MEMORY CONSOLIDATION

by

STEVEN SHOFNER

A THESIS

Presented to the Department of Psychology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

September 2021

THESIS APPROVAL PAGE

Student: Steven Shofner

Title: Memory Consolidation

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Psychology by:

Don Tucker	Chairperson
Michael Posner	Member
Phan Luu	Member

and

Andrew Karduna	Interim Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded September 2021

THESIS ABSTRACT

Steven Shofner

Master of Science

Department of Psychology

September 2021

Title: Memory Consolidation

This proposal synthesizes the work of other researchers (Hawkins et al. 2019) (Kim, Gulati, and Ganguly 2019) (Wei et al. 2018) (Tononi and Cirelli 2006) (Grossberg 2013) (Li et al. 2017) into a framework of memory consolidation by temporally ordered progression through sleep stages.

Drawing on the memory model of (Hawkins et al. 2019), wherein cortical columns receive limbifugal predictions and learn a decomposition into individual sequential expectations learned by recruiting newly developed synapses to dendrites of pyramidal cells that receive lateral input from nearby neurons which modulate the depolarization of the cell in order to allow neurons associated with the expectation to fire first. During sleep, the weights of these newly recruited neurons are adjusted ensuring salient memories are preserved and integrated with preexistent memories.

If sleep participates in memory consolidation as outlined, all stages of sleep must be essential to the successful consolidation and integration of new memories.

CURRICULUM VITAE

NAME OF AUTHOR: Steven Shofner

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Boise State University, Boise
Northwest Nazarene University, Nampa

DEGREES AWARDED:

Master of Science, Psychology, 2021, University of Oregon
Bachelor of Science, Computer Science, 2021, Boise State University

AREAS OF SPECIAL INTEREST:

Memory Consolidation
Neural Networks / Machine Learning

PROFESSIONAL EXPERIENCE:

Senior Software Engineer, Brain Electrophysiology Lab, 1 year

Senior Software Engineer, Electrical Geodesics, Inc., 6 years

ACKNOWLEDGMENTS

I wish to express sincere appreciation to all those who have assisted with the development of this thesis and drafting of this manuscript. My committee chair, Professor Don Tucker has provided considerable guidance and criticism throughout the process while offering consistent support. His probing questions and keen insights were essential in the formulation and refinement of these ideas. Professor Michael Posner provided invaluable feedback and suggestions. These suggestions are largely responsible for the inclusion of testable claims and the experiment proposed to test them. Phan Luu, PhD shared significant insights about cortical organization and feedback about the feasibility of some proposed descriptions here.

This thesis is dedicated to my mother, Eileen Shofner. In sometimes difficult times the home she made for her children was a refuge of stability. Her unshakable belief in me gave me the confidence to attempt challenging things while her work ethic provided an example of how to complete them.

To my wife, Ashley Shofner, whose pursuit of her dreams and determination to make them real expanded my conception of what was possible. Without her support this thesis would not exist.

And to my daughter, Olivia Shofner, whose curiosity and humor are a constant source of joy and inspiration.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Memory in neural networks both artificial and biological relies on the modification of synaptic weights in response to events sensed and perceived. Artificial neural networks are modeled to varying degrees of accuracy on biological neural networks and yet struggle to achieve some of the characteristic robustness of these networks commonly falling prey to catastrophic forgetting, where newly learned associations result in the loss of previous learning, and require significant training to learn basic percepts. While research and development with artificial neural networks is ongoing, these remain challenging deficits to address.

The thesis described here attempts to explain how biological neural networks achieve these contradictory goals. A system capable of one-shot learning, like biological neural networks, must exhibit significant capacity for plasticity, but this very plasticity is the means by which catastrophic forgetting occurs. How do brains accomplish this feat?

This thesis proposes that learning occurs simultaneously in the limbic core and neocortical circuits creating associations between higher-level and lower-level representations. The learning occurs on-line and in-place, such that the cortical synapses and neurons involved in the representations learn the associations directly rather than having this information broadcast in an off-line fashion. One of the functions of sleep, then, is the off-line consolidation of salient memories in a way that ensures maximal separation between representations and allows for reorganization.

What Is Memory?

It is important to define what is meant by memory before proceeding. There are a number of ways of understanding what memory is, ranging from colloquial to engineering specifications. Conversationally, we usually talk about memory as our ability to recall previously encountered events or information while in computer engineering, memory can be implemented as circuits of transistors, electrons stored in trapping layers, or any number of other technologies facilitating the persistence of information.

While related, these descriptions of memory are not what is described here. Instead, memory as represented by this model is sequential: we learn to expect sequences of inputs. These sequences are motivated, ordered, and encode a temporal dimension.

Memory is Motivated

Memory is invoked via limbifugal predictions driven by internal models of the relationship of the organism to the environment. Reading a line of text, for instance, requires sweeping the eyes over the page or screen and as the eyes move over the line information about the current state is encoded in compressed predictive representations which are propagated to relevant cortical columns. As long as the expected sequence is reflected by input the model generating the prediction is allowed to proceed without adjustment. Importantly, the next expected input is derived from the motivated action of the eyes over the line of text; reading, in this way, is similar to the whisking action of a mouses' whiskers or of moving the fingertip over a surface: without the motive action (which generates a sequence of inputs), very little can be perceived.

Memory is Ordered

Two sets of sequences composed of the same states, but with different orders may very likely represent very different information. For example, ***cat*** and ***act*** are composed of the same letters, but bear almost no common meaning. If memory represents sequences of expected states, these expected states must be disambiguated and ordered very specifically to convey accurate meaning.

Memory Encodes Temporal Sequences

Memory is not static and encodes sequences of ordered states as described above. Any individual state on its own is ambiguous and insufficient to describe perception which might rise to the level of consciousness. This thesis posits all memory, necessarily, represents this ordered and motivated sequence of states through time. Time, here, may be an abstract concept and does not necessarily equate to what is measured by clocks, but our memory for even abstract concepts encodes a temporal (though likely non-linear) dimension that allows us to predict sequences of states.

Cortical Columns as Simple Computational Units

The architecture of the neocortex displays a degree of uniformity across regions that suggests the repeated implementation of a common circuit throughout cortex. Others have proposed that these circuits which cortical columns are composed of may perform some kind of canonical computational operation capable of deriving useful information from a variety of inputs. (Hawkins et al. 2019), for example proposes that these columns (or mini-columns) learn full models of many different objects and are capable of predicting limbipetal inputs for many receptive fields (RFs) given an allocentric location specification which they suggest is a likely interpretation of the limbifugal prediction.

While I draw heavily on the Hawkins, et. al. model, this interpretation is more general. Rather than specifying allocentric locations, I suggest that limbfugal predictions represent the current state of the relevant part of the model propagated from the limbic core to cortical circuits. This expectation is a compressed representation of the sparse encoding of several sequential limbipetal inputs to the column. The canonical operation of the column is the decomposition of the compressed representation to the expected sequence.

Limbifugal prediction

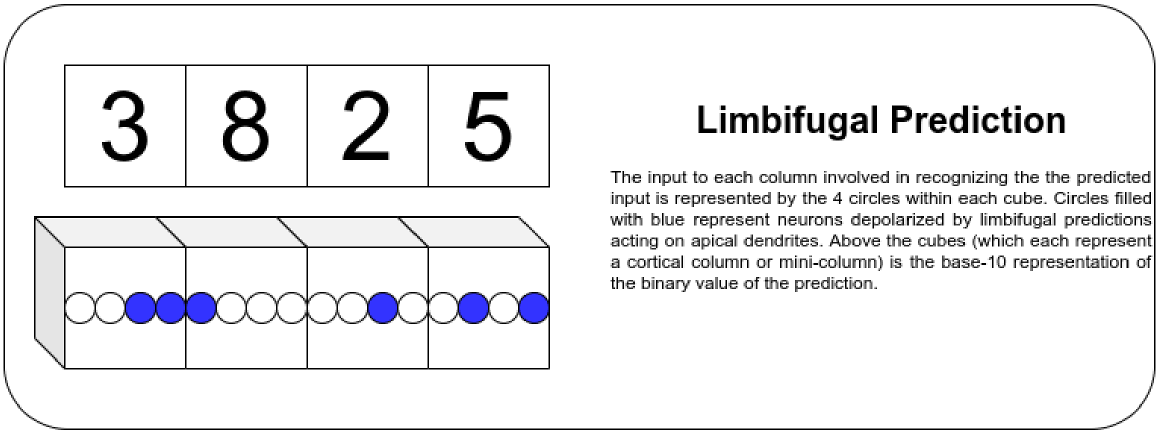
If it is true that cortical columns implement a repeated architecture capable of simple computation, the number and operation of these columns suggest a massively parallel system which must be loosely coupled to what seems to be the more serial operations of consciousness. Predictive coding theorists have suggested that higher level areas or the limbic core are actively modeling expected inputs and specifying these to cortical neurons which are tasked with comparing the expectation with the actual limbipetal input. The cortical feedback then must steer the model, correcting it as it deviates from the limbipetal input in a kind of Kalman filter.

As in the model proposed by Hawkins, et al. I suggest that this prediction is the temporally compressed expected sequence of inputs. The limbic core represents these sequential inputs at a different temporal scale, and part of the job of the cortex is to decompose the prediction to the correct sequence.

Decomposition

The limbifugal prediction, being temporally compressed can be thought of as a vector of binary values consisting of the composition of the expected limbipetal input patterns. The expectation can be treated as binary values because within the context of this model the neurons carrying the representation are activated (meaning they are experiencing rapid spiking) or not (no or very low level spiking activity).

The prediction to each column, then, being a vector of binary values can be represented by a real valued base-10 integer by a simple transformation from base-2 as shown in Figure 1. The mapping from base-2 to base-10 is merely performed here for the sake of convenience, but simplifies the description of operations of cortical columns. To that end, the value of the prediction is, as stated, a composition of multiple sequential elements and part of the work of the cortical column is to learn this decomposition.



The input to each column involved in recognizing the the predicted input is represented by the 4 circles within each cube. Circles filled with blue represent neurons depolarized by limbifugal predictions acting on apical dendrites. Above the cubes (which each represent a cortical column or mini-column) is the base-10 representation of the binary value of the prediction.

Figure 1: Limbifugal prediction.

After receiving the prediction, the associated neurons will be in a depolarized state allowing them to fire with fewer limbipetal inputs, and thus faster. Lateral connections from the activated neurons to neighboring neurons in the limbipetal output layer then inhibit other nearby neurons, ensuring the sparsity of the output encoding. In Figure 1, the predicted value of ***3825*** can be decomposed into an arbitrary (because of the size of the vectors - 4 neurons - there are only a few decompositions, but in actual cortical columns the number might be much larger) number of interstitial states. In this example, say ***3825*** should be decomposed into the components ***1024***, ***2800***, and ***0001*** as shown in Figure 2.

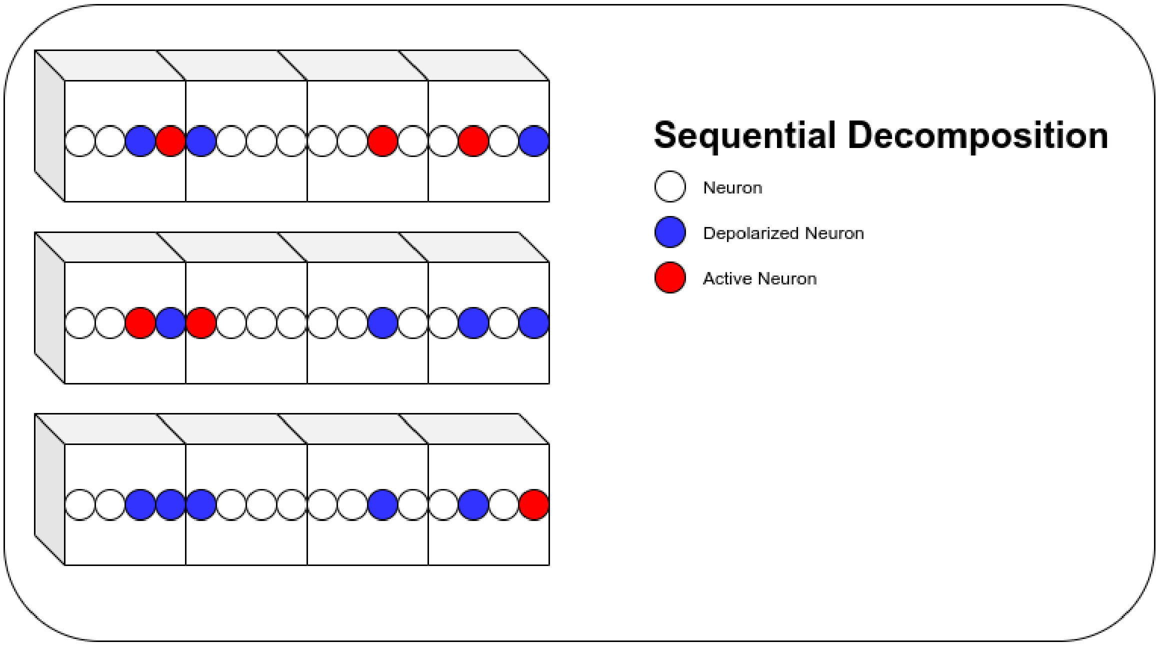


Figure 2: Sequential decomposition.

To accomplish this decomposition, neurons leverage distal dendritic synapses to predict the next state from the current state. Given the prediction of Figure 1 with the decomposition of Figure 2, if the relevant columns receive limbipetal input equivalent to ***1024*** (the first predicted state), the distal dendritic synapses of the second state will be activated by the activation of the limbipetal outputs of the first state, resulting in relative depolarization of the neurons of state 2 in anticipation of the next expected state as shown in Figure 3.

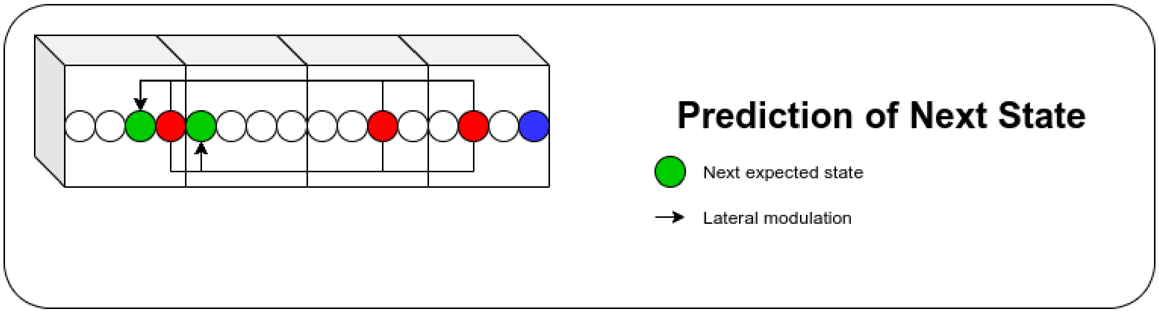


Figure 3: Prediction of next state.

What Happens if the Input Does Not Match the Prediction?

A mismatch between prediction and input can result in a reset which may update the model generating predictions. In this case the prediction would subsequently be updated to better match the current state. If no appropriate prediction can be found, a new percept may be initialized in the limbic core and the network may learn to associate the current sequence of inputs with this representation.

How Does Memory Arise in Neural Networks?

In conventional artificial neural networks (ANNs) like those commonly used to perform computation today, the network typically maintains a set of static synaptic weights specifying the response of a unit to the output of antecedent units. During training these weights are trained through back-propagation, typically by pursuit of global minima using some kind of gradient descent optimization. Gradient descent attempts to estimate the local gradient and predict the vector with the greatest negative vertical orientation, while back propagation takes the error in output of the network versus this vector and pushes it to the units responsible for the prediction. Over many iterations of this process the synaptic weights are adjusted so that the difference between the output of the network and the ground truth is minimized. Once training is complete, learning is typically disabled to protect the performance of the system on the current inputs/outputs and avoid overfitting. As described previously, this approach is both slow (it generally requires many iterations of training over the same inputs to reach convergence) and susceptible to catastrophic forgetting (taking a network trained on one set of data then training it over a second different set may degrade the performance of the network on the original).

While it is important to note that the ANNs described here are only loosely modeled on their biological counterpart and suffer from operational rigidity and susceptibility to catastrophic forgetting the representation of memory that powers them is derived from much of what we know about the function of cortical neurons. As in ANNs, cortical representation is distributed throughout the network in synaptic weights that allow assemblies to recognize learned inputs and to learn to recognize new inputs. Despite the generally uncanny effectiveness of these networks they lack the mechanism to model the input generators (for example, an image of a tree is generated by the tree which exists separate from the image) that is commonly believed to underlie predictive coding (see (Hawkins, Ahmad, and Cui 2017), for example). This lack of a model and the consequent predictions it facilitates may help explain why researchers are regularly able to compose images that look like nothing but noise and yet are recognized as real objects with high confidence by these networks (Nguyen, Yosinski, and Clune 2015).

Where Are Memories Made?

Cortical neurons in V1 have been shown to respond to inputs in their receptive fields. Unless these responses arise due to cortical circuits grown during maturation independent of the inputs to cortex, they must be learned. In fact, it has been shown in multiple studies carefully controlling for the inputs to primary sensory areas that without this input the response of neurons can be completely divorced from the usual response (Tanaka et al. 2009). The demonstrated effects of sensory deprivation demonstrates plasticity in these circuits facilitating adaptation to environments with particular (and peculiar, here) characteristics.

Predictive coding posits that higher cortical areas propagate predictions to lower cortical areas. A prediction implies knowledge of the expected input and this knowledge can only reasonable originate from the previous experience of these inputs. As the limbic core sits atop the cortical hierarchy, it is reasonable to expect that neurons in the limbic core also learn representations of these inputs. An example of these representations might be the grid and place cells in the thalamic reticular nucleus (TRN) that encode information about locations with a frame of reference external to the organism.

Why do Some Memories Persist?

So, memories must be encoded in the synaptic weights of neurons in the neocortex and limbic core. These weights affect the response of a cell to sets of inputs from antecedents and their weights have been suggested to be adjusted by a Hebbian learning rule where a presynaptic activation followed closely by a post synaptic activation results in a positive adjustment to the weight attributed to the synapse in question.

Some of these changes will persist for the life of the organism while others are ephemeral and will be lost in days if not sooner. Memory consolidation is the purported process of stabilizing these long-term memories. Through this process, certain memories are actively selected for consolidation while others are weakened and eventually lost.

Synaptic homeostasis theory posits that the effects of sleep on memory consolidation and forgetting are the passive effect of homeostatic processes (Tononi and Cirelli 2006). Alternatively, a number of researchers have presented evidence that both consolidation and forgetting are active processes (Kim, Gulati, and Ganguly 2019) (Li et al. 2017). while others have suggested consolidation is an active process and forgetting is a passive effect of competition driven by slow oscillations (Wei et al. 2018).

The putative action of morphological operations as described here participate in both active consolidation and active weakening or memories.

CHAPTER II

THESIS

This thesis describes a purported mechanism by which memory consolidation occurs during sleep and how the brain achieves fast on-line and in-place learning while avoiding cases of catastrophic forgetting. This process involves learning to decompose limbifugal predictions to the ordered sequence of expected inputs, how the generating model is updated by unexpected inputs, how new expectations are learned when no matching expectation can be found, and how these memories are actively pruned, maintained, and reorganized to facilitate integration with existing memories.

Intuitively, we recognize our ability to remember new information due to our usage of the facilities we have for doing so. Hebb proposed a framework describing a means by which synaptic weights can be modified by spike timing dependent plasticity (STDP) such that subsequent exposures to a common input benefit from the information encoded in these weights that has been successfully applied to artificial neural network (ANN) models and been shown to perform optimization of synaptic weights akin to gradient descent.

At the same time, as noted by (Tononi and Cirelli 2006), some evidence suggests that the overall level of activation grows over the course of waking and to ensure this process does not lead to the metabolic rate growing without bound some process must exist to maintain homeostasis and return global activation to a baseline level. This rescaling (or rebalancing) of synaptic weights must be able to protect learned memory that might not have been accessed recently while supporting the consolidation of new memories in order to avoid the potential of catastrophic forgetting while still allowing new information to be learned. Tononi and Cirrelli cite (Braun et al. 1997) as evidence for the purported metabolic effects of synaptic downscaling as posited by their theory, but other researchers have failed to find evidence for of lower metabolic rates following sleep than prior to it (Shannon et al. 2013).

Some new memories last for only a short period of time while others persist and can be accessed for years after the fact, so some process must be responsible for differentially protecting and pruning these new memories while preserving those previously consolidated.

Consolidation also seems to be responsible for extracting the gist of newly formed memories (Lewis and Durrant 2011) while simultaneously reorganizing them (Landmann et al. 2015). These processes taken together can facilitate the application of newly learned relationships to previously encountered problems.

To understand how sleep contributes to memory consolidation, it is important to understand how memories are initially encoded by waking conscious interaction with the environment. Once these new labile memories are formed, it becomes important to stabilize those which are important while pruning those that are not sufficiently salient to the sleeper: this is the work of sleep and the subject of most of the rest of this thesis.

The Operations of Sleep

As we sleep we move through an orderly progression of stages marked by characteristic changes in cortical dynamics as measured by EEG. It is the proposal of this thesis (drawing on the work of other theoreticians and researchers) that this progression is necessary and the order is essential to the consolidation and integration of newly learned memories. In brief, Non-REM stage 2 (N2), stage 3 (N3), and REM are directly involved in the long-term maintenance of memories which is achieved by the interaction of the different dynamics of each stage as described below.

N2 Sleep

Non-REM stage 2 is marked by the abundance of slow oscillations, δ-waves, and sleep spindles in the EEG of sleeping humans. Slow oscillations are the result of cycles of widespread cortical depolarization and hyperpolarization that seems to be mediated by subcortical - cortical interactions capable of eliciting widespread synchronization (Bernardi et al. 2018). δ-waves (also referred to as Type II slow waves), on the other hand are more focal, lower amplitude, and exhibit lower levels of synchronization. Intermixed with these are higher frequency bursts of activity known as sleep spindles.

Spindles, Slow Oscillations (SOs), δ-Waves,And The Adjustment of Cortical Synaptic Weights

In (Kim, Gulati, and Ganguly 2019), the researchers were able to demonstrate evidence supporting a model where the relative nesting of the sleep spindle to the SO vs. the δ-wave selected between strengthening and weakening of newly learned memories. Researchers have suggested that sleep spindles are involved in the reactivation of neocortical memories (Bergmann et al. 2012). If we take these two results together we might conclude that during N2 sleep, thalamocortically driven sleep spindles reactivate labile memories and that the relative nesting of these reactivations to SOs or δ-waves serve to strengthen or weaken the memory.

Drawing on the memory model described previously (and largely developed in (Hawkins et al. 2019)), limbipetal inputs of a cortical column received sequential excitatory inputs from lower level hierarchical areas. A "hidden layer" between the limbipetal input and limbipetal output learns a mapping from the input to the prediction. In this model, an unexpected input can cause a reset (like that described by Adaptive Resonance Theory (ART) (Grossberg 1987)). This reset will trigger a search for a new model which might end up finding no currently suitable model. If no suitable model can be found, the limbifugal prediction can reflect this by specifying a previously unused expectation. The hidden layers of the cortical column, then, must find a way to map the sequence of inputs to this new model. To do so, newly grown dendritic synapses are enlisted and the synaptic weights of these synapses are adjusted to map from the limbipetal input sequence to the limbifugal prediction.

New dendritic synapses are continually formed in the neocortex and are generally short lived. These synapses are leveraged in the learning of new mappings in the hidden layers of cortical columns as described above. During N2 sleep the weights of these new synapses are actively strengthened and weakened through reactivation driven by sleep spindles nested to the ongoing slow oscillations and δ-waves, a process separate from their initial growth (which is not considered here). Due to lateral inhibition of nearby pyramidal neurons during reactivation, the gradient of the synaptic weights across the layer activated is sharpened by the sleep spindle driven replay process when the spindle is nested to the slow oscillation (the weights of synapses to neurons participating in the activation are strengthened, but the weights of synapses to neurons which are inhibited are reduced)

Memory at the level of Cortical Columns

The previous discussion covered memory at the sub-cortical column level, at the level of cortical columnar response to limbipetal inputs, the neocortex exhibits systematic variation across areas. For instance, in area V1, the neocortex is organized retinotopically such that the distribution of receptive fields mimics the distribution of receptors in the eyes. This kind of systematicity has been shown in a number of cortical areas and might constitute an organizing principle for the cortex more generally.

Throughout the neocortex, in fact, we see organizational patterns that seem to reflect intuitive models of the principal components; the ***what*** and ***where*** pathways are an intuitively neat way to take an intractable problem and describe it in 2 primary dimensions. As we analyze the response of different cortical areas, we may discover a topographic distribution representative of the principal components of the problem being solved.

It is within this topology that I suggest the previously described operational characteristics of Non-REM sleep modulate columnar response in a fashion similar to the morphological operation applied to images within the field of computer vision.

Image Morphological Operations

In computer vision, morphological operations are context-aware functions that adjust pixel values in two dimensional images. To begin to understand why this might provide some insight into how the brain modulates memory during N2 sleep, consider that the neocortex is also organized in a pseudo two dimensional fashion with the third dimension being the cortical columns, but which might be thought of, here, as singular computational units. If, as suggested, cortical organization is systematic (and regardless of whether this systematicity reflects primary components) in all areas new memories must also be organized by this principal. Further, if the operation of a cortical column to a particular input can be abstracted such that we can treat them as atomic units memory consolidation may, in part, consist of the regularization of the responses of these columns.

Image morphological operation like dilation and erosion can regularize the pixels of input images which, if we accept the preceding, might serve as an analogue to cortical columns and may provide a framework for high level descriptions of the active strengthening and weakening of N2 sleep spindles previously described. The effects of applying morphological operations to images include filling holes, smoothing edges, and eliminating noise in object segmentations.

(Klinzing, Niethard, and Born 2019) propose that sleep contributes to the extraction of gist-like representations of the initial encoding of memories. Modeling the effects of spindle driven memory reactivation during sleep with morphological dilation and erosion suggests how this sleep might contribute to gist extraction.

Morphological Operations

Morphological operations in image analysis utilize a structuring element to select a set of neighboring pixels to consider during an update. In dilation, a considered pixel value is updated to match the pixel value of the highest neighboring pixel. In erosion, the reverse operation is performed, where the pixel under consideration is updated with the value of the lowest pixel in the neighborhood. These operations are performed over the whole image.

The result of a dilation is that high pixel values tend to expand outward, while for erosion they contract.

Morphological Operations on Cortical Columns

During N2 sleep, limbifugal predictions modulate the active dendrites of neurons within cortical columns. This modulation typically does not lead to activation, but because of the global state of the cortex (significant depolarization during the UP phase of SOs, for instance) during this period it can lead to firing. When this occurs synchronously with an SO UP state and the firing corresponds to the initial expected sequential input, lateral excitatory connections to the next predicted state are modulated and the global state leads to firing correlated to the next expected input state, strengthening these lateral modulatory connections and the associated memory. When the firing is instead nested to the lower amplitude and more local δ-waves, the firing does not exert sufficient modulatory influence to elicit the sequential activation, and failing this, the synaptic weights of the associated memory are actually weakened.

The limbifugal prediction, here, functions as the structuring element. The prediction selects the labile memory across an assortment of columns where the memory was originally learned. By nesting this activation to either the SO UP state or the δ-wave, something akin to dilation or erosion occurs.

While it is relatively simple to visualize this dilation and erosion operating on the retinotopically oriented columns of area V1 envisioning this operation over neocortical columns that receive input which is not so clearly delineated in two dimensions is challenging. Here we return to the idea of the organization of cortex according to principle components. It is a central claim of this thesis that cortical columns are organized over two dimensions in a systematic fashion, and one means of doing this for higher dimensional problems is by selecting the two principle components and orienting the mapping of these components into something approximating orthogonality over the surface of the neocortex.

Morphological Dilation as an Analogue of the Operation of Spindles Nested to SO UP States

By reinforcing columnar response to a stimulus through the nesting of sleep spindles to SO upstates, the neocortex performs a context-aware expansion of the regional response that can be approximated by morphological dilation.

Morphological Erosion as an Analogue of the Operation of Spindles Nested to δ-Waves

By weakening columnar response to a stimulus through the nesting of sleep spindles to δ-waves, the neocortex performs a context-aware reduction of the regional response that can be approximated by morphological erosion.

Modeling Gist Extraction with Morphological Operation

Why might it be appropriate to model the extraction of gist-like representations during sleeping with morphological operations? The application of dilation and erosion in image analysis serves similar purposes: smoothing object segmentation edges, separating object segmentations, filling holes in object segmentations, etc. The result of of erosion and dilation on the segmentation of an object is an approximation of the input with extraneous details omitted; conceptually, these operations function as a kind of low pass filter.

To understand how this might apply to cortical columns, consider the retinotopic organization of columns in area V1. Area V1 RFs are distributed in a systematic fashion determined by retinal organization.

Figure 4 is a well known image demonstrating this retinotopic organization of area V1 of macaque cortex (Tootell et al., 1982). The animal was injected with a dye, shown the image on the left, then the cortical tissue was imaged. The dye was drawn to the pattern of activation and a clear retinotopic mapping can be seen in the image on the right.

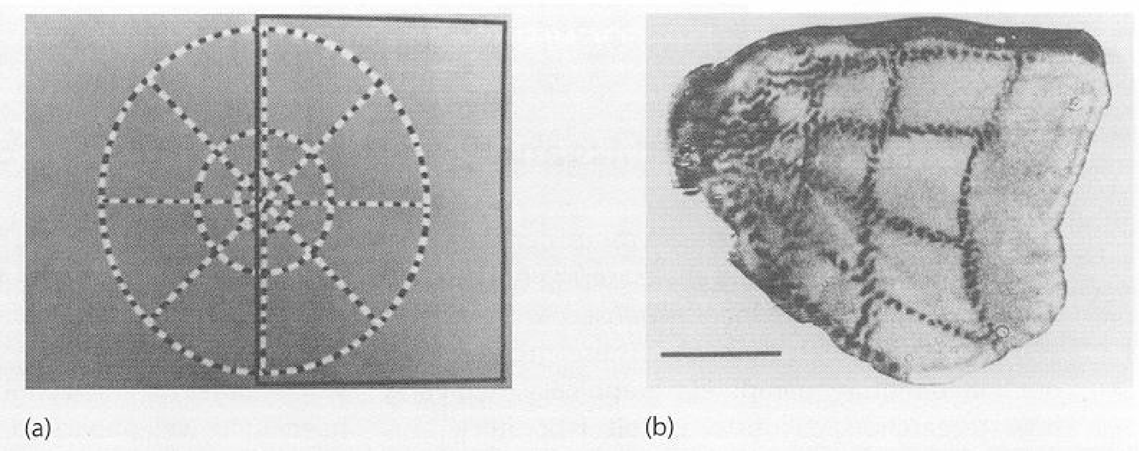


Figure 4: Illustration demonstrating retinotopic organization of macaque visual cortex.

If dilation and erosion are applied to this image which demonstrates the retinotopic organization of RFs, we can see that "gist" of the input pattern remains, but high frequency variations have been reduced.Figure 5, then, shows the effect of applying morphological dilation and erosion to the image from Figure 4.

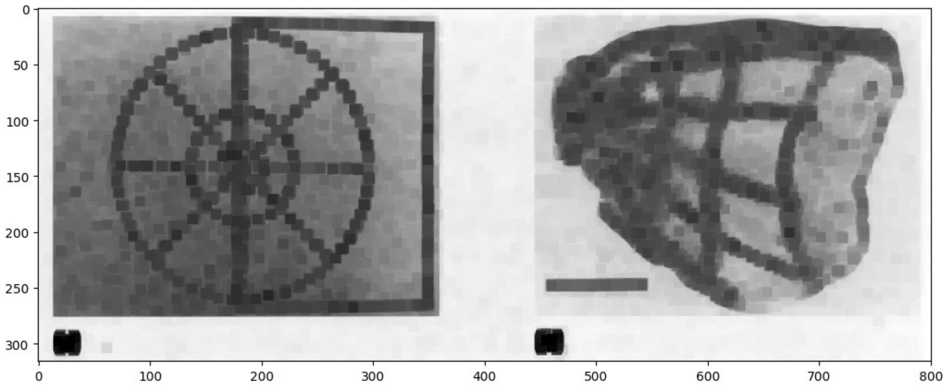


Figure 5: Illustration of the effect of 2 dilations followed by 2 erosions on the original retinotopic organization figure.

If the rest of neocortex shares this kind of systematicity to cortical columnar representation, modeling the effects arising from the action of sleep spindles with morphological operations might help to describe how these processes of consolidation accomplish gist extraction.

Non-REM Stage 3 Slow Oscillations and Synaptic Homeostasis

Following the active adjustment of cortical weights during N2 sleep, the unadorned slow oscillations of N3 perform the homeostatic task of renormalizing synaptic weights to achieve a global activational baseline adjustment. Because these SOs are not associated with limbifugal predictions in the same way spindles suggest the slow oscillations of N2 might be, no morphological control is exerted at this time and the adjustment to weights is more normally distributed and continues until reaching the baseline.

Because N3 typically follows N2, the overall outcome of the composition of the two stages is (over the course of several iterations of the stages throughout the night) a non-linear weight adjustment, allowing homeostasis to be preserved while protecting labile but salient memories.

REM Sleep

REM sleep has been suggested to selectively maintain or prune newly formed synapses (Li et al. 2017). The underlying memory model previously described, where new synaptic connections are enlisted in the formation of memory, requires a process such as this. By adjusting the weights of newly formed memories in the N2->N3 stages, these synapses are made ready for a process capable of strengthening and weakening or down-selecting the synapses associated with newly formed memories.

CHAPTER III

PREDICTIONS

Some aspects of this thesis are metaphorical in nature: demonstrating that the adjustment of cortical synaptic weights during N2 sleep approximates morphological erosion and dilation would be difficult to demonstrate, but at a higher level some of the claims may be falsifiable.

The Purported Active Adjustment of Synaptic Weights During N2 Sleep Supports Gist Extraction

N2 sleep, here, is claimed to support gist extraction by adjusting synaptic weights in a fashion similar to the image morphological operations. It would be difficult to demonstrate that weight adjustment in N2 sleep operates in this fashion, but by selectively interrupting N2 sleep or modulating the nesting of sleep spindles as described in (Kim, Gulati, and Ganguly 2019) gist extraction could be affected.

In the absence of N2 sleep, gist extraction should be minimized.

Materials and Methods

In (Djonlagic et al. 2009), the researchers report on an experiment that demonstrated a significant improvement on a probabilistic category learning task. The "Weather Prediction Task" presents subjects combinations of cards consisting of geometric shapes. The test can proceed in one of two modes: observational and feedback.

In the observational mode subjects are shown multiple combinations of cards along with the weather they predict (sun or rain) during training. The feedback mode asks subjects to make a prediction of the weather during training who are then provided feedback on the accuracy of their prediction. In either mode, testing presents the user with combinations of the cards used during training and asks them to predict the weather, but provides no feedback. Researchers have shown that the observational task engages the medial temporal lobe (MTL) and prefrontal cortex (PFC), while the feedback task engages the MTL and (later) the striatum.

In the sleep study, researchers asked the subjects to describe the model used to make their predictions and found significant differences between the no sleep and sleep groups. After sleeping, subjects were significantly more likely to attribute probabilities to individual cards versus assigning 0% and 100% probability values, demonstrating a more nuanced understanding of the underlying probabilities.

Drawing on these results, it might be possible to test the predicted link between the nesting of sleep spindles to slow oscillations and gist extraction.

Subjects

Subjects should be selected from a random pool of healthy adults and assigned to one of two groups; either a sham group to serve as a baseline or a treatment group. Individuals with neurological or sleep disorders will be excluded, as will those with a history of alcohol or narcotic abuse or regular use of sleep medication. For the 12 hours preceding the start of the initial session until the end of the final session subjects will be required to abstain from alcohol, caffeine, and other drugs known to affect sleep.

Experimental Design

All participants will have an initial sleep session prior to any training to collect EEG to serve as the individual baseline. After this initial session, subjects will complete the weather prediction task in the observational mode. The night following the performance of the task, subjects will again be monitored while sleeping, but the treatment group will also received transcranial electrical stimulation (TES) 180 degrees out of phase with the slow oscillations during all periods of N2 sleep. After sleeping, all subjects will be retested, and asked to describe their internal model of the rules of the game.

The observational task is ideal because of the engagement of the prefrontal cortex which is also a major source of sleep spindles. By targeting the PFC with an out of phase stimulation, the intent is to decouple spindle generation from slow oscillations by encouraging more spindle generation during SO DOWN states (which should, consequently, exhaust the relevant cortical cells making them less likely to fire during the UP state).

Statistical Analysis

Determination of the effectiveness of the decoupling is a major challenge, as it is difficult to recover EEG during TES. During the UP phase of the SO, though, no current will be injected and should allow for some collection of EEG. By comparing the number of spindles detected during these phases to the individual baseline, we can derive a ratio for both groups which should reflect greater in-phase spindle generation due to the effects of learning on spindle / SO nesting. If the treatment is successful in decoupling spindles from SO UP states, this should be reflected by a significantly lower ratio for the treatment group versus the sham group.

If the treatment is successful in reducing the nesting of spindles to SO UP states, individual pre-sleep / post-sleep scores will be calculated and the relative change in performance between groups over the two sessions will be compared.

Results

If the treatment is not shown to be effective in reducing the nesting of spindles to SO UP states, no conclusion can be drawn from the comparison of the sham group to the treatment group. However, if the treatment is successful, the results of this comparison might support or refute the prediction. If the treatment group performs significantly worse than the sham group in terms of relative improvement in the post-sleep results versus the pre-sleep results the prediction will be upheld. If the treatment group does not perform significantly worse than the sham group on the measure of relative performance post-sleep versus pre-sleep, the evidence would refute the prediction.

CHAPTER IV

CONCLUSION

This thesis attempts to describe a coherent framework for the consolidation of memory in sleep. In accordance with the purported operation of the three examined sleep stages (N2, N3, REM), it predicts that the order of the progression through these stages over the course of a period of sleep is essential to the successful consolidation of memory. Additionally it attributes to each of the stages unique contributions to the consolidation process.

N2 sleep is both protective of new and newly labile memories and involved in gist extraction because of the way the purported operations affect columnar responses. N3 sleep seems to be most important for homeostatic maintenance. Finally, N3 sleep protects or prunes the labile synapses thereby allowing for the reorganization, integration, and consolidation of new memories.

REFERENCES CITED

Bergmann, Til O, Matthias Mölle, Jens Diedrichs, Jan Born, and Hartwig R Siebner. 2012. “Sleep Spindle-Related Reactivation of Category-Specific Cortical Regions After Learning Face-Scene Associations.” *Neuroimage* 59 (3): 2733–42.

Bernardi, Giulio, Francesca Siclari, Giacomo Handjaras, Brady A Riedner, and Giulio Tononi. 2018. “Local and Widespread Slow Waves in Stable NREM Sleep: Evidence for Distinct Regulation Mechanisms.” *Frontiers in Human Neuroscience* 12: 248.

Braun, Allen R, TJ Balkin, NJ Wesenten, RE Carson, M Varga, Pl Baldwin, S Selbie, G Belenky, and P Herscovitch. 1997. “Regional Cerebral Blood Flow Throughout the Sleep-Wake Cycle. An H2 (15) o PET Study.” *Brain: A Journal of Neurology* 120 (7): 1173–97.

Djonlagic, Ina, Andrew Rosenfeld, Daphna Shohamy, Catherine Myers, Mark Gluck, and Robert Stickgold. 2009. “Sleep Enhances Category Learning.” *Learning & Memory* 16 (12): 751–55.

Grossberg, Stephen. 1987. “Competitive Learning: From Interactive Activation to Adaptive Resonance.” *Cognitive Science* 11 (1): 23–63.

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural networks*, *37*, 1-47.

Hawkins, Jeff, Subutai Ahmad, and Yuwei Cui. 2017. “A Theory of How Columns in the Neocortex Enable Learning the Structure of the World.” *Frontiers in Neural Circuits* 11: 81.

Hawkins, Jeff, Marcus Lewis, Mirko Klukas, Scott Purdy, and Subutai Ahmad. 2019. “A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex.” *Frontiers in Neural Circuits* 12: 121.

Kim, Jaekyung, Tanuj Gulati, and Karunesh Ganguly. 2019. “Competing Roles of Slow Oscillations and Delta Waves in Memory Consolidation versus Forgetting.” *Cell* 179 (2): 514–526.e13. https://doi.org/10.1016/j.cell.2019.08.040.

Klinzing, Jens G, Niels Niethard, and Jan Born. 2019. “Mechanisms of Systems Memory Consolidation During Sleep.” *Nature Neuroscience* 22 (10): 1598–1610.

Landmann, Nina, Marion Kuhn, Jonathan-Gabriel Maier, Kai Spiegelhalder, Chiara Baglioni, Lukas Frase, Dieter Riemann, Annette Sterr, and Christoph Nissen. 2015. “REM Sleep and Memory Reorganization: Potential Relevance for Psychiatry and Psychotherapy.” *Neurobiology of Learning and Memory* 122: 28–40.

Lewis, Penelope A, and Simon J Durrant. 2011. “Overlapping Memory Replay During Sleep Builds Cognitive Schemata.” *Trends in Cognitive Sciences* 15 (8): 343–51.

Li, Wei, Lei Ma, Guang Yang, and Wen-Biao Gan. 2017. “REM Sleep Selectively Prunes and Maintains New Synapses in Development and Learning.” *Nature Neuroscience* 20 (3): 427–37.

Nguyen, Anh, Jason Yosinski, and Jeff Clune. 2015. “Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 427–36.

Shannon, Benjamin John, Ronny A Dosenbach, Yi Su, Andrei G Vlassenko, Linda J Larson-Prior, Tracy S Nolan, Abraham Z Snyder, and Marcus E Raichle. 2013. “Morning-Evening Variation in Human Brain Metabolism and Memory Circuits.” *Journal of Neurophysiology* 109 (5): 1444–56.

Tanaka, Shigeru, Toshiki Tani, Jérôme Ribot, Kazunori O’Hashi, and Kazuyuki Imamura. 2009. “A Post-natal Critical Period for Orientation Plasticity in the Cat Visual Cortex.” *PLoS One* 4 (4): e5380.

Tononi, Giulio, and Chiara Cirelli. 2006. “Sleep function and synaptic homeostasis.” *Sleep Medicine Reviews* 10 (1): 49–62.

Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, *218*(4575), 902-904.

Wei, Y., Krishnan, G. P., Komarov, M., & Bazhenov, M. (2018). Differential roles of sleep spindles and sleep slow oscillations in memory consolidation. *PLoS computational biology*, *14*(7), e1006322.