .016 | osc | .22 |
| .017 | stationary | NA | .017 | stationary | NA |
| .018 | stationary | NA | .018 | stationary | NA |
| $0.009 < K_R < .02,\ overlap = 5, Ch = 3$ | | | $.009 < K_R < .019,\ overlap = 6,\ ch = 2$ | | |
| $K_R$ | state | Cosine | $K_R$ | state | Cosine |
| .009 | stationary | NA | .009 | stationary | NA |
| .010 | stationary | NA | .010 | stationary | NA |
| .011 | stationary | NA | .011 | stationary | NA |
| .012 | stationary | NA | .012 | stationary | NA |
| .013 | stationary | NA | .013 | stationary | NA |
| .014 | osc | .91 | .014 | osc | .92 |
| .015 | osc | .89 | .015 | osc | .90 |
| .016 | osc | .84 | .016 | osc | .84 |
| .017 | osc | .68 | .017 | osc | .75 |
| .018 | stationary | NA | .018 | osc | .73 |
| .019 | stationary | NA | .019 | osc | .69 |

Table 1: Parameter searches at length 20 for values of $K_R$ and overlap for *overlapping* sequences.$NA = Not\ Applicable$

| | $overlap = 2,\ ch = 6$ | | | $overlap = 5,\ ch = 3$ | | | $overlap = 6,\ ch = 2$ | |
|---|---|---|---|---|---|---|---|---|
| | $K_R$ values | | | | | | | |
| Length | .011 | .012 | .013 | .015 | .016 | .0185 | .015 | .016 |
| 20 | **.81** | stat. | .17 | **.77** | .65 | stat | **.80** | .76 |
| 40 | **.86** | stat | .71 | **.83** | .77 | .54 | **.78** | .75 |
| 60 | stat | **.71** | .63 | .72 | .74 | **.87** | .77 | **.85** |
| 80 | stat | **.58** | .55 | **.84** | .82 | .16 | .79 | **..81** |
| 100 | .40 | .12 | **.55** | **.78** | .77 | .63 | .77 | **.87** |
| 120 | - | - | - | jumps | stat | stat | .73 | **.87** |
| 140 | - | - | - | - | - | - | .71 | **.75** |
| 160 | - | - | - | - | - | - | **.65** | .40 |

Table 2: Bold shows the best result for a particular length. $K_R$ values were chosen from the parameters searches as the best 2 or 3 at length 20.

## Properties of Sequences

The properties of overlapping and shifted sequences are compared in figures 2,3 and 4. Overlapping sequence had a much wider spread of unit ON times i.e. the number of time steps a neuron might be on consecutively. Figure 5 shows that as overal activity increases, so does the standard deviation - this indicates the width of oscillation in simulation.
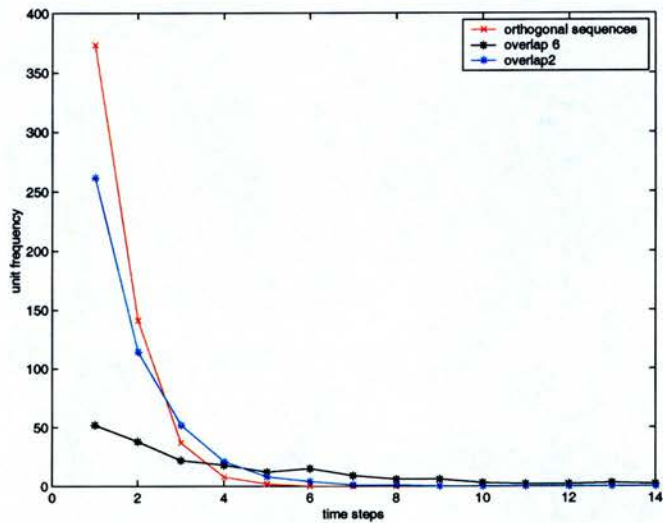
Figure 2: Distribution of unit ON times with in *overlapping* sequences with *overlap* = 2, 6 *and* 0
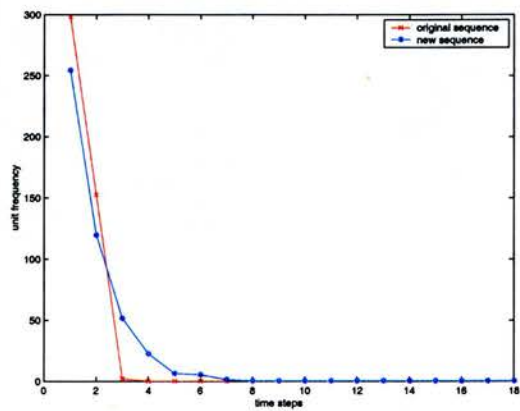


Figure 3: Distribution of unit ON times in *overlapping* sequences compared to *shifted* sequences, *overlap* = 2
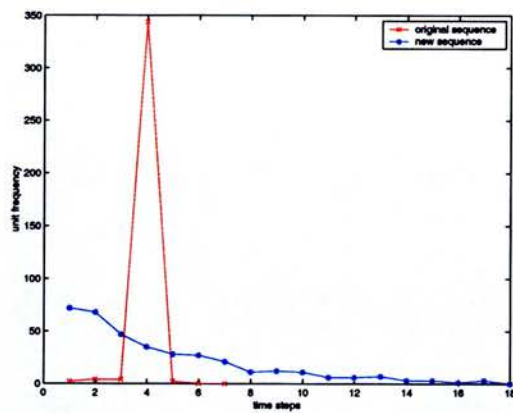
Figure 4: Distribution of unit ON times in *overlapping* sequences compared to *shifted* sequences, $overlap = 6$
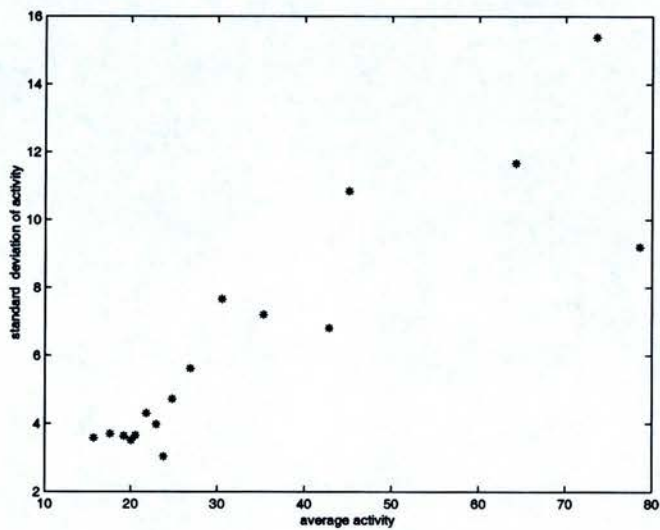


Figure 5: Standard deviation of activity against activity
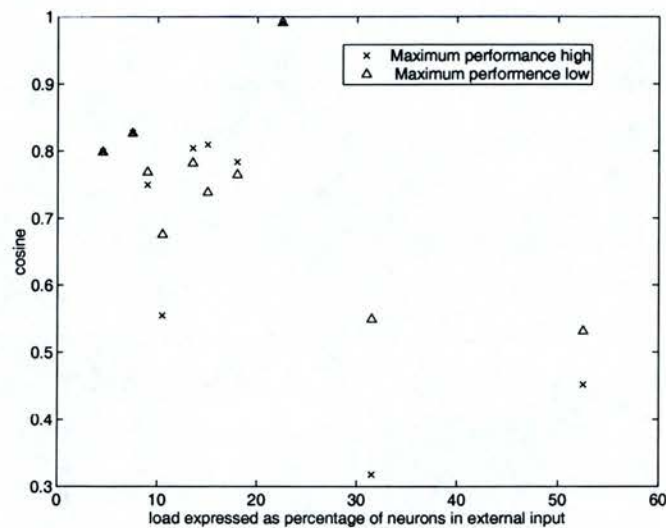
Figure 6: Maximum performance for a randomly selected set of sequences using 'high' and 'low ' thresholds.

## Thresholds

We compared two ways of setting the threshold for recall. The first way ('high'), set the threshold equal to the threshold during learning, by setting the contribution from the feedforward inhibition to level generated by the cue pattern. We compared this to allowing the feedforward contribution to be zero during recall ('low').
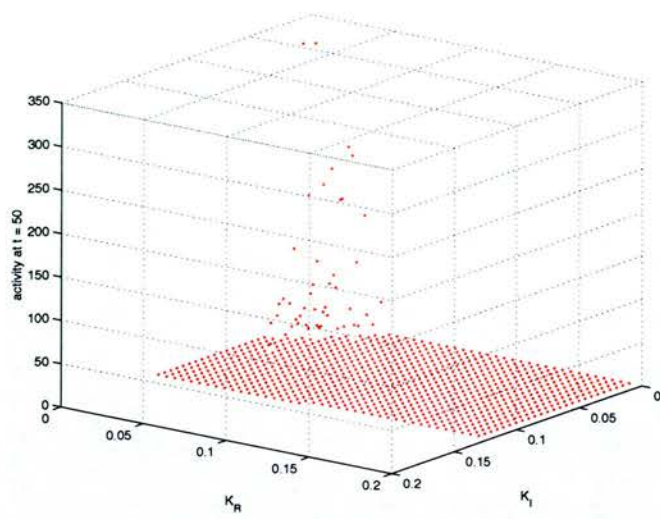
Figure 7: Activity against parameters showing the planar structure of the parameter space during learning. Best performance during recall occurs at the region of plane just before it breaks up

## Part IV: Additional data from chapter 7

Figure 7 shows parameter search for the one trial network, showing the settings of parameters that kept activity to equal external input.

# Bibliography

Acsady, L., Kamondi, A., Sik, A., Freund, T., and Buzsaki, G. (1998). Gabanergic cells are the major postsynaptic targets of mossy fibres in the rat hippocampus. *The Journal of Neuroscience*, **18**, 3386 –3403.

Aggleton, J. and Brown, M. (1999). Episodic memory, amnesia and the hippocampal -anterior thalamic axis. *Behavioural and Brain Sciences*, **22**, 471–489.

Alvarado, M. C. and Rudy, J. (1995). Rats with damage to the hippocampal formation are impaired on the transverse patterning problem but not on elemental discrimina- tions. *Behavioural Neuroscience*, **109**(2), 204–211.

Alvarado, M. C. and Rudy, J. W. (1992). Some properties of configural learning: An investigation of the transverse-patterning problem. *Journal of experimental psychol- ogy*, **18**(2), 145–153.

Amaral, D. and de Witter, M. P. (1995). *The rat central nervous system*, chapter The Hippocampal Formation. OUP.

Amarasingham, A. and Levy, W. B. (1997). Predicting the distribution of synaptic strengths and cell firing correlations in a self organising, sequence prediction model. *Neural Computation*, **10**, 25–57.

Anderson, P. and de Witter, M. (1989). The three dimensional organisation of the hippocampal formation. *Neuroscience*, **31**, 571–591.

Anderson, P., Bliss, T., and Skrede, K. (1971). Lamellar organisation of hippocampal excitatory pathways. *Experimental brain research*, **13**, 222–228.

205

Bannerman, D. M., Good, M. A., Butcher, S., 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.

Bannerman, D. M., Yee, B. K., Good, M. A., Heupal, M. J., Iversen, S., and Rawlins, J. (1999). Double dissociation of function within the hippocampus: a comparison of dorsal, ventral, and complete hippocampal cytotoxic lesion. *Behavioural Neuroscience*, **113**(6), 1170– 1188.

Baxter, M. and Murray, E. (2001a). Opposite relationship of hippocampal and rhinal cortex damage to delayed non-matching to sample deficits in monkeys. *Hippocampus*, **11**(1), 61–71.

Baxter, M. and Murray, E. (2001b). Effects of hippocampal lesions on delayed non match to sample in monkeys: A reply to Zola and Squire 2001. *Hippocampus*, **11**(1), 201–203.

Bertram, E. H. and Zhang, D. X. (1999). Thalamic excitation of hippocampal CA1 neurons : a comparison with the effects of CA3 stimulation. *Neuroscience*, **92**(1).

Best, P. J. and White, A. M. (1999). Placing hippocampal single - unit studies in a historical context. *Hippocampus*, **9**(4), 346–351.

Bi, G. and Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, **18**(24), 10464 –10474.

Bliss, T. and Lomo, T. (1973). Long lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of the perforant path. *Journal of Physiology*, **232**, 331–356.

Blum, K. I. and Abbott, L. F. (1996). Spatial map formation in the rat hippocampus. *Neural Computation*, **8**, 85–93.

Borisyuk, R., Denham, M., Denham, S., and Hoppensteadt, F. (1999). Computational models of predictive and memory related functions of the hippocampus. *Reviews in the Neurosciences*, **10**(3-4), 213–232.

Bose, A. and Recce, M. (2001). Phase procession and phase locking of hippocampal pyramidal cells. *Hippocampus*, **11**(204-215).

Buckingham, J. (1992). *Delicate Nets,Faint recollections*. Ph.D. thesis, University of Edinburgh.

Buckingham, J. and Willshaw, D. (1993). On setting thresholds in a incomplete associative network. *Network*, **4**, 441–459.

Buckingham, J. T. and Willshaw, D. J. (1992). Performance characteristics of the associative net. *Network*, **3**, 407–414.

Bunsey, M. and Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, **379**, 255–257.

Burgess, N., Recce, M., and O'Keefe, J. (1994). A model of hippocampal function. *Neural networks*, **7**(6/7), 1065–1081.

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

Cavus, I. and Teyler, T. J. (1998). Nmda receptor independent LTP in basal versus apical dendrites of CA1 pyramidal cells in rat hippocampal slice. *Hippocampus*, **8**, 373–379.

Chun, M. M. and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesiac subjects with hippocampal damage. *Nature*, **2**(9), 844–847.

Clark, R. and Squire, L. (1998). Classical conditioning and brain systems: the role of awareness. *Science*, **280**(5360), 77–81.

Clark, R., West, A., Zola, S., and Squire, L. (2001). Rats with hippocampal lesions are impaired on the delayed DNMS task. *Hippocampus*, **11**(2), 176–186.

Clifton, P., S.P.Vickers, and E.M.Somerville (1998). Little and often : Ingestive behaviour patterns following hippocampal lesions in rats. *Behavioural Neuroscience*, **112**(3), 502–511.

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

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

Cowan, W. M., Sudhof, T., and Stevens, C., editors (2001). *Synapses*. John Hopkins University Press.

Davidson, T. L. and Jarrard, L. E. (1993). A role for the hippocampus in the utilisation of hunger signals. *Behavioural and Neural Biology*, **59**, 161–171.

Dayan, P. and Willshaw, D. (1991). Optimising synaptic learning rules in linear associative memories. *Biological Cybernetics*, **65**, 253–265.

Desmedt, A., Garcia, R., and Jaffard, R. (1998). Differential modulation of changes in hippocampal-septal synaptic excitability by the amygdala as a function of either elemental or contextual fear conditioning in mice. *The Journal of Neuroscience*, **18**(1), 480–487.

Dusek, J. and Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the national academy of Sciences USA*, **94**, 7109–7114.

Dvorak-Carbonne, H. and Schuman, E. (1999a). Long term depression of temporoammonic-CA1 hippocampal synaptic transmission. *Journal of Neurophysiology*, **81**, 1036–1044.

Dvorak-Carbonne, H. and Schuman, E. M. (1999b). Patterned activity in stratum lacunosum moleculare inhibits CA1 pyramidal neuron firing. *Journal of Neurophysiology*, **82**(6), 3212–3222.

Eichenbaum, H. (1999). Conscious awareness memory and the hippocampus. *Nature: Neuroscience*, **2**(9), 775–776.

Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., and Tanilla, H. (1999). The hippocampus, memory and place cells. *Neuron*, **2**, 209–226.

Eldridge, L., Knowlton, B., Furmanski, C., Buckheimer, S., and Engel, S. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nature neuroscience*, **3**, 1149–1152.

Engel, A. ., Konig, P., Kreiter, A. K., and Singer, W. (1991). Inter-hemispheric synchronisation of oscillatory neural responses in the cat. *Science*, **252**(5009), 1177–1179.

Engert, F. and Bonhoffer, T. (1997). synapse specificity of long term potentiation breaks down at short distances. *Nature*, **388**, 279–284.

Fertinbeau, M. and McDonald, R. (2001). Dorsal/ventral hippocampus, fornix, and conditioned place preference. *Hippocampus*, **11**(2), 187–200.

Fodor, J. (1983). *The Modularity of Mind*. MIT press.

Freund, T. F. and Buzsaki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, **6**, 347–470.

Frey, U. and Morris, R. G. M. (1998). Synaptic tagging: implications for late maintenance of hippocampal long term potentiation. *TINS*, **21**(5), 181–188.

Gaffan, D. and Parker, A. (1996). Interaction of perirhinal cortex with the fornix-fimbria :memory for objects and " object in place " memory. *The Journal of Neuroscience*, **16**(18), 5864 –5869.

Gallagher, M. and Holland, P. (1992). Preserved configural learning and spatial learning impairment with hippocampal lesions. *Hippocampus*, **2**, 81–88.

Good, M., DeHoz, L., and Morris, R. (1998). Contingent versus incidental context processing during conditioning: Dissociation after excitotoxic hippocampal plus dentate gyrus lesions. *Hippocampus*, **8**, 147–159.

Goswami, U. (1998). *Cognitive Development*, chapter 7. Psychology Press Ltd.

Gothard, K., Skaggs, W., Moore, K., and NcNaughton, B. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *The journal of Neuroscience*, **2**(16), 823–835.

Graham, B. and Willshaw, D. (1995). Improving recall from an associative memory. *Biological Cybernetics*, **72**, 337–346.

Graham, B. and Willshaw, D. (1997). Capacity and information efficiency of the associative network. *Network*, **8**(1), 35–54.

Granger, R., Wiebe, S. P., Taketani, M., and Lynch, G. (1996). Distinct memory circuits composing the hippocampal region. *Hippocampus*, **6**, 567–578.

Greene, A., Prepscius, C., and Levy, W. (2002). Primacy versus recency in a quantitativeodel: Activity is the critical distinction. *Learning and Memory*, **7**, 48–57.

Guazzelli, A., Bota, M., and Ardbib, M. (2001). Competitive hebbian learning and the hippocampal place cell system. *Hippocampus*, **11**(3), 216–239.

Gurden, H., Tassin, J. P., and Jay, T. (1999). Integrity of the mesocortical dopaminergic system is necessary for complete expression of in vivo hippocampal - prefrontal cortex LTP. *Neuroscience*, **94**(4), 1019–1027.

Hampson, R. E., Hedberg, T., and Deadwyler, S. (2000). Differential information procesing by hippocampal and subicular neurons. *Annals of the New York Academy of Sciences*, **911**.

Han, J. S., Gallagher, M., and Holland, P. (1998). Hippocampal lesions enhance configural learning by reducing proactive interference. *Hippocampus*, **8**(2), 138– 146.

Hartley, T., Burgess, N., Lever, C., Caccuci, F., and O'Keefe, J. (2000). Modelling place fields in terms of cortical inputs to the hippocampus. *Hippocampus*, **10**, 369–370.

Hasselmo, M., Wyble, B. P., and Wallenstein, G. (1996). Encoding and retrieval of episodic memories: Role of cholinergic and gabanergic modulation in the hippocampus. *Hippocampus*, **6**, 693–708.

Heilman, K. and Valenstein, E., editors (1979). *The Handbook of Clinical Neuropsychology*. OUP, first edition.

Heilman, K. and Valenstein, E., editors (1985). *The Handbook of Clinical Neuropsychology*. OUP, second edition.

Heilman, K. and Valenstein, E., editors (1993). *The Handbook of Clinical Neuropsychology*. OUP, third edition.

Henze, D. A., Urban, N. N., and Barrionuevo, G. (2000). The multifarious hippocampal mossy fibre pathway: A review. *Neuroscience*, **98**(3), 407–427.

Hock, B. J. and Bunsey, M. D. (1998). Differential effects of dorsal and ventral hippocampal lesions. *The Journal of Neuroscience*, **18**(17), 7027–7032.

Hoffman, D. and Johnston, D. (1999a). Neuromodulation of dendritic action potentials. *Journal of Neurophysiology*, **81**(1), 408–411.

Hoffman, D. and Johnston, D. (1999b). Down regulation of transient k+ channels in dendrites of hippocampal CA1 pyramidal neurons by activation of pka and pkc. *The Journal of Neuroscience*, pages 3521–3528.

Hoffman, D., Magee, D., , Colbert, C., and Johnston, D. (1997). K+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature*, **387**, 869–875.

Hollup, S. A., Molden, S., Donnett, J., Moser, M. B., and Moser, E. I. (2001). Place fields of rat pyramidal cells and spatial learning in the water maze. *European Journal of Neuroscience*, **13**, 1197–1208.

Holscher, C., Anwyl, R., and Rowan, M. (1997). Stimulation on the positive phase of hippocampal theta rhythm induces long term potentiation that can be de potentiated by stimulation on the negative phase in area CA1in vivo. *The journal of Neuroscience*, **17**(6), 6470–6477.

Hopfield, J. (1982). Neural networks with emergent computational abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554–2558.

Hopfield, J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature*.

Hoz, L. D. (2000). *Memories along the temporal axis in rodent hippocampus*. Ph.D. thesis, University of Edinburgh.

Hoz, L. D. and Morris, R. G. M. (1999). Rats with a small amount of temporal spared hippocampus can learn a reference memory task in the water maze. *Society for Neuroscience abstracts*, **25**, 624.

Huang, Y. Y. and Kandel, E. (1995). D1/ d5 receptor agonists. *PNAS*, **92**(7), 2446–2450.

Ishizuka, N., Weber, J., and Amaral, D. (1990). Organisation of intra hippocampal projections originating from CA3 pyramidal cells in the rat. *The Journal of Comparative Neurology*, **295**, 580–623.

Ishizuka, N., Cowan, W. M., and Amaral, D. (1995). A quantitative analysis of the dendritic organisation of pyramidal cells in the rat hippocampus. *The Journal of Comparative Neurology 362: 17-45*, **362**, 17–45.

Izaki, Y., Takita, M., and Nomura, M. (2000). Comparative induction of ltd between dorsal and ventral hippocampal CA1 in the anaesthetised rat. *Neuroscience Letters*, **294**(3), 171–174.

Jakab, L. (1995). *The central nervous system of the Rat*, chapter The Septum. OUP.

James, W. (1890). *The Principles of Psychology*. Havard University Press, reprinted 1981 edition.

Jarrard, I. E. (2001). Retrograde amnesia and consolidation: anatomical and lesion considerations. *Hippocampus*, **11**(1), 43–45.

Jung, N. W., et.al., S. I. W., and McNaughton, B. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *journal of Neuroscience*, **14**, 7347–7356.

Kapur, N. and Brooks, D. (1999). Temporally specific retrograde amnesia in two cases of discrete bilateral hippocampal pathology. *Hippocampus*, **9**(3), 247–254.

Kirchhoff, B. A., Wagner, A. D., Maril, A., and Stern, C. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience*, **20**(16), 6173–6180.

Kobayashi, T., Nishijo, H., Fukuda, M., Burres, J., and Ono, T. (1997). Task-dependent representations in rat hippocampal place neurons. *Journal of Neurophysiology*.

Koutstaal, W., Schacter, D. L., Verfaellie, M., Brenner, C., and Jackson, E. (1999). Perceptually based false recognition of novel objects in amnesia: effects of category size and similarity to category prototypes. *Cognitive Psychology*, **16**, 317–341.

Kowlton, B. and Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Exp. Psych. Learn. Mem. Cog.*, **20**(1), 79–91.

Kudrimoti, H., Barnes, C., and B.L.McNaughton (1999). Reactivation of hippocampal cell assemblies: Effects of behavioural state, experience, and EEG dynamics. *The journal of Neuroscience*, **19**(10), 490–4101.

Kumar, S. S. and Faber, D. S. (1999). Plasticity of first-order synapses: interactions between homosynaptic long-term potentiation and heterosynaptically evoked dopaminergic potentiation. *Journal of Neuroscience*, **19**(5), 1620–1635.

Larson, J. and Lynch, G. (1989). Theta pattern stimulation and the induction of LTP : The sequence in which synapses are stimulated determines the degree to which they potentiate. *Brain Research*, **489**, 49–58.

Larson, J., Wong, D., and Lynch, G. (1986). Patterned stimulation at the theta frequency is optimal for LTP induction. *Brain Research*, **368**, 347–350.

Lavaneaux, P. and Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, **10**, 420–430.

Lenck-Santini, P., Save, E., and Poucet, B. (2001). Evidence for a relationship between place -cell firing and spatial memory performance. *Hippocampus*, **11**, 377–387.

Levy, W. and Steward, O. (1983a). Temporal contiguity requirements for long term associative potentiation /depression in the hippocampus. *Neuroscience*, **8**(4), 791–797.

Levy, W. B. (1996). A sequence predicting CA3 is a flexible associator. *Hippocampus*, **6**, 579–590.

Levy, W. B. and Steward, O. (1983b). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, **8**(4), 791–797.

Levy, W. B. and Wu, X. (1996). Context codes versus capacity. *Network: Computation in Neural systems 7*, pages 371–384.

Levy, W. B., Brassel, S. E., and Moore, S. D. (1983). Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience*, **8**(4), 799–808.

Li, X.-G., Somogyi, P., Ylinen, A., and i, G. B. (1994). The hippocampal CA3 network:an in vivo intracellular labeling study. *The Journal of Comparative Neurology*, **339**, 181–208.

Lisman, J. (1999). Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate CA3 interactions. *Neuron*, **22**, 233–242.

Lisman, J. and Otmakhova, N. (2001). Storage ,recall, and novelty detection of sequences by the hippocampus: elaborating on the SOCRATIC model to account for the normal and aberrant effects of dopamine. *Hippocampus*, **11**, 551–568.

Lowe, W. and McDonald, S. (2000). The direct route : mediated priming in semantic space. *Proceedings of the 22nd annual meeting of the Cognitive Science Society*.

Ludvig, N. (1999). Place cells can flexibly terminate and develop their spatial firing . a new theory for their function. *Physiology and behaviour*, **67**(1), 57–67.

Lynch, J. and Larson, G. (1989). Role of NMDA receptors in the induction of synaptic potentiation by burst stimulation patterned after the theta rhythm. *Brain Research*, **441**, 111–118.

Maccaferri, G. and McBain, C. (1996). Long-term potentiation in distinct subtypes of hippocampal nonpyramidal neurons. *The Journal of Neuroscience*, **16**(17), 5334–5343.

Maguire, E. and Mummery, C. (1999). Differential modulation of a common retrieval network revealed by positron emission tomography. *Hippocampus*, **9**, 54–61.

Manns, J. R., Clark, R. E., and Squire, L. R. (2000). Awareness predicts the magnitude of single -cue trace eye-blink conditioning. *Hippocampus*, **10**, 181–186.

Maren, S. and Fanselow, M. S. (1997). Electrolytic lesions of fimbria/fornix,dorsal hippocampus or entorhinal cortex produce anterograde amnesia in contextual fear conditioning. *Neurobiology of Learning and memory*, **67**, 142–149.

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

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical transactions of the Royal Society of London*, **262**, 23–81.

Martens, U., Capito, B., and Wree, A. (1998). Septotemporal distribution of [3hmk801,[3h]ampa and [3h] kainate binding sites in the rat hippocampus. *Anatomy and Embryology*, **198**, 195–204.

Maruki, K., Izaki, Y., Hori, K., Nomura, M., and Yamauchi, T. (2001a). Effects of rat ventral and dorsal hippocampus temporal inactivation on a delayed alteration task. *Brain Research*, **895**(1), 273.

Maruki, K., Izaki, Y., Nomara, M., and Yamauchi, T. (2001b). Difference in paired pulse facilitation and long term potentiation between dorsal and ventral CA1 regions in anaesthetised rats. *Hippocampus*, **11**(6), 655–661.

McClelland, J. and Goddard, N. (1996). Considerations arising from a complementary learning systems perspective on hippocampus and neocortex. *Hippocampus*, **6**(6), 654–665.

McDonald, R., Murphy, R., Guarraci, F., Gortler, J. R., White, N. M., and Baker, A. (1997). Systematic comparison of the effects of hippocampal and fornix-fimbria lesions on the acquisition of three configural discriminations. *Hippocampus*, **7**, 371–388.

McKoon, G. and Ratcliffe, R. (1986a). Automatic activation of episodic information in a semantic memory task. *Journal of Experimental psychology*, **12**(1), 108–115.

McKoon, G. and Ratcliffe, R. (1986b). More on the distinction between episodic and semantic memory. *Journal of Experimental Psychology: Learning and Cognition*, **12**(2), 212–314.

McKoon, G., Ratcliffe, R., and Dell, S. (1986). A critical evaluation of the semantic-episodic distinction. *Journal of Experimental Psychology: Learning and Cognition*, **12**(2), 295–306.

McNaughton, N. and Gray, J. (2000). *The neuropsychology of anxiety*. OUP.

Melo, B., Winocur, G., and Moscovitch, M. (1999). False recall and false recognition. *Cognitive Neuropsychology*, **16**, 343–359.

Milner, B., Corkin, S., and Teuber, H.-L. (1968). Further analysis of the amnesiac syndrome: 14 year follow up study of H.M. *Neuropychologia*, **6**, 215–234.

Minai, A. (1997). Covariance learning of correlated data in neural networks. *Neural Computation*, **9**(3), 667–681.

Minai, A. and Levy, W. B. (1993). The dynamics of sparse random networks. *Biological Cybernetics*, **70**, 177–187.

Morris, R. (1981). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience methods*, **11**, 47–60.

Moser, E., Moser, M., and Andersen, P. (1993). Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *J. Neuroscience*, **13**, 3916–3925.

Moser, M. B. and Moser, E. I. (1998a). Functional differentiation in the hippocampus. *Hippocampus*, **8**, 608–619.

Moser, M. B. and Moser, E. I. (1998b). Distributed encoding and retrieval of spatial memory in the hippocampus. *The Journal of Neuroscience*, **15**, 7535–7542.

Moser, M. B. and Moser, E. I. (2000). Pretraining and the function of hippocampal long term potentiation. *Neuron*, **26**, 559–561.

Moser, M. B., Moser, E. I., Forrest, E., Andersen, P., and Morris, R. (1995). Spatial learning with a minislab in the dorsal hippocampus. *PNAS*, **92**, 9697–9701.

Muller, R. and Kubie, J. (1989). The firing of hippocampal place cells predicts the future position of freely moving rats. *The Journal of Neuroscience*, **9**(12), 4101–4110.

Muller, R. and Kubie, J. L. (1987). The effects on changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of neuroscience*, **7**(7), 1951–1968.

Muller, R. U., Stead, M., and Pach, J. (1996). The hippocampus as a cognitive graph. *The Journal of General Physiology*, **107**(6), 663–694.

Murray, E. A. and Mishkin, M. (1998). Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *The Journal of Neuroscience*, **18**(16), 6568–6582.

Nader, K., Schafe, G., and le Doux, J. (2000). Fear memories require synthesis require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, **406**(6797), 722–726.

Naito, M. (1990). Repetition priming in children and adults: age related dissociation between implicit and explicit memory. *Journal of Experimental Child Psychology*, **50**, 174–177.

Norman, D. and Rumelhart, D. (1975). *Explorations in cognition*. Freeman.

Norman, K. A. and O'Reilly, R. (2001). Modelling hippocampal and neocortical contributions to recognition memory : A complementary learning systems approach. Technical report, University of Colorado.

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

O'Keefe, J. and Dostrovsky, T. (1971). The hippocampus as a spatial map. *Brain Research*, **31**, 171–175.

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

O'Reilly, R. and Rudy, J. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, **10**, 389–397.

Otmakhova, N. A. and Lisman, J. E. (1996). D1/ d5 dopamine receptor activation increases the magnitude of early long term potentiation at CA1 hippocampal synapses. *Journal of Neuroscience*, **16**(23), 7478–7486.

Otmakhova, N. A. and Lisman, J. E. (1998). D1/ d5 dopamine receptors inhibit depotentiation at CA1 synapses via camp dependent mechanism. *Journal of Neuroscience*, **18**(4), 1270–1279.

Otmakhova, N. A. and Lisman, J. E. (1999a). Dopamine selectively inhibits the direct cortical pathway to the CA1 hippocampal region. *Journal of Neuroscience*, **19**(4), 1437–1453.

Otmakhova, N. A. and Lisman, J. E. (1999b). Dopamine , serotonin, and noradrenaline strongly inhibit the direct perforant path- CA1 input but have little effect on the schaffer collateral input. *Annals of the New York Academy of Sciences*, pages 462–463.

Papatheodoropoulos, C. and Kostopoulos, G. (2000a). Decreased ability of temporal hippocampal region CA1 to produce long term potentiation. *Neuroscience Letters*, **286**(1), 57–60.

Papatheodoropoulos, C. and Kostopoulos, G. (2000b). Dorsal -ventral differentiation of short term synaptic plasticity in rat CA1 hippocampal region. *Neuroscience Letters*, **279**(3), 177–180.

Paulsen, O. and Moser, E. I. (1998). A model of hippocampal memory encoding and retrieval: Gabanergic control of synaptic plasticity. *TINS*, **21**(7), 273–278.

Penttonnen, M., Kamondi, A., Ascady, L., and Buzsaki, G. (1998). Gamma frequency oscillation in the hippocampus of the rat : intracellular analysis in vivo. *European Journal of Neuroscience*, **10**, 718–728.

Philips, R. and LeDoux, J. (1992). Differential contribution of the amygdala and hippocampus to cued and conditioned freezing. *Behavioural Neuroscience*, **106**(2), 274–285.

Philips, R. and LeDoux, J. (1994). Lesions of the dorsal hippocampus affect the background but not the foreground of contextual conditioning. *Learning and Memory*, **1**, 34–44.

Pike, F. G., Meredith, R. M., Olding, A., and Paulsen, O. (1999). Postsynaptic bursting is essential for ' hebbian' induction of associative long-term LTP at excitatory synapses in rat hippocampus. *Journal of Physiology*, **518**(2), 571–576.

Poe, G., Nitz, D., McNaughton, B., and Barnes, C. (2000). Experience-dependent phase reversal of hippocampal firing during rem sleep. *Brain Research*, pages 176–180.

Quirk, G. J., Muller, R., and Kubie, J. L. (1990). The firing of hippocampal place fields in the dark depends on the rats recent experience. *J Neuroscience*, **10**(6), 2008–2017.

Rainer, G., Rao, S., and Miller, E. (1999). Prospective coding for objects in primate prefrontal cortex. *The Journal of Neuroscience*, **19**(13), 5493–5505.

Ramos, J. (1998). Retrograde amnesia for spatial information,: a dissociation between intra and extra maze cues following hippocampus lesions in rats. *European Journal of Neuroscience*, **10**, 3295–3301.

Reed, J. M., Squire, L., Patalano, A. L., Smith, E., and Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioural Neuroscience*, **113**(3), 411–419.

Richmond, M. A., Yee, B. K., Pouzet, B., Veenman, L., Rawlins, J., Feldon, J., and Bannerman, D. (1999). Dissociating context and space within the hippocam-

pus:effects of complete, dorsal,and ventral excitotoxic hippocampal lesions on conditioned freezing and spatial learning. *Behavioural Neuroscience*, **113**(6), 1189–1203.

Riedel, G., Micheau, J., Lam, A. ., Roloff, E., Martin, S., Bridge, H., de Hoz, L., Poeschel, B., McCulloch, J., and Morris, R. (1999). Reversible neural inactivation reveals hippocampal participation in several memory processes. *Nature Neuroscience*, **2**(10), 898–905.

Rolls, E. T. (1996). A theory of hippocampal function in memory. *Hippocampus*, **6**, 601–620.

Rotenberg, A. and Muller, R. U. (1997). Variable place-cell coupling to a contiuously viewed stimulus: Evidence that the hippocampus acts as a perceptual system. *Philosophical Transactions of the Royal Society of London, series B*, **352**, 1505 –1513.

Ryle, G. (1949). *The concept of mind*. Hutchesin, reprinted Penguin.

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

Saucier, D. and Cain, D. P. (1995). Spatial learning without NMDA receptor dependent long term potentiation. *Nature*.

Schacter, D. L., Verfaillie, M., and Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesiac patients. *Journal of Memory and Language*, **35**, 319–334.

Seaman, J. K. and Durstwitz, D. (2001). Dopamine d1/d5 receptor modulation of excitatory synaptic inputs to layer v prefrontal cortex neurons. *Proc. Natl. Ac. USA*, **98**(1).

Shanks, D. and John, M. S. (1994). Characteristics of dissociable learning systems. *Behavioural and Brain sciences*, **17**, 367–447.

Shapiro, M. L., Tanilla, H., and Eichenbaum, H. (1997). Cues that hippocampal place cells encode : Dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, **7**, 624–642.

Sharp, P. E. (1999a). Subicular place cells expand or contract their spatial firing pattern to fit the size of the environment in an open field but not in the presence of barriers. *behavioural Neuroscience*, **113**, 643–662.

Sharp, P. E. (1999b). Complementary roles for hippocampal versus subicular / entorhinal place cells in coding place, context, and events. *Hippocampus*, **9**, 432–442.

Skaggs, W. and McNaughton, B. (1998). Spatial firing properties of two hippocampal CA1 populations in an environment containing two visually identical regions. *The journal of Neuroscience*, **18**(20), 8455–8466.

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

Small, S., Nava, A., Perera, G., R.DeLaPaz, Mayeux, R., and Stern, Y. (2001). Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nature Neuroscience*, **4**(4), 442.

Smith, A. C. and Levy, W. B. (2000). Controlling activity fluctuations in a large, sparsely connected random network. *Network: Computation in Neural Systems*, **11**, 63–81.

Sohal, V. and Hasselmo, M. (1998). GABAb modulation improves sequence disambiguation in computational models of hippocampal regions CA3. *Hippocampus*, **6**, 171–193.

Sperber, D. and Wilson, D. (1986). *Relevance 1st ed*. Blackwell's.

Sperber, D. and Wilson, D. (1995). *Relevance 2nd ed*. Blackwell's.

Squire, L. and Knowlton, B. (1995). Learning about categories in the absence of memory. *PNAS*, **92**, 12470–12474.

Squire, L., Clark, R., and Knowlton, B. (2001). Retrograde amnesia. *Hippocampus*, **11**(1), 50–53.

Squire, L. R. and Zola, S. M. (1998). Episodic memory,semantic memory and amnesia. *Hippocampus*, **9**, 205–211.

Strang, B., Fletcher, P., R. Hensen, K. F., and Dolan, R. (1999). Segregating the functions of the human hippocampus. *Proceedings of the National academy of Sciences U.S.A*, **96**, 40934–4039.

Sutherland, R. and Rudy, J. (1989). Configural association theory: the role of the hippocampal formation in learning memory and amnesia. *Psychobiology*, **17**(2), 129–144.

Sutherland, R., Weisand, M., Mumby, D., *et al.* (2001). Retrograde amnesia after hippocampal damage: Recent versus remote memories in two tasks. *Hippocampus*, **11**(1), 27–42.

Suzuki, W. A. and Eichenbaum, H. (2000). The neurophysiology of memory. *Annals of the new York academy of sciences*, **911**.

Swanson-Park, J. L., Coussens, C. M., Mason-Parker, S., Raymond, C., E.L.Haergreaves, Dragunow, M. ., Cohen, A., and Abraham, W. (1999). A double dissociation within the hippocampus of dopamine D1/D5 receptor and beta-adrenergic receptor contributions to the persistence of LTP. *Neuroscience*, **92**(2), 485–497.

Tamamaki, N. and Nojyo, Y. (1991). Crossing fibre arrays in the rat hippocampus as demonstrated by three-dimensional reconstruction. *The Journal of Comparative Neurology*, **303**, 435–442.

Tanila, H. (1999). Hippocampal place cells can develop distinct representations of two visually identical environments. *Hippocampus*, pages 235–346.

Thierry, A., Gioanni, Y., Degentais, E., and Glowinski, J. (2000). Hippocampal - prefrontal cortex pathway: Anatomical and electrophysiological characteristics. *Hippocampus*, **10**(4), 411–419.

Thompson, L. T. and Best, P. J. (1990). Long term stability of place -field activity recorded from the dorsal hippocampus. *Brain Research*, **509**, 299–308.

Traub, R., Sprunston, N., Soltesz, I., Konnerth, A., Whittington, M., and Jefferys, J. (1998). Gamma frequency oscillations : a neuronal population phenomenon, regulated by synaptic and intrinsic cellular processes, and inducing synaptic plasticity. *Progress in Neurobiology*, **55**, 563–575.

Traub, R. D., Jefferys, J. G. R., and Whittington, M. A. (1999). *Fast Oscillations in Cortical Circuits*. MIT Press.

Tulving, E. (1983). *Elements of Episodic memory*. Oxford: Clarendon press.

Tulving, E. (1986). What kind of hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning and Cognition*, **12**(2), 307–311.

Tulving, E. and Markowitsch, M. (1998). Episodic and declarative memory:role of the hippocampus. *Hippocampus*, **8**, 198–204.

Urban, N. N. and Barrionuevo, G. (1998). Active summation of excitatory postsynaptic potentials in hippocampal CA3 pyramidal neurons. *Proceedings of the National Academy of sciences*, **95**, 1140–11455.

Vargha-Khadem, F., Gadian, D. G., Watkin, K. E., *et al.* (1997). Differential effect of early hippocampal pathology on episodic and semantic memory. *Science*, **277**, 376–380.

Verney, C., Baulac, M., Berger, B., Alvarez, C., Vigny, A., and Helle, K. (1985). Morphological evidence for a dopaminergic terminal field in the hippocampal formation of the young and adult rat. *Neuroscience*, **14**(4), 1039–1052.

Vertes, R. P. and Kocsis, B. (1997). Brainstem-diencephelo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience*, **81**(4), 893–926.

Vinogradova, O. S. (1995). Expression , control and probable functional significance of the neuronal theta rhythm. *Progress in Neurobiology*, **45**, 523–583.

Vizi, E. and Kiss, J. (1998). Neurochemistry and neuropharmocolgy of the major hippocampal transmitter systems: synaptic and non synaptic interactions. *Hippocampus*, **8**, 566–607.

Wallenstein, G. V. and Hasselmo, M. E. (1997). Gabanergic modulation of hippocampal population activity: Sequence learning, place field development, and phase precession effect. *Journal of Physiology*, **78**, 393–408.

Wallenstein, G. V., Eichenbaum, H., and Hasselmo, M. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neuroscience*, **21**(18), 317–323.

Wan, H., Aggleton, J., and M.Brown (1999). Differential contributions of the hippocampus and perirhinal cortex to recognition memory. *The Journal of Neuroscience*, **19**(3), 1142 –1148.

Warrington, E. and Weiskrantz, L. (1970). The amnesiac syndrome: consolidation or retrieval? *Nature*, **228**, 628–630.

Weiskrantz, L. (1997). *Consciousness lost and found: A neuropsychological perspective..* OUP.

Wickliffe, C. A., editor (1991). *Memory Mechanisms. A tribute to G. V. Goddard*, chapter 1-5, pages 3–105. Lawrence Erlbaum Associates.

Willshaw, D. (1971). *Models of associative memory*. Ph.D. thesis, University of Edinburgh.

Willshaw, D. and Dayan, P. (1990). Optimal plasticity in matrix memories: what goes up must come down. *Neural Computation*, **2**, 85–93.

Willshaw, D., Hallam, J., Gingell, S., and Lau, S. (1996). Marr's theory of the neocortex. *Neural Computation*, **9**(4), 911–936.

Willshaw, D. J. and Buckingham, J. T. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil Trans Roy Soc Lond B*, **329**, 205–215.

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, **222**, 960–962.

Wilson, M. and McNaughton, B. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, **265**, 676–679.

Wilson, R. C., Levy, W. B., and Steward, O. (1981). Changes in translation of synaptic excitation to dentate granule cell discharge following LTP. *Journal of neurophysiology*, **46**, 339–355.

Winocur, G., McDonald, R., and Moscovitch, M. (2001). Anterograde and retrograde amnesia in rats with large hippocampal lesion. *Hippocampus*, **11**, 18–26.

Witter, M. (2000). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus*, **10**, 398–410.

Wood, E., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, **613-615**.

Wood, E., Dudchenko, P., *et al.* (2000). Hippocampal neurons encode information about different types of memory episodes occuring in the same location. *Neuron*, **27**, 623–633. Author list incomplete.

Wu, X., Baxter, R. A., and Levy, W. B. (1996). Context codes and the effect of noisy learning on a simplified hippocampal CA3 model. *Biological Cybernetics*, **74**, 159–165.

Wu, X., Tyrcha, J., and Levy, W. B. (1998). A neural network solution to the transverse patterning problem depends on repetition of the input code. *Biological Cybernetics*.

Wyble, B. P., Linster, C., and Hassellmo, M. (2000). Size of CA1-evoked synaptic potentials is related to theta rhythm phase in rat hippocampus. *Journal of Physiology*.

Xu, L., Amwyl, R., and Rowan, M. (1998). Spatial exploration induces a persistent reversal of long term potentiation in the rat hippocampus. *Nature*, **394**, 891.

Yang, S. N. (2000). Sustained enhancement of AMPA receptor and NMDA receptor mediated currents induced by dopamine d1/d5 repecptor activation: an esentail role of postsynaptic Ca2+. *Hippocampus*, **10**(1).

Yeckel, M. and Berger, T. (1998). Spatial distribution of potentiated synapses in hippocampus: Dependence on cellular mechanisms and network properties. *The Journal of Neuroscience*, pages 438–450.

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, **10**(2), 403–430.

Zola, S. M., Squire, L. R., *et al.* (2000). Impaired recognition memory in monkeys after damage limited to hippocampal region. *The Journal of Neuroscience*, **20**(1), 451–463.