COORDINATED MEMORY TRACE REACTIVATION ACROSS DISTRIBUTED NEURAL ENSEMBLES IN THE PRIMATE NEOCORTEX

by

Kari Lee Hoffman

A Dissertation Submitted to the Faculty of the GRADUATE INTERDISCIPLINARY PROGRAM IN NEUROSCIENCE In Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY In the Graduate College THE UNIVERSITY OF ARIZONA

2003
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Karl Lee Hoffman entitled Coordinated Memory Trace Reactivation Across Distributed Neural Ensembles in Primate Neocortex and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Bruce McNaughton
Date 4/25/03

Carol Barnes
Date 4/25/03

Lynn Nadel
Date 4/24/03

Lee Ryan
Date 4/24/03

Peter De Weerd
Date 4/24/03

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director Bruce McNaughton
Date May 14, 2003
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
ACKNOWLEDGEMENTS

The project I undertook for my dissertation was an unruly beast. Fortunately, I was not alone in this journey, and the help and support I received allowed me to complete the project.

I thank my advisor, Dr. Bruce McNaughton, for his fearlessness and creativity, for sharing with me glimpses of his genius, and for sticking his neck out in support of this project. Thanks to my co-advisor, Dr. Carol Barnes, who knew exactly what I needed to hear in order to press on; without her courage and strength I would not have finished. The lab environment they have fostered is remarkable, and I thank my lucky stars that I was able to be a part of it.

I’m grateful for the helpful comments and guidance of my other committee members: Lynn Nadel, Peter De Weerd, Lee Ryan, and Rich Zemel.

The tireless efforts of a fleet of helpful hands and brains made the project possible. I’m grateful for the assistance I received from: Chad Anderson, Francesco Battaglia, Kate Chemodurow, Jannice de Dios, Tim Ellmore, Andy Fuglevand, Kati Gothard, Kimmey Hardesty, Nihit Kaul, Peter Lipa, Kevin Spitler, Casey Stengel and the Neuralynx crew.

I will forever think about the brain, our society, and the universe in different ways thanks to the numerous conversations that took place in the hallowed halls of NSMA. To Kati, Andy, Francesco, Stephen, Jim, Geeta, Doug (Wellington and Nitz), David (Redish and Euston), Lauren, F, Marsha, Tim, Mark, Jason, Peter: thanks for keeping it fun and interesting.

Mom, Dad: you set me up for a life of exploration. Wherever I go, you are with me. Thank you for everything.
TABLE OF CONTENTS

LIST OF ILLUSTRATIONS ........................................................................................................... 1

LIST OF TABLES ........................................................................................................................ 8

ABSTRACT ..................................................................................................................................... 9

1 INTRODUCTION .......................................................................................................................... 10

1.1 MEMORY CONSOLIDATION ................................................................................................. 12

1.1.1 Retroactive Interference .......................................................................................... 12

1.1.2 Electroconvulsive therapy and epilepsy ................................................................... 14

1.1.3 Medial temporal lobe damage: the hippocampal connection .................................... 19

1.1.4 Theories of memory consolidation ........................................................................... 44

1.1.5 Conclusion .................................................................................................................. 58

1.2 SLEEP AND MEMORY CONSOLIDATION ......................................................................... 60

1.2.1 Introduction ................................................................................................................ 60

1.2.2 REM ............................................................................................................................ 63

1.2.3 Non-REM .................................................................................................................... 66

1.2.4 Dual process and serial theories .............................................................................. 71

1.3 CORTICAL NETWORK ACTIVITY ..................................................................................... 76

1.3.1 Neocortical networks ............................................................................................... 76

1.3.2 Hippocampal networks ............................................................................................. 85

1.3.3 Hippocampo-cortical interactions .......................................................................... 88

1.4 MEMORY TRACE REACTIVATION ..................................................................................... 92

1.4.1 Reactivating sequences ............................................................................................. 96

1.4.2 Sleep state: SWS versus REM ............................................................................... 98

1.4.3 Beyond the hippocampus ........................................................................................ 99

1.4.4 Persistent activity in the primate brain .................................................................... 100

1.4.5 Conclusion ................................................................................................................ 102

2 METHODS ............................................................................................................................... 104

2.1 ELECTRODE DRIVE DEVELOPMENT ........................................................................... 104

2.1.1 Circuit board development ....................................................................................... 105

2.1.2 Electrodes .................................................................................................................. 109

2.2 IMPLANT PROCEDURES ................................................................................................. 111

2.2.1 Pre-operative preparation ....................................................................................... 111

2.2.2 Implant surgery ......................................................................................................... 117

2.3 ORGANIZATION OF DAILY SESSIONS ......................................................................... 120

2.3.1 Head restraint system .............................................................................................. 120

2.3.2 Electrode advancing ................................................................................................. 122

2.3.3 Recording procedures .............................................................................................. 125

2.3.4 Behavioral procedures .............................................................................................. 127

2.4 CELL ISOLATION AND LOCALIZATION ....................................................................... 130
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 RESULTS</td>
<td>132</td>
</tr>
<tr>
<td>3.1 GENERAL</td>
<td>132</td>
</tr>
<tr>
<td>3.1.1 Multiple single unit activity</td>
<td>134</td>
</tr>
<tr>
<td>3.1.2 Relation between neural activity and behavior</td>
<td>139</td>
</tr>
<tr>
<td>3.2 MEMORY TRACE REACTIVATION</td>
<td>141</td>
</tr>
<tr>
<td>3.2.1 Overall memory trace reactivation</td>
<td>144</td>
</tr>
<tr>
<td>3.2.2 Stability of cells and ensemble states during rest</td>
<td>145</td>
</tr>
<tr>
<td>3.2.3 Within and across arrays</td>
<td>147</td>
</tr>
<tr>
<td>3.3 REACTIVATION OF SEQUENCES</td>
<td>152</td>
</tr>
<tr>
<td>3.4 VIGILANCE STATE AND MEMORY TRACE REACTIVATION</td>
<td>156</td>
</tr>
<tr>
<td>3.5 MEMORY TRACE REACTIVATION OVER TIME</td>
<td>157</td>
</tr>
<tr>
<td>3.6 OPTIMIZATION OF MEMORY TRACE REACTIVATION ANALYSES</td>
<td>161</td>
</tr>
<tr>
<td>3.7 CONCLUSION</td>
<td>164</td>
</tr>
<tr>
<td>4 DISCUSSION</td>
<td>165</td>
</tr>
<tr>
<td>4.1 TECHNOLOGICAL AND PROCEDURAL DEVELOPMENTS</td>
<td>165</td>
</tr>
<tr>
<td>4.2 MEMORY TRACE REACTIVATION</td>
<td>168</td>
</tr>
<tr>
<td>4.2.1 Dorsal prefrontal cortex</td>
<td>168</td>
</tr>
<tr>
<td>4.2.2 Reactivation of a distributed memory trace</td>
<td>172</td>
</tr>
<tr>
<td>4.2.3 The nuts and bolts of memory trace reactivation</td>
<td>175</td>
</tr>
<tr>
<td>4.3 MEMORY CONSOLIDATION</td>
<td>180</td>
</tr>
<tr>
<td>4.3.1 Trace replay theory as a mechanism for memory consolidation</td>
<td>180</td>
</tr>
<tr>
<td>4.3.2 Reconsolidation</td>
<td>181</td>
</tr>
<tr>
<td>4.4 CONCLUSION</td>
<td>192</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>193</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

Figure 1.1 Possible patterns of retrograde amnesia ................................................................. 21
Figure 2.1 54-Channel electrode drive ..................................................................................... 106
Figure 2.2 Smallest configuration of 4, first-generation 144-channel electrode drives .......... 108
Figure 2.3 Second-generation 144-channel drive ................................................................. 109
Figure 2.4 Orthoview renderings of the pre-operative MRs .................................................. 111
Figure 2.5 Target array coordinates superimposed on a magnetic resonance rendering of the brain ................................................................. 113
Figure 2.6 Drive encasement design ....................................................................................... 116
Figure 2.7 Illustrations of the head restraint-implant interface ............................................. 121
Figure 2.8 Pusher software display ........................................................................................ 124
Figure 2.9 Cheetah display during data acquisition ............................................................... 126
Figure 2.10 Virtual Alley Maze .............................................................................................. 129
Figure 2.11 BBClust output from one electrode ................................................................. 130
Figure 2.12 MR and CT co-registration ................................................................................ 131
Figure 3.1 Physiological classification of cells ...................................................................... 135
Figure 3.2 Average waveforms of 99 simultaneously recorded cells ...................................... 137
Figure 3.3 Firing rate histograms for cells in each array during each epoch ......................... 138
Figure 3.4 Task-related neural activity .................................................................................. 140
Figure 3.5 Explained variance measure of memory trace reactivation overall and across sessions ................................................................. 144
Figure 3.6 Explained variance using different 10-minute blocks of Rest 1 data .................. 146
Figure 3.7 Total count of pairs of cells eligible for spike-train correlations ......................... 147
Figure 3.8 Explained variances for pairs of cells within and across each array .................... 148
Figure 3.9 Distribution of cell pair correlations for PFC and non-PFC cell pairs across behavioral epochs ................................................................. 150
Figure 3.10 Normalized cell sparsity index distributions for non-PFC and PFC cells .......... 151
Figure 3.11 Example cross-correlograms of two cells from the task and the flanking rest epochs ....................................................................................... 153
Figure 3.12 Preservation of temporal bias within and across arrays ..................................... 154
Figure 3.13 Distribution of bias values for PFC and non-PFC cell pairs across behavioral epochs ................................................................. 155
Figure 3.14 Memory trace reactivation during sleep and quiet wakefulness ........................ 156
Figure 3.15 Explained variance above control levels as a function of Rest 2 delay ............... 157
Figure 3.16 Distribution of cell pair correlations during Rest 1 and Task for Novel and Familiar tasks ....................................................................................... 159
Figure 3.17 Explained variances after removing sparse population vectors from the task epoch ....................................................................................... 162
Figure 4.1 Coordinated memory trace reactivation in three of four brain regions ............... 173
Figure 4.2 Reconsolidation as explained by trace replay theory ........................................... 187
LIST OF TABLES

Table 1.1 Summary of prospective studies of retrograde amnesia (RA) in animals with damage to the hippocampal formation. ................................................................. 33

Table 1.2 Relationship between different sleep stages and learning of declarative or procedural memories.............................................................................................. 75

Table 1.3 Functional classification of neocortical cells.......................................................................................................................... 77

Table 2.1 Cortical connectivity of targeted brain regions....................................................................................................................... 114

Table 3.1 Summary of cell counts and behavioral task for eligible sessions...................................................................................... 132

Table 3.2 Time spent with eyes closed during 10-minute rest epoch blocks...................................................................................... 133
ABSTRACT

The process of forming a long-lasting memory may involve the selective linking together of neural representations stored widely throughout neocortex. The successful binding together of these disparate representations may require their coordinated reactivation while the cortex is 'offline' i.e., not engaged in processing external stimuli. This hypothesis was tested through simultaneous extracellular recording of 28-99 cells over four sites in the macaque neocortex. The recordings were conducted as the monkey performed repetitive reaching tasks, and in rest periods immediately preceding and following the task. In motor, somatosensory and parietal cortex (but not prefrontal cortex), the task-related neural activity patterns within and across regions were similar to the activity patterns seen afterwards, during the rest epoch. Moreover, the temporal sequences of neural ensemble activity that occurred during task performance were preserved in subsequent rest. The preservation of correlation structure and temporal sequencing are consistent with the reactivation of a memory trace and not merely the persistence of a fixed activity pattern. The observed memory trace reactivation was coordinated over large expanses of neocortex, confirming a fundamental tenet of the trace replay theory of memory consolidation.
1 INTRODUCTION

Musicians are taught that the cadences, or breaks, in music are as important as the notes. Playwrights and actors are revered for their use of pauses in dialogue, for resisting the temptation to fill up spaces with words. Paradoxically, our appreciation of a performance may be based on what isn’t there. Even outside the domain of art appreciation, we benefit from the occasional absence of input: allowing breaks between bouts of learning helps us to remember that material later. Athletes and musicians experience performance improvements ‘for free’ after a night’s break in activity, during sleep. Moving to the biology inside our heads, the absence of experience during sleep produces as much cortical reorganization as continued experience. What could be the benefit of such pauses? How is it that ‘not practicing’ makes perfect?

Over a century of research on memory consolidation has established that, after learning, time free of interruption is required for the memory to become firmly established. Sleep represents an exaggerated period ‘free of interruption’, and is a time of heightened consolidation. What is striking about sleep is that, contrary to intuition, the brain is still quite active, revealing an important clue to the benefit of ‘breaks’ in learning. The activity during sleep is different from that of wakefulness, demonstrating a remarkable level of synchronized activity throughout the cortex. One implication of synchrony in neural networks is the potential for Hebbian learning in the participating neurons. Although it is not clear which neurons participate in the episodes of neocortical
synchrony, there is evidence that neurons encoding a previously experienced event are synchronously reactivated in sleep.

The significance of the last observation becomes clear after considering the sleep and memory consolidation literature in more detail, which is the aim of the introduction that follows. But to reduce the guiding principle that will emerge to a phrase: *out of sight is not out of mind.*
1.1 MEMORY CONSOLIDATION

Mueller and Pilzecker (1900), were the first to report a behavioral procedure that interfered with memory consolidation. They discovered that when subjects were required to learn a second list of nonsense syllables immediately after learning the first list, memory for the first list was impaired. The authors speculated that some type of neural 'fixation' continues even in the absence of the stimuli to be learned. They further suggested that new material given during this preservation period would disrupt fixation, impairing memory for the original material.

Burnham (1903) noted that the preservation-consolidation hypothesis also explained previous reports that cerebral trauma led to a greater loss of recent than remote memories (Ribot 1882), since the recent memories would not have had sufficient 'fixation' time. This time-limited, or graded, retrograde amnesia (RA) is therefore sometimes referred to as a Ribot gradient. Much of the evidence for memory consolidation is based on observations of graded retrograde amnesia that follow any of several types of interfering manipulations, described below.

1.1.1 Retroactive Interference

The least invasive way to disrupt the memory for one task is to immediately follow its acquisition with a second, similar task. This procedure results in an effect known as retroactive interference. It was used by Mueller and Pilzecker (1900) at the turn of the century, but was not popularized until decades later, by psychologists interested in
the process of forgetting (Jenkins and Dallenbach 1924; Underwood 1948; Briggs 1954).

In the standard protocol, experimental and control groups are exposed to learning and recall stages in the following order:

Experimental group: Learn Set A  Learn Set B  Time passes  Recall Set A
Control group: Learn Set A  Do nothing  Time passes  Recall Set A

In this study the experimental group learns a second word list immediately afterwards whereas the control group completes some nonverbal activity to prevent rehearsal or learning similar verbal material. After a delay, both groups are asked to recall the original list. The result is better recall in the control group than in the group that experienced the intervening (Set B) list.

Interference can also result when a distractor list precedes the target list (proactive interference), leading to the supposition that presentation order is irrelevant, and that both interference effects are due to competition among similar tasks. Melton and Irwin (1940) addressed this issue by calculating the degree of competition based on the number of intrusion responses from the original list during learning of the second list (which varied in the number of repetitions given). They then compared the degree of competition to the level of retroactive interference. The result was a large increase in retroactive interference incongruent with the effects of competition, establishing that retroactive interference involved a temporal component above and beyond the effect of competition. Further experiments on retroactive interference support the conclusion that learning a second set of material can interfere with critical processing of the original material, if the second set is presented within a limited time window of the original material. Time-limited interference has even been observed in motor skill learning, suggesting interference may
be a fundamental feature of memory processing (Brashers-Krug, Shadmehr et al. 1996; Shadmehr and Brashers-Krug 1997).

1.1.2 Electroconvulsive therapy and epilepsy

Normal electrical activity of the brain is important for memories to consolidate. Duncan (1949) and Gerard (1949) were the first to use electroconvulsive (EC) shock to interfere with memory consolidation. In the first EC experiments, shocks were delivered to rodents at various time points from seconds to hours following a daily active avoidance trial. After 18 days of treatment, retrograde amnesia was evident in the shorter delay groups, suggesting that the memory required some time after learning to consolidate, after which it was resistant to shock treatments. The delays between acquisition and disruptive EC shocks vary, depending on the task and, in particular, on the criteria used to measure memory strength. But the overall observation that EC shocks delivered shortly after acquisition produce amnesia is remarkably consistent (see Glickman 1961; McGaugh and Petrinovich 1966; Miller 1968, for reviews).

Humans rarely are subjected to electrical shock; however, the electrical induction of a series of grand mal-type seizures, known as electroconvulsive therapy (ECT), has been used for therapeutic purposes in some cases of severe depression. The most dire consequence of this treatment is retrograde amnesia (Zubin and Barrera 1941; Janis 1948; Janis and Astrachan 1951; Squire 1986). Specifically, events that occurred days or weeks prior to ECT are frequently forgotten, and occasionally memory impairments can extend
back months or years (Squire 1975; Squire, Slater et al. 1975; Squire, Slater et al. 1981; Squire and Slater 1983; Frith, Stevens et al. 1987).

It has been suggested that autobiographical memories are more susceptible to disruption (Janis and Astrachan 1951), but a more recent study found the opposite pattern of results (Lisanby, Maddox et al. 2000). A Personal Impersonal Memory Test was constructed to compare autobiographical from impersonal memories on equal terms. Prior to ECT treatments, several categories of personal and of impersonal events were presented to subjects, who recalled as many examples as possible from the preceding four years. Subjects then estimated the date each memory occurred, and were asked to provide as many details as possible for the most recent (within 3 months old) and most remote (at least 3 years old) personal and impersonal memory. Thus, both the quantity and quality of memories could be assessed, and the quality of <3 month-old and >3-year old memories could be compared. After ECT, or an equivalent time interval, retest performance was compared to the pre-treatment baseline. The experimental group of clinically depressed patients (n=55) recalled about 20% fewer personal and 30% fewer impersonal memories after ECT than in baseline testing, whereas control subjects (n=36) showed no change. Patients also described fewer details overall, and proportionately fewer details of recent than of remote events relative to controls, indicating a graded retrograde amnesia.

One mechanism by which ECT could disrupt memories is by directly changing the electrical firing characteristics of neurons, though the biological effects of ECT may be far more extensive. ECT may alter the concentration of neuromodulators or metabolic activity, which may, in turn, affect neural activity. It is important to consider that memory
disruption could be the result of any number of interacting sequellae of the seizures induced by ECT, and not simply the result of direct modification of electrical changes in neural activity.

Transcranial magnetic stimulation (TMS) procedures, which do not induce seizures, may provide a more restricted means of altering the electrical properties of neurons. Tests of motor learning after an interfering TMS procedure provide converging evidence for a memory consolidation period (Muellbacher, Ziemann et al. 2002). The result of TMS applied over motor cortex was a disruption in performance of an elementary motor task learned immediately prior to stimulation. In contrast, no performance deficits were observed when a 6-hour delay was imposed between acquisition and TMS, indicating a time-dependent susceptibility of the memory. The evidence from ECT and TMS both support the idea that acute electrical activity delivered to the brain can interfere with memory consolidation. Based on the ECT and TMS results, yet another, ‘organic’ alteration of neural electrical activity might be expected to disrupt memory consolidation: epileptic seizures.

Memory experiments on epileptics typically test short-term learning and retention, in the absence of intervening seizures. They frequently report no acquisition or recall impairments. If, however, epilepsy selectively alters consolidation, leaving acquisition and recall abilities intact, one would expect to observe impairments only after the occurrence of intervening seizures. Indeed, the few reports that test recall at longer time periods consistently show that epileptics suffer retrograde amnesia.
A review of transient epileptic amnesia (TEA) describes not only the characteristic episodes of transient anterograde amnesia associated with seizure activity, but also a graded RA for events that predate the onset of the TEA condition (Zeman, Boniface et al. 1998). One case study of an epileptic with a temporal lobe focus showed graded retrograde amnesia apparently surrounding epileptic attacks (Kapur, Young et al. 1989). Case studies (Kapur, Millar et al. 1997; O'Connor, Sieggreen et al. 1997) and a group study of 23 epileptics (Blake, Wroe et al. 2000), show that memory recall within 30 minutes is unimpaired (presumably with no intervening seizures), but memory tested at 1, 6 or 8 weeks was impaired for word lists, story and design recall, and story recall, respectively. The group study reported an average of 19 seizures per month, so it is assumed that dozens of seizures occurred between learning and retention.

In the two case studies, retrograde amnesia was also reported for famous personalities and news events occurring over one or two decades in the respective cases. In the latter study (O'Connor, Sieggreen et al. 1997), a 3-year follow-up test revealed worse recall specifically for material learned in the last decade, reflecting the ongoing temporal specificity of RA with continued seizures. These studies all report temporal lobe seizure foci, giving some indication that memory may be particularly sensitive to dysfunction in this brain region.

Not all evidence suggests that seizures selectively interfere with memory consolidation. A second group study tested autobiographical memory in 25 unilateral temporal lobe epileptics, 12 of whom had undergone medial temporal lobe removal (Viskontas, McAndrews et al. 2000). No differences were reported for pre- versus post-
operative groups, so the results are presumably representative of deficits in medial temporal lobe epileptics. Using the Autobiographical Memory Interview (Kopelman, Wilson et al. 1989) subjects showed memory impairments for personally-experienced events but not for autobiographical facts. Recall of three personal events was scored up to 3 points each for “descriptive richness and specificity in time”, for a total of 9 possible points in each of the 3 time periods (age 0-18, 18-30, and within the last 5 years). A flat impairment for personal event memories was observed, with patients scoring roughly 5/9 points for each time period, and controls scoring 8/9. The use of only four levels of scoring per episode, and only three events per time period, may limit the sensitivity of the test. Although clear differences between control and patient groups emerged, more subtle differences across time periods and between pre- and post-operative patients may have simply gone undetected with this measure.

The collective results of ECT and TMS studies suggest that electrical interference can cause graded retrograde amnesia without lasting impairments of subsequent learning. In epileptics, graded RA is reported in three case studies and two group reports, and flat RA is reported in a group study in which the experimental group contains some former epileptics with medial temporal lobe resections. Since the epileptics showing graded RA all had temporal lobe foci, these reports further suggest retrograde amnesia is associated with medial temporal lobe dysfunction.
1.1.3 **Medial temporal lobe damage: the hippocampal connection**

The effects of medial temporal lobe (MTL) damage in humans has become a major focus of memory research; patients often suffer from the inability to form new memories (anterograde amnesia), but also from the disproportionate loss of recently formed memories, or graded RA (Scoville and Milner 1957; Walker 1957; Penfield and Milner 1958; Zola-Morgan, Squire et al. 1986; Beatty, Salmon et al. 1987; Salmon, Lasker et al. 1988; Squire, Haist et al. 1989; Squire 1992; Yoneda, Yamadori et al. 1992; Rempel-Clower, Zola et al. 1996; Nadel and Moscovitch 1997; Reed and Squire 1998; Kapur and Brooks 1999; Spiers, Maguire et al. 2001 ). In non-human animals, damage restricted to the hippocampus, located in the heart of the medial temporal lobe, also produces graded RA, though there are considerable differences in the tasks given to animals and those given to humans (Winocur 1990; Zola-Morgan and Squire 1990; Kim and Fanselow 1992; Kim, Clark et al. 1995; Anagnostaras, Maren et al. 1999; Winocur, McDonald et al. 2001). These results have been interpreted as evidence of a time-limited role of the hippocampus in forming memories (Scoville and Milner 1957; Buzsaki 1989; Milner 1989; Squire 1992; McClelland, McNaughton et al. 1995; Squire and Alvarez 1995; Knowlton and Fanselow 1998).

But whereas some studies show graded RA, others show no RA, and still others show that retrograde amnesia is just as strong for events in the distant past (‘flat’ RA) (Sanders and Warrington 1971; Albert, Butters et al. 1981; Cermak and O'Connor 1983; Beatty, Salmon et al. 1988; Tulving, Schacter et al. 1988; Warrington and McCarthy...

The interpretation of these results has led to a variety of conclusions. Take, for example, a recent issue of the journal Hippocampus, in which the following four opinions were expressed:

"...the idea that has come out of retrograde amnesia studies, namely, that memory is reorganized as time passes and gradually becomes independent of the hippocampus is rather vague...Nevertheless, the evidence that something like this occurs is substantial. In the case of patients with damage limited to the hippocampal formation, temporally graded retrograde amnesia has been observed repeatedly in the clinic and the laboratory." (Squire, Clark et al. 2001, p.54)

"...there is no clear-cut answer to the question, 'Does the MTL play a time-limited role in information storage?'" (Murray and Bussey 2001, p.6)

"...it has been assumed for some time that early, fragile memory depends on the hippocampal system, while later, stable memory does not...Although this view is widely accepted, the data in support of it are equivocal, at best." (Nadel and Bohbot 2001, p.56)

"...experiences that occur long before hippocampal damage will be remembered better than events closer to the time of damage, resulting in a temporally graded retrograde amnesia. This pattern is expected on the principle that the consolidation process will be complete for old memories, and access to them does not depend on the hippocampus...this approach has received considerable support in the animal and human literature." (Winocur, McDonald et al. 2001, p.18)

Confounding the issue, certain results are considered evidence for graded RA by some, and evidence of flat RA by others. For the sake of clarity and consistency within this document, Figure 1.1 describes several possible patterns of results and the type of RA they indicate.
Figure 1.1 Possible patterns of retrograde amnesia. The right side of each figure represents the longest delay between learning and brain damage in the lesioned group, whose performance is indicated by the red line. Black lines indicate control group performance. A. No difference between groups at any time point. B. Similar performance deficits of the lesioned group at all time points, with forgetting in both groups. C. Similar performance deficits at all time points, but no forgetting. D. Diminishing deficits with increasing memory age in the lesioned group and no forgetting in the control group. E. Diminishing deficits in the lesioned group and forgetting in the control group. The dotted line represents chance levels of performance. F. Diminishing deficits over time, with significant impairments even at the most remote time point, indicated by the arrowheads. Plots B and C show flat RA, whereas plots in the bottom row show graded RA, with the provision that in E, the lesioned group’s performance must be above chance. F indicates graded RA and a persistent dependence on the lesioned structure. Note that graph D would show the same results as F if the most remote time point in D was removed. Thus, one explanation for results resembling those in F is that the range of time points sampled was too limited.

What determines whether hippocampal damage leads to no RA, time-limited RA, or flat RA? One answer may lie in the type of memory being encoded. For example, there are no systematic deficits in memory for many types of skills or habits following hippocampal or medial temporal lobe damage in non-human primates (Salmon 1987) or humans (Scoville and Milner 1957; Penfield and Milner 1958; Squire, Cohen et al. 1984; Beatty, Salmon et al. 1987; Gabrieli, Corkin et al. 1993; Squire and Zola 1997).
Repetition priming is also spared (Milner, Corkin et al. 1968; Gabrieli, Milberg et al. 1990). Patients may report that they have never practiced the tasks before, even when they show clear improvements. These tasks are therefore called implicit (or procedural) memory tasks, to contrast with explicit (or declarative) tasks which require conscious, or explicit, recall, and which are sensitive to MTL damage.

Not all types of implicit memories are spared following MTL damage, however. One class of implicit tasks involves the non-conscious use of contextual cues, where context includes relational information about the spatial arrangement of stimuli in the environment. When normal subjects view a scene and report no noticeable changes, they nevertheless spend more time examining regions in which the relation among objects has been altered (Ryan, Althoff et al. 2000). Amnesics fail to show this implicit effect. Similarly, detection of a target embedded among distractors is normally facilitated when the target is presented in the same spatial location for a given configuration of distractors (Chun and Phelps 1999). Facilitation occurs even though subjects are unaware of the repetition. In contrast, amnesics with MTL damage fail to show facilitation on this implicit task, despite no impairments on other implicit memory tasks. It would appear that when performance depends on the integration of spatial relationships among cues, medial temporal lobe function is necessary. Typically, this kind of processing falls in the domain of explicit memory, but one class of implicit tasks shares this dependence on relational information. With the exception of these two studies, implicit memory performance does not benefit from relational information, thus when tasks are referred to as 'implicit', only non-relational implicit tasks are considered.
When damage occurs outside the medial temporal lobe, patients are impaired on implicit but not explicit memory tasks (Keane, Gabrieli, et al. 1995; Fleischman, Vaidya et al. 1997; Vaidya, Gabrieli et al. 1998). Although there are reports of amnesia covering both explicit and implicit material, there is no evidence indicating that this broad effect can result from damage restricted to the medial temporal lobe (Andrews, Poser et al. 1982; Rousseaux, Delafosse et al. 1984). The dissociations between MTL and extra-MTL damage, and between explicit and implicit tasks, memory impairments suggest separate neural systems are involved in each type of recall. Specifically, hippocampal damage leads to no RA for non-relational implicit tasks, but to some type of RA for explicit tasks. By extension, explicit memory might consist of classes of memory that result in either flat or graded RA following MTL damage.

**Episodic versus semantic memory**

Memory research has led psychologists to make a distinction between episodic explicit memories and semantic explicit memories (Tulving 1972). *Episodic memory* is "the conscious recollection of contextually bound personal experiences that occur at a particular time and place." (Rosenbaum, Winocur et al. 2001, p. 184). *Semantic memory* is defined as “knowledge about the world that is divorced from the context in which it was acquired...just about any kind of memory that is neither episodic nor procedural qualifies as semantic.” (Nadel and Moscovitch 1998, p. 435).
Specific anterograde deficits as evidence for separate types of memory

Although much semantic knowledge may be learned quickly and efficiently through episodic learning, it is possible to gain semantic knowledge in the absence of episodic memories. For example, in Kinsbourne and Wood (1975) it was reported that amnesic patients had deficits in autobiographical (episodic) memory in single word cue tests but that they had no trouble retrieving general (semantic) information from the cues. Semantic memories have been observed during infantile amnesia, when creation of lasting episodic memories is not yet possible (Nadel and Zola-Morgan 1984), as well as in cases of childhood damage to the hippocampus (Vargha-Khadem, Gadian et al. 1997; Gadian, Aicardi et al. 2000), in which episodic but not semantic memory deficits are observed. Another dissociation includes the observation of priming for semantic but not episodic autobiographical information (Tulving, Schacter et al. 1988; van der Linden and Schils 1992; Squire and Zola 1998). Moreover, even when episodic memory of a learning session is impaired, factual information can be learned, albeit at a slower rate (Hamann and Squire 1995). Once learned, amnesics can use the memories as flexibly as controls. Semantic and episodic memories appear to be supported by separate neural systems, such that episodic memory can facilitate, but is not required for, the formation of semantic memories.
Retrograde amnesia in humans: are episodic and semantic memories differently affected by hippocampal damage?

There are numerous reports of autobiographical memory impairments that extend back 20-40 years prior to damage of the medial temporal lobe (see Nadel and Moscovitch 1997; Spiers, Maguire et al. 2001 for review). Often, memory for public events is also impaired, and seems to parallel the extent of deficits in memory for famous people. Several reports comparing autobiographical memory, memory for famous people, and public event memory, find differences in the temporal extent of memory loss according to the material tested (Cermak and O'Connor 1983; Barr, Goldberg et al. 1990; Rempel-Clower, Zola et al. 1996). Memory impairments are reported for autobiographical events dating back 20-40 years, and are often considered examples of flat RA, whereas memory for public events and for famous people gets progressively better for more remote events (graded RA) over the same or a slightly shorter range of time. Despite the trend for the MTL to be more permanently involved in processing episodic memories, there are numerous counter-examples in which MTL damage produces more extensive memory impairments for famous people and public events than impairments for episodic memory (Spiers, Maguire et al. 2001). The degree of conflict in these reports suggests that if any such dissociation exists, it is not very robust. Unlike the subtle differences in memory for personal events, public events, and famous personalities, knowledge of grammar, words, and objects (i.e. general semantics) is almost always unimpaired (Scoville and Milner 1957; Cermak and O'Connor 1983; Damasio, Eslinger et al. 1985; Barr, Goldberg et al.
1990; Victor and Agamanolis 1990). Possible confounds to these results are the variety and extent of brain damage.

Finding patients with circumscribed lesions is a recurring challenge to neuropsychologists. Even when histological verification is possible, demonstrating restricted regions of neural damage does not establish that other brain regions functioned normally. Be that as it may, there are a few reports of damage restricted to the hippocampus: RB (Zola-Morgan, Squire et al. 1986), GD (Rempel-Clower, Zola et al. 1996), BE and LC (Kapur and Brooks 1999), VC (Kartsounis, Rudge et al. 1995; Cipolotti, Shallice et al. 2001), LJ (Reed and Squire 1998), YK (Hirano and Noguchi 1998), and one case by Victor and Agamanolis (1990). Of the eight patients described, five had retrograde amnesia extending beyond 10 years, though in two cases this extended RA was selective for famous faces and public events, not episodic memory. Of the three cases with extended episodic RA, two also presented with seizures and either anoxia or ischemia, all of which are associated with extensive RA but do not necessarily result in cell death in all affected brain regions (Kapur, Young et al. 1989; Kapur, Millar et al. 1997; O'Connor, Sieggreen et al. 1997; Zeman, Boniface et al. 1998; Viskontas, McAndrews et al. 2000). It is difficult to determine whether the hippocampal damage, more diffuse, undetected damage, or both is responsible for the extent of retrograde amnesia. That leaves one case in which damage restricted to the hippocampus produces episodic amnesia extending beyond two years.

By including patients with damage to the hippocampus and adjacent entorhinal cortex (but without reports of epilepsy), two more examples of extensive RA appear,
including flat loss of episodic memory over 30 years (Rempel-Clower, Zola et al. 1996), and graded loss of episodic memory over 10-15 years (Schnider, Bassetti et al. 1995). In both cases, memory for famous people was impaired over 15 years. When more extensive damage of the medial temporal lobe occurs, there are numerous reports of extended RA (Nadel and Moscovitch 1997; Spiers, Maguire et al. 2001). This could be the result of the extra-hippocampal damage, or simply because the hippocampal damage is more complete.

The issue of whether episodic and semantic memories are differentially dependent on hippocampal function remains largely unresolved. It has been suggested that the autobiographical memories are more extensive, described as a “flat RA”, as illustrated in Figure 1.1B,C, or F, whereas semantic memories can reveal graded RA, as shown in Figure 1.1 D, E, or F. This distinction is not supported, however, based on the deficits of hippocampal and medial temporal lobe patients. In several cases, memory for famous people, presumably a semantic kind of memory, was often as temporally extensive as memory for personal and public events. Perhaps the tendency for graded memory loss of famous people and public events indicates a separate memory process, however, given the limits of retrospective studies, it may be impossible to measure reliably subtle differences in RA slope.

The most obvious evidence for RA selectively disrupting episodic memories was reviewed in the epilepsy section, and included 12 subjects who had temporal lobe resections (Viskontas, McAndrews et al. 2000). Based on the Autobiographical Memory Interview (Kopelman, Wilson et al. 1989), personal semantic memory was unimpaired at
all time points, but personal episodic memory was impaired at all time points, including childhood. There are some concerns that the episodic test may lack sensitivity to the degree of details recalled, and, more importantly, the resections were not restricted to the hippocampus. In fact, there was no relationship between memory impairment and the level of hippocampal sclerosis. Yet, this is a clear example of differential impairment in episodic and semantic memories resulting from medial temporal lobe damage. Whether this effect will be replicated in other studies remains to be seen.

There remains the possibility that the tests used to determine the length of RA may be too permissive to distinguish true episodic retrieval from 'semanticized' versions. For example, when graded RA is observed for spatial memories, the nature of the memories may fall into a more simple associative or semantic category rather than one incorporating richly detailed, relational spatial information, analogous to the kind of detail present in episodic memories. One patient, K.C., with anterograde and retrograde autobiographical amnesia revealed an intact internal spatial representation of the neighborhood in which he'd lived for 40 years. Despite his ability to draw a map of the neighborhood and accurately estimate distances between points and describe routes, he showed severe deficits in recognizing buildings from that neighborhood from other buildings, and he could not identify the appropriate location of these buildings in his neighborhood (Rosenbaum, Priselac et al. 2000). His performance on many basic spatial memory tests might lead one to conclude he has spared remote spatial memory, when, in fact, more demanding tests demonstrate that he lacks the detailed spatial maps present for controls. In fact, reports of graded RA in other amnesics may have resulted from a failure
to use tests that are sensitive to the spatial memory deficits seen in K.C. (Teng and Squire 1999). If this principle applies beyond the domain of spatial memory, there may be serious underestimates the prevalence of flat RA for episodic memories.

On the whole, there is some evidence to suggest that medial temporal lobe damage alters even the most remote episodic memories, whereas remote semantic (or 'semanticized') memories may be preserved. The few cases supporting this distinction further suggest that other, more common, tests of memory may not be stringent enough to distinguish between true episodic memory and recall of semantic elements of an episode. Even more consistent is the evidence for a distinction between general semantic memory and factual memory relating to events. General semantic memory includes items learned over a broad range of time, possibly recalled in the absence of any accompanying episodic context, and are often unaffected by medial temporal lobe damage. In contrast, items learned with few exposures, over a restricted period of time, may be subject to more extensive RA. Because the personal and public event and famous people tests described above were designed to test the strength of memories over restricted periods of time, the items were necessarily limited in their length of exposure compared to items testing general semantic memory. In some sense, that makes these memories 'more episodic' than general semantic memories. In fact, while it may be useful to treat episodic and semantic memories as distinct, they may reflect the extremes of a continuum, varying in their degree of dependence on medial temporal lobe systems.

Detecting which neural systems contribute to the formation of explicit memories over time is difficult in humans. The studies reviewed here are hampered by the
unpredictable extent of damage or dysfunction to the brain, and by the inability to design prospective studies that can control the content and age of memories. These are precisely the advantages of using animal models.

Anterograde deficits in animals

Distinguishing episodic from semantic memory in non-human animals may be impossible (but see Morris 2001). Yet even in animals, there is a relationship between the task to be learned and the pattern of deficits associated with hippocampal damage. This relationship can help define types of memories in animals that may be analogous to episodic and semantic memory.

In humans, hippocampal damage can prevent the formation of new episodic memories; in animals, new tasks requiring recall of the environmental context or requiring integration of numerous spatial relations are dependent on hippocampal function. Without a hippocampus, rats have difficulty finding a goal when only a constellation of room cues are available, such as when the goal is not marked by a local cue (Morris, Garrud et al. 1982; Sutherland, Kolb et al. 1982; Morris 1990). Conditioning to context is also sensitive to hippocampal damage, depending on the conditioning parameters and type of lesion (Good and Honey 1991; Kim and Fanselow 1992; Phillips and LeDoux 1992; Phillips and LeDoux 1994). The hippocampus in animals may be necessary to represent the relations among contextual cues, just as the hippocampus in humans may be necessary to bind together the bits of spatio-temporal elements comprising episodic or relational implicit memory. If hippocampal-dependent tasks
reveal an extended, flat RA in animals following damage restricted to the hippocampus, it would lend support to the claim that episodic memory in humans always requires a hippocampus.

**Retrograde deficits in animals**

Tests of retrograde amnesia in animals have produced the full spectrum of possible results (compare Figure 1.1 with Table 1.1). What variables might be responsible for the apparent inconsistencies? The species involved and the use of within or between subject designs do not explain the results. The choice of time points, however, is critical to detecting graded RA: if the most remote time point is still within the RA period, a flat effect may be observed, if the first time point begins after RA has dissipated, no effect would be observed. In addition, the number of time points sampled may affect the accuracy in estimating the duration of RA, but may also trigger 'learning set' in within-subject studies (Murray and Bussey 2001). Also, the number of animals per group can dramatically affect the statistical power of multiple-factor designs. Several studies report mean values consistent with graded RA (resembling those of Figure 1.1D and E), but with large error bars, and thus no significant interaction.

The difficulty of recall is another critical variable that is determined by the duration between learning and recall and by the method of measuring recall. In some tests, control animals show no forgetting; in others, they are near or at chance at the longest delays. Some tests use a discrete measure of recall or forgetting (e.g. 2-choice discrimination), whereas other measures are continuous (e.g. latency to reach a goal).
Finally, the tasks themselves may require very different types of learning, and therefore may involve different neural systems. Habit or skill memory has already been described as one type of memory independent of hippocampal function in animals. The following overview of animal studies is therefore grouped by task, keeping in mind some of the other variables that might explain the pattern of results. Reports of the related phenomenon of reconsolidation are considered separately in Section 4.3.2.
Table 1.1 Summary of prospective studies of retrograde amnesia (RA) in animals with damage to the hippocampal formation. The columns list the Study, the species of animal used, where and how lesions were created, whether the experimental design was within- (W) or between- (B) subjects, the task and/or apparatus, Rcnt—the shortest delay period, Rmt—the first delay period where no RA occurred, or, in the case of flat RA, the longest delay measured, Normal Forget?—whether control animals showed forgetting, Remote H IN?—whether the longest delay showed no difference between controls and lesioned animals, and whether Graded RA? was observed. The RA shown in Figure 1F would be reported as ‘No’ in Remote H IN? column and ‘Yes’ in the Graded RA column. Abbreviations: H-hippocampus, dH-dorsal hippocampus, EC- entorhinal cortex, Fx-fornix, Sb- subiculum, A-amygdala, Ibo-Ibotenic acid, NMDA-N-methyl-D-aspartate, Electro-electrolytic, Aspir-Aspiration, RF-radio frequency, Tsx-transection.
Turning first to tasks thought to require hippocampal function, the results of hidden-platform water maze tasks are reported as showing a flat RA, whereas dry water maze and plus maze tasks produced graded RA. In the first water maze study listed in Table 1.1 (Bolhuis, Stewart et al. 1994), only two time points were tested, and, indeed, hippocampal animals were at chance at both 3-day and 14-week intervals, presenting a flat RA. The control animals, however, only performed slightly above chance, even at the recent, 3-day time point. While this may have been sufficient to detect differences between lesion and control groups, detection of changes over time would have been unlikely with both groups performing so poorly. The second water maze study is free of this problem at the most recent, 2-week time point, when control animals found the platform 3 times as fast as lesioned animals, however, control animals show significant forgetting by the 14 week time point, taking just as long to find the platform as lesioned animals (Mumby, Astur et al. 1999). The results resemble Figure 1.1E, with the caveat that the authors provide no estimate of ‘chance’ latency. Consistent with graded RA, control groups had significantly shorter latencies than lesioned animals for the recently learned task, but both groups performed equivalently for the task learned 14-weeks prior to lesioning. Despite this, a repeated-measures ANOVA failed to indicate a main effect of task recency or a group x task recency interaction, showing only a significant main effect of group, indicating a flat RA. Moreover, using the results of a probe trial, controls spent a greater proportion of time swimming in the correct quadrant than lesioned animals, for tasks learned at 2 or 14 weeks prior to lesion. Again, the ANOVA showed no main effect
for task and no interaction. Thus, depending on the measure used, results may resemble graded or flat RA, but the statistics suggest flat RA.

In a partial replication of Mumby et al. (1999), rats were trained in water maze and object discrimination tasks, but in this case, the object discrimination task involved swimming in the same water maze apparatus, in the same room, and on the same day as the hidden platform task (Sutherland, Weisend et al. 2001). A given rat learned these tasks at one of the following time points: 1 day, 1, 2, 4 or 15 weeks prior to lesioning. Lesioned animals took longer to find the platform than did control animals, and longer intervals between learning and recall resulted in longer latencies for both groups, indicating forgetting. In addition, lesioned animals took longer than did control animals at the 1 day, 1 week and 2 week delays, but not at 4 or 15 week delays, suggesting the hippocampus is important for recall only in the first few weeks after learning a task; however, there was no statistically significant interaction between group and task recency. Thus, this is the second water maze study in which a statistically flat RA is described despite the nonsignificant appearance of a graded RA.

Other hippocampal-dependent spatial memory tasks show a clear graded RA. Using a dry water maze in which rats searched in a cylindrical arena for a hidden goal, rats who learned the task 14 weeks before lesioning were over twice as fast as rats who learned the task 1 week before lesioning (Kubie, Sutherland et al. 1999). The control animals were just as fast as the 14-week hippocampal rats, regardless of when the task was learned, thus results resembled Figure 1.1D. In another task, the rats were released in one of three arms of a plus maze, and had to go a goal arm that was in a consistent
location in room coordinates (Ramos 1998). Control animals performed progressively worse with increasing delays between learning and recall (i.e. forgetting), whereas animals with hippocampal damage got progressively better. There was a statistical main effect of group and a group x task recency interaction, indicating a graded RA in which the hippocampus is no longer required 7 weeks after learning the task.

Hippocampal lesions also impair context memory in recently, but not remotely, learned fear conditioning tasks. Anagostaras et al. (2001) presented compelling evidence that the memory of the context of the environment itself, and not some other aspect of the conditioning task, is critical to the time-limited RA following hippocampal damage. They merely exposed one group of rats to an environment one month before conditioning began. Both groups then underwent fear conditioning in the environment with a tone CS, and were lesioned one day later. The lesioned group for whom the environment was new showed amnesia when exposed to the context alone, expressed as reduced freezing, whereas rats exposed to the environment one month earlier showed perfect retention. Thus, a 1-month old memory of an environment prevented subsequent conditioning from being susceptible to hippocampal damage.

Of the various other tasks used to test RA, only two showed a flat RA. The first was the platform discrimination task learned at the same time, in the same room, on the same apparatus as the hidden platform water maze. Subsequent recall on this task requires that the rat ignore the room and maze, and alter behavior based on the presence of 2 visual cues (the platforms). It is known that the hippocampal representation of an environment can completely change, or remap, based on changes in task requirements,
despite no or very little change in the environment (Markus, Qin et al. 1995). It follows that the hippocampal neurons contain information that could allow an animal to discriminate between the very similar environments of the hidden platform and visible platform tasks. Thus, the hippocampal lesioned animals may fail to perform well, not because of the object discrimination task per se, or of the absence of some previously stored memory, but because an animal without a hippocampus may not distinguish between the two very similar tasks.

The other task showing flat RA required object discrimination in monkeys following the combined aspiration lesions of the hippocampus, amygdala, and entorhinal cortex (Salmon, Zola-Morgan et al. 1987). This is, by far, the most extensive damage of the experiments listed in Table 1.1, but was included because of the paucity of primate RA studies. It is of considerable interest that the clearest example of flat RA involved the most extensive damage in the medial temporal lobe, analogous to the majority of flat RA cases reported in humans. Of the two other primate studies, one showed graded RA (Zola-Morgan and Squire 1990), and the other has been interpreted as showing graded RA by some, and flat RA by others, though neither conclusion is justified (Gaffan 1993).

The Gaffan study (Gaffan 1993), cannot be assessed properly for RA, since both non-lesioned and lesioned groups were subject to a retroactive interference task, leaving no true control group to use as a baseline for memory performance. The non-lesioned ‘control’ group of monkeys learned a discrimination task on one set of scenes (Set A). Six months later, a second set of scenes was learned (Set B) and immediately afterwards, recall was tested on both sets. Thus, the recall test caused Set A to become an intervening
set that may have interfered with consolidation of the newly learned Set B. Consistent with retroactive interference literature, subsequent memory for Set B was impaired (in this case, relative to memory for Set A, whose acquisition was not followed immediately by an intervening set). The lesion group underwent the same interference conditions, and then received fornix transections. The lesioned monkeys were worse, overall, than non-lesioned monkeys, yet they, too, showed better recall for the remote Set A than the recent Set B. Without a proper control group, however, the apparent 'graded RA' in both interference and interference + lesion groups cannot be verified.

Finally, there are several reports of pharmacological and genetic manipulations, which can be thought of as either temporary or permanent 'lesions' restricted to certain receptor types. In most of the studies designed to examine consolidation processes, the manipulation occurs within an hour after the task is completed. These studies may be tapping into initial molecular cascades necessary for consolidation to commence, but do not explain the observation that RA periods extend over days and weeks. Three studies stand in exception, providing converging evidence for the involvement of the hippocampus in the hours or days following acquisition.

Izquierdo and colleagues (1997), tested the necessity of several structures for retrieval of an experience acquired, 1, 31, or 60 days prior to testing. They delivered the AMPA-blocker CNQX to the hippocampus, amygdala, entorhinal cortex or parietal cortex prior to testing recall in a step-down inhibitory avoidance task. The hippocampus and amygdala injections disrupted memory when given at the one-day interval, but not at 31 or 60 day intervals. The entorhinal and parietal injections disrupted memory when
given at 1 and 31 days and at 1, 31, and 61 days, respectively. This suggests that memory for the task requires normal excitatory transmission among all these structures within a day of acquisition, but is only critical in neocortical structures for more remote memories. In the same study, AP5 was used to interfere with any NMDA-dependent plasticity that may occur after an experience. Here, the amnesic effects were seen on a much shorter time scale (from 0 to 3 hrs), and neocortical injections continued to be effective at longer delays than the hippocampal injections. This suggests plasticity shortly after an experience may be critical for memory consolidation, which is expressed in the activity patterns of cortical structures, and less critically in the activity patterns in the hippocampus, over weeks.

Riedel et al. (1999) show findings complementary to those of Izquierdo et al. (1997), namely, that normal hippocampal synaptic transmission is important for memory consolidation in the days following acquisition. Chronic AMPA receptor blockade with LY326325 lasting either 1-7 or 5-12 days after learning the Morris water maze task impaired performance measured 16 days post-acquisition. An acute dose of LY administered one day after acquisition did not impair later performance, possibly because the acute dose only alters activity for 4-6 hours, compared to the >160 hours of treatment in the other two conditions. Because recall was not tested under the influence of LY in this experiment, one cannot directly compare the results of the two studies, even at the common 1-day time point. Assuming a similar role for the hippocampus in both tasks, the two studies suggest the hippocampus is important for recall 1 day after acquisition, but
only neocortical regions are necessary by the 31st day. Continued hippocampal activity over the intervening days is necessary to maintain the memory.

NMDA-dependent plasticity is one mechanism by which memories could be consolidated, and which could depend on continued hippocampal activity as suggested by the previous two studies. The importance of hippocampal NMDA-receptor function after acquisition was tested using inducible, reversible, and CA1-specific NR1 knockout mice (Shimizu, Tang et al. 2000). Using a task design similar to that of Riedel et al. (1999), the NMDA receptors in CA1 were effectively knocked out from 1-7 or 9-14 days after water maze learning, and mice were retested after 15 days. The mice in the 1-7 day group were impaired upon retesting, but the 9-14 day group was not, suggesting NMDA-dependent activity in the week after learning is important for memory retention. This matches well with the experiments described in Riedel (1999), which reported an even longer dependence for hippocampal excitatory transmission. Consider that it is possible to have continued neural activity without plasticity in those synapses, but not vice versa. For example, it may be that the hippocampus undergoes plastic change for a restricted time during and after learning, so that, even with time, it can continue to recall the appropriate memory, expressed as a pattern of excitatory transmission, in order to facilitate plasticity in other structures.

Mice were also tested on a fear conditioning task like the one described by Kim and Fanselow (1992). The NR1 receptor was knocked out from 0-14 or from 21-29 days after conditioning and a retention test was given on the 29th day. Only the 0-14 group had impaired performance in the conditioned context, and no group showed impaired
conditioning to the tone CS. Again, these results are similar to those found following hippocampal lesions, suggesting that the CA1 NMDA receptor plays a critical role in consolidation, at least initially. The Izquierdo (1997) study reported a much shorter time-dependence of memory on the NMDA receptor in the hippocampus. The species, technique and task of the Izquierdo (1997) study was different from the present study (Shimizu, Tang et al. 2000), accounting for differences in the temporal extent of deficits. It may be that the step-down inhibitory avoidance task requires less continued plasticity in the hippocampus than do contextual fear or water maze tasks. Taken together, these studies show further evidence of graded retrograde amnesia following hippocampal dysfunction, and that hippocampal NMDA-dependent plasticity after learning appears to be critical to the consolidation process.

Based on animal studies, the case for flat RA, independent of time, resulting from hippocampal damage is weak. The vast majority of results indicate graded RA, with only 3 out of 21 cases in Table 1.1 reporting a flat RA following lesions restricted to the hippocampus. These three cases each have their own weaknesses, making it difficult to deny that hippocampal damage generally leads to graded RA. There is still the possibility that the prevalence of graded RA is due to the recall measurements. Just as the human memory tests may not have been sensitive to the details of episodic memories, the tasks used in the animal studies may not have required a great degree of relational/contextual processing. There is always the possibility that tasks will be created that demonstrate the permanent necessity of hippocampal function. The present studies, however, show that
memory formation for a variety of tasks is temporarily dependent on hippocampal function.

This does not mean that the hippocampus plays no part in recalling remote memories. Although animals with hippocampal lesions had better recall for remote memories than recent memories (relative to controls), the remote memories of control animals were still better than those of lesioned animals in several studies. See Figure 1.1F for an example of graded RA in the face of some degree of permanent impairment. In Table 1.1, the second to last column indicates whether remote memory recall is independent of hippocampal function. Almost twice as many experiments report some continued degree of impairment as those reporting no remote impairment. One could conclude that the memories tested were not remote enough, therefore never reaching hippocampal independence. The other interpretation is that the hippocampus may always facilitate recall of some memories, a hypothesis that is not mutually exclusive of a time-limited role of the hippocampus in memory consolidation. Whatever the exact role of the hippocampus may be, the evidence from both human and animal lesion literature suggest that regions outside the hippocampus become *better able* to support recall over time.

For cases in which flat RA, independent of time, is observed, perhaps structures outside the hippocampus are involved (Squire and Alvarez 1995). In the Viscontas et al., (2000) study no relationship was reported between hippocampal damage and the flat RA for episodic material in patients with temporal lobe resections. Monkeys with combined aspiration lesions of the amygdala, hippocampus, and entorhinal cortex showed a rare example of a flat RA in animal studies. Flat RA in humans is more common when
damage extends beyond the hippocampal formation, though this could also be explained by more complete damage of the hippocampus. If so, it may be that all declarative memories are stored in the hippocampus, but recent memories are more easily lost when the hippocampus is partially damaged. Evidence against the latter statement is found in a patient who showed flat RA for public and personal events, with skill learning intact (Kapur, Ellison et al. 1992). Magnetic resonance scans revealed anterior temporal lobe but not hippocampal pathology. An imaging study showed a relationship between entorhinal cortex activity and memories dating back up to 20 years (Haist, Bowden Gore et al. 2001). Furthermore, animal studies report that virtually complete hippocampal loss is accompanied by graded RA, with no relationship between extent of hippocampal damage and extent of RA (Winocur, McDonald et al. 2001; Clark, Broadbent et al. 2002). The appearance of extensive and/or flat RA seems more closely related to the extent of medial temporal lobe damage, than to damage restricted to the hippocampus.

Even after accepting a time-limited role of the hippocampus in memory consolidation, the case has been made that the more the 'binding' of event elements is required, the greater the extent of amnesia following hippocampal damage. In habit learning, by definition, the environment is irrelevant and no RA is seen following hippocampal damage. On the other extreme, some types of memories consist of relational information about the context, that is, the particular arrangement of stimuli in the environment. If performance requires memory for relational information, then damage to the hippocampus may cause deficits regardless of the age of the memory. In support of this, rats typically use relational strategies to solve some spatial tasks, whereas rats with
hippocampal lesions tend to learn based on direct cue associations (Winocur and Brekenridge 1973; Winocur 1982; Ramos 1998). But perhaps the most interesting situation lies between these two extremes: memories that ultimately can be described by a limited set of contextual cues, but whose encoding is facilitated by relational contextual information. This continuum of the degree of relational encoding may be analogous to an episodic–semantic memory continuum in humans. The fewer the contextual elements that need to be associated, the sooner the memory will be unaffected by damage to the hippocampus. Unfortunately, this is difficult to test, yet the idea that the hippocampus provides some ‘linking’ information during memory formation is prevalent in memory consolidation theories, as is discussed below.

1.1.4 Theories of memory consolidation

The early theories of memory consolidation attempted to explain the retroactive interference data. Following the preservation-consolidation hypothesis offered by Mueller and Pilzecker (1900), Hebb (1949) developed a more mechanistic, two stage theory of memory consolidation. According to this theory, the memory trace initially exists as reverberatory electrical activity in neural circuits while a permanent structural change is affected. This also accounted for the ECT results, prominent at the time. The massive electrical activity of the shock followed by near cessation of neural activity would disrupt the reverberatory circuit before a structural change had taken place. Longer intervals between acquisition and ECT would provide more opportunity for structural changes to develop. The two-stage theory shaped the experiments and theories
that followed, pushing theories of memory consolidation to describe not only ‘what’ the phenomenon is, but also ‘how’ the brain might accomplish this. Evidence obtained since the description of the two-stage theory, however, discounts the role of an uninterrupted reverberatory circuit that is responsible for the temporary maintenance of the memory trace. In addition, more recent theories of memory consolidation have provided better accounts of how a “permanent structural change” would manifest itself in the brain.

The time-limited role of hippocampus in memory consolidation: “standard” models

Memory consolidation theories became focused on a time-limited role of medial temporal lobe structures based on reports of graded RA following MTL damage (Scoville and Milner 1957; Marr 1971; Squire, Cohen et al. 1984; Teyler and DiScenna 1986; Squire 1992; Treves and Rolls 1994; McClelland, McNaughton et al. 1995; Squire and Alvarez 1995). Most of these theories are variations on the following theoretical theme:

1. Initially, the hippocampus is required to sustain the memory; the neocortex alone is not sufficient.

2. Over time, the neocortex becomes capable of sustaining the memory, possibly resulting from hippocampal-neocortical interactions.

One group of theories suggests that medial temporal lobe structures, or the hippocampus in particular, serves as a temporary memory store, and that memories are transferred to the neocortex over time (Alvarez and Squire 1994; Treves and Rolls 1994; Cho, Kesner et al. 1995). This may be accomplished through ‘fast-learning’ in the
hippocampus, which can sustain the memory while the ‘slow-learning’ neocortex undergoes a more gradual synaptic modification necessary to sustain the memory trace (Alvarez and Squire 1994; Treves and Rolls 1994; McClelland, McNaughton et al. 1995). Why might the neocortex be a slow learner? David Marr was perhaps the first to describe how the neocortex, unlike the hippocampus, would be capable of organizing, or categorizing, mnemonic information (Marr 1970; Marr 1971). Later formal models showed that slow, interleaved learning is a necessary process for successful category learning in the neocortex (McClelland, McNaughton et al. 1995).

In some of these theories, the hippocampus is assumed to run into the limits of its storage capacity with time, causing the traces of old memories to be overwritten. When recall is attempted for such memories, either the hippocampus shows less activity, because it fails to bring up any memory trace, or it “pattern completes” by recalling the most overlapping (but incorrect) memory trace. In either event, successful recall would only result from extra-hippocampal traces, since the hippocampus would no longer be able to contribute through it’s overwritten traces. Thus, one outcome predicted by this type of theory is decreased hippocampal activity for remote memories.

Another group of theories suggests the hippocampus is important not because it stores the memory temporarily, but because it somehow connects or re-establishes the memory trace that is distributed in the neocortex, until the neocortex is able to accomplish this on its own (Teyler and DiScenna 1986; Buzsaki 1989; Squire and Alvarez 1995). For example, the hippocampal code may represent an ‘index’ of neocortical activity during an experience. The later reactivation of the index could, in
turn, cause retrieval of the distributed neocortical memory trace (Teyler and DiScenna 1986). Other related theories suggest that the hippocampus links together geographically separate brain areas supporting the memory (Squire, Cohen et al. 1984; Squire and Alvarez 1995; Rolls 2000). These theories posit that eventually, retrieval independent of hippocampal function is made possible by the binding together of the neocortical regions supporting the memory trace. In a mechanistic extension of these theories, trace replay theory posits that hippocampal binding of neocortical memory traces causes coincident activity in neurons participating in the distributed trace, which, in turn, strengthens the connections between these neurons through Hebbian plasticity. The hypothesis of this dissertation was designed to test a major tenet of trace replay theory, therefore this theory will be discussed in more detail, after considering one final twist on consolidation theory.

Multiple Trace Theory

Not all researchers are convinced that graded retrograde amnesia is caused by the increasing hippocampal independence of explicit memories over time. The cause of graded RA is only one point of contention Nadel and Moscovitch raise against what they call ‘standard consolidation theory’ (Nadel and Moscovitch 1997). They answer with the alternative, multiple trace theory (MTT). According to the latter theory, the hippocampus is always necessary for episodic or relational-contextual memory processing. If the hippocampus is damaged, a flat RA for these types of memories results. What changes over time is not the neocortical connectivity, but the number of traces in the hippocampus. MTT states that each recall of an episode results in the addition of a new,
sparse-coded memory trace distributed among hippocampal neurons, thus remote episodic memories have more traces in the hippocampus than recent ones. The theory states that the reason graded RA is a common outcome of hippocampectomized animals is not because the neocortex becomes capable of supporting the memory, but because of incomplete damage to the hippocampus. The remaining functional hippocampal tissue is more likely to contain some of the many remote memory traces than of the few recent memory traces, yielding better memory for remote material.

Although MTT provides an alternate explanation of the retrograde amnesia data, the 'multiple trace' part of multiple trace theory has no empirical basis. But because the theory specifies the (hypothetical) neural mechanisms underlying hippocampal memory formation, the basis of MTT readily testable. If new traces are created within the hippocampus with each iteration of the episode/context, then repeated entry into an environment should cause different groups of neurons to become active. In addition, the sum hippocampal neural activity should be low for new memories, but increase with time. Independent of the specifics of implementation, MTT predicts that complete hippocampal damage will destroy all episodic/relational memories and prevent their formation. Further, the degree of hippocampal damage should be proportional to the temporal extent of RA.

What evidence, if any, supports these predictions? It is well-known among hippocampal physiologists that, when an animal re-enters an environment, the same set of hippocampal cells becomes active (O'Keefe and Nadel 1978; Thompson and Best 1990; Barnes, Suster et al. 1997). There is no evidence that other cells are consistently recruited
upon re-entry, nor that an independent set of cells becomes active. On the other hand, if cells are thought to be recruited in addition to the original set of cells, and if coding is very sparse, then the small proportion of cells sampled in electrode recordings may make it difficult to detect the additional traces.

The ideal test would be an unbiased measure indicating both the active and the inactive cells in an environment. RNA of the immediate early gene Arc appears in the nucleus of CA1 pyramidal cells within 5 minutes after cellular activity, but leaves the nucleus, entering the cytoplasm, within 30 minutes. Thus, fluorescence in situ hybridization of Arc RNA allows the identification of cells that were active at either of two different time points. Thus, if an animal explores two environments at the corresponding time points, it is possible to determine the hippocampal ensemble or ‘map’ for each environment (Guzowski, McNaughton et al. 1999). When this experiment was conducted, hippocampal cells showed only the independent probability that a cell active in one environment was active in the next, measured as the presence of both cytoplasmic and nuclear Arc. This is just as predicted by MTT. But when rats re-enter the same environment, almost all (90%) previously active cells were active again upon re-entry. Apparently, the same ‘map’ was retrieved. The proportion of CA1 cells active in an environment is roughly 40%, indicating a very distributed representation; however, when returning to the same environment, only 5% of CA1 cells were newly activated, far below what would be expected for a new trace. Moreover, there were more cells shutting off than becoming activated upon re-entry, indicating that this 5% is not only insufficient, it’s actually exceeded by results in the opposite direction. Perhaps this is because it was
only the second entry into the environment, occurring in a short timeframe. Yet whether repeated re-entry occurs on a daily basis over many days, or all in one day, no differences are seen in the proportion of Arc-labeled cells.

These results do not support the claim that multiple traces accrue with repeated recall in CA1 populations. But, in this case, more definitive measures of neural activity throughout CA1 would provide stronger evidence against multiple trace theory. Currently, the exact relationship between Arc activation and neural activity is not known. Although there is a strong positive correlation, the Arc measurements are binary, thus there is presumably some threshold of neural activity necessary to elicit Arc activation. Other biophysical firing characteristics may also differentially affect Arc activation, such as whether or not activity occurs as part of a 'burst'. There remains the possibility, however unlikely, that neural activity during formation of a new 'multiple' memory trace is different from activity during formation of a new trace or during recall of old traces. This difference in activity would need to be of the sort that would not be detected by the small sampling of extracellular electrophysiology, nor by Arc activation. In conclusion, there is evidence against the formation of multiple traces in the hippocampus, but no evidence at present in support of this hypothesis.

Even though the specific implementation of MTT is not supported by data, in developing this theory the authors raise thoughtful concerns about what it means for the brain to form a memory. Their specific issues with standard consolidation theory are, by most proponents of the theory, only constraints for certain versions of consolidation theory. For example, stating that a memory can be recalled independent of hippocampal
activity in no way implies that the hippocampus stops in its tracks, ceasing all activity. This result is only predicted as one possible outcome of those consolidation theories that state that the hippocampus must reflect a temporary, limited-capacity memory store. The distinction among memory consolidation theories must be emphasized to prevent more statements like the following:

"If the HPC were to have relinquished its role to the neocortex as memories age, as suggested by traditional consolidation theory, this would have been expressed as changes in activation across time period, such that the HPC would be less active and the neocortex more so as one recounts older memories relative to new ones." (Rosenbaum, Winocur et al. 2001, p.186)

As a result of this logic, imaging studies showing hippocampal fMRI activation that was equivalent for recent and remote memories have been interpreted as contradicting a time-limited role of the hippocampus in memory consolidation (Haist, Bowden Gore et al. 2001; Maguire, Henson et al. 2001; Ryan, Nadel et al. 2001). This statement is only correct for a subset of consolidation theories; other consolidation theories do not require any change in hippocampal activity. In contrast, MTT does predict a change in hippocampal activity as memories age. According to MTT, net hippocampal activity should increase as a memory becomes more remote, although this may be difficult to detect based on the relationship between neural activity and the fMRI signal (Scannell and Young 1999). Consistent hippocampal activation, if anything, favors
memory consolidation theories over MTT. The fMRI activity observed for remote memories suggests that the hippocampus may always recall its “map” for a given episode, and, whether as a result of hippocampal feedback or independent of it, the neocortex should show recall-related activity, no greater or less with time.

Several of the features of memory formation that Nadel and Moscovitch sought to incorporate in MTT are not inconsistent with consolidation theory. Much of the purported conflict is resolved if one does not require that the hippocampus either facilitate new memories or be perennially involved in the memory trace. Figure 1.1F illustrates how the hippocampus may be particularly important in the hours and days after an event, and still confer some advantage even as that memory becomes remote. What role could the hippocampus be playing, to produce these results? The hippocampal code is thought to reflect high-order, abstract associations of information that may enrich the content of the memory, consistent with a permanent role. Yet the very connectivity that supports these representations may also enable the coordinated strengthening of the connections in the neocortical memory trace, indicating a time-limited role. Both functions of the hippocampus are accounted for in a theory of hippocampal-dependent memory consolidation that will be termed ‘trace replay theory’.

**Trace Replay Theory**

When compared with the body of memory consolidation theories, the three defining features of trace replay theory are 1) that the hippocampus links together parts of the memory trace from separated regions of neocortex, producing coordinated activity in
the relevant cells from those regions, 2) that this coordinated activity results in Hebbian plasticity, and 3) that such coordinated activity is made possible through the patterns of connectivity within and across hippocampo-neocortical ensembles.

The first feature is well-described in several theories of memory consolidation (Squire, Cohen et al. 1984; Teyler and DiScenna 1986; Squire and Alvarez 1995). The importance of the coordinated replay of neural activity, supporting Hebbian learning, helps to explain the gradual strengthening of the neocortical memory trace, and the continued, but not necessary, activity within the hippocampus.

The second feature has been implicated in a few theories of memory consolidation. Although Hebbian learning rules have been used in models of memory consolidation (Alvarez and Squire 1994; Treves and Rolls 1994), only synaptic re-entry reinforcement (SRR) theory incorporates a cellular/molecular implementation of Hebbian plasticity to describe the memory consolidation process (Wittenberg and Tsien 2002). SRR theory expands on the idea that the hippocampus links and potentiates neocortical activity (Squire, Cohen et al. 1984; Teyler and DiScenna 1986; Buzsaki 1989; Treves and Rolls 1994). In addition, SRR theory specifies a role for the NMDA receptor, a neural coincidence detector, based on evidence that NMDA receptors in CA1 are important for memory consolidation (Shimizu, Tang et al. 2000). Specifically, SRR theory suggests that multiple iterations of NMDA-receptor-dependent plasticity in the hippocampus are necessary for consolidation of the recently acquired memory through Hebbian synaptic modification. The strengthened synapses in the hippocampal network elicit coincident activity of the neocortical network, causing NMDA-dependent plasticity in the cortico-
cortical connections of the memory trace. As the connections of the memory trace across different sensory modules become strengthened, the hippocampus is no longer required to generate coincident activity in the neocortex.

The third feature of trace replay theory simply makes explicit the connectivity patterns in cortical networks (Figure 1.2). Neurons are more interconnected within 'modules' (e.g. sensory modalities, subregions within modalities, or cortical columns) than across modules. Some regions of cortex receive convergent projections from modules, and provide feedback projections, serving as an indirect route for connecting regions. These regions of convergence represent the upper level of a hierarchy, and may themselves send projections that converge in a still higher level. The result is illustrated in Figure 1.2.A: a multi-tiered hierarchy in which primary sensory cortex is at the lowest level, and the medial temporal lobe represents one of the highest levels, receiving highly convergent projections from association areas of neocortex (based on anatomical evidence in the rat: Coogan and Burkhalter 1993; cat: Scannell, Blakemore et al. 1995; and monkey: Felleman and Van Essen 1991; for functional interpretations see Fuster 1995; Mesulam 1998; Hilgetag, Burns et al. 2000).

Initially, a distributed memory trace can be recalled only through feedback from the higher-level areas, causing coincident activity even among poorly connected elements of the trace. SRR theory predicts NMDA-dependent synaptic modification will strengthen connections in the neocortex. With enough repetition, the memory trace will be elicited in the neocortical networks independent of hippocampal feedback connections (Figure 1.2 B, C).
Figure 1.2 Schematic representations of the hierarchical organization of neocortex, and its role in trace replay theory. A. Anatomically-based hierarchy of visual (green), somatosensory (blue), and polysensory/associative cortex (red), based on previously reported connectivity patterns (Felleman and Van Essen 1991; Mesulam 1998). Each line indicates connections between brain regions (black nodes). Reciprocal connections have been demonstrated in over 40 of the 52 connections listed. The hippocampus sits atop the hierarchy, receiving associational input, and area 7a (among other brain regions) receives a convergence of visual and somatosensory projections. B. “Top-down” facilitation of memory trace formation. Two levels of cortical hierarchy are depicted. In both levels, colored polygons represent neural modules (e.g. cortical columns, brain regions, sensory modalities). Black circles in each module represent active units for a given experience. Notice that within a module, units have dense connectivity whereas lateral connections among modules are sparse, thus the cortex as a whole cannot support arbitrary, widely-distributed associations. Ascending projections reflect convergence of signals from lower level modules, thus when lower level units across modules are active, associations within the upper level module are made. In addition, return projections from the upper to the lower levels are strengthened. Subsequently, if the appropriate upper level units are activated, they may elicit coincident activity in the lower level components of the experience across modules. C. With repetition, this top-down reactivation may help generate direct connections among lower level modules (represented as thick black lines), which may support associative retrieval independently of the upper module. The connections from the upper level are colored gray for clarity, and do not indicate the absence of activation in the upper level.
Models built upon some aspects of this architecture have been successful in terms of either the storage capacity or the dynamics of complete pattern retrieval (i.e., memory recall). The ability of one such modular network to store a sufficient number of patterns (i.e., memory traces) is dependent on the intra- and inter-modular connectivity, the non-random connections defining the hierarchy, and the sparsity of modules used for a given pattern (O'Kane and Treves 1992; Fulvi Mari and Treves 1998). In another model, complete pattern retrieval across modules is possible when the relevant connections across modules is sufficiently strengthened via a Hebbian learning rule, and provided that some “cue” pattern is presented to one module (Renart, Parga et al. 1999; Renart, Parga et al. 1999). Moreover, when one module has retrieved the correct pattern, retrieval in another module becomes more resistant to the effects of noise. In contrast, weak connections allow modules to retrieve patterns independently. The above models illustrate the importance of the observed neocortical architecture in retrieving distributed memory traces created through Hebbian plasticity.

There may be limits to how directly any two neocortical neurons subserving the memory trace may be connected. In some cases, direct lateral connections may be strengthened or established with very few rounds of coincident activity; in other cases a sequential cascade of plasticity traveling down the hierarchy may be necessary. Some degree of indirect connection via the upper level regions may always be necessary for some memory traces.

These conditions fit nicely with those that led to the formulation of MTT. There is a limit to the number of arbitrary, cross-regional connections that are formed. The fewer
the new cortico-cortico connections required, or the closer the upper level region of convergence, the faster the neocortex will be able to support the memory trace. This is consistent with the idea from MTT that the neocortex should be capable of sustaining semantic memories more quickly, given their independence from contextual ties. Put in terms of the hierarchical framework, the more 'semantic' a memory, the greater the proportion of intra-module connections, and the fewer the novel cross-module connections (i.e., arbitrary elements) to be formed. In contrast, the more episodic the memory, the more numerous the novel cross-module connections required, and the more likely it will be that cueing will require some 'top-down' feedback to retrieve the entire memory trace.

According to this view, the consolidation process would show the most change initially, when the newly reinforced hippocampal synapses cause the hippocampal network to activate spontaneously, eliciting recall in the neocortex. With time, the occasional cued or externally induced recall provides another bout of coincident activity, and may also increase the probability that these patterns will re-emerge offline. Moreover, some modules (involving the most 'arbitrary' or infrequent associations) may converge only at the highest-level modules, with numerous synapses intervening, thus may take the most repetition to connect directly. As a result, there may be ongoing shifts to lower levels of the hierarchy long after the memory was acquired.

This explains the temporal extension of RA, the possibility that the hippocampus may facilitate recall of even remote memories, the increasing amnesia with damage extending to associative cortex in the MTL, and the differences of RA between episodic,
time-specific semantic, and general semantic memories. It supports the idea that semantic memories can be formed without hippocampal function, that learning will be slower, but that once learned, the memories are retained as well as in individuals with intact hippocampi (Gaffan 1993; Vargha-Khadem, Gadian et al. 1997).

One prediction of this theory is that when elements of an episode are highly familiar and have been frequently associated in the past, the neocortical trace should become independent of the hippocampus more rapidly than when the elements are novel or are in novel combinations (e.g., arbitrary associations). In support of this prediction, a task learned only one day prior to hippocampal lesions can be recalled, but only when learning occurred in a familiar context (Anagnostaras, Gale et al. 2001). Another prediction of the theory is that, upon cueing, the neocortex will take longer to retrieve the memory trace of a recent memory that still requires hippocampal completion than of a remote, more directly connected, memory. This experiment has yet to be conducted.

1.1.5 Conclusion

In sum, implicit and explicit memories alike become resistant to disruption with time, and are particularly susceptible in the minutes to hours after an experience. Previous reports suggest medial temporal lobe dysfunction selectively impairs explicit, but not implicit memories, although it may be the relational aspect common to explicit memories that is critically dependent on medial temporal lobe structures. Specifically, hippocampal integrity is necessary for recalling many types of memories shortly after they are acquired, and may continue to be necessary for weeks and months (in animals)
or for months and years (in humans). The most episodic or relational memories may always benefit from hippocampal function.

Theories stressing the initial hippocampal-dependence of declarative memories describe some role of the hippocampus in binding together or ‘teaching’ otherwise disconnected neocortical components of the memory. Many theories of memory consolidation do not offer a mechanistic description of how the brain accomplishes this, but some theories speculate that the hippocampus may help generate synchronized repeated activity in the distributed array of neurons encoding the memory trace. This repeated activity could produce stronger connections within the neocortex via Hebbian plasticity, making the memory trace more neocortically interconnected and less dependent on the hippocampus for retrieval. This type of mechanistic description begs for evidence that 1. the brain undergoes ‘offline’ periods during which the hippocampus and neocortex can develop a dialogue, and 2. hippocampal and neocortical neurons encoding an event selectively discharge together afterwards.
1.2 SLEEP AND MEMORY CONSOLIDATION

1.2.1 Introduction

Reports of sleep-related memory enhancement go back at least as far as Jenkins and Dallenbach (1924). Why might sleep be related to memory consolidation? In some respects, sleep can be thought of as the opposite extreme to interference tasks: sleep prevents the active learning of new environmental stimuli. Indeed, naps prevent the decline in perceptual discrimination performance seen when subjects are not allowed an intervening nap (Mednick, Nakayama et al. 2002). But evidence suggests that whether in the form of a nap or full night-long episode, sleep reflects more than the absence of intervening stimuli, taking an active role in memory consolidation (Benson and Feinberg 1977; Mednick, Nakayama et al. 2002). The results from decades of research underline the importance of differentiating sleep into component stages and using multiple approaches to infer the functions of each sleep stage. After briefly describing sleep stages distinguished on the basis of global changes in brain activity, some common techniques for isolating the effects of particular sleep stages will also be discussed, and finally, the role of each sleep stage in memory consolidation will be considered.

Sleep Stages

Two cardinal types of sleep, rapid eye movement (REM) and non-REM, alternate during sleep. For humans, four or five non-REM/REM cycles occur in one night’s sleep, with each cycle lasting roughly 90 minutes. Although the cycle duration is kept constant, the relative duration of REM sleep to deep non-REM sleep increases with each
successive sleep cycle. Thus, the second half of the night contains more REM sleep than the first half.

REM sleep, also called paradoxical, active, or desynchronized sleep, has high-frequency, low-amplitude EEG, virtually indistinguishable from the EEG recorded during wakefulness. But unlike the muscle activity associated with wakefulness, REM sleep is characterized by clusters of rapid eye movements in the EOG, and muscle atonia in the EMG.

Non-REM sleep is divided into four stages, in increasing order of sleep depth. The EEG in stage 2 sleep is characterized by spindle oscillations and K-complexes superimposed on slow, synchronized oscillations. As sleep progresses to stages 3 and 4 sleep, also called slow-wave sleep (SWS), the slow oscillations continue, but the spindle waves and K-complexes diminish as delta frequency oscillations increase. The origins and significance of these oscillations will be described in more detail in Section 1.3, but they are mentioned here to help differentiate among sleep stages.

Techniques

The literature suggests REM and non-REM may play special roles in memory consolidation, but often it is difficult to isolate the effects of different stages of sleep. For example, one common method for evaluating sleep-related consolidation is to have subjects learn a task, and then look for changes in certain sleep parameters above control levels, whether based on within subject baselines, or compared to subjects who did not engage in the learning task. Typically, changes in the proportion of time spent in each
stage of sleep constitute the dependent variable in these experiments. Unfortunately, longer sleep stages may not be related to more or better memory processing. But recently, other sleep-related parameters have been explored, such as the amount and strength of the oscillatory activity in non-REM sleep, which may be more directly related to mechanisms of consolidation. Another variation on this method is to measure the correlation between a parameter of each sleep stage (e.g. duration) with performance levels measured before or after sleep. When performance covaries with the duration of a stage of sleep, it is suggestive of a common phenomenon, though causality cannot be assessed.

A second approach uses selective sleep deprivation after learning a task to induce memory impairments, visible upon retesting. With this method, causality can be attributed to the manipulation (deprivation), making it a popular method for testing hypotheses about memory consolidation during sleep. But this particular manipulation involves repeated wakening throughout the night, and the deprivation may interfere with a huge variety of processes other than the sleep stage *per se*. For example, disruption of a sleep phase may interfere with proper functioning in other sleep stages, or with other biological processes such as stress or metabolism, which may, in turn, alter memory consolidation. Moreover, some sleep stages are easier to disrupt selectively than others, making it difficult to design an experiment that can show a dissociation of consolidation with different sleep stages.

Yet another technique minimizes the disruption of function associated with deprivation techniques, while still manipulating the relative amounts of certain sleep stages. In this approach, subjects learn a task either before a whole night's sleep or after
being awakened halfway through their night’s sleep, then allowing them to rest for the second half of the night. Neither group is sleep deprived, but one group’s memories may benefit from a full night’s sleep whereas the other group’s memories only benefit from the REM-rich second half of sleep. If the full night group shows better subsequent performance, it would appear to be due to the early part of sleep, possibly resulting from the greater time spent in SWS. In contrast, if both groups show equal performance improvements, the second, REM-rich half of sleep would appear to be the critical factor. This technique capitalizes on the relative preponderance of REM sleep in the second half of the night, making the assumption that the small amount of non-REM sleep is functionally negligible, and vice versa for the first four hours of sleep. This assumption has been called into question in at least one study, which found that late-night, non-REM Stage 2 sleep, was important for consolidation (Walker, Brakefield et al. 2002), thus caution must be used in interpreting the half night interruption technique.

Clearly, all of the techniques described have serious limitations; however, because their weaknesses differ, a combination of these techniques may nonetheless converge upon a clearer picture of the relationship between sleep and memory consolidation.

1.2.2 REM

Perhaps because of the early discovery that dreams occur during REM sleep, because the EEG resembles that of waking activity, or because it is more difficult to remove non-REM sleep selectively, early studies investigated the role of REM sleep in memory formation (Hobson and McCarley 1977; Hobson, Pace-Schott et al. 2000). In
fact, dreams also occur during non-REM sleep, but at lower frequency and of a less hallucinatory and more realistic quality (Cavallero, Cicogna et al. 1992; Nielsen 2000; Fosse, Stickgold et al. 2001). That, in addition to animal studies of synchronized activity during SWS, suggested a role for non-REM sleep in memory consolidation (Steriade 2001). Evidence for the involvement of non-REM and REM sleep in memory consolidation will be described separately, and then considered collectively in evaluating current theories of the function of sleep in memory formation.

Several investigators have supported the idea that memory is consolidated during REM sleep (Fishbein 1970; Pearlman 1971; Fishbein and Gutwein 1977; Bloch, Hennevin et al. 1979; Hars, Hennevin et al. 1985; Smith 1985; Hennevin, Hars et al. 1995; Smith 1995; Fishbein 1996; Smith 1996). In rats, a long history of experiments suggests that heightened or enriched waking experience leads to increases in REM sleep, and that REM sleep deprivation produces memory deficits for recently learned material (see Hennevin, Hars et al. 1995; Smith 1995; Smith 1996 for reviews). Strong criticisms have been laid against these experiments, from the stressful methods of selective REM deprivation to the nonspecific increases in total sleep time following heightened experiences, to the circular logic of using REM sleep increases to determine what constitutes an enriched experience. Moreover, after a wealth of studies across numerous labs, some investigators conclude that there is no consistent relationship between REM and memory consolidation (McGrath and Cohen 1978; Horne and McGrath 1984; Smith 1985; Vertes and Eastman 2000).
In humans, the use of selective sleep deprivation is complemented by half-night learning (interruption) and correlative techniques. Although each method has its limitations, one emerging pattern is the involvement of REM sleep within 24 hours of learning, particularly for tasks involving procedural memory. These procedural memory tasks include word priming, perceptual skill learning and motor skill learning. With deficits in the range of 20-50% in the absence of REM sleep, procedural memory consolidation may be closely tied to REM sleep (Smith 1996).

**Sensory-motor skills**

Performance of a finger opposition sequence is correlated with the amount of REM sleep ($r=0.61$, $P<0.004$) but not with any stage of non-REM sleep, experienced within 24 hours of learning (Fischer, Hallschmid et al. 2002). Even when a sleep deprived group is allowed to rest the following night, their performance will still lag behind the group given sleep after acquisition, suggesting consolidation of these tasks is greatest during the first night of sleep. In sleep following learning of a prism adaptation task, the proportion of REM sleep increased (De Koninck and Prevost 1991). Increases in REM were also observed for subjects learning a novel set of movements (trampolining), but not for equally active subjects performing familiar exercises (Buchegger and Meier-Koll 1988; Buchegger, Fritsch et al. 1991). Finally, among several tasks administered, Conway and Smith (1994) and Plihal and Born (1997) found REM deprivation impaired memory for a Corsi block tapping task, the tower of Hanoi task, and a mirror tracing task.
Both deprivation and correlation studies suggest sensorimotor skill learning is associated with REM sleep.

Perceptual skills

Only a handful of studies have explored the benefits of sleep for perceptual memories. Performance of tasks requiring visual and auditory learning resulted in altered proportions of time spent in REM, including increased REM bursts for visual and decreased bursts for auditory learning (Verschoor and Holdstock 1984).

Karni et al. (1994) also reported impairments following REM deprivation using a perceptual texture discrimination task (TDT). Two articles have since described the effects of sleep on this task, both reporting results that are partially inconsistent with those of Karni et al (1994). Gais et al. (2000) used the ½ night interruption technique, showing a critical involvement of non-REM sleep early in the night, and a small degree of late REM involvement only after early SWS-rich sleep preceded. This suggests that impairments should be greatest following non-REM deprivation, with lesser impairments due to REM deprivation, contrary to the results reported in Karni et al. (1994). In the third study, overnight performance improvement was compared with a combination of SWS in the first ¼ of the night and with REM in the last ¼ of the night, producing a correlation of .89 (p< 0.0001), greater than the correlation using either sleep factor alone (Stickgold, Whidbee et al. 2000). This striking result is consistent with those of Gais et al. (2000), but does not explain why the original study found impairments following only REM deprivation. The pattern of results underlines the importance of using multiple
approaches and warns against drawing strong conclusions from only one source of data. Although the details remain unresolved, what all three studies have in common is some improvement in discrimination performance associated with REM sleep.

**Priming**

Word-stem completion tasks were given to subjects who were part of selective REM deprivation (Conway and Smith 1994) or \( \frac{1}{2} \) night interruption protocols (Plihal and Born 1999). The pattern of observed deficits was consistent with REM involvement in improved performance of this task, particularly late in the night.

**Exceptions**

Thus far, eleven studies have been described, all of which show some role for REM sleep in procedural memory formation. These data are consistent with the dual process theory of memory consolidation during sleep, in which REM sleep is thought to be necessary for procedural memories, and non-REM sleep is thought to be necessary for declarative memory (Smith 1995; Stickgold, Hobson et al. 2001). But not all procedural memory studies show REM sleep facilitation. Performance on a motor pursuit task was sensitive to deprivation of stage 2 sleep only, not REM (Smith and MacNeill 1994). Likewise, the proportion of stage 2 sleep was the only sleep stage correlated with performance improvements of a sequential finger tapping task (Walker, Brakefield et al. 2002). Thus, most, but not all, evidence suggests REM sleep facilitates procedural memory formation.
Implicit memory tasks may show the strongest and most consistent influence of REM sleep, but they are not the only tasks to benefit from REM sleep. When given stories to recall, REM- but not non-REM deprived subjects show impairments (Tilley and Empson 1978). The concerns that REM sleep deprivation causes more general dysfunctions in sleep apply to this study, but there is yet another example of a relationship between REM sleep and declarative memory that did not involve deprivation. Students taking a French immersion class were monitored four nights before the class, during the class term, after completion of the class. Better performance in the class was related to greater proportions of REM sleep during the class term (De Koninck, Lorrain et al. 1989). Language learning is classified as declarative memory, though it may involve the recruitment of a dedicated neural system not otherwise involved in declarative processing. Nevertheless, this study stands out as an exception to the division of REM/non-REM sleep by procedural/declarative types of memories that does not involve sleep interference.

1.2.3 Non-REM

Having just described two accounts of REM-facilitated declarative memories, the question at hand is whether there is sufficient evidence that NREM is a better candidate for supporting declarative memory formation.

Whereas total sleep deprivation results in impaired memory for verbal paired associates (VPA), selective REM deprivation has no effect, suggesting it is the non-REM sleep that facilitates VPA memory consolidation (Chernik 1972; Ekstrand 1972).
Memory for word lists or word clusters is also unimpaired by REM deprivation (Lewin and Glaubman 1975). An alternative explanation to these studies is that some sleep is better than no sleep, irrespective of sleep stage.

To address this issue, a ½ night interruption method was used. Verbal paired-associate (VPA) recall was better for subjects given the first half of one night’s sleep, than for those who were sleep deprived or those given 4 hr sleep, then trained, then given the second half of the night’s sleep (4hr). This suggests a role for early sleep, most notable for its high proportion of SWS (Yaroush, Sullivan et al. 1971; Barrett and Ekstrand 1972; Plihal and Born 1997). Yet another study compared sleep changes following VPA learning to sleep changes using the same stimuli processed for letter shape categorization (Gais, Molle et al. 2002). Aside from the careful control condition, this study analyzed several sleep parameters, including duration, EEG power, and density of sleep-related oscillations. The results were striking: spindle wave density was greater in the VPA group, and was strongest earlier in sleep. The only EEG power difference was in the delta band, suggesting either more or stronger oscillatory activity in SWS. Finally, they observed no change in sleep stage duration. This study suggests “quality not quantity” of non-REM sleep is critical to memory formation. It also illustrates that the brain activity specific to certain sleep stages may selectively implicate some sleep stages in memory processing.

Other tasks involving declarative memories provide further evidence for either stage 2 spindle wave or SWS involvement in memory formation. Subjects exploring a virtual reality maze for 8 hours showed longer time spent in stage 2 sleep as well as
enhanced spindle activity in the following sleep period (Meier-Koll, Bussmann et al. 1999). Using the interruption method before testing for improvements on a declarative mental spatial rotation task and on a procedural word priming task, Plihal and Born (1999) found performance on the rotation task improved specifically with early (SWS-rich) sleep, whereas performance for the priming task improved specifically with late (REM-rich) sleep.

Exceptions

The two exceptions to the dual process theory showing a REM involvement in declarative memory were described in the REM section. The other type of exception would be cases of non-REM involvement in procedural tasks. There are numerous examples of this type, in fact, five of the 14 studies of procedural memory listed in Table 1.2 fall into this category. Note that evidence of non-REM sleep involvement in procedural tasks is not always to the exclusion of a role for REM sleep. By comparison, only two of the ten listed studies of declarative memories are related to REM sleep, and at the exclusion of non-REM sleep. Thus, non-REM sleep appears to be more likely to influence both declarative and procedural memories.

Although the effects on perceptual learning were not assessed, one study of sleep and cortical plasticity also is consistent with a role of non-REM sleep in procedural task consolidation (Frank, Issa et al. 2001). The study in question used a monocular deprivation procedure, known to bias neuronal responses towards stimuli given to the non-deprived eye. Sleep following the deprivation procedure produced a greater bias
effect than an equivalent period of wakefulness in a darkened room (free from monocular deprivation). In fact, the enhancement in the bias is as great following sleep as following an equivalent amount of continued deprivation experience. In this study, the effects were almost entirely explained by non-REM sleep, rather than by REM sleep. Although the relationship between the cortical reorganization and behavior were not addressed, this study represents a link between sleep and the neural mechanisms that may underlie learning.

1.2.4 Dual process and serial theories

What do these results suggest about the dual process theory, which posits that REM sleep processes procedural memories and non-REM sleep processes declarative memories? With only two exceptions, whose merits are questioned above, the following statements are consistent with the data reviewed:

1. When REM sleep is involved in memory processing, it facilitates procedural memory formation.

2. Declarative memories benefit from non-REM sleep.

The classification of memory types is, thus far, a useful distinction to make when considering the vast quantities of studies on sleep and memory consolidation. The dual process theory is certainly not the only one that has been proposed, but only one other, related, theory of the relationship between sleep and memory consolidation comes close to fitting the data described here.
Serial, double-step theories and beyond

The sleep experiments described thus far have been divided by effects on REM and non-REM, and a possible relationship between REM and implicit and non-REM and explicit memory. Several groups have espoused some version of a serial theory of sleep-dependant memory consolidation (Giuditta, Ambrosini et al. 1995; Smith 1995; Plihal and Born 1999; Stickgold, Hobson et al. 2001). Based largely on the ½ night interruption studies, the strong interpretation of the serial theory is that memories are processed in SWS early in the night, then in REM sleep late in the night, regardless of the kind of memory in question. Only one of the seven interruption studies reported here describe both early and late involvement for a memory task, as predicted by the serial theory. By adding the stipulation that declarative tasks should be more sensitive early in the night and procedural tasks should be sensitive late in the night, the number increases to seven out of seven consistent studies. But interruption tasks cannot test which stage of sleep is important early in the night versus late in the night. The assumption is that the predominance of SWS early in the night and of REM late in the night is indicative of their increased function. Several correlation studies have tested this assumption directly.

By far the strongest and most convincing result supporting the serial theory is described in Stickgold et al. (2000). The closest relationship between sleep stage and performance was obtained by combining the SWS in the first ¼ of the night and REM in the last ¼ of the night, producing a correlation of .89 (p<.0001) with task improvement. Of the few other correlation-type sleep studies that have analyzed both sleep timing and
sleep stage as separate factors, none found an early-SWS/late-REM relationship with performance.

In a finger tapping task, Fischer et al. (2002), tested specifically for a relationship between motor learning and a combination of early SWS and late REM, but found only a flat temporal relationship between REM sleep and performance (r = 0.61, p< 0.004).

The same lab that reported the early SWS/late REM effect in a perceptual task failed to find the effect in a simple motor skill task (finger tapping sequence comparable to the task used in Fischer et al. (2002). Walker et al. (2002) reports late-night stage 2 sleep but no other sleep stage is important in motor skill learning, explaining 52% of the inter subject variance in performance (r = 0.72 p < 0.008). This result suggests that late-night benefits may not always be the result of increase REM sleep, as assumed. It also stands against the theory that REM is a necessary sleep stage for memory consolidation, assuming that duration is a sensitive parameter. The authors suggest the contrast between this finding and those reporting SWS and REM involvement may be due to differences in task complexity or perceptual/motor differences, and no additional attempt is made to account for this finding within the framework of the dual process theory.

Despite a large quantity of data, the results do not lead to a clear account for what might be happening in sleep to facilitate memory formation. There is a trend suggesting procedural/declarative differences in the contributions of REM and non-REM sleep. And, though the exact behavioral correlates may not be understood, each sleep stage can be involved independently of the other stages. This issue is unlikely to be resolved until the characteristic neural activity patterns of each sleep stage are better understood. This
would provide fodder for hypotheses of *mechanisms* of memory consolidation during sleep. For example, the few reports analyzing the density and strength of stage 2 spindle oscillations support the idea that the underlying neural activity patterns in a sleep stage determines its role in memory consolidation. The global synchrony of spindle oscillations may be relevant to the rewiring of cortical ensembles. As such, an understanding of how cortical network dynamics (e.g. oscillations) lead to plastic changes will impose important constraints on the evolving theories of both general memory consolidation, and sleep-dependent consolidation.
Table 1.2 Relationship between different sleep stages and learning of declarative or procedural memories. The study, task, sleep manipulation, and sleep stage are listed below. Abbreviations: E, early sleep; L, late sleep; + indicates that stages plays a role in consolidation, - indicates no role in consolidation. For added clarity, the + cases are given green highlights in the non-REM sleep column, and yellow highlights in the REM sleep column.

<table>
<thead>
<tr>
<th>Task Manipulation</th>
<th>Task</th>
<th>Manipulation</th>
<th>non-REM</th>
<th>REM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stage 1</td>
<td>Stage 2</td>
<td>SWS Stages 3, 4</td>
</tr>
<tr>
<td>Priming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conway &amp; Smith 1994</td>
<td>Word Stem</td>
<td>REM Dep</td>
<td>-E</td>
<td>+L</td>
</tr>
<tr>
<td>Plihal &amp; Born 1999</td>
<td>Word Stem</td>
<td>Intrpt</td>
<td>-E</td>
<td>+L</td>
</tr>
<tr>
<td>Perceptual skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karni et al. 1994</td>
<td>TDT</td>
<td>Dep</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Stickgold et al. 2000</td>
<td>TDT</td>
<td>Corr</td>
<td>+E</td>
<td>+L</td>
</tr>
<tr>
<td>Gais et al. 2000</td>
<td>TDT</td>
<td>Intrpt</td>
<td>+E</td>
<td>+/-L</td>
</tr>
<tr>
<td>Kattler et al. 1994</td>
<td>SSstim</td>
<td>Corr</td>
<td>+E</td>
<td>n/a</td>
</tr>
<tr>
<td>Verschoor &amp; Holdstock 1984</td>
<td>Vis/Aud Lm</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Motor skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plihal &amp; Born 1997</td>
<td>MirrorTrace</td>
<td>Intrpt</td>
<td>-E</td>
<td>+L</td>
</tr>
<tr>
<td>DeKoninck &amp; Prevost 1991</td>
<td>Prism Adaptation</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Buchegger et al. 1991</td>
<td>Trampoline</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Fischer et al. 2002</td>
<td>FOS</td>
<td>Corr</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Conway &amp; Smith 1994</td>
<td>Corsi, Hanoi</td>
<td>REM Dep</td>
<td>-E</td>
<td>+L</td>
</tr>
<tr>
<td>Smith &amp; MacNeill 1994</td>
<td>Pursuit Rotor</td>
<td>Dep</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Walker et al. 2002</td>
<td>Fgr Tap Seq</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Explicit/Declarative</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castaldo et al. 1974; Chernik 1972</td>
<td>VPA</td>
<td>REM Dep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yarouch et al. 1971</td>
<td>VPA</td>
<td>Intrpt</td>
<td>+E</td>
<td>-L</td>
</tr>
<tr>
<td>Barrett &amp; Ekstrand 1972</td>
<td>VPA</td>
<td>Intrpt</td>
<td>+E</td>
<td>-L</td>
</tr>
<tr>
<td>Plihal &amp; Born 1997</td>
<td>VPA</td>
<td>Intrpt</td>
<td>+E</td>
<td>-L</td>
</tr>
<tr>
<td>Gais et al. 2002</td>
<td>VPA</td>
<td>Corr</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Plihal &amp; Born 1999</td>
<td>VPA</td>
<td>Corr</td>
<td>+E</td>
<td>-L</td>
</tr>
<tr>
<td>Meier-Koll et al. 1999</td>
<td>Spatial</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tilley &amp; Empson 1978</td>
<td>2°''Language</td>
<td>R/St4 Dep</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>DeKoninck et al. 1989</td>
<td>2°''Language</td>
<td>Corr</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
1.3 CORTICAL NETWORK ACTIVITY

The activity of neural ensembles is based on the intrinsic membrane properties of the constituent neurons and on their synaptic interactions, which may also include neuromodulatory influences. Neocortical and hippocampal networks will be described separately, owing to the fundamental differences in their characteristic patterns of ensemble activity. Despite the differences, however, the picture that emerges illustrates overarching similarities between hippocampal and neocortical network activity that may contribute to the formation of distributed memory traces.

1.3.1 Neocortical networks

Neocortical neurons have been classified according to four distinct firing classes: regular spiking (RS), intrinsic bursting (IB), fast rhythmic bursting (FRB, also ‘chattering’), and fast spiking (FS). Initially, firing classes were based on the results of in vitro studies, using a slice preparation (Connors, Gutnick et al. 1982; McCormick, Connors et al. 1985) but evidence has also been found in vivo, in a chronic preparation of unanesthetized cats (Steriade, Timofeev et al. 2001).

Early investigations led researchers to believe that pyramidal cells belonged to different classes than interneurons. Although both “fast” classes typically contain inhibitory interneurons, excitatory pyramidal cells have been observed among their ranks. Moreover, it is possible for cells to change classes, depending on the experimental setup or the animal’s vigilance state. For example, excitatory corticothalamic and inhibitory local circuit basket cells can transition from RS to FRB to FS in response to increasing membrane depolarization (Steriade, Timofeev et al. 2001). Similarly, as synaptic activity
increases, IB neurons are transformed to RS neurons. These transitions may be controlled by modulatory neurotransmitters (Mason and Larkman 1990), possibly explaining why fewer IB and more FS neurons are reported in the awake animal versus the slice. The awake animal has greater levels of synaptic activity and modulatory neurotransmitters (i.e., a more depolarized membrane) thus IBs are converted to RSs and FRBs are converted to FSs. These class distinctions may appear to convey little information, since, for example, an FRB cell may be excitatory or inhibitory and a given cell is not bound to remain an FRB. On the contrary, class transitions reflect the level of depolarization and excitability of the cell and how it may respond to EPSPs. These factors vary with vigilance state and determine the types of network activity that may ensue. Thus, an understanding of how and when class transitions occur provides information about the network activity and about the vigilance states of the animal, from waking, to slow-wave sleep, and ultimately into REM sleep.

Table 1.3 Functional classification of neocortical cells. RS - regular spiking, IB - intrinsic bursting, FRB - fast rhythmic bursting, FS - fast spiking, ISI - inter-spike interval.

<table>
<thead>
<tr>
<th>Name</th>
<th>RS</th>
<th>IB</th>
<th>FRB</th>
<th>FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burst activity</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Maintained activity?</td>
<td>No, adapts</td>
<td>No, inactivates</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ISI (ms)</td>
<td>20-100</td>
<td>3 (in burst)</td>
<td>1.5-3.5 (in burst)</td>
<td>20-50</td>
</tr>
<tr>
<td>Spike width (ms)</td>
<td>0.6-1</td>
<td>&gt;0.5-1</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Proportion (wake)</td>
<td>51%</td>
<td>&lt;5%</td>
<td>21%</td>
<td>24%</td>
</tr>
<tr>
<td>Morphology</td>
<td>typically pyramidal</td>
<td>typically Layer V pyramidal</td>
<td>Interneurons, Layer II/III &amp; corticothalamic pyramidal</td>
<td>typically GABA interneurons</td>
</tr>
</tbody>
</table>
Waking activity

Although wakefulness is characterized by depolarized membrane potentials and high firing rates in the neocortex, the input resistance of these cells is higher than in other vigilance states (Steriade, Timofeev et al. 2001). Rather than indicating decreased synaptic activity, high input resistance is probably elevated due to increased levels of neuromodulators such as ACh, NE, and 5-HT. Among other actions, these neuromodulators reduce K⁺ currents, increasing membrane resistance while also making cells more depolarized. The related conditions of a depolarized membrane, increased neuromodulator levels, and increased input resistance (all signatures of the waking state) provide the necessary conditions to elicit fast oscillations.

Fast Oscillations

The most salient features of fast oscillations are their low amplitude field potentials and their lack of global synchronization. Fast oscillations occur in the sigma and gamma frequency bands (20-80Hz) during wakefulness, particularly during periods of focused attention. Unlike the slower oscillations of slow-wave sleep, cortical fast oscillations are in phase throughout the cortical depth (Steriade and Amzica 1996), and are only locally coherent in space (<1mm) and duration (100-500ms) (Eckhorn, Bauer et al. 1988; Gray and Singer 1989; Steriade and Amzica 1996; Destexhe, Contreras et al. 1999). Synchronization of fast oscillations in two separate regions of neocortex has been observed, but this synchronization is transient and is not part of a global synchronization
of the neocortex (Gray 1994). The EEG recorded during wakefulness is therefore relatively flat, containing low amplitude fast oscillations that vary from region to region.

A case could be made that global synchronization is not necessary for fast oscillations to be useful in coordinated trace replay, only that the relevant ‘modules’ be synchronized; however, the limited temporal synchrony makes it unlikely that fast oscillations would provide sufficient bouts of replay for consolidation. The ability of fast oscillations to facilitate Hebbian plasticity over distributed circuits remains to be demonstrated.

**Slow-wave sleep**

The progression of sleep is classified into non-REM sleep, specifically Stages 1-4, occurring in that order, followed by REM sleep. The patterns of EEG activity in non-REM sleep consist of several types of large-amplitude oscillations, all of which are of slower frequency than the oscillations seen in wake or REM sleep, hence the moniker ‘slow-wave sleep’. The slower oscillations are most notable in Stages 2-4 sleep, and are closely linked to thalamic activity. As sleep begins, the thalamus cuts off its widespread sensory input, disconnecting the cortex from the external world. In place of sensory relays, the thalamus develops and propagates its own oscillations to the neocortex, constituting two of the three characteristic oscillations of slow-wave sleep. These three oscillations, spindles, delta waves, and slow waves will be described in descending order of frequency.
Spindle oscillations

Spindle oscillations are encountered at the onset and in the early stages of sleep. Perhaps because of their striking symmetrical pattern of waxing and waning, or perhaps because they mark sleep onset, spindle waves have been described in the human EEG literature as early as Berger (1933) and Loomis et al. (1935). Spindles occur at 12-18Hz in humans, and 7-14 Hz in rats and cats, and typically last 1-5s (Steriade and Deschenes 1984). A hallmark of stage 2 sleep onset, spindles are diminished in stage 3 sleep and disappear in stage 4 sleep, REM sleep, and upon arousal (Jankel and Niedermeyer 1985). Spindles are nearly simultaneous throughout neocortex in vivo in cats and humans (Contreras, Destexhe et al. 1996; Contreras, Destexhe et al. 1997). Whereas the thalamus generates spindles, corticothalamic cells are responsible for the widespread synchronization of spindles.

Unlike fast oscillations, spindle waves undergo phase reversal from superficial to deep layers of cortex, meaning that a superficially recorded spindle peak corresponds to a depth recorded spindle trough (making it necessary to specify whether peaks and troughs are recorded superficially or with depth electrodes). Neocortical pyramidal cells fire in the depth trough of spindles, the result of an EPSP/IPSP sequence that increases Ca$^{2+}$ entry into dendrites (Contreras, Destexhe et al. 1997). The dramatically decreased membrane resistance is due to both strong glutamatergic conductances in the dendrites but also GABAergic conductances near the soma, producing shunting inhibition. The result is greater synaptic activity and Ca$^{2+}$ entry than would be expected based solely on the observed action potential output of that cell. In the extreme, the shunting inhibition
could prevent excessive firing that might otherwise follow the increased levels of Ca\(^{2+}\).

Several molecular cascades associated with plastic changes are sensitive to Ca\(^{2+}\) entry, within the frequency ranges provided by thalamocortical inputs during spindle activity (Gu and Spitzer 1995; De Koninck and Schulman 1998; Li, Llopis et al. 1998). Thus, spindles may themselves elicit synaptic change or play a preparatory role, recruiting the molecular machinery that will be needed for modification of subsequent synaptic events.

Evidence that spindles may be involved in synaptic plasticity include the potentiated responses of cortical association neurons observed following spontaneous or evoked spindle frequency oscillations (Steriade, Timofeev et al. 1998). In addition, once spindles are triggered, they become self-sustained (i.e. independent of a triggering stimulus), reminiscent of reverberatory circuits (Bazhenov, Timofeev et al. 1998; Steriade 2000). Spindle-frequency reverberation is confined to SWS, when thalamic synaptic inhibition occurs, effectively cutting the neocortex off from sensory inputs. These occurrences are all conducive to the offline replay of neural patterns, when repeated synchronous oscillation in the absence of incoming stimuli could help establish a neocortical memory trace.

**Delta waves**

The thalamus also generates delta oscillations (1 or 2 - 4 Hz) (Steriade, McCormick et al. 1993), which appear in neocortical Layers II/III and V (Steriade, Jones et al. 1990) and are susceptible to neocortical modulation and synchronization. Delta waves predominate in the late stages of SWS.
Two currents in thalamocortical cells interact to generate intrinsic delta waves: the hyperpolarization-activated cation current ($I_h$) and the transient low-threshold Ca$^{2+}$ current ($I_t$) (Steriade 2001). The thalamocortical cells must undergo hyperpolarization for the $I_h$ current to deinactivate, and the delta wave to arise.

Like spindle oscillations, delta waves contain phases of hyperpolarization that can deinactivate Ca$^{2+}$ channels, allowing greater Ca$^{2+}$ influx during the depolarizing phase. The Ca$^{2+}$ increases could occur as a direct result of T-type Ca$^{2+}$ channel opening, but also by the consequent opening voltage-dependent NMDA receptors. Synaptic modification could, therefore, be the end result of high dendritic Ca$^{2+}$ levels occurring during delta waves, though this has yet to be verified experimentally.

_Slow oscillations_

The slow oscillation of 0.02 – 1Hz was first described in detail in intracellular recordings of anesthetized cats (Steriade, Contreras et al. 1993; Steriade, Nunez et al. 1993; Steriade, Nunez et al. 1993) but had also been observed in cortical slice preparations and in the EEG of humans during slow-wave sleep (Achermann and Borbely 1997; Sanchez-Vives and McCormick 2000). It is present throughout all stages of SWS (Amzica and Steriade 1997; Steriade and Amzica 1998) occurring at the onset of SWS but not visible during waking (Steriade, Timofeev et al. 2001). Slow and delta oscillations, though similar in frequency ranges, are distinct, as the latter can occur embedded in the former (Steriade, Nunez et al. 1993) and delta waves decline in
subsequent bouts during synchronized sleep, in contrast to slow waves, which are maintained (Achermann and Borbely 1997).

If further classification is to be made, spindles and delta waves can be thought of as the “faster” SWS oscillations, often embedded in slow oscillations. This classification extends beyond frequency, however. Whereas delta and spindle waves are generated in the thalamus and synchronized through corticothalamic feedback, slow wave oscillations are generated in the neocortex but synchronized over large neocortical areas via thalamic inputs. The synchrony associated with slow oscillations is far greater than that associated with spindle oscillations (Contreras and Steriade 1997). Moreover, some, but not all, cell types tend to be more active in the depth negative phase of spindle and delta waves. During slow oscillations, all four neuron types are active in the depolarizing phase (depth EEG troughs) and silent in the hyperpolarization phase (depth positive field potentials) (Destexhe, Contreras et al. 1999). This striking modulation in global neocortical activity led to the description of slow oscillation activity as “up-“ and “down-states”.

The hyperpolarizing part of the slow oscillation is accompanied by a substantial increase in membrane resistance, a level twice as much as that seen in the depolarizing phase, and similar to levels seen in wakefulness. This, plus the absence of GABAergic fast-spiking activity (FS), during this phase, indicate that GABA is not responsible for the hyperpolarization. According to Steriade (2001) disfacilitation and Ca^2+ dependent K+ currents must be the main mechanism for the hyperpolarization. Slow waves, therefore are driven by very different cellular mechanisms than delta or spindle oscillations, the
former independent of GABA inhibition, and the latter subject to oscillatory GABA shunting.

Release from the slow oscillation hyperpolarization phase leads to a fast transition into the depolarized up state. The transition often triggers delta or spindle bouts. In fact, one of the most recognizable EEG signatures during SWS, the K complex, is an example of slow-wave and spindle grouping. The typical K complex is composed of an initial sharp, transient peak (depth positive) followed by a depth negative slow wave lasting >0.5 sec, on which a train of spindles often rides. This is consistent with the early reports that K complexes occur primarily in stage 2 sleep, when spindles are most common (Jankel and Niedermeyer 1985).

**REM Sleep**

The oscillatory and single unit activity seen in the neocortex during rapid eye movement (REM) sleep is nearly identical to the pattern seen in wakefulness. To summarize the key features, both REM and wakefulness are characterized by low-amplitude, locally coherent fast oscillations. Increased levels of ACh, increased membrane resistance, more depolarized resting membrane potentials, and fewer bursting cells are observed. The observation that fast oscillations can become synchronized briefly in two spatially separate areas, leaves open the possibility that some type of plasticity may occur, although the synchronization is more spatially and temporally restricted than that seen during SWS oscillations (Steriade and Amzica 1996). Thus, just as fast oscillations in waking activity may be important for some neural processes, so, too, may
be fast oscillations in REM sleep; however, current evidence does not suggest that fast oscillations are the ideal candidate for Hebbian plasticity in widely distributed neural ensembles. One point of speculation: REM sleep has previously been implicated in procedural learning. If procedural learning is somewhat restricted to local ensembles and does not require a widely distributed trace, then the local synchrony of fast oscillations may be sufficient for consolidation. More research will be needed to test this supposition.

1.3.2 Hippocampal networks

Network activity in the hippocampus is unique from that of the neocortex, but shares some features such as the dependence on behavioral state, and the nesting of fast oscillations in slower ones. The two hallmark slower oscillations around which both behavioral state and faster oscillations can be grouped are theta rhythms and sharp waves.

Packing it in: theta, gamma, and exploration

One of the two hallmark oscillations in the hippocampus is the theta rhythm, also called rhythmic slow activity (RSA). Theta waves are present in human and non-human primates, and are also seen in the rat during voluntary movements such as walking, rearing, and other exploratory behaviors as well as during REM sleep (Vanderwolf 1969; Arnolds, Aitink et al. 1979). For a recent review of theta production, see Buzsaki (2002).

The hippocampus, known to be important for spatial memory, shows place-selective activity only during theta periods. Procedures that diminish hippocampal theta impair spatial memory, suggesting it may contribute to memory formation. Theta may
facilitate encoding by effectively ‘binning’ input signals into time windows more suitable for synaptic modification. Long-term potentiation (LTP) in the hippocampus is produced following pairs of pulses separated by around 200ms (5Hz), or stimulus trains delivered on the peaks of theta in stratum radiatum or the hilus, when cells are maximally excitable (Larson and Lynch 1986; Greenstein, Pavlides et al. 1988; Huerta and Lisman 1995; Orr, Rao et al. 2001). Subsequent stimulation delivered on the troughs of theta produces depotentiation (Huerta and Lisman 1995; Holscher, Anwyl et al. 1997). Consistent with the role of theta in grouping input signals, the hippocampus receives inputs from neurons in superficial entorhinal cortex during theta periods, in phase with theta recorded in the dentate gyrus (Mitchell and Ranck 1980; Boeijinga and Lopes da Silva 1988; Stewart, Quirk et al. 1992).

Theta also groups or modulates the amplitude of oscillations in the gamma frequency range (40-100 Hz). Gamma is strong and uninterrupted in the hilar region of the dentate gyrus, resulting from synchronous discharge in hilar interneurons. As with theta, superficial layer neurons in the entorhinal cortex are active in phase with local gamma. A given projection neuron may not fire with each cycle of gamma, but as a population, they are capable of producing strong volleys that are received by hippocampal neurons entrained to the same frequency (Bragin, Jando et al. 1995; Chrobak and Buzsaki 1998).

In sum, as an animal is exploring, taking in information about its environment, the hippocampus produces rhythmic time windows of increased excitability. Inputs from neocortex become similarly grouped, producing a summation of excitation that arrives in
precisely the time window most effective at eliciting hippocampal responses. Preliminary evidence is consistent with a role for theta and embedded gamma oscillations to facilitate synaptic change within hippocampal circuits (Buzsaki 1989).

The role of theta during REM sleep is less clear, but by extension from the waking functions described above, it may involve the encoding of the neocortical activity patterns prominent during REM sleep.

Shipping it out: sharp waves, ripples, and 'pausing for reflection'

The other critical oscillations in the rat hippocampus are sharp waves and the ripple oscillations embedded in them. Sharp waves, also referred to as EEG spikes or large-irregular activity (LIA), have been described in rats and primates, during SWS, periods of quiet wakefulness, and during consummatory behaviors (Freemon, McNew et al. 1969; Freemon and Walter 1970; Vanderwolf 1990). Sharp waves are large-amplitude waves lasting 50-150ms, and repeating aperiodically every 1-2 sec. During exploratory behaviors and REM sleep, sets of highly interconnected CA3 neurons are actively inhibited. In quiet wakefulness and SWS, they are freed from inhibition, thus the Shaffer collateral projections to CA1 produce a massive depolarization in pyramidal cell dendrites (Buzsaki 1986; Chrobak and Buzsaki 1998). The additional activation of basket cells in CA1 produces rapid, rhythmic IPSPs at the soma of the CA1 pyramidal cells. This EPSP/IPSP interaction is the basis for the 125-300Hz ‘ripple’ oscillation that occurs during sharp waves (Ylinen, Bragin et al. 1995), seen in a variety of species (O'Keefe and Nadel 1978; Kanamori 1986; Bragin, Engel et al. 1999). As is typical of
pyramidal cell responses to IPSP-based oscillations, the cell may not spike on each depolarized phase of the membrane oscillation, but when threshold is reached, it will be during the depolarizing phase. There are 5-10 waves in a typical ripple, and whereas each sharp wave may recruit roughly 10,000 CA1 cells, each individual ripple wave appears to recruit select subsets of those cells (Chrobak and Buzsaki 1996; Chrobak and Buzsaki 1998). Subiculum and deep layer entorhinal cortex fire concurrently with CA1 population bursts, suggesting a possible mechanism for hippocampal ensembles to influence neocortical activity (Chrobak and Buzsaki 1994; Chrobak and Buzsaki 1996).

During 'offline' periods of behavior, such as quiet wakefulness and SWS, it is not necessary for the neocortex to process information about the external world, making sleep an ideal time for 'firming up' memory traces (Marr 1970; Marr 1971). It is precisely during these periods that the hippocampus appears to be in 'output' mode. Coincidentally, in SWS, the neocortex undergoes an unparalleled period of synchronous activity, possibly facilitating plastic change. But is there any relationship between the hippocampal and neocortical oscillations seen in SWS, or are they independent phenomena?

1.3.3 Hippocampo-cortical interactions

Theories of memory consolidation must account for an initial hippocampal-dependence for memory formation that decreases over time. Typically, the hippocampus is credited with the facilitation of neocortical memory consolidation, by teaching, indexing, or binding the appropriate neocortical network. It is possible, however, that the
hippocampus does not interact with the neocortex at all – that the neocortex has all the necessary machinery for memory formation, requiring only time to consolidate. One clue to the role of the hippocampus in neocortical memory trace formation is the synchronization of network activity between the hippocampus and the rest of the neocortex.

The ‘output’ oscillations of the hippocampus, sharp wave/ripple complexes, appear to co-occur with spindle oscillations, with ripples tending to precede spindle onset by <1 sec (Siapas and Wilson 1998; Sirota, Csicsvari et al. 2003). As a group, CA1 pyramidal cell unit activity shows a slight tendency to fire within a 1 sec window of spindle onset, and neocortical cells showed a small modulation to ripples, visible within one second following ripple onset. The modulation of neocortical cells was similar in magnitude and delay for both ripples and spindles. The size of the effects, with correlation coefficients between .01 and .04, may seem insufficient to comprise a strong synchronizing signal across areas. On the other hand, these correlations were based on activity from 39 hippocampal and 27 neocortical single units. If ripple-spindle activity is responsible for coordinated recruitment of a memory trace, then only a small fraction of the available cells should be recruited. More cells, or specifically those cells thought to contribute to a memory trace, would need to be studied before determining the significance of ripple-spindle interactions.

When over 2600 single units in neocortex were examined, over 20% of the cells showed significant increases in activity during sharp wave events (F. Battaglia and B. McNaughton, submitted). This suggests the results of the previous study were merely the
tip of the iceberg (Siapas and Wilson 1998). Consistent with these results, increases in the delta and spindle frequency bands were observed in neocortical power spectra aligned to ripple periods (Sirota, Csicsvari et al. 2003).

Interestingly, in the second preceding sharp waves, neocortical cells were more likely to decrease their firing rate, with roughly 20% showing significant decreases in activity. This global dip in activity extended up to 5s prior to sharp wave onset. Global changes in neocortical activity fluctuating every few seconds are characteristic of the neocortical slow oscillation, known to contain spindles. Indeed, the transition from down to up state in the slow oscillation was somewhat predictive of sharp wave onset, with sharp waves tending to occur slightly before or at the time of transitions to the up state (F. Battaglia and B. McNaughton, submitted). The silence in the neocortex associated with down states may cause disinhibition of CA3 networks, eliciting sharp waves in CA1. This may, in turn, contribute to an up state transition in neocortex. Such transitions are often accompanied by the onset of faster delta and spindle oscillations in the neocortex, and, consistent with this pattern, spindle and delta activity is far greater in the seconds following sharp waves. It therefore appears that the neocortex responds following hippocampal output activity, and that the two regions share state fluctuations in the timeframe of seconds, either through direct interactions or through a third shared input.

There is also evidence that the hippocampus responds to the faster group of SWS oscillations in the neocortex. Spindles and theta oscillations are associated with increased ripple and single unit activity in the 20-100ms timeframe, though the implications of these interactions are less clear (Sirota, Csicsvari et al. 2003).
It is known that the CA1 pyramidal cells participating in ripples vary over each ripple wave. There is some indication that the distribution of neocortical cells activated following sharp waves also varies. Similarly, the variable spatial distribution of spindle power in successive spindle events may be another correlate of the same phenomenon (Destexhe and Steriade 1998).

One possible scenario is that memory trace consolidation results from the synchronized activation of appropriate ensembles across hippocampo-cortical networks. The spindle participants could determine the sharp wave participants, as suggested by Csicsvari (2000). Given the propensity of CA3 for pattern completion (Mizumori, McNaughton et al. 1989; Nakazawa, Quirk et al. 2002), the enriched CA1 output could 1. help select a more distributed range of cells contributing to the memory trace, or at least 2. provide a means of binding hippocampal output to those neocortical cells active in the spindle, strengthening the cortical network. Each spindle wave is yet another opportunity to refine the selection of neurons involved in the memory trace, producing yet another burst of dendritic calcium influx to promote plasticity in those cells that contributed to the spindle activity. Each new spindle bout may begin with activity in only a local region of interconnected neocortical cells but, perhaps through hippocampal feedback, expands, or is refined, to include those cells constituting the distributed memory trace. This broad speculation requires evidence that those cells active during an event tend to be the cells that are active during subsequent oscillatory periods.
1.4 MEMORY TRACE REACTIVATION

One of the foremost researchers of memory consolidation recently wrote, “The 1990s marked the period in which the phenomenon of temporally graded retrograde amnesia was documented in experimental animals. Perhaps the new decade will yield the first clues about its mechanisms.” (Squire, Clark et al. 2001 p.54)

Although much ground remains to be covered, we now have the first glimpses of neural activity that may underlie the process of memory consolidation. Studies using positron emission tomography (PET) describe “reactivation” in humans (Maquet, Laureys et al. 2000; Laureys, Peigneux et al. 2001); however, the resolution of these imaging methods precludes the identification of neurons belonging to a given memory trace (particularly for distributed memory traces). The “reactivation” detected may reflect merely the continued correlates of the experience such as elevated body temperature or metabolic recovery that may be differential across brain regions. While the results are consistent with the timing of consolidation, and provide some converging evidence across species, more tangible evidence of the reactivation of a neural memory trace is needed, consistent with theories of memory consolidation.

According to trace replay theory, the neural mechanism of memory consolidation involves the repeated, coincident activity of a distributed memory trace, in order for the trace to be strengthened. To obtain coincident activity throughout the network of a new memory trace, top-down feedback is necessary, namely, from the hippocampus to the neocortical ‘modules’. If the theory is correct:
1. The pattern of neural ensemble activity specific to an event will be reactivated during subsequent periods of behavioral inactivity, a phenomenon henceforth referred to as 'memory trace reactivation'.

2. Reactivation of an event-related hippocampal ensemble will elicit coordinated reactivation of the neocortical ensemble, which is distributed throughout various regions of neocortex.

3. The neocortex will undergo selective synaptic strengthening among neurons representing the memory trace.

Evidence for the first prediction is growing, primarily based on recordings from the hippocampus of rats. The first indication of offline replay came from the increased firing rate specific to hippocampal cells that were active in previous behavior, but not in cells that were inactive (Pavlides and Winson 1989). Because only active/inactive pairs of cells were analyzed, the effect could have reflected merely an increased propensity for a previously active cell to fire (a necessary but not sufficient result), and says nothing about selective reactivation of neural ensemble activity patterns from previous behavior.

Hippocampal neurons become active for particular locations, or 'place fields', as a rat traverses a track or arena. Because cells have different place fields, a map of the environment can be reconstructed from the ensemble activity of hippocampal neurons (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978). The spike trains of cells with overlapping place fields have correlated activity during waking behavior, but also a greater tendency for correlated activity during SWS after behavior, both in CA1 (Wilson
and McNaughton 1994) and dentate gyrus (Shen, Kudrimoti et al. 1998) principal cells. Conversely, pairs of cells with non-overlapping place fields have uncorrelated spike train activity during behavior and in sleep after behavior. These higher sleep correlations in the ‘overlap’ group cannot be explained by non-specific increases in firing rates of individual cells, since cells in both groups were active during behavior. Moreover, the preservation of correlations appears to be task-related rather than merely reflecting pre-existing patterns of activity: the spike-train correlations during the task were more closely related to those of sleep following the task than those of sleep preceding the task. The most parsimonious explanation is that the cells active together during the task reflect part of a memory trace that is reactivated during sleep after the task.

Rather than partition cells pairs into two groups based on their activity during behavior (overlapping and non-overlapping), later studies compared the full distribution of correlations between all cell pairs during behavior to the distribution of correlations in subsequent sleep (Qin, McNaughton et al. 1997; Kudrimoti, Barnes et al. 1999). To control for any variance in sleep that was not task-related, they factored out the contribution of the pairwise spike-train correlations during preceding sleep to the correlations of both task and subsequent sleep epochs. The result revealed a small but sometimes significant correlation between preceding sleep and behavior, but the effect was never large enough to eliminate a significant explained variance between behavior and subsequent sleep. The explained variance measure revealed further evidence that task-related changes in ensemble activity recur during periods of inactivity following experience. Because the task was familiar to the animals at the time of recording, the
small correlations between task and preceding rest may reflect continued reactivation of the previous day’s session.

The smaller magnitude of reactivation seen in sleep preceding a familiar task is consistent with observations of decay over time. In a 30-minute rest session following behavior, each successive 10-minute block of time shows less explained variance between behavior and later sleep, factoring out preceding sleep (Kudrimoti, Barnes et al. 1999). The experimental design of that study included two maze experiences, flanked by rest sessions. The final rest session showed reactivation related to both mazes, though the most recently experienced maze is associated with the greatest explained variance. The interleaved reactivation of the first maze argues against reverberation or continuing activity of behavioral sequences; rather, it demonstrates that neural activity patterns re-emerge during sleep or quiet wakefulness.

The spike-train correlation method of measuring reactivation is robust in that it can be reliably detected using hippocampal neural ensembles of at least 25 cells, it is relatively unaffected by changes in the bin sizes used to calculate the spike-train correlations, at least between 20 and 100 ms, and the correlations are independent of changes in firing rate (S. Cowen, unpublished observations). Using this method, memory trace reactivation has now been detected in several different studies. There are limitations, however, in what aspects of reactivation the spike-train correlation method can measure.
1.4.1 Reactivating sequences

It is unclear from spike-train correlations whether the order of activity patterns is preserved during reactivation. Although the results described above are consistent with a replay of ensemble activity in the same order as that observed during behavior, it is also consistent with “strobes” of ensemble activity, occurring irrespective of the order observed during behavior. Evidence for the former comes from a study in which the order of activity between cell pairs during behavior was compared to the order of activity during sleep before and after. Results showed that if activity in one cell consistently preceded activity in the other cell during maze running, the cells tended to show the same temporal bias in subsequent sleep, but not in preceding sleep (Skaggs and McNaughton 1996).

The temporal bias measure of reactivation provides the only evidence to date that there is a connection between reactivation and memory. Once age is factored out, there was a significant relationship between rats’ spatial memory performance on the Morris water maze and the temporal bias correlations obtained from a different maze-running task and in rest afterward ($R^2 = 0.388$, $p < 0.01$; Gerrard, 2002).

It has been suggested that poor cell isolation could produce an erroneous temporal bias in cells with no overlapping activity (Quirk and Wilson 1999). The extreme conditions used to generate the artifact suggests it is unlikely to be the sole source of the temporal bias results reported in Skaggs and McNaughton (1996). This point was reinforced by the report of reduced preservation of temporal bias in aged compared to adult rats despite the same cell isolation procedures (Gerrard 2002).
As further evidence of sequence reactivation, the spike trains of 4-8 CA1 pyramidal cells were simultaneously recorded and repeating spike sequences (of around 200ms duration) were counted, using a template matching algorithm (Nadasdy, Hirase et al. 1999). Additionally, the occurrences of spike triplets were counted, the joint probabilities of each pair in the triplet were factored out, and the resulting “unexpected triplets” were compared to those occurring from shuffled spike trains. Both methods detected more sequences for real spike trains than for that of shuffled data. Sequences common to behavior and subsequent sleep occurred more frequently than to behavior and preceding sleep.

Both the template-matching and “unexpected” triplet sequence measures of reactivation have been expanded upon by another research group (see Louie and Wilson 2001; Lee and Wilson 2002). A sliding-window template-matching algorithm was applied to rat hippocampal ensemble recordings during bouts of sleep and activity (Louie and Wilson 2001). When compared to shuffled spike train matches, 44% of the bouts showed a significant match to a period of the behavior. This is further evidence that the order of activity is preserved from behavior to reactivation. In a second study, the relative order of neural ensemble activity in a SWS burst was scored according to the degree of match with the order of activity as the rat repeatedly traversed a track (Lee and Wilson 2002). Using combinatorics, the probability of obtaining a match as good or better than that observed was used to establish chance levels of matching. The results were clear-cut: spike sequences obtained in subsequent sleep exhibited significant similarity to behavioral sequences; the sequences from preceding sleep were not significantly similar.
For this result, a minimum of four cells was included in the spike sequence, producing the strongest indication to date that the sequences of the neural code of an event are faithfully replayed.

Although in an entirely different species and neural system, it was discovered that cells whose activity predicted motor articulation of birdsongs "replayed" in subsequent sleep sessions (Dave and Margoliash 2000). This spontaneous reactivation of the song pattern exceeded chance levels using several shuffling procedures. Furthermore, when songs were actually played back, the cellular activity closely matched the birdsong, with the exception that the neural activity continued after the song had stopped, producing activity corresponding to what would have been the next syllable in the song. That replay of waking activity is seen across species with diverse neural systems suggests it may represent a fundamental mechanism of plasticity in neural ensembles.

1.4.2 Sleep state: SWS versus REM

Several studies have detected reactivation in SWS following behavior (Wilson and McNaughton 1994; Nadasdy, Hirase et al. 1999; Lee and Wilson 2002), but also in quiet wakefulness, in both cases revealing greater reactivation during sharp wave/ripple complexes than during the inter-ripple periods (Kudrimoti, Barnes et al. 1999; Gerrard, Kudrimoti et al. 2001). Reactivation during REM in the hour after behavior was analyzed in one of these studies, but no reactivation was detected (Kudrimoti, Barnes et al. 1999). In contrast, only one study to date reported reactivation specifically during REM sleep episodes, when it occurred primarily preceding familiar behaviors rather than following
them (Louie and Wilson 2001). The ensemble and temporal structure of activity appears to be coherent during REM reactivation, although the timing appears to be expanded slightly. This is interesting in that SWS reactivation may constitute a temporally coherent, but considerably compressed pattern, when compared to waking patterns (Skaggs and McNaughton 1996; Lee and Wilson 2002). The onset of reactivation in different sleep stages is consistent with the serial theories of consolidation during sleep, in that reactivation may occur during SWS early in the night and during REM sleep late in the night. At present, the safest conclusion is that reactivation occurs during SWS, regardless of its role in REM sleep. The ripple activity in CA1 during SWS and quiet wakefulness provide physiological conditions conducive to synaptic modification, and also appears to be a period during which the hippocampus and neocortex share correlated oscillations (Siapas and Wilson 1998; Sirota, Csicsvari et al. 2003 Battaglia and McNaughton, in preparation).

1.4.3 Beyond the hippocampus

If hippocampal reactivation is a principle mechanism subserving memory consolidation, as proposed by trace replay theory, then not only should there be reactivation in neocortex, but the cross-region activity patterns from behavior should be reactivated together in both regions. There is only one reactivation study including neocortical recordings or involving multiple structures (Qin, McNaughton et al. 1997). Cells from CA1 and the posterior parietal cortex (PPC) of rats bear out the predictions of trace replay theory: in sleep following behavior, pairwise correlations both within and
between the hippocampus and PPC show significant explained variance with activity during behavior, even after accounting for the covariances in prior sleep. Moreover, temporal asymmetry of cross-correlograms was preserved from behavior to subsequent sleep within, but not between, structures. This suggests that temporal order is preserved within a brain region, but the order from hippocampus to PPC is altered during sleep following behavior. Such a pattern is expected to occur if the hippocampus orchestrates (i.e., leads) reactivation in neocortical structures, whereas the flow of information during acquisition is the reverse. Some theories have suggested that the CA3 region of the hippocampus could initiate retrieval of pattern that would be passed on to, and completed in, neocortical circuits (e.g., Treves and Rolls 1992). In practice, a more flexible dialogue of reactivation may exist between the hippocampal and neocortical structures, and possibly beyond, as evidenced by the observation of memory trace reactivation in the dentate gyrus (Shen, Kudrimoti et al. 1998), which projects exclusively to, but receives minimal projections from, CA3, and by reciprocal modulation of hippocampal and neocortical cells (Siapas and Wilson 1998; Sirota, Csicsvari et al. 2003; Battaglia and McNaughton, submitted). Memory trace reactivation has also been observed in the nucleus accumbens, which also is modulated by hippocampal sharp waves, thus the phenomenon is not limited to cortical networks (Pennartz, Geurtsen et al. 2001).

1.4.4 Persistent activity in the primate brain

Although reactivation as described above has not been studied in humans or monkeys, there is some indication that the neural representation of a stimulus can be
activated in its absence through the processes of cued recall. During the delay period between two related events, monkeys associate the events by predicting the second event, rather than simply keeping the first event in mind until the appearance of the second. The process by which a preceding stimulus elicits prediction of a subsequent stimulus is termed prospective coding, and is one example of cued recall seen in monkeys.

Typically, prospective coding is seen using some variant of the delayed paired associate (DPA) task. For example, in Colombo and Graziano (Colombo and Graziano 1994), monkeys were trained to associate one of several antecedent tones (cues) with one of several objects (matches). During the delay between the associates, objects, but not tones, interfered with performance, suggesting the monkeys were recalling the appropriate object during the delay period.

Neural activity in numerous brain structures is consistent with prospective coding, that is, cells which are selective for a given match are also active in the delay following the cue. In a haptic-visual DPA task, visually responsive cells in V4 are active during the delay following presentation of the associated haptic stimulus. In anterior inferior temporal cortex (Sakai and Miyashita 1991; Tomita, Ohbayashi et al. 1999), perirhinal cortex (Erickson and Desimone 1999), and lateral prefrontal cortex (Rainer, Rao et al. 1999), delay activity often reflects the upcoming image, and the appearance of this prospective activity roughly corresponds with the time course of associate learning (Erickson and Desimone 1999; Messinger, Squire et al. 2001). Although the “recall” seen during the delay has only been described at the single-cell level, it is likely that the cue
stimulus elicits a recall of the ensemble representation of the stimulus, given the high proportions of cells which show this effect (Rainer, Rao et al. 1999).

Learning in the paired associate task appears to be one example of top-down processing consistent with trace replay theory. Whereas visual responses to the cue are evident in area TE before they appear in perirhinal cortex, indicating bottom-up processing, 'memory retrieval' responses appeared in perirhinal cortex before it appeared in TE, suggesting top-down processing (Naya, Yoshida et al. 2001). Plasticity may occur in the perirhinal cortex as a result of paired-associate learning, indicated by clusters of cells in the perirhinal cortex showing upregulation of brain-derived neurotrophic factor (BDNF) and zif268 (Tokuyama, Okuno et al. 2000; Tokuyama, Okuno et al. 2002). Moreover, the neurons in TE do not develop associated activity for the pairs following lesions to the entorhinal and perirhinal cortex (Higuchi and Miyashita 1996). These developments suggest that, for one type of learning, the medial temporal lobe structures may provide feedback facilitating association learning in the neocortex. Although reactivation in the primate has not yet been established, cued recall may be a useful, if not necessary, tool for eliciting replay.

1.4.5 Conclusion

Cells in the rat hippocampus show signs of memory trace reactivation during sharp wave/ripple complexes in quiet wakefulness or SWS. These oscillations are correlated with synchronized activity in the neocortex, possibly providing the optimal conditions for Hebbian plasticity in the neocortex. Indeed, in the sole study of
neocortical reactivation, both hippocampal and posterior parietal cells active together during behavior were co-active afterwards, consistent with trace replay theory.

The study of memory trace reactivation is still in its infancy; fewer than a dozen articles on memory trace reactivation have been published, all within the last decade. Almost all studies of memory trace reactivation are constrained to the rat hippocampus, leaving two major gaps in the evidence for trace replay theory: (i) there are no studies of memory trace reactivation in human or non-human primates, and (ii) trace replay theory hinges on coincident activity in the distributed memory trace, yet there are no reports of memory trace reactivation occurring throughout the neocortex in a coordinated fashion. These two issues are the subject of investigation described in this dissertation.
2 METHODS

2.1 ELECTRODE DRIVE DEVELOPMENT

Due to the distributed nature of memories, neurons active in one stored pattern are shared with other stored patterns, thus as the number of recorded neurons increases, the likelihood of confusing one pattern with another pattern decreases. Consequently, the proposed experiment is dependent on the ability to record from as many cells as possible. The sparse coding in the neocortex only adds to the importance of attaining cell yields large enough to contain some cells responsive to a given stimulus pattern. To this end, 54- and 144-channel were developed in conjunction with B. L. McNaughton and implemented by K. Stengel at Neuralynx, Inc. The fundamental design common to both drives is as follows: the uninsulated back of each electrode makes contact with a 30ga stainless steel guide cannula, which is electrically isolated from the other electrodes’ cannulae (the latter design principle has been described previously, deCharms, Blake et al. 1999). Each cannula is press-fit into a plated hole on a custom-made circuit board (Neuralynx, Inc., Tucson, AZ) which, in turn, connects to a preamplifier. An electrode is lowered by gradually advancing a wire down the back of the appropriate cannula, pushing the electrode tip through an insulated barrier of silicon rubber, before penetrating the brain (Figure 2.1A).
2.1.1 Circuit board development

The 54-channel drive prototype allowed the design principles described above to be tested with minimal additional resources. Complete recording systems for the 54-channel hyperdrive were already available, including the 54-channel ‘hyperdrive’ circuit boards that connect to two 27-channel preamplifiers. In the prototype, a bundle of 30ga cannulae was made, in which each cannula was electrically isolated from the others, and each cannula had a separate wire connected to it externally, leading to one of the hyperdrive board input holes and secured with a cactus needle. This drive was tested successfully in the rodent, leading to the design of a new circuit board with traces routed from the cannulae to the preamplifier connector pins, eliminating the need for external wire connections (Figure 2.1B, C).

In the 54-channel design, the array was located in the center of the 3.5 cm circuit board, thus the center-to-center distance of the targeted brain areas could also be no less than 3.5 cm, making it difficult to sample more than two brain regions simultaneously. Because of the desire for greater sampling than two 54-channel drives could provide, the 144-channel drive was developed.
Figure 2.1 54-Channel electrode drive. A. Diagram of the electrode drive and pusher. At surgery, the electrode is contained in the guide cannula, isolated from the brain by a layer of silicon rubber (silastic). By inserting the 0.005" pusher rod down the back of the cannula, eventually the rod will contact the back of the electrode, pushing it through the silastic and into the brain. Electrical continuity is achieved by bending the uninsulated back of the electrode, so that it contacts the stainless steel cannula. The cannula, which never enters the brain, is press-fit into a hole on the circuit board. Everything except the pusher device remains chronically implanted. B. Top: Soft x-ray image of a rat implanted with a prototype 54-channel drive (courtesy Frank Houston and Brad Barber). The dark rectangle is the bundle of cannulae, with fine stainless steel electrodes emerging from the bottom. Bottom: The 54-channel drive. The cannula array is in the center of the board, with visible traces to the preamplifier connector pins. The end of the pusher device appears from the top of the picture.
The 144-channel drive consisted of a 12 x 12 square lattice of electrodes with a 650 μm inter-electrode spacing, for an electrode footprint of ≈ 8 x 8 mm. The size of the array grid was designed to maximize the number of probes in a brain region, yet provide ample inter-electrode space to minimize the brain compression between electrodes. In this configuration, two rows of cannulae lead to one 27-channel preamplifier, for a total of 6 preamplifiers per 144-channel drive. A second modification from the 54-channel design was the placement of the cannulae in the far corner of the circuit board, allowing the arrays from up to four boards to sit within a few millimeters of each other, if necessary (Figure 2.2).

Although four drives could be configured to reach a variety of brain regions in the primate brain, the minimum space occupied by the four circuit boards would be an unwieldy 7 x 11 cm. The drive design dictates that the minimum cannula length (i.e. array height) is the distance from the back of the electrode tip inside the cannula to the target brain region, reaching up to 3-5 cm. This height represents unused space in the original design, but was capitalized upon in the second generation of 144-channel drives.
Figure 2.2 Smallest configuration of 4, first-generation 144-channel electrode drives. The circuit board layout is seen in the lower left quadrant. The 12 x 12 grid containing holes for cannulae is visible in the upper right corner of the circuit board. Traces link pairs of rows in the grid to one of the six preamplifiers connectors, seen here as large paired columns of dots. By rotating and/or inverting the board, the other three configurations are possible, producing a 576-channel system 7 cm long by 11cm wide, with an 8 x 8 mm electrode footprint for each drive. The smallest region that can contain all 576 electrodes is shown by the blue rectangle, measuring 20 mm x 26 mm, the same size as one of the second-generation 144-channel drives. Presently, single first-generation drives are used only in chronic, behaving rodent preparations.

In the second version, each 144-channel drive consists of six circuit boards stacked along the length of the cannulae, with each board containing contacts for two rows of cannulae, and connecting to one preamplifier (Figure 2.3). The cannula arrays are still located at the edge of the boards, but the reduced circuit board size of 20 x 26 mm can be arranged to a minimum space of 4 x 5.3 cm, roughly one-quarter the size of the original
144-channel configuration. Four such 144-channel drives were used in the present study, one of which had a centrally located array (the PFC drive), to minimize space.

Figure 2.3 Second-generation 144-channel drive. A. Using the same principles as the original 144-channel drive, in this version each preamplifier attaches to its own board, but any one board only contains traces and plated holes for two rows of cannulae. Stacking these boards dramatically reduces the space occupied; each board is only 20 x 26 mm. B. Configuration of four drives arranged according to actual target implant coordinates. The drives are held in place by a delrin encasement (base plate shown) attached to a plastic skull model with a fast-curing resin.

2.1.2 Electrodes

In the initial 54-channel implant, 50 μm teflon-coated stainless steel electrodes were used. Although the recording properties of the electrodes were acceptable, there were two concerns about continuing with these small-diameter electrodes: first, the electrodes might not be able to travel several centimeters through the primate brain
without bending, and second, the gap between the back of the electrode and the inner diameter of the cannula could cause the pusher rod to get wedged, preventing electrode from advancing. For these reasons, larger electrodes were tested.

Four variants of stainless steel electrodes were used in an acute preparation in the rodent. The electrodes (FHC, Bowdoinham, ME) were either 75 or 100 µm in diameter, tapered to either a rounded fine or rounded medium tip of ~3 µm. Each electrode was lowered (one at a time) through the dorsal surface of the brain until the hippocampus was reached. As each electrode was lowered, the quantity and amplitude of neural waveforms was monitored. No differences were observed among the different types of electrodes. Electrodes were also tested for their ability to pass through dura mater, using a fresh macaque cadaver. Subsequent recordings of these electrodes in an acute preparation in the rodent revealed no difference in recording characteristics.

Results of the electrode tests suggested that it would be possible to record cells in the monkey using 75 µm and 100 µm electrodes, without removing the dura mater. The electrodes used in the present study were 75 µm in diameter, tapered to a 3 µm tip, made of epoxy-coated stainless steel, and gold-plated down to 0.8 - 2 MΩ impedance (FHC #UESMCHSE).
2.2 IMPLANT PROCEDURES

2.2.1 Pre-operative preparation

MRI-based selection of implant coordinates

In order to determine the stereotaxic coordinates of target areas, magnetic resonance images (MRIs) were obtained. Facilities and imaging assistance were made available courtesy Art Gmitro and the Flynn Biological MRI/MRS Program. For the procedure, the monkey was anesthetized under Ketamine (10 mg/kg) initially, and isoflurane (1-3%) for the remainder of the imaging procedure. The monkey's head was immobilized in a stereotaxic plane while 124 slices of 1mm thickness were made using T1-weighted SPGR, 4NEX (TR=26, TE=3) scans with a 25 cm FOV.

Figure 2.4 Orthoview renderings of the pre-operative MRs. Axial, sagittal, and coronal views (L-R) of the subject monkey's head in a stereotaxic plane. The pink crosshair in each image determines the current slice for each plane, and can be selected on any of the three views. The anterior-posterior, medial-lateral, and dorsal-ventral coordinates relative to intra-aural 0 are listed in yellow at the top of each image. For this figure, the crosshair is located on the midline at AP 0, at the height where the implant oval should meet the skull. The targeted range of the four arrays can be displayed along with the encasement dimensions, as seen in the coronal view, which is shown in the plane of the parietal array. The orthoview software is run in Matlab and was created by T. M. Ellmore.
The resulting MR slices were merged into a three-dimensional reconstruction, in stereotaxic coordinates, using ANSI and custom Orthoview software (developed by T. Ellmore). Example images from Orthoview are shown in Figure 2.4.

Coordinates for the arrays and the encasement were taken from these renderings. The right hemisphere arrays would be implanted over somatosensory and dorsal prefrontal cortex. The dorsal prefrontal cortex target center was AP 31.5 ML 8.5, at the rostral tip of the arcuate sulcus, sampling areas 8B and adjacent area 9L (Figure 2.5C). This region is not a common target of study, but lesion and stimulation experiments implicate this area in eye and ear orientation (Mitz and Godschalk 1989; Bon and Lucchetti 1994). The somatosensory target center was AP 8.5 ML 13, sampling areas 3a, 3b, 1, 2 5 and possibly some VIP (Figure 2.5E). These regions of somatosensory cortex should receive cutaneous input from the hand and arm, as well as more complex sensorimotor information in area 5 (see Romo Hernandez et al 2002 for recent review).

In the left hemisphere, arrays would be implanted over motor cortex and posterior parietal cortex. The motor cortex target was centered at AP 14.5 and ML -16, sampling area 4 of primary motor cortex and possibly adjacent area 3a (Figure 2.5D), regions containing information about instantaneous reaching velocity (reviewed in Schwartz and Moran 2000). Relative to intra-aural zero and the midline, the posterior parietal target area was AP 0 and ML -11 centered between the post-central dimple and the IP sulcus, and sampling areas 5, LIP, and VIP and some 7a (Figure 2.5F). Neural activity in these areas predicts the movement required for reaches to visual targets, as described by
Cavada (2001). The major afferent and efferent cortical pathways of the targeted regions are listed in Table 2.1.

Figure 2.5 Target array coordinates superimposed on a magnetic resonance rendering of the brain. A., B. Anterior and lateral views, respectively, of the reconstructed brain slices. C-F. Coronal slices of the view shown in A., reflecting the middle of the anterior-posterior extent of each array. Slices are shown from left to right in anterior to posterior order. (C. dorsal prefrontal cortex, D. motor cortex, E. somatosensory cortex, F. posterior parietal cortex). The left hemisphere of the brain appears on the right side of each image.
<table>
<thead>
<tr>
<th>Implant target</th>
<th>Afferent connections</th>
<th>Efferent projections</th>
<th>Level in hierarchy</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>8B/9L</td>
<td>PO, 30, 23, 19, 7a, 9/46d, 46, 6, 47/12, 13, 24a</td>
<td>8a, 9, 9/46</td>
<td>high</td>
<td>Petrides and Pandya 1999</td>
</tr>
<tr>
<td>3a</td>
<td>1, 2, 5, 6, Cin, SMA, Ins</td>
<td>1, 2, 6</td>
<td>lowest</td>
<td>Jones et al. 1978; Felleman and Van Essen 1991; Darian-Smith et al. 1993</td>
</tr>
<tr>
<td>3B</td>
<td>1, 2, Ins</td>
<td>1, 2, S2</td>
<td>lowest</td>
<td>Jones et al. 1978; Darian-Smith et al. 1993; Kaas 1993</td>
</tr>
<tr>
<td>1</td>
<td>3a, 2, 5, S2</td>
<td>3a, 2, 5, S2, 7b, 6</td>
<td>low</td>
<td>Jones et al. 1978; Felleman and Van Essen 1991</td>
</tr>
<tr>
<td>2</td>
<td>5, S2, 4,</td>
<td>5, S2, 7b, 4, 6, SMA,</td>
<td>low</td>
<td>Jones et al. 1978; Felleman and Van Essen 1991</td>
</tr>
<tr>
<td>5</td>
<td>7b, 6, Ri, S2, 7b, 6, 4, SMA</td>
<td>mid</td>
<td>Jones et al. 1978; Felleman and Van Essen 1991</td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>V2, MT, MST, PO, FEF</td>
<td>TEO, MST, PO, 7a, PMv, FEF</td>
<td>mid-high</td>
<td>Matelli et al. 1986; Cavada and Goldman-Rakic 1989; Felleman and Van Essen 1991</td>
</tr>
<tr>
<td>LIP</td>
<td>MST, 7a, FEF</td>
<td>V3, TEO, MST, PO, TF, 7a, FEF, 46, PMv</td>
<td>mid-high</td>
<td>Matelli et al. 1986; Cavada and Goldman-Rakic 1989; Felleman and Van Essen 1991</td>
</tr>
<tr>
<td>7a</td>
<td>TEa, MST, PO, STP, TF, TH, LIP, VIP, MIP, 7b, FEF, 46</td>
<td>TEa, MST, PO, STP, TF, TH, LIP, FEF, 46, PMv</td>
<td>high</td>
<td>Matelli et al. 1986; Cavada and Goldman-Rakic 1989; Felleman and Van Essen 1991</td>
</tr>
<tr>
<td>4</td>
<td>2, 5, 6, SMA, Ins, Cin</td>
<td>2, 6, SMA</td>
<td>high</td>
<td>Jones et al., 1978; Felleman and Van Essen 1991; Darian-Smith et al. 1993</td>
</tr>
</tbody>
</table>

Table 2.1 Cortical connectivity of targeted brain regions. Implant targets, listed on the far left, are based on the pre-operative MRs; primary implant targets are in bold print. The major cortical inputs and outputs of each target region are listed, as is the putative level in the cortical hierarchy, where primary sensory areas receiving thalamic input constitute the lowest level, and regions receiving increasingly convergent, 'associative', connections from other neocortical regions would be at higher levels. Abbreviations:

Cin = cingulate cortex  
FEF = frontal eye field  
Ins = insular cortex  
LIP = lateral intraparietal area  
MIP = medial intraparietal area  
MST = medial superior temporal area  
MT = medial temporal area  
PMv = ventral premotor area (6a)  
PO = parieto-occipital area  
Ri = retroinsular area  
S2 = secondary somatosensory cortex  
SMA = supplementary motor area  
TEa = anterior temporal area  
TF = medial temporal lobe area  
TH = medial parahippocampal gyrus area  
V2 = second visual area  
V3 = third visual area  
VIP = ventral intraparietal area
Encasement design and fabrication

The chronic implantation of four 144-channel electrode drives required the design of the smallest and lightest possible encasement that would still protect the drives. To save space, the encasement also served as part of the head restraint system, which was necessary to support the weight of the cables and minimize movement artifacts during recordings. The encasement was fabricated based on the target brain coordinates determined pre-operatively with MRIs, therefore the encasement also served as a template for aligning the arrays vertically and to the appropriate AP/ML coordinates during surgery.

The encasement is illustrated in Figure 2.6. The two main pieces of the encasement, the base plate and the cover, were fabricated out of delrin plastic (SDIA, Tucson, AZ). The bottom of the base plate contains an oval lip roughly matching the skull circumference at the level where the implant will rest. The inside of the oval was milled out, leaving spaces for the implant stabilizing screws, the bone screws and the shaft of the cannula arrays. The remaining spaces were filled in with a fast-curing resin at the time of surgery. The top of the base plate is a platform accommodating all four drives with additional room for the cover. For each 144-channel drive there is one ½" diameter circular hole in the base plate to accommodate the shaft of each cannula array, and two threaded holes for bolts used to secure the drives to the base plate. Also within the perimeter of the oval are holes for the implant stabilizing screws, and small alignment holes marking AP/ML zero and bregma with respect to the center of the cannula grids. At
each corner of the top of the base plate are threaded holes used to secure the base plate to the head restraint plate during recording sessions, and to attach the cover when recording sessions are not taking place.

The lateral edges of the base plate are beveled, matching the angle of the inside of the head restraint plate. This allows the head restraint plate to be slid in place from behind, and subsequently secured to the chair. See Section 2.3 for details on the head restraint system.

**Figure 2.6 Drive encasement design.** The components of the encasement include milled pieces of delrin plastic (blue) metal fasteners (silver), and the drives (green). Both upright and inverted types of bone screws were used to anchor the encasement to the skull. The resin (not shown) is poured into the cavity created when the base plate oval reaches the skull. Both bone and implant stabilizing screws are partly protruding into this cavity, thus the resin bonds the skull and the base plate via the respective screws. The arrays are aligned by the array bolts to the correct location on the base plate and secured. The cover contains a rubber gasket to seal the encasement once the cover screws are in place. The beveled edges on the base plate guide the positioning of the head restraint system.
2.2.2 Implant surgery

A male 5.5 kg rhesus macaque (Macaca mulatta) was implanted aseptically according to NIH guidelines. The monkey was initially anesthetized with ketamine (10 mg/kg) combined with atropine (0.5cc) followed by administration of isoflurane (1-3%). After the monkey underwent placement in the stereotaxic apparatus and a sterile surgical field was prepared, an oval was traced out on the scalp using a template slightly smaller than the encasement oval. After making a superficial incision around the oval, a combination of lidocaine (15 mg/mL) and epinephrine (5μg/mL) was injected subcutaneously. After making a complete incision, the oval “cap” was removed and the remaining bone surface cleaned and dried.

The base plate, attached to two stereotaxic carriers, was positioned so that its reference points were aligned with AP 0 and bregma on the skull. Additional presurgically estimated coordinates were compared to the observed coordinates at various stages of surgery to ensure correct base plate positioning. Beveled delrin cylinders were dropped down the base plate array holes, marking the center and circumference of the array craniotomies. The circumference of the base plate oval was marked on the skull, and bone screw positions were marked in the remaining spaces on the skull. Array and bone screw craniotomies were made. After ground screws were inserted into the skull, two additional types of bone screws were used. Upright screws were threaded into the skull (Synthes, Inc., Paoli, PA), and slotted stainless steel screws (Crist Instrument Co.,
Inc., Hagerstown, MD) were inverted, such that the flat head of the screw sat between the skull and the dura mater, secured with nuts and a washer at the surface of the skull. EMG wires were inserted into the appropriate muscle bellies, tunneled subcutaneously into the region of the implant, and up through an unoccupied base plate screw holes.

The craniotomy space was sealed with bone wax, and the lubricated cylinders were lowered over each craniotomy. Dental acrylic was placed around the edge of the craniotomies and cylinders, sealing them off from the remaining space under the base plate. The resin containment band was added to the bottom of the base plate oval, and lowered until it rested on the skull. Any gaps between the band and the skull were filled in with a dental impression material (Vinylpolysiloxane, light body, regular set, Henry Schein, Inc., Melville, NY). The implant stabilizing screws and the array screws were added to the base plate. The bottom of the implant was formed by injecting a fast-curing resin into any unoccupied screw opening in the base plate (Ace Resin and Hobbies, Tucson, AZ). This resin was selected because of its curing speed and its low curing temperature, even in large quantities.

Once the resin cured, the resin containment band and delrin cylinders were removed, and uncured silastic (Dow Corning 3140 RTV coating, Midland, MI) was added to the newly-formed tunnels over the craniotomies. The drives were slowly lowered onto the drive bolts on the base plate, and secured in place. Because the tips of the arrays had a layer of cured silastic, there would be a continuous seal of silastic between the bone wax over the dura, the cannulae, and the environment. The base plate was removed from the stereotaxic carriers, and the skin was sutured to meet the oval, as
much as possible. Using fixed-length lowering rods, some of the electrodes were advanced into the silastic at the time of surgery, though none were advanced far enough to penetrate the dura (i.e. the impedance between electrode and ground screws remained high). Finally, the cover, with greased gasket, was screwed in place and wrapped with bandaging tape for cushioning.

A prototype version of the encasement was implanted on a fresh monkey cadaver (in conjunction with the electrode testing described in Section 2.1), to verify the surgical protocol, and to test the strength of the encasement as a head restraint system. Both aspects were successful: the encasement was implanted according to the protocol, and was strong enough to survive repeated attempts at removal through impact and torque tests.

**Post-operative recovery**

After surgery, the monkey received the analgesic buprenorphine initially, and later motrin. The day after surgery, the monkey was fully alert and mobile, consuming his full ration of food. In the first week, the monkey was anesthetized with ketamine for wound cleaning and electrode advancement, but was not involved in behavioral testing. Within two weeks, the monkey was habituated to head restraint and willing to perform reaching tasks for food reward. The first recording session meeting cell count requirements took place 13 days post-operatively.
2.3 ORGANIZATION OF DAILY SESSIONS

Daily sessions began by bringing the chaired monkey into the behavioral testing room and applying head restraint. This was followed by a period of assessing cell yields and lowering electrodes to further increase cell yields. Provided sufficient progress was made, a recording session followed.

2.3.1 Head restraint system

Head immobilization was achieved by fastening the implant to a plate that, in turn, was secured to the primate chair (see Figure 2.7). The bevel on the base plate conformed to the head restraint plate, which acted as a "rail" to guide the base plate into position. Two skewers were then inserted through both the chair posts and the head restraint plate, supporting the weight of the head restraint plate and encasement. The cover of the implant was removed, and small screws locked the base plate to the head restraint plate through the screwholes that had been used to attach the cover. The head restraint was left in place for the remainder of the session.

After securing the encasement, 24 headstage preamplifiers were attached to the drives (6 per drive), and ultimately connected to the acquisition system. A preliminary cell count on each of the four systems indicated the degree of electrode stability and the drives that had the most cells. Typically, there were several hours of electrode advancing. If a sufficient number of new cells were obtained within 4-6 hours of advancing, a recording session would follow. If there were insufficient new cells, the remainder of the daily session would be spent advancing more electrodes.
Figure 2.7  A. Top and front-view illustrations of the head restraint-implant interface. A. Head restraint plate handle. B. Preamplifier cables. C. Skewers attaching the head restraint plate to the primate chair. D. The base plate. E. Stack of 6 preamplifiers. F. 144-channel drive. G. Primate chair posts to which the skewers attach (not drawn in top diagram). B. Photograph of head restraint system. Letters are the same as in A. Two drives are connected to preamplifiers, and the base plate is outlined in red. The removable brass bars on the head restraint plate are used to stabilize the preamplifiers and cables to minimize movement artifact while recording.
2.3.2 Electrode advancing

Lowering the electrodes requires head immobilization, therefore to prevent strain on the implant in the first week after surgery, the monkey was lightly anesthetized with up to 0.3cc Ketamine during the first advancing sessions.

Pusher apparatus

Two devices were used to lower the electrodes. The "one-shot" pusher was a 0.005" piano wire protruding from a 30ga cannula by a precise length, and crimped in place. A 23ga cannula was added to the back for stability and ease in handling. The wire was placed down the back of a guide cannula, pushing the back of the electrode until the pusher cannula contacted the guide cannula. The one-shot pusher is easy to make and use, and can advance a large number of electrodes in a short period of time, however, there is a limit to the minimum increment lowered and the speed of movement.

Gradual lowering in increments as small as one micron was accomplished using a micromanipulator attached to an assembly that pushes a 0.005" piano wire. The manipulator can be attached to the chair, and a flexible shaft allows precise placement of the tip onto a guide cannula. The length of the pusher was monitored and automatically updated via a serial link to a PC.

The one-shot pushers were used in the first two post-operative weeks, to lower electrodes quickly through the silastic barrier, while monitoring the electrode impedance to the ground screw. When the impedance with reference to the ground screw drops, it is assumed that the electrode tip passed through the silastic and into the dura. The length of
the pusher at the point of impedance drop provides an estimated brain surface depth, or zero point for that electrode, used in subsequent depth calculations. Once the pusher lengths at the zero points are provided, the pusher software can be used to calculate ongoing brain depth estimates during advancing sessions.

**Pusher software**

The pusher program (run in MATLAB, developed by F. Battaglia) queried the manipulator length and subtracted the length recorded at the brain zero point for that electrode, to estimate the depth of the electrodes in the brain. The previous distance pushed was also recorded, providing a “delta” estimate of the distance pushed since last entering that cannula. Other parameters measured were the time when a given electrode was lowered, and any other parameters definable on a scale of 0-10, could also be entered manually (e.g. cell count or relative spike amplitude). The program could highlight electrodes to lower based on any of these parameters. For example, it was possible to select only those electrodes that were less than 2mm into the brain, had not been moved in 24 hours, and currently had no cells (Figure 2.8). Although the pusher wire had to be inserted into each cannula manually, the set of electrodes to be lowered, and the manipulator depth at which the electrode would be contacted, was provided by the program, expediting the process of lowering electrodes substantially. In addition to the pusher software, an oscilloscope, audio monitor, and a PC running the Cheetah software were used to assist in advancing sessions.
Typically, the push wire engaged 20-140 cannulae in one hour, with the number varying considerably depending on the strategy chosen for lowering, and on the amount of time devoted to listening to and characterizing neural responses.

Figure 2.8 Pusher software display. The main window is in the upper right corner. Based on the previous lowering of the selected electrode (L1), the push wire has not advanced the electrode any further since its most recent selection (Delta this round), but it was previously lowered 681 μm beyond the estimated surface of the brain. The current distance between the pusher tip and the back of the electrode is 3361 μm (Micrometer gap). As the micromanipulator is moved, the depth record updates, and is saved in a text file (Depth History). The window in the upper left can display any of several color plots, in this case, the depth of each electrode. Automated ordering of the electrodes to be pushed is made possible through the window in the lower right. This can be set according to depth, cell quality, or spatial parameters, and prevents the user from having to manually select the next electrode, allowing uninterrupted operation of the manipulator. When selected, the window to the left will indicate which electrodes are queued for lowering, and which electrode is currently selected.
2.3.3 Recording procedures

The data acquisition system is based on a 144-channel module. Each of the 144 signals per drive was passed through a 27-channel unity-gain headstage preamplifier, for a total of 6 preamplifiers per drive, or 24 in total (Neuralynx, Inc., Tucson, AZ). Signal from each group of 6 preamplifiers was fed via tether cables to one of four 144-channel patch panels. At the patch panel, up to two of the 144 channels could be selected as references for differential single unit recording, and up to 16 of the 144 channels could be selected and filtered separately as continuous recording channels (CSCs). For logistical reasons, only four electrode channels were split off into the CSCs for each patch panel.

From the patch panel, the 144 single unit signals were band pass filtered between 600Hz and 6kHz using 18, 8-channel external amplifiers. The digitally programmable gain and filter settings on the amplifiers were controlled by one of four 160-channel ‘Cheetah’ data acquisition machines, which thresholded and digitized the signals at 30kHz (Neuralynx, Inc., Tucson, AZ). Signals split to CSCs were sent to one of two, 8-channel external amplifiers, filtered between 1Hz and 3kHz, and sampled at 2kHz. Four PCs, one per 144-channel system, were used to record and display the data being collected (Figure 2.9). In addition to the neural CSC inputs, the audio output of the alley maze task, described below, occupied one continuously sampled channel. Manual and automated event codes were also timestamped with the Cheetah acquisition system.
Figure 2.9 Cheetah display during data acquisition. Each rectangle plots the supra-threshold waveforms from one electrode (time versus differential voltage amplitude). The layout of the plots corresponds to the position of the respective electrodes in the 12 x 12 electrode array. Four such displays were used in the present study, one for each array.

Synchronization was achieved offline by aligning the timestamps of the four systems in one of two ways. Headstage power was toggled at the beginning of each epoch, causing transients simultaneously on all four systems, and at the same time causing the headstage diodes to flicker, for video synchronization. The precise time of the transients was identified on a continuously sampled channel of each system, for each epoch, providing the offset and expansion factor necessary to align the timestamps across systems. In addition, later recording sessions used a four-channel counter that
incremented with the clock pulses of each system, allowing offline alignment of all systems' timestamps (P. Lipa, MATLAB program). As an additional measure, sine waves from a stimulus generator were fed into two CSC channels on all four systems, thus the peaks across systems could be precisely aligned.

2.3.4 Behavioral procedures

Prior to implantation, the monkey had been trained to touch objects and touchscreen cues for treats, and was familiar with rest sessions preceding and following a task. The monkey resumed behavioral sessions after habituating to head restraint, almost two weeks following implant surgery. During these sessions, the chaired, head restrained monkey was first placed in a small, darkened chamber for at least 30 minutes. The room outside the chamber was dark, quiet and unoccupied, and low-level background noise was played in the chamber.

During task epochs, the monkey was moved outside the chamber and a portal was opened allowing reaching movement outside the chair. If the touchscreen task was involved, a juice spout and lever were attached to the chair. A description of the tasks involved is described below. After the monkey stopped performing the task, or after 40 min., whichever occurred first, the external devices were removed from the chair, if applicable, the portal was closed, and the monkey was returned to the chamber for a second rest session in the chamber.
Behavior during rest sessions was monitored with an infrared security camera, and typically consisted of long periods of immobility during which time the eyes were closed, and the jaw was slack, interrupted by brief periods of wakefulness.

Tasks were designed to elicit activity in the four target brain regions by providing repeated somatosensory, motor, and visual targets in multiple locations for juice or fruit reward. Tasks included: alternating left and right reaches to two serially-presented objects, (Sessions 1,2), bimanual foraging for a hidden treat in alternating left and right locations (Session 3), reaches through a grating for treats in one of four locations (up, down, left, right; Sessions 4-6), and navigation through a virtual environment (Sessions 7-9) requiring lever pulls and reaches to a touchscreen (Figure 2.10). The virtual environment was created in Java by N. Kaul and C. Anderson. The program received the lever and touchscreen activation as inputs, and sent a pulse that triggered a solenoid, delivering juice for a specifiable duration. The program also sent event codes to the Cheetah acquisition system, indicating virtual location, cue location, touchscreen activation, timing of juice delivery, and activation of the soundfiles elicited during hallway traversal.
Figure 2.10 Virtual Alley Maze. A. The beginning of a lap. A lever pull advances position in a fixed trajectory towards the end wall, with sound cues elicited as some objects pass out of view of the screen. When the end of the hall is reached, the lever is disabled, and a cue appears on the end wall. B. One end of the alley. When the cue (blue square) on the touchscreen is selected, it disappears, juice reward is delivered, and the navigation lever is enabled. The return journey begins when the lever is pulled, and the same process repeats at the opposite end wall.
2.4 CELL ISOLATION AND LOCALIZATION

Cells were isolated offline using in-house semi-automated and manual clustering software (semi-automated: BBClust, P. Lipa; manual: MClust, A.D. Redish) based on parameters such as waveform amplitude, energy (related to the area under the waveform), and principle components of energy (Figure 2.11). Cells with signal amplitude fluctuating below threshold during any session epoch, or with fewer than 100 spikes, were excluded from analysis. For all spike train correlation analyses, only pairs of cells from different electrodes were used, to reduce possible contamination resulting from inadequate cell isolation.

Figure 2.11 BBClust output from one electrode. A. The left figure shows the BBClust nearest-neighbor decision tree, with two peaks aside from the main one, each indicating a separate cluster. When plotted according to several different waveshape parameters (right), the identified clusters emerge from the noise activity. B. The mean waveshape (top) and log inter-spike interval histogram in ms (bottom) are plotted for the two isolated cells. The difference in waveshape and ISI distribution, as well as the absence of ISI points to the far left, suggest these are two, well-isolated cells.
The location of the electrodes was determined by transforming the pre-operative MRIs into the same coordinate frame as the post-operative CTs, based on aligning structural features such as bone and sinus points in all three dimensions. This was accomplished after reslicing the MRIs to a 0.5 mm voxel size to match the CTs (Figure 2.12).

**Figure 2.12 MR and CT co-registration.** A. Axial CT of the implanted monkey. Electrode cross-sections are visible as white dots inside the indicated regions (dashed white squares). At this dorsal level, the medial electrodes are in brain, whereas the lateral electrodes are still above or at the level of the skull, visible as a white, interrupted oval. Skull screws appear as bright spots with a radial artifact. B. Top panel. Coronal slices of CT renderings, registered to stereotaxic coordinates. In the left, anterior image, electrodes from the PFC array are visible. The right image is the one plane in which both motor and somatosensory electrodes are visible. After the recording sessions but prior to CT scanning, many electrodes were lowered 5-30mm. Bottom panel. Coronal slices of pre-operative MR renderings, co-registered with the CTs.
3 RESULTS

3.1 GENERAL

Recording sessions were interspersed throughout the three-month period of implantation, occurring roughly every third day. The remaining daily sessions were devoted entirely to electrode advancing, or to training in the case of the two weeks preceding alley maze recording sessions.

Task performance

For all sessions included in the following analyses, reaching sequences were repeated between 13 and 56 times, over 15-40 minutes. Two sessions were excluded on the basis of inadequate behavior.

Table 3.1 Summary of cell counts and behavioral task for eligible sessions

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Cluster count</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4/17</td>
<td>28</td>
<td>Alt. L/R reach to objects</td>
</tr>
<tr>
<td>2</td>
<td>4/19</td>
<td>38</td>
<td>Sequential limb grooming: l, r arm; l, r leg</td>
</tr>
<tr>
<td>3</td>
<td>4/25</td>
<td>90</td>
<td>Alt. L/R reach to objects</td>
</tr>
<tr>
<td>4</td>
<td>5/1</td>
<td>76</td>
<td>Bimanual alt. L/R foraging for hidden treat</td>
</tr>
<tr>
<td>5</td>
<td>5/9</td>
<td>67</td>
<td>Sequential reach to objects: L/C/R</td>
</tr>
<tr>
<td>6</td>
<td>5/10</td>
<td>97</td>
<td>Sequential reach to objects: L/C/R</td>
</tr>
<tr>
<td>7</td>
<td>5/11</td>
<td>98</td>
<td>Sequential reach U/D/L/R through grating</td>
</tr>
<tr>
<td>8</td>
<td>5/15</td>
<td>63</td>
<td>Sequential reach U/D/L/R through grating</td>
</tr>
<tr>
<td>9</td>
<td>5/16</td>
<td>83</td>
<td>Sequential reach U/D/L/R through grating</td>
</tr>
<tr>
<td>10</td>
<td>6/25</td>
<td>61</td>
<td>Navigation through virtual environment</td>
</tr>
<tr>
<td>11</td>
<td>6/26</td>
<td>64</td>
<td>Navigation through virtual environment</td>
</tr>
<tr>
<td>12</td>
<td>6/27</td>
<td>99</td>
<td>Navigation through virtual environment</td>
</tr>
<tr>
<td>13</td>
<td>6/28</td>
<td>56</td>
<td>Navigation through virtual environment</td>
</tr>
</tbody>
</table>
Rest behaviors

Although there was no EMG or EOG signal to measure vigilance state, the monkey's face was monitored with an infrared camera and videotaped during rest epochs, providing an indirect measure of sleep and waking states. In a 30-minute rest epoch, the monkey typically closed his eyes within 1-2 minutes of entering the chamber, and spent the majority of the time with eyes closed. In general, each epoch included at least one period when eyes remained shut for 5-15 minutes without interruption. During these bouts facial muscle tone appeared to diminish, giving the monkey a 'slack-jawed' appearance. When EEG was monitored, clear behavioral correlates were seen: relatively flat, oscillatory and large-amplitude activity corresponded to eyes open, 'drowsy' closing eyes, and continued closure. Because only the last 10 minutes of Rest 1 and the first 10 minutes of Rest 2 were used in reactivation analysis, the time spent with eyes closed was compared during these restricted blocks of time (Table 3.2).

<table>
<thead>
<tr>
<th></th>
<th>Mean closed (sec)</th>
<th>% closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest 1</td>
<td>48 ± 25</td>
<td>75 ± 17</td>
</tr>
<tr>
<td>Rest 2</td>
<td>40 ± 17</td>
<td>63 ± 16</td>
</tr>
<tr>
<td>P value</td>
<td>0.33</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 3.2 Time spent with eyes closed during 10-minute rest epoch blocks. Means ± SEMs from eight sessions are listed. Mean closed is the average uninterrupted period of closed eyes, i.e., mean ‘bout’ duration. Percent closed reflects the proportion of time spent with eyes shut. Both mean closed and % closed measures are calculated from the 10-minute blocks that were used for reactivation analyses. Thus, the monkey typically spent around 7 of the 10 minutes with his eyes closed, and once closed, eyes remained shut for an average of over 40 seconds (though this duration varied greatly). Sleep state as indicated by mean duration and proportion of time spent with closed eyes was similar across rest epochs (paired t-tests, p > 0.05).
3.1.1 Multiple single unit activity

Of all the sessions in which neural data was acquired, only sessions estimated to contain greater than 20 cells underwent offline cell isolation procedures.

In previous studies of reactivation, only pyramidal cells, not interneurons, were included, based on their sparse, location-specific firing characteristics. Neocortical cells have also been functionally classified, therefore sorting based on these classes could be essential to detecting memory trace reactivation.

Classification

The action potentials belonging to one neuron can be isolated by the clustering of that neuron's spikes in plots projecting various waveform parameters. The same technique used for cell isolation was applied to the physiological classification of cells. It was expected, based on previous descriptions of classification, that the spike size and width, symmetry of peak and valley, firing rate and burstiness of a cell would be relevant parameters to use in differentiating cell types (Connors, Gutnick et al. 1982; McCormick, Connors et al. 1985). These features were projected in two-dimensional scatterplots, similar to the cell isolation plots. Burstiness and other firing characteristics were measured as the principal components of the log inter-spike interval, which has a distinct early peak for bursty cells.
The result of the classification procedure was the isolation of poor, noisy waveforms, but no classification among good waveforms. Each cell's average waveform and spiking activity was calculated from the entire recording session, since this procedure was intended as a filter prior to reactivation analysis. This may be one reason that no
clear divisions could be used to classify cells. It is known that cells can change classes with transitions in vigilance state, thus some cell averages over waking and sleep may have "smeared" any existing clusters. Alternatively, the electrode shape and impedance, the areas of cortex recorded, and the lowering procedures may have biased the sampling of towards one class of cell. Whatever the cause, no classes of cells emerged from the plotting procedure. The benefit of this procedure was the clustering of small, poor-quality waveforms, thought to be noise or multiple-unit signal rather than isolated neural signal. After removing the poor-quality units, two additional sessions were excluded based on insufficient cell counts.

**Neural activity by session, epoch, and array**

A total of 800 cells were isolated and used in further analysis. The motor cortex (M), somatosensory cortex (SS), and dorsal prefrontal cortex (PFC) arrays yielded 253, 243, and 284 cells, respectively. In the parietal array, only a small fraction of the 144 electrodes penetrated the dura, thus this array yielded only 20 cells. Simultaneous recordings in these sessions varied from 28-99 cells, and typically exceeded 50 cells that were stable for the duration of a session (Figure 3.2). Relatively good overnight stability was suggested by the fact that waveforms from previous days appeared to be maintained on about 1/4 – 1/3 of electrodes.
Figure 3.2 Average waveforms of 99 simultaneously recorded cells from Session 8. Colors indicate cells recorded from parietal, motor, somatosensory, and dorsal prefrontal arrays, respectively.

There was no significant difference in the mean firing rates of cells by brain region or by behavioral epoch, and no significant interaction (two-way ANOVA), although firing rates tended to be slightly higher during the task (Figure 3.3).
Figure 3.3 Firing rate histograms for cells in each array during each epoch. The firing rate is given in Hz, and counts are expressed as a proportion of the total cells for that array in each epoch, using cells from all 9 sessions. The vast majority of cells fire at less than 1 Hz, regardless of brain region or behavioral epoch. PP, posterior parietal area (N = 20); M, motor cortex (N = 253); SS, somatosensory area (N = 243); PFC, prefrontal cortex (N = 284).
3.1.2 Relation between neural activity and behavior

Functional characterization

Due to the large cell yields and limited recording time, functional characterization was not conducted systematically each day; however, the few cells that were characterized from the prefrontal, parietal, and somatosensory arrays had responses that were consistent with the estimated location of their respective electrodes. For example, one fairly superficial, anterior probe from the parietal array recorded a cell with a vigorous direction-sensitive response to light touch of the right (contralateral) forearm. The somatosensory array recorded a cell responding to light touch on the left (contralateral) ulnar forearm and another cell responding to touch on the dorsal surface of the hand and middle two fingers. Another cell responded to both eccentric movement of the left arm and touch on top of hand or ventral wrist. The prefrontal cortex had one cell that appeared to respond as the monkey visually tracked large and/or salient objects passing proximally to his upper left side (e.g. when a foot was raised and passed to the left of his head, or when B.L.M., but not K.L.H., walked past the monkey's left side).

Task-related activity

The activity of cells during the task indicated some degree of selectivity to aspects of the task. In some sessions, groups of simultaneously recorded cells showed similar periodic activity corresponding to repetitions of the task. Peri-event time histograms (PETHs) revealed cells in all four brain areas that exhibited firing modulation to task events (Figure 3.4).
Figure 3.4 Task-related neural activity. A. Cell by time firing rate matrix from somatosensory cortex within the first five minutes of one task. The task repeated approximately every 40 sec, matching the response periodicity of several cells. Firing rates are truncated at 10 Hz for illustrative purposes. B. Peri-event time histograms of neural firing related to 52 juice reward events from one alley maze session. Juice delivery was preceded by a successful touch of the touchscreen cue. After a correct touch, the monkey retracted his arm and consumed the juice before resuming lever pulling. Task-related activity was present in all four brain regions, with a variety of response types, both excitatory and inhibitory. Approximately 20% (41/219) of the cells examined had significant PETH modulation around juice delivery or reach events, indicating that neurons were in some way responsive to elements of the task. Reprinted with permission from Hoffman & McNaughton, Coordinated Reactivation of Distributed Memory Traces in Primate Neocortex, Science 297, 2070 (2002). Copyright 2002 American Association for the Advancement of Science.
3.2 MEMORY TRACE REACTIVATION

If reactivation occurs, cells that were active together during the task should tend to be active together afterwards, and cells active at different times during the task should not be co-active afterwards. The reemergence of neural activity patterns can be quantified by computing how much of the variance in the distribution of spike train correlations during Rest 2 can be explained statistically by the pattern of correlations during Task after factoring out the distribution of correlations already present in the baseline period (i.e., Rest 1).

 Reactivation levels have been shown to decay with each subsequent block of 10 minutes of rest following the task (Kudrimoti, Barnes et al. 1999), therefore the main analysis focused on the 10 minutes of rest immediately before and after the task. Pairs of cells from the same electrode were ineligible, to prevent potential contamination due to poor cell isolation. For each epoch of each session, spike trains of all eligible pairs of cells were binned into $T$ intervals of 300ms, producing sequences of spike counts $f_i[t]$. Similar results were obtained for sessions tested at bin sizes of 50, 80, 100, or 300ms.

The normalized correlation $C$ between each pair $(ij)$ of spike trains was computed using the equation:

$$ C_{ij} = \frac{1}{T} \sum_{t=1}^{T} f_i[t] \cdot f_j[t] - \mu_i \cdot \mu_j }{ \sigma_i \cdot \sigma_j }, \text{ derived from Pearson's correlation coefficient}$$

(Perkel, Gerstein et al. 1967) where $\mu$ is the mean and $\sigma$ is the standard deviation of $f[t]$.

Thus the same population of cells produces three sets of firing-rate correlations, one set for Rest 1, one set for Task, and one set for Rest 2. According to the hypothesis
in question, the Rest 2 correlations should be more similar to the cell-pair matched correlations of Task than those from Rest 1. This hypothesis is a prediction of the theory that the set of ensemble activity patterns during Rest 2 resemble those of Task because of some process of memory reactivation. The set of patterns in Rest 1 is predicted to bear less resemblance to the set of patterns in Task, because many of the patterns have not yet occurred, at least not recently. Rest 1 is therefore a control period. To test the hypothesis, the explained variance (EV) was calculated for correlations from Rest 2 and Task, given the correlations of Rest 1 (Qin, McNaughton et al. 1997; McNaughton 1998; Kudrimoti, Barnes et al. 1999).

The EV was calculated based on the square of the partial correlation coefficients:

\[ EV = r^2_{T,R2|R1} = \left( \frac{r_{T,R2} - r_{T,R1}r_{R2,R1}}{\sqrt{(1-r^2_{T,R1})(1-r^2_{R2,R1})}} \right)^2, \]

where \( T, R1 \) and \( R2 \) represent Task, Rest 1, and Rest 2 epochs. The EV (Task, Rest 2 | Rest 1) reflects only reactivation seen immediately after a task; any Rest 1 replay of task activity (from a previous session) would be eliminated. This experimental EV reflects how well the correlations in Rest 2 predict the correlations in Task after factoring out any explanation obtained from the correlations in Rest 1.

To better estimate the changes in Rest 2 correlation structure attributable to reactivation of task-related activity, a "control" EV is calculated between Rest 1 and Task, factoring out Rest 2. The control EV provides a baseline measure of changes in correlation structure over time that cannot reflect reactivation of the representation of that day's task (since a memory cannot precede its formative occurrence). If the experimental
EV (Rest2, Task |Rest1) exceeds the control EV (Rest 1, Task |Rest 2), memory trace reactivation is said to occur. This is a more stringent criterion for reactivation than simply testing whether the experimental EV exceeds zero, since the difference of EVs will always be smaller than the experimental EV alone.

In summary, memory trace reactivation is measured as the difference between the experimental and control EVs, which are based on the observations of cell pair spike train correlations. The null hypothesis is that there is no difference between the two EVs because the Rest 2 correlations do not explain more of the variance in Task correlations than do those of Rest 1.

The bootstrapping method of resampling was used to estimate the true distribution of the test statistic (the difference in EVs) by generating many random samples from the observed quantities (cell pair firing rate correlations). Each random sample must be of the same size as the original, therefore the sampling is done with replacement, producing repetitions of some observations and omissions of others in each sample. After bootstrapping with 500 repetitions, the 95% confidence limits of the EV difference were computed.

The resampling needs to be applied at the level of cell-pair correlations rather than to the cells themselves. Otherwise, sampling with replacement would select either all or none of a given cell’s correlations, which are the observations from which the statistic is computed. Thus, the sample would no longer be random, containing complete sets of cell-pairs belonging to the selected cells. Moreover, if the cells themselves are sampled with replacement, many of the resulting cell-pair correlations will be autocorrelations,
since many cells will be represented multiple times. These correlations will have a value of 1.0, distorting the distribution. Removal of autocorrelations would produce samples of differing size, in violation of the bootstrapping design; therefore, bootstrapping of cell-pair correlations is the appropriate method for estimating the EV distribution.

3.2.1 Overall memory trace reactivation

Pooling the correlations from all nine sessions by epoch, a total of 21,288 cell pairs were eligible for correlation analysis, producing an overall explained variance significantly greater than the epoch-swapped control levels (p< 0.05). Substantial explained variance was apparent across most sessions (Figure 3.5B), but not across all brain regions.

Figure 3.5 Explained variance measure of memory trace reactivation overall and across sessions. A. Total explained variance over all pairs of cells (Rest 2, Task | Rest 1), and the epoch-swapped Control explained variance (Rest1, Task | Rest 2) pooled across sessions and arrays. Error bars show the SEM. B. Explained variance of each of the nine sessions.
3.2.2 Stability of cells and ensemble states during rest

Waking levels of body temperature, neuromodulatory activity, or metabolism may change only gradually, thus their influence on neural activity during wakefulness also may apply to periods of sleep immediately following wakefulness. This suggests the activity of neurons occurring in early but not late rest may bear a greater similarity to waking activity due to some general property of the waking state and not the task, per se. By analyzing only late Rest 1 and early Rest 2 periods, one might observe a significant Task-Rest 2 relationship due to these slowly-changing factors that are not due to memory trace reactivation.

For clarity, it is not sufficient for a waking factor to show its effects through alteration of the firing rates of cells, as the reactivation measure is independent of firing rate. Moreover, the pattern of activity cannot simply persist uninterrupted from wakefulness, or there would be no variation in activity from which to calculate firing rate correlations. If, however, there is an increased tendency for certain brain states to re-emerge as a result of the residual ‘waking’ factors, a possible confound exists. This alternative explanation can be tested by calculating the explained variance results using the early part of Rest 1, and comparing this to the standard method of calculating the EV using the late part of Rest 1. The hypothesis would predict that the beginning of rest epochs tend to evoke activity patterns more similar to waking than those seen late in rest. Accordingly, the Early Rest 1 cell-pair correlations should be more similar to those of Task than Late Rest 1 correlations are to those of Task. The control EV, which reflects the relationship between any Rest 1 epoch and Task, factoring out any Rest 2
relationship, should therefore be higher when calculated for Early Rest 1 than for Late Rest 1. More importantly, if the experimental EV was strictly the result of some waking factor, and not memory trace reactivation, then the experimental EV calculation using early Rest 1 should show no significant explained variance.

The opposite pattern of results would be expected if there is electrode drift over time: Late Rest 1 cell-pair correlations should be more similar to Task than those of Early Rest 1. Other measures suggest that electrode drift was not an issue in these recording sessions, but this analysis serves as an additional check.

![Bar chart](image)

**Figure 3.6 Explained variance using different 10-minute blocks of Rest 1 data.** The standard measure (left) includes the last 10 minutes of Rest 1 and the first 10 minutes of Rest 2. Both experimental (dark) and control (light) EVs are shown. The right plot shows the EV when the first 10 minutes of both Rest epochs is used. There are no significant differences between late and early Rest 1 conditions, suggesting neither temporal proximity to wakefulness nor electrode drift explain the observed results.
3.2.3 Within and across arrays

Having established that the observed explained variance appears to be attributable to memory trace reactivation and not some non-specific modulation following the waking state, the foremost question of coherent reactivation was addressed. Correlations based on all combinations of cell pairs from the same array (i.e. within-area correlations) produced a significant explained variance for three of the four arrays: PP ($p = 0.023$), M ($p = 0.0002$), and SS ($p = 0.0006$). Across-array explained variance is based on correlations in which one cell in the pair is from a different array than the other cell in the pair. The cell pair correlation distribution between any two arrays is the result of cycling through all possible such pairings including cells from both arrays. Significant explained variance across arrays was observed from correlations of PP-M pairs ($p = 0.017$), PP-SS pairs ($p = 0.029$), and M-SS pairs ($p = 0.001$). In contrast, the activity of PFC cells paired within or across areas did not lead to significant explained variance above control levels (PFC $p = 0.468$; PP-PFC $p = 0.1841$; M-PFC $p = 0.095$; SS-PFC $p = 0.8810$).

Figure 3.7 Total count of pairs of cells eligible for explained variance analysis. Note that there are few pairs of cells involving the parietal array, but similar quantities within and across the other arrays. Abbreviations:

- PP - posterior parietal cortex,
- M - motor cortex,
- SS - somatosensory cortex,
- PFC - prefrontal cortex.
Figure 3.8 Explained variances for pairs of cells within and across each array.
*Top:* In most cases, the experimental explained variances (Task, Rest 2 | Rest 1, in black) are greater than the control explained variances (Task, Rest 1 | Rest 2, in gray).
*Bottom:* The explained variance in excess of control values reveals significant effects in all combinations not including prefrontal cortex cells. Cell pair combinations in red are significant (p<0.05) and in blue are nonsignificant. Error bars represent 95% confidence intervals of the bootstrapped data.
Given the striking absence of reactivation when the PFC was involved, one might suppose that the task failed to activate PFC cells. The observation of task-related activity in a few PFC cells, and the similarity of firing rates across areas suggest that the PFC was active; however, that activity may not have resulted in a similar distribution of cell-pair correlations. For example, the activity in PFC cells may have been more independent of the task, in general, and more independent the activity of other cells, as well, producing more cell-pair correlations of small magnitude. Indeed, when comparing the distribution of all PFC-related cell pair correlations to those that did not include PFC cells, there are more correlations near zero (Figure 3.9).

Cells that are task-selective are often silent until an adequate stimulus is presented, at which time the firing rate increases dramatically. Such a cell may have the same mean firing rate as a cell that constantly fires at the same intermediate rate, independent of the task, but the task-selective cell would have high sparsity whereas the task-independent cell would have low sparsity. Thus, despite similar mean firing rates, the activity of cells in PFC may be less sparse than that of cells in other regions, indicating their insensitivity to task variables. The sparsity index, \( a \), for each cell is defined as:

\[
a = \frac{1}{N_b} \left( \sum f_i \right)^2 / \left( \sum f_i^2 \right)
\]

where \( N_b \) is the number of bins in the spike train, and \( f_i \) is the firing rate in the \( i \)th bin. Thus a maximally distributed firing rate has a sparsity index of 1, whereas a maximally sparse firing rate has an index of \( 1/N_b \). Thus if PFC cells are active, but not fluctuating based on the task, they will have a higher sparsity index value than non-PFC cells.
Figure 3.9 Distribution of cell pair correlations for PFC and non-PFC cell pairs across behavioral epochs. The top row shows the normalized histograms of cell pair correlations for non-PFC cell pairs in Rest 1, Task, and Rest 2 epochs. The middle row shows the same plots for cell pairs containing at least one PFC cell. The bottom row plots the cumulative distributions of the same data; the red line plots the top, non-PFC distribution, and the blue line plots the middle, PFC distribution. Note the tendency for PFC cell pairs to have lower correlation values than non-PFC cell pairs.
Figure 3.10 Normalized cell sparsity index distributions for non-PFC and PFC cells. A. Sparsity index distribution of non-PFC cells (i.e., cells from posterior parietal, motor, or somatosensory cortex) calculated during the task epoch. B. Sparsity index distribution of cells from the prefrontal cortex calculated from the task epoch. C. Cumulative distributions for both PFC (blue line) and non-PFC (red line) cells. The PFC distribution was significantly different from the non-PFC distribution (Kolmogorov-Smirnov test, p < 0.05). PFC cells tended to have greater sparsity (i.e., lower values) than non-PFC cells, with index values in the range of 0 – 0.1.

The results of the sparsity measure run counter to the hypothesis that PFC cells show less firing rate fluctuation than non-PFC cells (Figure 3.10). The sparsity index revealed that, when comparing to the non-PFC distribution, there are more PFC cells in the 0-0.1 range, fewer in the intermediate ranges, but no striking difference in the high range. Since low index values indicate the greatest firing rate sparsity, the PFC sample tended to include cells with greater sparsity, rather than less, as predicted. This suggests that the activity of PFC cells fluctuates, though the cause of the fluctuation may or may not be related to the task.
3.3 REACTIVATION OF SEQUENCES

The extent to which sequences of activity were preserved during reactivation was also assessed. When one cell tends to fire before another cell as a consequence of the task, temporal bias appears as an offset, or asymmetry, in the peak of the cross-correlogram (CCG) of those two cells. If the temporal biases of Task CCGs are more similar to those of Rest 2 CCGs than those of Rest 1 CCGs, this indicates the presence of some degree of sequence reactivation (Figure 3.11).

For each epoch, cross-correlograms (CCGs) were calculated by grouping spikes into 10 ms bins, and calculating the correlation histogram between all cell pairs over ±1 sec time lags, similar to that described previously (Skaggs and McNaughton 1996). In this analysis the window around the CCG was expanded and the center of mass was used rather than area under the CCG within a fixed window of time lag, as the former measure is sensitive to the variability of peak latencies present in the current CCGs. CCGs with fewer than a total of 300 counts were eliminated, as were the corresponding CCGs from the other epochs. This threshold was determined empirically as the lowest threshold that removed spurious peaks in the CCGs due to insufficient spikes. As a result, some analysis groups had too few bias values for further analysis (e.g. within-parietal cell pairs). The bias \( B \) for the CCG of cell-pair \( i \) and \( j \) is calculated as:

\[
B_{ij} = \frac{\sum_{t=1}^{N} C_{ij}(t) \cdot t}{\sum_{t=1}^{N} C_{ij}(t)},
\]

where \( C_{ij}(t) \) is the cross-correlation of cell pair \( ij \) at time lag \( t \) and \( N \) is the number of time lags in the cross-correlogram.
Based on the relationship of cell-pair biases across epochs, significant reactivation of sequences of neural ensemble activity was observed within M and SS, but not within PFC (Figure 3.12). Reactivation of sequences was also seen between PP and M, but not between M and SS nor any combination that included PFC cells. Because the PFC cells consistently fail to show memory trace reactivation, they are excluded from subsequent tests that further explore the parameters of memory trace reactivation.
Figure 3.12 Preservation of temporal bias within and across arrays. Each plot shows the temporal bias for cell pairs during task versus the difference Rest 2 and Rest 1 bias. Each point represents the bias of two cells from the respective arrays. Within-array combinations are shown in the top row, across-array combinations on the bottom. The black lines are the regressions for each scatterplot, and the proportion of cell pairs falling in the predicted quadrants (white) are listed in the upper right of each plot. Both within motor and somatosensory, and across parietal and motor arrays, temporal bias was preserved. Neither cell pairs across motor and somatosensory cortex nor any combination of cells from the prefrontal array reached significance. Correlations can be heavily influenced by outliers and, therefore, may not accurately reflect a trend in the distribution. If the present correlations were only the result of a few outliers, then the number of points should be equally distributed in each quadrant (leaving only the magnitude to influence the correlation). When significant correlations were observed, there were greater proportions of points in the quadrants that are consistent with preserved temporal bias (white squares). In contrast, points were equally distributed for the (non-significant) PFC groups, indicating that the temporal bias effect is not exclusively the result of the magnitude of a few outliers.
Figure 3.13 Distribution of bias values for PFC and non-PFC cell pairs across behavioral epochs. The top row shows the normalized histograms of bias values for non-PFC cells in Rest 1, Task, and Rest 2 epochs. The middle row shows the same plots for bias values of PFC cell pairs. The bottom row plots the cumulative distributions of the same data; the red line plots the top, non-PFC distribution, and the blue line plots the middle, PFC distribution. As in Figure 3.9, there is a tendency for PFC cell pairs to have low bias values relative to the non-PFC cell pairs.
3.4 VIGILANCE STATE AND MEMORY TRACE REACTIVATION

Formal assessment of vigilance state including classification of REM and non-REM sleep requires EEG, and EMG or EOG. Although these measurements were not consistently available due to technical difficulties, the global changes from waking to sleep could be indirectly assessed by monitoring times when eyes were open or shut during the rest epoch.

Segments for Rest 1 and Rest 2 were collected and binned to match those used in spike train correlations (300ms). The durations for each condition were equalized by randomly selecting bins without replacement from the longer-duration condition (typically 'closed'), until both times had equal numbers of bins (i.e. equal durations). There was no need to consider the continuity of bins, because bin order is irrelevant for the spike-train correlation analysis. No PFC cells were used in this analysis.

Both open-eye and closed-eye conditions showed significant explained variance, though there was no difference across conditions. Thus memory trace reactivation occurs both during sleep and quiet wakefulness.

Figure 3.14 Memory trace reactivation during sleep and quiet wakefulness. The explained variance derived from periods in which eyes remained closed for at least four seconds is shown on the left; explained variance based on all periods in which eyes were open is shown on the right. Error bars indicate SEM of bootstrapped data. The explained variance was significant for both closed and open conditions (p<0.01), but there was no difference between these groups.
3.5 MEMORY TRACE REACTIVATION OVER TIME

Previously it was shown that the relationship between correlations from the task and in subsequent rest decreased with each passing 10 minute block of rest (Kudrimoti, Barnes et al. 1999). In the present data, the same trend emerged, with the third block of 10 minutes showing no significant explained variance above control levels.

![Figure 3.15 Explained variance above control levels as a function of Rest 2 delay. The first (0-10 minute) bar reflects the time period used in all other reactivation analyses. Only the third block of 10 minutes failed to show significant reactivation. Error bars indicate the 95% confidence intervals of the bootstrapped data.](image)

Although decay is observed within an hour following behavior, it must be reiterated that this measure shows only “new” reactivation of the immediately preceding task. In previous reactivation studies, some relationship existed between the correlations in Rest 1 and Task. This relationship has been deliberately factored out, such that pre-existing correlations are not erroneously attributed to the recall of a memory trace; however, in some cases the preceding correlation structure may indeed reflect reactivation of a
familiar task. In all of the previous studies showing a significant relationship between preceding rest and task correlations, the task apparatus and behavior were exclusive and presented daily. It is possible that placement of the animal into the rest setting elicited recall of the memory trace of that task, predicting the event to follow. The one study in which both novel and familiar mazes were used only reported the control EV, and not the correlation between Rest 1 and Task, thus the Rest 2 correlations are always factored out (Gerrard, Kudrimoti et al. 2001). This makes it impossible to assess the strength of Rest 1 reactivation of the familiar task, since presumably the same neural ensemble of the memory trace in Rest 1 also reflects this memory trace in Rest 2 and is therefore factored out. Indeed, the novel and familiar control EVs are all near 0.

In the present study, the alley maze tasks (Sessions 7-9) were similar to the conditions of the rat tasks. The behaviors and apparatus were highly familiar; during the month of training that preceded the recording sessions, the alley maze was given exclusively as the daily task, and therefore was predictable during preceding rest epochs.

In contrast, Sessions 3 and 4 involved novel reach tasks for bimanual foraging and reaches through a grating. Although the location and general procedures were familiar and unchanging, the required use of both hands and foraging in the Session 3 task was novel, and the grating apparatus, the use of reaches in the vertical plane, the consequent novel reaching sequence, as well as the use of different grips for vertical and horizontal reaches were all novel in Session 4.
These unique task elements could not have been predicted in the preceding rest epoch, leading to the question, would Rest 1 and Task correlations be stronger for familiar, predictable tasks, than for novel tasks?

Figure 3.16  Distribution of cell pair correlations during Rest 1 and Task for Novel and Familiar tasks. Each cell pair firing rate correlation is one point in the scatterplot. The epoch correlation between Task and Rest 1 cell pair correlations was greater in familiar than in novel sessions (p<0.001). This was true even when the familiar cell pair correlations were subsampled to equal the number of novel cell pair correlations. The linear regression line is shown in black, and the epoch correlations are visible in the upper right of each plot.

Caution must be used in interpreting the novel and familiar results. The experiments were not designed for this test, and both novel sessions occurred before the familiar sessions. Also, all three familiar sessions were from the same virtual reality task. Although both novel and familiar tasks involved sequential reaching, the considerable
differences in visual input, the use of a computer screen, and juice reward rather than treats, are just a few possible differences across groups. How these differences might lead to the specific results observed is unknown, but they reflect potential confounding factors, nonetheless.
3.6 OPTIMIZATION OF MEMORY TRACE REACTIVATION ANALYSES

Overall, the activity of cortical networks is fairly sparse, observed as low overall firing rates. Given that only a small sample of cortical cells is recorded, the result would be a disproportionately sparse population vector. In the extreme case, some events will fail to elicit any activity in the cells recorded, producing 'empty' population vectors. In such a case, reactivation of that event could not be detectable from the sampled cells. In the cell pair correlation analysis, the undersampled population vectors are expected to add noise to the cell-pair correlations (expressed as more occurrences of bins in which neither cell fires), tending to reduce the absolute magnitude of the correlation. By restricting the analysis to the less sparse population vectors during task, there should be an increase in the proportion of bins in which at least one member of a cell pair is active, increasing the absolute correlation value and providing more task-related variance in the correlation distribution on which the EV analysis is based. This could be one way to increase the signal-to-noise in the explained variance analysis.

The sparsity index, $a$, is defined as:

$$a = (1/N_c)(\sum f_i) / (\sum f_i^2)$$

where $N_c$ is the number of cells in the ensemble, and $f_i$ is the average firing rate of the Ith cell in that time bin. Thus, sparsity ($a$) can be calculated for each population vector during the task, and the cell pair correlations recalculated from only the least sparse (most active) population vectors. The results are shown in Figure 3.17.
Figure 3.17 Explained variances after removing sparse population vectors from the task epoch. A. Normalized sparsity distribution for population vectors from the task epoch of one session. The sparsity index (a) is plotted on the x-axis. Five threshold values are illustrated as vertical lines; all values to the right of a line are included in the explained variance analysis. The green and yellow lines are the threshold values used in B. B. Normalized cell pair correlation distributions calculated from population vectors exceeding the green and yellow thresholds, respectively. As predicted, removal of the sparse vectors (low index values) tends to reduce the proportion of small correlations. C. Explained variances based on increasingly stringent sparsity thresholds. The black bars show the resulting explained variance when only vectors exceeding the respective threshold are included. Gray bars show the resulting explained variance when a random subsample of the same number of vectors is included. Since the far left bar includes all population vectors, the gray and black bars are the same.
Filtering the time bins during task by the corresponding population vector sparsity had mixed results. The number of time bins used to calculate the cell pair correlations can have an impact on the resulting explained variance. At the 0.1 sparsity index threshold, nearly half of the time bins were removed, yet the explained variance was unaffected. But when greater proportions of the distribution are removed, the explained variance drops off. The lack of increase in explained variance suggests this filtering process is not an effective means of increasing the signal-to-noise post hoc. However, when comparing to a random subsample of equal size, the relative benefits increase with increasing sparsity thresholds. This suggests that, although removing vectors may be more harmful than helpful, attempts at obtaining maximally distributed (i.e., highly active) population vectors during a behavioral session could increase the ability to detect reactivation.
3.7 CONCLUSION

Four 144-channel arrays were implanted in a monkey, allowing chronic, multi-channel single unit recording in the behaving animal. Yields varied day-to-day, but typically provided neural ensembles of over 50 neurons across several brain regions stable for the duration of the recording session.

Memory trace reactivation was observed for a variety of tasks, when sampling cells from different brain regions, including across hemispheres. The results are consistent with most of the preceding literature on memory trace reactivation. Although sleep state could not be directly assessed, indirect measures indicate significant explained variance during sleep and quiet wakefulness. Decay of reactivation over the 30 minutes of subsequent rest was also observed. Reactivation in parietal cortex was also replicated, in this case in the primate brain.

The foremost issue of coordinated memory trace reactivation throughout the neocortex was demonstrated by significant explained variance within and across motor, somatosensory, and parietal cortex. Further, the sequence of activity between cells during the task tended to be preserved during the rest epoch afterwards, even across two of the brain regions recorded. In notable exception to the coordinated memory trace reactivation seen in three areas, the dorsal prefrontal cortex failed to show memory trace reactivation in all measurements taken.
4 DISCUSSION

4.1 TECHNOLOGICAL AND PROCEDURAL DEVELOPMENTS

The first use of the 576-channel electrode implant yielded stable, simultaneous recordings of single unit ensembles and EEG data. In the three month period of implantation, enough data were obtained to complete several different analyses. In this respect, the implant was a success; however, the data were collected in spite of several major setbacks that should be improved upon in future implants.

First and foremost, the implant became unstable and had to be removed. A layer of vascular tissue infiltrated and expanded the region between the resin and the skull. In addition, the bone that had been removed for insertion of the inverted anchor bone screws had not grown in, and may have even eroded. The tissue growth apparently caused the bone screws to lift along with the implant, leading to instability. One factor that may have contributed to the tissue growth and bone decay was the presence of a strip of exposed bone where there was insufficient skin remaining to pull up against the implant. The bone exposure may have led to infection, which, in turn, led to tissue regrowth. Alternatively, uncured monomer in the resin may have been biologically incompatible, causing an immune response leading to tissue growth that would isolate the skull from the monomer.

Both of these factors can be addressed in the next implant. During surgery, darts can be made in the circumference of skin, allowing further retraction, and, therefore, a small circumference oval that will mate with the implant. If the resin monomer was a problem, several solutions are possible. An intervening layer of inert material could be
added to the skull surface prior to adding the resin. The implant could be made of a biocompatible material other than resin. Finally, bone growth factors applied around the inverted screw craniotomies would encourage bone regrowth, increasing the holding power of those screws. The stability of chronic electrode implants is a problem faced by many electrophysiologists; these and other possible techniques are being discussed and tested across laboratories to find the ideal procedure to ensure a long, healthy, stable implant.

Another obvious problem with the implant was the failure of some of the electrodes to penetrate the dura. Aside from removing the dura, stiffer electrodes could be used, and less silastic, ensuring that the electrode is minimally exposed between the cannula and the dura. In addition, the simultaneous cell yield per electrode was low, typically only one cell. These, along with other known problems with the electrodes such as bending and shunting capacitance across the insulation, lead to the conclusion that the electrodes used were not optimal, and other types should be considered. Nevertheless, these electrodes provided sufficient neural signal for the present experiment.

Depending on the questions addressed in future experiments, different spatial distributions of the array may be desired. Closer inter-electrode spacing would allow greater sampling of small brain regions, and some brain regions would be better suited to rectangular strips rather than a square grid. A new 240-channel drive has been designed that allows some flexibility in the array shape, though the minimum inter-electrode spacing is still \(=600 \, \mu m\) (Neuralynx, Inc.).
Wires for EMG were implanted in facial muscles during the electrode implantation surgery for later discrimination of sleep stages. Unfortunately, the material used to seal the fluid resin adhered to the EMG wires, which were inadvertently removed along with the seal. Adding a liberal coating of lubricant to the exposed EMG wire should prevent this problem in future surgeries.

The greatest limit to larger cell yields was the time required for electrode advancing. Typically, 4-6 hours of lowering occurred on recording days, which were often interleaved with days consisting entirely of electrode advancing. By automating the alignment and insertion of the pusher rod, and particularly the fast but precise movement of the rod down to meet the back of the electrode, time and labor would be dramatically reduced.

Improvements are an anticipated part of new technology. The data reported in this study would not have been obtained without the development of the high-density recording system (i.e. the 144-channel drives and 160-channel acquisition system). As design improvements are made, even more ambitious questions about the dynamics of neural ensembles may become technologically feasible.
4.2 MEMORY TRACE REACTIVATION

4.2.1 Dorsal prefrontal cortex

One of the most surprising results from the present experiment was the lack of memory trace reactivation in the dorsal prefrontal cortex array. Despite cell yields (Figure 3.6) firing rates (Figure 3.3) and task-related activity (Figure 3.4) similar to those obtained with the other arrays, no measure of reactivation produced an effect when PFC cells were involved.

Given that the PFC cells were selective during the task, and just as active during the task as during Rest epochs, and as active as the other three brain regions, the simple explanation that PFC cells weren't active is unfounded. The narrow distribution of cell-pair correlations involving PFC cells during Task suggests that the relationship between the activity of PFC cells and that of other cells (including within PFC) may fluctuate within behavioral epochs.

Perhaps other tasks would require different PFC engagement, more in line with memory trace reactivation. It is known that the distribution of neurotransmitters across brain regions can vary with the task being performed. The prefrontal cortex shows dissociations in the glutamate and dopamine concentrations present during sensory-guided versus working memory tasks (Kodama, Hikosaka et al. 2002). The neurotransmitter composition could determine which cells participate in an ensemble, what patterns that ensemble can generate, and the temporal duration over which the ensemble is bound together, factors that could affect the likelihood of reactivation in that brain region. More data using various tasks would be required to address this possibility.
This dynamic relationship of PFC cells may be a consequence of the hierarchical position of this region relative to those of the other arrays. Based on the co-registered CT and MR images, the electrodes were lowered into 8B and adjacent 9L, at the superior tip of the arcuate sulcus. Projections from this area are unique from those of other prefrontal areas, including those of area 9/46 in the principal sulcus. In one study (Petrides and Pandya 1999), fast blue retrograde tracer was injected into two animals, with an injection site restricted to the area just dorsal to the tip of the superior arcuate, in area 8b, matching the implant location of the dorsal PFC electrode array. Both animals in that study showed similar results. Specific frontal regions had the most extensive labeling: area 6 just caudal to the injection, areas 9, 9/46d, and 46; and orbitofrontal areas 47/12, 10, 14, and 13. The medial surface also had extensive labeling in the prefrontal regions and 8B, 9, 10, and 32, but also more caudally including area 23, PGm, 19, 31, and retrosplenial area 30. There was also labeling in the caudal, medial part of the IP lobule in visual areas Opt, caudal PG and POa. Unlike dorsolateral PFC, area 8B received no projections from LIP or VIP in the region targeted by the parietal array. Moreover, no label was found in primary somatosensory cortex, primary motor cortex, or in the IP sulcus. In contrast, the remaining three targeted regions are interconnected. The areas around the IP sulcus are reciprocally connected to motor cortex (Marconi, Genovesio et al. 2001; Tanne-Gariepy, Rouiller et al. 2002), somatosensory cortex has access to motor areas and vice versa, either directly or via area 6 (Jones, Coulter et al. 1978), and the path from somatosensory to parietal areas was illustrated in Figure 1.2 (Cavada and Goldman-Rakic 1989; Felleman and Van Essen 1991).
The PFC region labeled (Petrides and Pandya 1999) was not connected to primary sensory cortices, and, in fact, was most strongly connected to other PFC regions. Moreover, the other three regions were more closely interconnected than was PFC with these regions, possibly affecting the likelihood of achieving coherent replay.

Another interpretation is that the prefrontal region in question is functionally unique from other cortical regions, possibly serving a role that is incompatible with memory trace reactivation. Trace replay theory treats all cortex as equipotent, defined only in terms of connectivity within and across modules. But the neocortex can vary greatly in its laminar distribution. Some models of cortical hierarchy include laminar profiles of projections in determining feedback from feedforward connections, but even then, some regions of neocortex stand out in violation of the heuristics used (Felleman and Van Essen 1991). Indeed, area 8B and adjacent parts of area 9 are cytoarchitectonically distinct, notable for their poorly developed layer IV (to the point of being considered dysgranular). This area is also distinct in its sparse but prominent, deeply stained, large pyramidal cells in layer Va, even more than in layer III (Petrides and Pandya 1999). The regions surrounding the 8B/9 area do not share these characteristics, nor do many regions of sensory and association cortex. The projection pattern of axon collaterals in this area is also distinct from other areas of cortex. Many local collaterals, which form ‘stripes’, terminate on GABA-positive dendritic shafts, and lead to both excitatory and di/polyisynaptic inhibitory responses. In contrast, distal projections terminate on dendritic spines and lead to exclusively excitatory responses. These results indicate that the intrinsic connections of PFC show a unique pattern and
that within-stripe feedback inhibition may be more prominent than feedforward inhibition between stripes or across regions. The near absence of an 'input' layer IV in area 8b, its primarily prefrontal afferents, and the prevalence of local-circuit feedback inhibition may lead to circuit dynamics somewhat unique to this region.

Functionally, the dorsal PFC differs from dorsolateral PFC, and both differ in some respects from non-prefrontal regions. Lesions to area 8B/9 produced no impairments for the working memory tasks that require an intact dorsolateral PFC (Levy and Goldman-Rakic 1999). Area 8B/9 and dorsolateral PFC show differential changes in glutamate and dopamine concentrations across two different tasks (Kodama, Hikosaka et al. 2002). Area 8B/9 shows increased glutamate but not dopamine concentrations for sensory-guided tasks, but no neurotransmitter fluctuations for delayed response tasks. There is also indication that area 8B/9 is part of a broader stretch of prefrontal cortex selective to ear and eye orientation, though this region is distinct from frontal and supplementary eye fields (Mitz and Godschalk 1989; Bon and Lucchetti 1994). In contrast to other areas of neocortex, prefrontal cortex neurons have been shown to reflect previous or future performance rather than track current performance (Hasegawa, Blitz et al. 2000).

The failure to observe reactivation in dorsal prefrontal cortex seems to contrast with the evidence from human neuroimaging literature that right dorsal PFC is active during episodic memory retrieval in humans. The role this region plays in memory tasks may be peripheral, as part of a cognitive set associated with intentional retrieval, rather than retrieval per se (Lepage, Ghaffar et al. 2000). Consistent with this role, the right
dorsal PFC activity in human imaging studies has been associated with a variety of other tasks requiring working memory and/or monitoring functions (MacLeod, Buckner et al. 1998). The human and monkey literature together suggest the PFC may facilitate intentional memory retrieval only through its role in short-term mnemonic, executive, or monitoring functions, which may not be stored as components of an episodic memory.

In sum, area 8B/9 contains a distinct pattern of morphological and functional characteristics, potentially affording it a unique role in neural function that may be mutually exclusive with memory trace reactivation. The alternative explanations are that the specific tasks selected, and/or the location of this area in the hierarchy relative to the other targeted areas, led to the present results. Further research will be required to understand the failure to observe memory trace reactivation in the dorsal prefrontal cortex.

4.2.2 Reactivation of a distributed memory trace

The major finding of the present study is that the activity pattern of a distributed neural ensemble during task performance coherently re-emerges afterwards. The within- and across- array cell pair correlations in subsequent rest were predictive of the correlations during the task, even after factoring out those of a preceding baseline period. The control variance explained by correlations in task and the baseline period, factoring out subsequent rest was significantly lower than the variance explained by task and subsequent rest correlations given those of the baseline period. Even the order of neural activity during task tended to be preserved in rest afterward.
The analyses used in the present study match the most common tests of reactivation in previous studies; however, they may not be optimal for detecting memory trace reactivation. It would be useful to examine the activity of all simultaneously recorded cells as an ensemble, rather than in all possible pairwise combinations. Other techniques have been developed to address this issue, but they all require considerable task selectivity in the sampled neurons. Cellular activity from the hippocampus as rats repeatedly traverse an environment provides ideal data for this purpose. In the present study, more cells, or, more to the point, greater proportions of task selectivity would be needed in each region in order to obtain data of similar quality. Future monkey studies should be able to approach this goal with greater cell yields and modified tasks.

The observed explained variance is a mere fraction of the total variance among correlations across epochs. The current methods cannot distinguish whether the low values of explained variance and bias preservation result from imperfect recall of recent patterns or an interleaving of recent memories with other activity states, possibly
including other memory traces. Neither explanation would be unexpected, given the repeated attempts that may be required to recruit the entire memory trace, and the numerous possible neural ensemble patterns that may be elicited during the rest epoch. The small amounts of reactivation seen in the present studies, however, should encourage the pursuit of experiments designed to elicit a stronger effect.

Restriction of data to periods of time when reactivation should be concentrated is a start. The ripple-oscillation periods in the hippocampus are one example, but one could even restrict the task data to periods in which the most consistent, differentiable neural responses are seen. Focusing on cued rather than spontaneous reactivation is another possibility that should yield stronger effects, though this procedure would probably require an awake animal. It may be possible to elicit reactivation even in a resting animal by selectively stimulating the electrodes recording a subset of cells constituting the memory trace. A stimulation procedure may not be entirely biologically realistic, but a stronger case for reactivation would be made if the remaining neurons in the ensemble were recruited when a subset, but not a random selection, of neurons was stimulated. Moreover, stimulation procedures could test the prediction from trace replay theory that old memory traces should be faster to ‘pattern complete’ than new memory traces. Finally, if reactivation is part of a memory consolidation process, and if neural patterns from activity can be reliably elicited, then massive, repeated retrieval should confer better subsequent memory. These are just a few directions that future experiments may take.
4.2.3 The nuts and bolts of memory trace reactivation

In addition to the principal finding of coordinated memory trace reactivation, some additional characteristics of reactivation were explored.

Vigilance state

Consistent with previous observations, significant reactivation was observed during periods of quiet wakefulness, when eyes were open, and during rest, when eyes remained closed (Kudrimoti, Barnes et al. 1999; Gerrard, Kudrimoti et al. 2001). In the hippocampus, quiet wakefulness and slow-wave sleep are characterized by the presence of sharp waves and ripple complexes. These oscillations have been associated with periods of increased synchrony with output areas of entorhinal cortex, and also with spindle oscillations in the neocortex (Chrobak and Buzsaki 1994; Siapas and Wilson 1998). They are also the periods during which reactivation is seen in hippocampal ensembles (Kudrimoti, Barnes et al. 1999; Gerrard, Kudrimoti et al. 2001).

In the neocortex, however, little mention is made of oscillations in the quiet wakeful state; only non-REM periods contain global, synchronous oscillations coherent in space and time. Yet the primate neocortical data show reactivation even during quiet wakefulness. One possibility is that synchronous oscillations, in fact, do occur in quiet wakefulness, mirroring the hippocampal oscillations. Typically, an experiment involving sleep only analyzes the oscillations occurring during sleep, and wakeful oscillations are detected as an animal completes a task. Thus, there may be a lack of focus on non-attentive, quiet wakeful periods. Alternatively, the quiet wakeful periods in the present
study may be filled with volitional task recall: periods in which the monkey ‘thinks
about’ the task he just completed. This is not the sort of explanation amenable to testing,
but it is a possibility nonetheless. Future studies should explore the relationship between
synchronizing oscillations and memory trace reactivation, to attempt to clarify the
observation of reactivation across vigilance states.

**Reactivation as an automatic and obligatory process**

The explained variance measure factors out the relationship between task
correlations and baseline correlations of rest preceding the task. Thus, the presence of
significant EV for familiar tasks indicates that ‘new’ reactivation occurs in subsequent
rest, in excess of the variance explained by correlations preceding the task. Why would
the brain reactivate an event that has already been learned? Although the parameters
guiding the selection of a pattern to reactivate are unknown, one possibility is that
reactivation is an automatic consequence of whatever neural pattern was recently active.
There may be no neural marker of ‘old’, ‘new’, or even ‘important: remember me’ as
part of the reactivation process: it may simply reiterate the most recent or most salient
patterns. Consistent with this notion, experiments in which reward magnitude was varied,
failed to find differences in the degree of reactivation (Cowen, Kudrimoti et al. 2000).

Even if reactivation reflects replay of the unique episode that just transpired, the
neurons affected may be nearly the same as those that were involved in previous episodes
of task performance, and may therefore benefit from the new ‘statistics’ that could
reinforce the appropriate connections. This interpretation emphasizes the continued
function of the whole cortical hierarchy even for well-established memories that may not require the whole network.

If reactivation is indeed obligatory, one might predict the network could saturate—that synapses would meet their maximum strength. First, the quantity of spontaneous reactivation is clearly small, and possibly interleaved with other patterns, making saturation unlikely. Moreover, in the week to month timescale, there is one piece of evidence suggesting synapses that were saturated can recalibrate their range to allow even further potentiation (where none had previously been possible; (Rioult-Pedotti, Friedman et al. 2000; Rioult-Pedotti and Donoghue 2002).

Thus repetition long after an initial exposure could allow a greater synaptic bang for the reactivation buck. Such an effect would match well with memory data showing that repetition strengthens memories, and that the more distributed the repetitions, the greater the benefit. Where synaptic range recalibration may allow further strengthening over time, other mechanisms may decrease the plasticity in remote traces. Although the molecular mechanisms are unknown, the immediate early gene zif268, implicated in synaptic plasticity, is expressed in CA1 of the hippocampus upon retrieval of recent but not remote contextual fear memories (Hall, Thomas et al. 2001). Regardless of the specific mechanisms, the effect of reactivation on memory consolidation may be adjusted ‘downstream’, at the level of synaptic plasticity. Before becoming too concerned with these details, however, a relationship between memory trace reactivation, synaptic plasticity, and memory consolidation needs to be established.
Reactivation diminishes with time, until the next recall

The amount of reactivation detected in the present study, as in previous studies, decreased with each 10-minute block of time following the task (Kudrimoti, Barnes et al. 1999; Gerrard, Kudrimoti et al. 2001). This does not suggest that reactivation ceases within 30 minutes, only that the current analysis methods are not able to detect it. This is particularly true given that any reactivation of a familiar task prior to that day’s session is factored out of the EV analysis. Indeed, Rest 1 and Task correlations for familiar sessions are more related than those of novel sessions. The relationship between preceding rest and behavior for familiar tasks was seen with rat hippocampal data, as well, suggesting either residual reactivation from the previous day’s session or an invariant relationship in the activity between some cells (Wilson and McNaughton 1994; Kudrimoti, Barnes et al. 1999).

The temporal gradient observed in retrograde amnesia varies widely, and can extend back for years in humans, yet reactivation shows relatively fast decay. These two observations are not incompatible, if one considers that reactivation may occur anew with each recall, as suggested in the previous section. Combining the evidence of obligatory reactivation and decay rate, it may be that the first several hours after learning are the most important. Beyond that, it is not time per se, but the number of repetitions of event recall that determines the strength of the memory trace. Some memories may require many repetitions to form the most direct connections among elements of the memory trace, thus one night’s reactivation may be insufficient. With enough repeated recall,
however, the memory may shift to more direct connectivity, rendering it impervious to damage in upper-level (i.e. hippocampal) regions. These hypotheses remain to be tested.
4.3 MEMORY CONSOLIDATION

4.3.1 Trace replay theory as a mechanism for memory consolidation

The emerging characteristics of memory consolidation are well described by trace
replay theory.

1. Recall of the memory trace requires the hippocampus less with time.
   TRT: The top of the hierarchy is necessary to elicit coincident lower level
   activity, until more direct lower level connections are established.

2. Hippocampal function persists, nevertheless.
   TRT: The top of the hierarchy will still be recruited with the memory trace,
   although it may not be required.

3. Consolidation takes longer or may not be possible when damage extends into the
   medial temporal lobe.
   TRT: Extensive damage to upper levels in the hierarchy will slow or prevent the
   formation of direct connections.

4. Consolidation can be achieved in minutes, but may take years.
   TRT: Consolidation takes less time when fewer novel details need to be
   associated in forming the memory trace.

5. Consolidation is concentrated during sleep.
   TRT: Sleep includes periods of increased global synchrony, well suited for
   Hebbian synaptic modification of the memory trace.

There is no direct evidence that memory trace reactivation is part of the process of
memory consolidation (but see Gerrard 2002 for the first indirect evidence). Of the
characteristics listed, memory trace reactivation is in agreement with #1., though it does
not address the necessity of high-level regions, with #2., inasmuch as reactivation of the
distributed trace occurs coherently, and hippocampal reactivation occurs for familiar tasks, and #5., given that the strongest reactivation occurs during sharp wave/ripple periods in the hippocampus.

Clearly, much territory remains to be explored. The first step in establishing the feasibility of trace replay theory was showing that hippocampal ensembles replay after a task, even for familiar tasks (2). The second step began with the report of hippocampal and parietal cortex reactivation by Qin and colleagues (1997), and continued with this study, showing that ensembles from multiple neocortical areas reactivate together, as a unified memory trace (1). The remaining issues are dependent on these observations.

The results of the present study indicate that the memory trace can act as a unitary entity, despite the distribution of its component neurons throughout the neocortex. There exists one final ‘feature’ of consolidation not mentioned previously. This phenomenon has been described in one light, but may now be considered in a new light, one that stresses the importance of coordinated reactivation of the complete memory trace for proper memory consolidation.

### 4.3.2 Reconsolidation

According to trace replay theory, as a memory is consolidated, the connections among neurons participating in the memory trace are strengthened. Thus, once consolidated, the memory should be impervious to manipulations that initially would have interfered with the trace, as connections were still being formed. The phenomenon of reconsolidation, however, would appear to call into question the stability of remote
memory traces. When a cue that elicits recall of a remote memory is immediately followed by one of several manipulations, subsequent recall is impaired. In such studies, controls reveal that the manipulation is only disruptive when memories were recently recalled. Moreover, recall in the presence of sham manipulations is unimpaired. This line of investigation has led to the conclusion that the process of recall makes old memories become “new” again, that is, susceptible to disruption (see Nadel and Land 2000; Sara 2000; Myers and Davis 2002, for review). But the data do not necessarily indicate that reconsolidation involves the reversion of a memory to a ‘younger’ state; rather, it may indicate some unusual conditions under which new memories are formed, and it is the new, unconsolidated memory that is sensitive to interference, just as described by consolidation theory.

Most of the reconsolidation literature uses fear conditioning as the learning paradigm. Animals learn that a tone or light conditioned stimulus (CS) predicts a shock unconditioned stimulus (US), leading them to respond to the CS as they had initially responded to the US (e.g., by freezing or avoiding the location in which the US occurred). The recall test, called a ‘reminder’ in reconsolidation literature, involves presentation of the CS alone. Importantly, this procedure is called a probe trial or an extinction trial in other experiments, indicating that animals eventually will stop responding to the CS as they learn that the CS does not predict the US. Immediately after one such reminder trial, in which animals show a conditioned response, ECS is given. Upon retesting the following day, animals given ECS no longer show a conditioned response to the CS (Misanin, Miller et al. 1968; Mactutus, Riccio et al. 1979). The
memory does not appear to have been erased, however, as the presentation of the US alone can make later presentation of the CS alone effective once again (Miller and Springer 1972). The memory recovery effect has been described following extinction trials, as well, leading to the suggestion that extinction trials create a new, competing memory that does not lead to the conditioned response (Miller and Springer 1972).

Similar results are obtained when the hippocampus is lesioned within 24 hours of recall. Whereas rats demonstrate a conditioned response if the hippocampus is lesioned 30 days after learning, they fail to show a response if a reminder trial immediately preceded the lesion (Land, Bunsey et al. 2000). Lesions following recall are effective up to, but not exceeding 24 hours. In one study of memory consolidation, animals that recalled a previously learned context the day before lesioning failed to show post-operative reconsolidation effects (i.e., memory for the remotely learned material was unimpaired; (Anagnostaras, Gale et al. 2001). But another consolidation study in monkeys showed reconsolidation effects when recall of remotely learned scene stimuli was followed by fornix transection (Gaffan 1993). In this study, there was no group receiving the lesion without having experienced the reminder, therefore the amnesia may reflect consolidation rather than reconsolidation effects. By systematically varying the interval between the reminder and the lesion surgery, it was demonstrated that at 4 or 24, but not 48 hours, reminders led to post-operative performance deficits (Debiec, LeDoux et al. 2002). Thus, reconsolidation appears to have a temporal gradient similar to consolidation.
An alternate explanation for these effects arises from trace replay theory, in a demonstration of coincidence gone wrong. It is known that ECS causes massive fluctuations in the activity of neurons throughout the cortex. The effects on neural activity in the hours following ECS are largely unknown, but are sufficient to disrupt memory formation. Consider the evidence that a memory trace, regardless of age, is automatically reactivated following an experience. The co-activity of the memory trace with neurons unrelated to the trace, induced by the ECS, could lead to the strengthening of a new ensemble containing some mixture of task-related and task-irrelevant neurons. This new, "noise" trace, may not be associated with the conditioned response, thus if the post-treatment CS triggers recall of the new trace, performance would reflect successful recall of an incorrect memory trace. This interpretation is supported by the retrieval of the original trace when a less ambiguous cue is given. The US, which was present in the learning, but not the reminder trial, should be more strongly associated with the original trace, therefore a post-treatment US should elicit a conditioned response. This process is illustrated in Figure 4.2A.

Lesions may have a similar effect on neural activity. The delicate balance of excitation and inhibition in one area can be altered radically by removal of afferent activity. It is known that the effects of lesions in one brain region can dramatically affect activity in 'downstream' regions, even when no cell death is observed outside the lesioned area. For example, the immediate early gene c-fos, a marker for neural activity, shows increased expression in widespread cortical areas following electrolytic lesions of the perirhinal cortex (Glenn, Woodside et al. 2002).
It is conceivable that hippocampal lesions cause instability of the activity in the entorhinal cortex, which, due to its position in the hierarchy, could lead to a cascade of activity throughout the neocortex that is unrelated to any memory trace. If this occurs during a period of reactivation then, once again, co-activity may bind ‘irrelevant’ groups of neurons to the reactivating memory trace, effectively creating a new trace that is later recalled (Figure 4.2A). This process would only occur if the old memory trace was being reactivated, e.g. immediately after recall. In the absence of a recent reminder, no new, interfering trace is created; the original memory is the only one associated with the cues and, therefore, the only one recalled afterward.

The results of other reconsolidation studies are more ambiguous, making it difficult to arrive at one explanation for the variety of contradicting results. The primary technique in these studies is to administer anisomycin, a protein synthesis inhibitor, after the reminder trial. Since protein synthesis is necessary for plasticity, and has been implicated in memory consolidation, it may prevent the ‘re-formation’ of an old memory during reconsolidation (Teyler and DiScenna 1987; Larkman and Jack 1995; Nicoll and Malenka 1995). Indeed, several but not all studies report performance impairments following anisomycin treatment.

Although no one explanation accounts for all the data, trace replay theory predicts a distinction between localized and systemic administration that is born out in most of the studies. According to the theory, the coordinated replay of the entire distributed trace causes Hebbian plasticity among the participating synapses. But should synaptic strengthening be prevented in a module or portion of the trace, while the remaining
ensemble continued to reactivate and, hence, strengthen, then the memory trace would gradually come to exclude that module. The result would be the creation of a new, stronger trace from a subset of the original trace (Figure 4.2B). With one exception, studies using localized anisomycin treatment show results consistent with this model (Nader, Schafe et al. 2000; Taubenfeld, Milekic et al. 2001; but not Vianna, Szapiro et al. 2001; Debiec, LeDoux et al. 2002).

In contrast, systemic anisomycin should prevent further plasticity to the entire reactivating trace, leaving the trace no stronger or weaker than it was prior to drug treatment (Figure 4.2C). Indeed, studies of systemic anisomycin treatment report the absence of impairment (Lattal and Abel 2001). One study, however, found impairments restricted to recently learned (2- and 7-day-old), but not remote, memories (Milekic and Alberini 2002). One might suppose this is due to the early, labile state of the memory traces; however, administration of anisomycin 2 or 7 days after learning, in the absence of the reminder, produced no impairment. Thus, the recall of the memory trace seems to have been necessary for anisomycin to be effective. There may be subtleties in how strong the memory trace was in each of these conditions, and more research is necessary to explore this intriguing result. Nevertheless, most of the data fit the predictions of trace replay theory.
Figure 4.2 Reconsolidation as explained by trace replay theory. Icons from left to right depict the state of the memory trace over time. When a context or cue is presented, successful recall of the original memory trace is indicated below the icon at that time point. A. During learning, a memory trace is formed and consolidated through reactivation (arrow). Presentation of a (CS) cue after consolidation leads to successful retrieval of the memory trace. If recall is immediately followed by ECS (or lesions), reactivation results in a mixture of active elements, creating a new trace. Subsequent CS presentation leads to recall of the new trace, whereas US presentation followed by CS presentation biases selection processes in favor of the original trace. B. As in A., but reactivation under local anisomycin during reactivation causes the selective strengthening of part of the memory trace, effectively degrading the non-reactivating portion of the trace. The result is a new trace that is recalled upon CS cue presentation. C. Systemic anisomycin treatment during reactivation prevents further strengthening of the memory trace, but does not cause any decay, therefore the memory trace is maintained in its pre-reactivation state, and subsequent recall of that trace is unimpaired. D. Treatment with MK801 or a β-adrenergic antagonist may selectively depotentiate active synapses. Thus, when given during reactivation, the memory trace being reactivated would suffer selective decay, and subsequent recall would be impaired.
Other cases of ‘reconsolidation’ may ultimately be viewed as evidence linking memory trace reactivation and synaptic plasticity, as predicted by trace replay theory. Rats that had learned to locate food in a radial arm maze were susceptible to an NMDA-receptor blocker administered systemically within two hours following recall (Przybyslawski and Sara 1997). The NMDA receptor, which allows Ca$^{2+}$ influx when the cell membrane is depolarized, is thought to encourage synaptic change by selectively increasing the dendritic Ca$^{2+}$ concentration of active synapses. Post-synaptic Ca$^{2+}$ chelators prevent long-term potentiation (LTP), whereas releasing caged Ca$^{2+}$ can mimic the effects of LTP (Lynch, Larson et al. 1983; Malenka, Kauer et al. 1988; Yang, Tang et al. 1999).

Interestingly, low or modest increases in Ca$^{2+}$ concentrations protracted in time are associated with decreases in synaptic efficacy (Yang, Tang et al. 1999; Mizuno, Kanazawa et al. 2001). In fact, when calcium chelators are applied to neocortical pyramidal cells, stimulation that ordinarily produces LTP can actually produce long-term depression (Hansel, Artola et al. 1996). Thus, a critical concentration of Ca$^{2+}$ must be exceeded for synaptic strengthening, otherwise, the synapse may actually weaken. When NMDA receptors are blocked, Ca$^{2+}$ entry may be decreased, but not prevented, leading to depotentiation of active synapses. If the blocker is present while a trace is being reactivated, specifically those synapses involved in the trace may be weakened, leading to decay of the memory trace. Little is known about whether Ca$^{2+}$ concentrations in vivo under NMDA receptor blockade would place active synapses within the range for synaptic weakening. Nevertheless, the implication is that a mechanism theoretically
associated with memory formation may have been used as a tool to selectively remove a
memory trace.

Amnestic effects have also been seen when recall trials were followed by β-
adrenergic antagonists (Lewis, Bregman et al. 1972; Lewis and Bregman 1973; Bregman,
Nicholas et al. 1976). In fact, the effective time course is the same for consolidation and
reconsolidation: between one and five hours after recall (Sara, Roullet et al. 1999). Beta
adrenergic receptors are thought to activate cAMP pathways that can lead to the synthesis
of proteins important in synaptic plasticity (Mayford, Abel et al. 1995). Thus β-
adrenergic antagonists may act to disrupt a similar pathway as NMDA-receptor blockers,
but in a later phase of plasticity, selectively decreasing the strength of the reactivating
memory trace. If these cases are interpreted correctly, the implications are startling. It
may be possible to erase specific memories simply by allowing them to reactivate under
the influence of certain drugs.

If the hypothesized mechanism is correct, this final observation could even be
used as a tool to investigate the relationship between LTP and learning. In describing a
hypothetical experiment that would determine whether LTP is a learning mechanism, one
investigator states,
"The key to this particular fictitious experiment is a new "designer drug" agent X, that selectively erases LTP in an activity-dependent manner...In the presence of agent X, the stimulation that normally induces LTP should result in the erasure of any pre-existing LTP, without affecting either any baseline component of transmission or the LTP on an orthogonal set of synapses that remain unstimulated while agent X is present. Having established the physiological selectivity of agent X, the complementary behavioral experiment would involve training the animal on two discrimination problems (e.g. two different spatial memory tasks). After learning is complete, the animal is given agent X and then re-exposed to one of the problems. Performance on this problem should deteriorate rapidly. After the drug has worn off, the animal is retested on both tasks. A persistent performance deficit is observed on the experimental task, but perfect retention is found on the control task that was not attempted in the presence of agent X. Retraining on the experimental task in the absence of agent X should proceed as in naive rats." (Barnes 1995, p.754).

Part of the fictitious behavioral test has already been conducted (Przybyslawski and Sara 1997). In theory, the secret to the effectiveness of MK-801 or β-adrenergic antagonists is not the mere presence of the compound, but that memory trace reactivation selects the appropriate cells for memory erasure.

Other researchers have suggested the selective weakening of a memory trace could be useful in treating phobias or traumatic memories (Nadel and Land 2000).

In sum, the predictions of trace replay theory suggest reactivation in the midst of extraneous activity, such as that which might follow ECS and lesions, can lead to inappropriate recall (Figure 4.2A). But even when reactivation is unhindered, the prevention of ensuing synaptic plasticity selective to some, but not all, of the memory trace also leads to inappropriate recall (Figure 4.2B). In fact, the neural specificity of memory trace reactivation that may enable the selective strengthening of synapses could
be used in conjunction with certain drugs to selectively erase a memory trace (Figure 4.2 D).

The predictions above are, at present, pure speculation, given the utter absence of evidence suggesting memory trace reactivation leads to synaptic plasticity. The experiment described in the present dissertation provides the only empirical evidence that neocortical neurons reflecting a widely distributed memory trace generate the kind of activity necessary for Hebbian plasticity after an event. Although this is a necessary starting point, numerous experiments will need to follow up on the predictions of trace replay theory, hopefully leading to better accounts of the dynamics of memory formation in the brain.
4.4 CONCLUSION

Trace replay theory accounts for a wide variety of characteristics of memory consolidation. If 'offline' periods are prevented after learning, memory is impaired. Learning material with intervening breaks is more effective than learning in one continuous block, i.e. spaced learning is better than massed learning. Repetition, which may be viewed as conscious or volitional trace replay, produces better recall. Proper hippocampal function is critical for recent memories, but its importance wanes as memories age. Even the peculiar impairments following reconsolidation procedures may be explained by trace replay theory. Yet, only a decade ago, there was no evidence that neural ensembles representing an event reactivated afterwards – the fundamental basis of trace replay theory.

At present, there is emerging evidence that memory traces reactivate, and do so in ways consistent with trace replay theory. Reactivation occurs during 'offline' periods of quiet wakefulness and sleep that is characterized by global synchronous oscillations. Reactivation is strongest immediately after an event, but can occur even following intervening events. Reactivation is seen in the hippocampus, even for familiar events, indicating continued hippocampal activity for remote memories. This is true in both adult and aged rats, though aged rats show reduced preservation of sequences during memory trace reactivation. Rats with stronger sequence preservation tend to show better memory recall, albeit in a separate task. Finally, reactivation occurs together in hippocampal and parietal cortex ensembles.
The present study revealed the first evidence that primates also show memory trace reactivation. As in the rat hippocampus, the primate neocortex shows memory trace reactivation during behavioral periods of sleep and quiet wakefulness. Reactivation among these neocortical ensembles also diminishes with time after an experience, but can be observed in rest preceding familiar tasks.

More importantly, a critical tenet of trace replay theory was supported by the observation of coordinated replay of distributed neocortical ensembles reflecting a preceding event. This effect included the preservation of temporal order among reactivating ensembles, even across hemispheres. Despite the coordinated replay among three different brain regions, the dorsal prefrontal cortex did not show signs of memory trace reactivation. More experiments will be necessary to explain this interesting pattern of results.

Although the relationship between memory trace reactivation and memory consolidation remains to be clarified, the results described in this dissertation are a critical prerequisite for reactivation to function as a mechanism for memory consolidation. One major avenue of research would explore the connection between memory trace reactivation and plasticity in reactivated circuits. Another issue to investigate is the relationship between the amount of reactivation and subsequent memory strength. These topics will surely lead to more detailed studies of the neural mechanisms of memory formation. Yet even at this early stage of exploration, we now
have some idea of why "gaps" are so important in learning: memories begin when the inputs end.
REFERENCES


De Renzi, E. and F. Lucchelli (1993). "Dense retrograde amnesia, intact learning
capability and abnormal forgetting rate: a consolidation deficit?" Cortex 29(3):
449-66.

Debiec, J., J. E. LeDoux, et al. (2002). "Cellular and systems reconsolidation in the


Destexhe, A., D. Contreras, et al. (1999). "Spatiotemporal analysis of local field
potentials and unit discharges in cat cerebral cortex during natural wake and sleep

single neurons, and large-scale networks organize sleep oscillations.

Comparative Physiology and Psychology 42: 32-34.

linking in the visual cortex? Multiple electrode and correlation analyses in the

Ekstrand, B. R. (1972). To sleep perchance to dream (about why we forget). Human
memory: Festschrift in honor of Benton J Underwood. C. P. Duncan and L.

during and after visual stimulus association learning." Journal of Neuroscience
19(23): 10404-16.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed hierarchical processing in the


