UNIVERSITY OF CALIFORNIA, SAN DIEGO

A unified anatomical theory and computational model of cognitive information processing in the mammalian brain and the introduction of DNA reco codes

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in
Engineering Sciences (Mechanical Engineering)

by

Soren Solari

Committee in charge:

Professor Robert R. Bitmead, Chair
Professor Robert Hecht-Nielsen, Co-Chair
Professor Thomas Bewley
Professor Glenn Northcutt
Professor Virginia de Sa

2009

The dissertation of Soren Solari is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____
Co-chair

_____
Chair

University of California, San Diego

2009

Dedicated to my parents, Ray and Mieke, and my wife, Ruthi, for making me who I am.

I pass this way but once,
any good therefore that I can do,
or any kindness that I can show
to any fellow creature,
let me do it now,
let me not defer or neglect it,
for I shall not pass this way again.
(modified from William Penn)

TABLE OF CONTENTS

LIST OF SYMBOLS

$\mathbf{X}$	Bold uppercase typically specifies a matrix.
$\underline{x}_i$	Underline lowercase typically specifies a column vector i of $\mathbf{X}$.
x_{ji}	j^{th} row, i^{th} column element of $\mathbf{X}$.
k	Typically indexes a source neural field k.
y	Typically indexes a target neural field y.
i	Typically indexes a source cell assembly (symbol) i.
j	Typically indexes a target cell assembly (symbol) j.
a	Typically indexes a source neuron a.
b	Typically indexes a target neuron b.
$[k]$	As a superscript, designates neural field k.
$[y,k]$	As a superscript, designates connection from neural field k to y.
N_k	Number of neurons in neural field k.
L_k	Number of symbols (cell assemblies) in neural field k.
M_k	Average number of neurons per cell assembly in neural field k.
r_k	Average number of cell assemblies per neuron in neural field k.
$\mathbf{X}^{[k]}$	Binary matrix of cell assembly organization of k, where Rows = neurons, columns = cell assemblies.
$\underline{x}_i^{[k]}$	Column vector i of the matrix $\mathbf{X}^{[k]}$, $x_{ai}^{[k]} = 1$ specifies neuron a is part of cell assembly i.
$\mathbf{L}^{[y,k]}$	Binary matrix of associations between k and y.
$l_{ji}^{[y,k]}$	If $l_{ji}^{[y,k]} = 1$, row j and column $i = 1$, then, source cell assembly i in k is associated with target cell assembly j in y.
$\lambda_c^{[y,k]}$	Average number of target associations in y per source cell assembly in k,
$\lambda_r^{[y,k]}$	Average number of source associations from k per target cell assembly in y,
$\mathbf{U}^{[y,k]}$	Binary matrix of the existing unstrengthened synapses from k to y.
$\mathbf{W}^{[y,k]}$	Binary matrix of the existing strengthened synapses from k to y.
$s^{[k,y]}$	average(mean) synaptic strength between neural field k and y.
$\widetilde{\eta}_b(t)^{[y]}$	Neuronal excitation vector.
$\eta_a(t)^{[k]}$	Neuronal activity in k, If $\eta_a(t)^{[k]} = 1$ then neuron a fired an action potential at time t.
$\widetilde{\alpha}_i(t)^{[y]}$	Symbol excitation, linear sum of fractionally active associations.
$\alpha_j(t)^{[y]}$	Cell assembly(symbol) activity, fraction of neurons in cell assembly i active in y.

LIST OF FIGURES

ACKNOWLEDGMENTS

To my research advisor, Professor Robert-Hecht Nielsen, without whom this dissertation and my journey would not have been possible. You have been much more than a research advisor, you have been a mentor, and the lessons I have learned from our many wonderful discussions stretch far beyond anything I could have asked for coming into graduate school. I will always look back fondly on our times together. I hope that my path in life will reflect well on the insights I have gleamed from you. And of course I must thank you for having the courage to invest the years you did to discover and introduce Confabulation Theory, I believe history will one day reflect back kindly on your discovery.

To my co-advisor Robert Bitmead, for your guidance along the way and for stepping in and helping when I had few to turn to in my graduate path. And to my committee members Tom Bewley and Virginia de Sa, for providing the support and guidance for my interdisciplinary research topic.

To my neuroanatomist guru committee member, Glenn Northcutt. I owe you a great deal of gratitude for introducing me to neuroanatomy. When I first took your class I didn't even really know what the cerebral cortex or thalamus was. You provided a great set of shoulders to attempt to stand on.

To my current labmates, Rupert Minnett and Andrew Smith for all our discussions, your insights and your help proof reading much of my writing. I hope there will be many more days talking confabulation theory. I wish the three other next generation students in the lab who will carry the torch my best as well.

To my early tensegrity labmates, J.P. Pinaud and Wai Leung Chan for their support, friendship, and guidance in the early stages of my graduate career. I will fondly remember the many days spent in the lab with you two.

To Stephen Larson, and the Friday crew. Thanks for being an instigator of open discussion and sharing ideas, I enjoyed the many Friday lunches we spent together.

To Jonathan Yu. My graduate experience would not have been the same without you. Your friendship has been invaluable to me on many levels. Thanks for all our discussions and keeping all my lofty theoretical talk in check by bringing some experimentalist to the table.

To my San Diego friends/family. Your love and support was felt all the way through this processes and I couldn't ask for a better family away from home.

To my Boys back home. You may not have a clue what I've been doing all these years, but knowing you always have brothers there supporting you no matter what you are doing will always be a part of my success.

To Shannon Kawika Phelps, thank you for your mentorship and sharing the art with me. The art you taught me has guided me well through many of the challenges that the Phd process presented.

To all my other friends and family for your love encouragement and support, you spirits always seem to be with me.

And more than anyone, I have to thank my parents, Ray and Mieke Solari, for always encouraging me to climb the highest mountain. My accomplishments are simply a reflection of the wonderful parents you have been and the sacrifices that you have made so that I could be where I am. I love you dearly.

Finally, maybe even more than my parents, I have to thank my mazing wife Ruthi Belle Bozman-Moss Solari;) You are my soulmate, earthmate, and reflection. My research and this dissertation could not have been written without your unconditional love and support. I met you when I started this whole process and now I am excited to end this process with you and for the next journey in our lives.

Chapter 2, in full, is a modified reprint of the material as it appears in Physics of Life Reviews, Solari, S., Smith, A., Minnett, R., Hecht-Nielsen, R., Elsevier, 2008. The dissertation author was the primary author of this material.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Solari, Soren. The author was the primary investigator and author of this material.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Network, Solari, S. and Hecht-Nielsen, R. 2009. The dissertation author was the primary investigator and author of this material.

Chapter 5, in part, is currently being prepared for submission for publication of the material. Solari, S. and Hecht-Nielsen, R. The dissertation author was the primary investigator and author of this material.

Chapter 6, in part, is currently being prepared for submission for publication of the material. Solari, Soren. The dissertation author was the primary investigator and author of this material.

Chapter 7, in part, has been submitted for publication of the material and is in the editorial review processes. Solari, S., Tsigelny, I., Sharikov, Y., Hecht-Nielsen, R. The dissertation author was the primary investigator and author of this material.

VITA

1999	B.A./B.S, Electrical Engineering, University of San Diego
1999	Who's Who Among Students in American Universities and Colleges, University of San Diego
2000-2002	Senior Test Engineer, Cadence Design Systems
2003-2004	Electrical and Computer Engineering Best Teaching Assistant Award, University of California, San Diego
2004	M.S, Engineering Sciences (Mechanical Engineering), University of California, San Diego
2007	C.Phil., Engineering Sciences (Mechanical Engineering), University of California, San Diego
2003–2009	Research Assistant University of California, San Diego
2009	Ph.D., Engineering Sciences (Mechanical Engineering), University of California, San Diego.

PUBLICATIONS

Journal Papers

1. Soren Solari, Andrew Smith, Rupert Minnet, Robert Hecht-Nielsen
   Confabulation Theory
   Physics of Life Reviews, **5**: 106-120, (2008).

Conference Papers

1. Wai Leung Chan, Soren Solari, Robert E. Skelton
   Shape Control of Tensegrity Plate - An Experimental Study
   Proceeding of ASME 2004 International Mechanical Engineering Congress and Exposition, Anaheim, CA (2004).

2. J.P. Pinaud, Soren Solari, Robert E. Skelton
   Deployment of a Class 2 Tensegrity Boom
   SPIE's 11th Annual International Symposium on Smart Structures and Materials, San Diego, CA (2004).

FIELDS OF STUDY

Studies in Integrative Neuroscience
Professor Robert Hecht-Nielsen
Professor Glenn Northcutt
Professor Virginia de Sa

Studies in Dynamic Systems and Control
Professor Robert R. Bitmead
Professor Thomas Bewley

ABSTRACT OF THE DISSERTATION

A unified anatomical theory and computational model of cognitive information processing in the mammalian brain and the introduction of DNA reco codes

by

Soren Solari
Doctor of Philosophy in Engineering Sciences (Mechanical Engineering)
University of California, San Diego, 2009
Professor Robert R. Bitmead, Chair
Professor Robert Hecht-Nielsen, Co-Chair

This dissertation presents a comprehensive unified anatomical theory in conjunction with computational models that serve to provide a complete working explanatory framework for cognitive information processing in the mammalian brain. Our model provides sufficient detail such that we are able to hypothesize the function of individual populations of neurons as they correlate to psychological observation.

We first introduce our working hypothesis, confabulation theory, on the fundamental cortical organization and information processing operations underlying cognition in the mammalian brain. We present a comprehensive neuroanatomical review, designed to uncover the blueprint of primate non-primary cortical neuroanatomy in conjunction with the thalamus and basal ganglia. More than a review, we synthesize hundreds of original neuroanatomical experiments into a single viewpoint of the basic functional circuits (i.e. blueprint) underlying all cognitive information processing. We propose that there are 8 basic pyramidal neural fields, and only 3 types of thalamocortical projections, all having a prototypical function. We explicitly hypothesize their function in relation to confabulation theory. We present a new mathematical model, termed a multi-associative memory, of the implementation of multi-modal associations in the cerebral cortex. The model serves as a basis for understanding the genomic implementation and utilization of associations in randomly wired brains of varying sizes. We present a biologically plausible thalamocortical attractor network capable of the necessary and sufficient conditions related to the controlled application of confabulations in cognitive information processing. Finally, we propose a unified explanatory model of cognitive information processing hypothesizing the neural mechanisms underlying cortically represented perceptions, cortically represented behaviors, working memory, hippocampal short term memory, cortically consolidated long term memory, procedural memory, attention, cognitive information processing, which we believe provides enough detail to begin designing testable neuroscience experiments.

As part of our effort to understand the mammalian brain, we discovered several unknown properties of DNA, and present novel evidence regarding the symmetry and genomic diversity of exactly repeated DNA reverse complimentary (RECO) codes.

1

Introduction

1.1 Why

Why in the year 2009, with all of our technology and knowledge, do we, as a human species, not fundamentally understand the question 'how does our brain work'? With this dissertation, I have fundamentally chosen to ask that basic question. Specifically, how is information stored and processed in the mammalian cerebral cortex and thalamus? Two structures without which, cognitive information processing does not exist.

Our methodology is simple. Begin with an established hypothesis of the basic information processing operations proposed as the basis for cognition, then test that hypothesis against the known facts of neuroscience through anatomical and computational modeling. The clear distinction of my work as opposed to others is the absolute rigorous adherence to the known (and lesser known) facts of neuroanatomy. When our theory or model appeared inconsistent with the neuroanatomy, we changed the theory or model to be consistent with the neuroanatomy. When I hit a road block to questions about brain function, I looked deeper into the neuroanatomy for answers. As was stated over 130 years ago by Bernard Gudden:

> Faced with an anatomical fact proven beyond doubt, any physiological result that stands in contradiction to it, loses all its meaning...So, first anatomy and then physiology; but if first physiology, then not without anatomy. From (Brodmann, 1909).

The importance of the adherence to known neuroanatomy and the inclusion of that knowledge into any model of cognition cannot be understated in the light of history. Too often have theoretical ideas or preconceived notions trumped known neuroanatomy. Central nervous systems at the simplest description are composed of neurons and the connections between

them, where significant damage to either results in well established physical or cognitive disabilities(Geschwind, 1965). Therefore, neuroanatomy provides the ultimate clues and final answers to the question of how the brain works.

1.2 Motivation: a brief history of our understanding of the cerebral cortex

In Greek times, the location of cognitive information processing was a debated topic, ranging from the earliest suggestion by Hippocrates(c. 400 B.C.E), that the brain is responsible for cognition, to Aristotle's cardiocentric view that cognitive processes occur exclusively in the heart(Finger, 1994; Clarke and Dewhurst, 1972). The debate continued until the 4th century, when the western scientific view of cognition was established by loosely blending some neuroscience work done by Galen with Christian theology. In the 4th and 5th centuries, the Church Fathers created the idea of ventricular functional localization, stating that all cognitive processes resided in the neuron-free, fluid filled ventricles of the brain. The religious community further railed against human dissection and thereby shunted any progress toward fundamentally understanding the human brain through neuroanatomy. The ventricular theory explicitly stated that all cognitive processes occurred in the ventricles of the brain and that all brain nuclei had no function. This was the prevailing scientific viewpoint for over 1400 years until it was finally overcome in the 1800's. The earliest challenge to the ventricular doctrine actually occurred in 1664 by Willis (a neuroanatomist) who published a major work introducing an anatomically accurate description of the brain, and in it, suggested that the cerebral cortex had function. His work sparked debates and new ideas that continued to evolve for 200 years. The accurate modern day view of nervous systems was finally set into being by one of the great neuroanatomists, Ramon y Cajal, in the late 1880's. Ramon y Cajal was the principle player in determining that the nervous system and its function was the result of individual neurons communicating with each other, thus forcing all hypothesized brain function into the neuron filled gray matter of the brain. From this point forward, neuroscience was able to progress with improving experimental techniques in understanding the function of various brain nuclei.

As we just described, we as a species only began understanding our brain a little over 150 years ago. Hopefully, this fact is both surprising, disturbing, and encouraging. Surprising that we have only recently begun to make real progress into understanding ourselves despite millennia of pontificating about the mind, consciousness, and the soul. Disturbing that sufficient evidence existed in Greek times (400 B.C.E) to begin to make enormous progress into understanding the

brain, and that the prevention of doing neuroanatomical work by those propagating a religious belief stunted our understanding for millennia. Encouraging that now in the 21st century, we are at the cusp of history, with the combination of technology and grounded science, to finally unlock for the first and last time in history, the mysteries of our brain.

The historical perspective is critical in understanding the factors that have limited our understanding of the brain and the necessity for the approach taken by this dissertation. In looking at the historical facts, we gain insight into the most fruitful directions for research into the brain. The primary and recurring phenomena continually limiting our understanding is the shift of neuroscience(the study of the brain) away from explaining the function of the neuroanatomy and toward exploring abstract notions of brain function. Treating the brain as a black box and describing what happens in a stimulus response paradigm is not in itself capable of explaining anatomical function, and is in essence what has been done for millennia. The recurring danger in limiting our understanding today is no different. And despite all the technology today (and maybe because of it) the same recurring danger is looming. One only needs to question many neuroscientists about detailed neuroanatomy, or to look at the number of neuroanatomy papers being published to see that the shift has once again occurred away from a principled focus on explicitly understanding and explaining neuroanatomical function.

History tells us that focusing on the function of the brain's neuroanatomy has lead to virtually all the paradigm shifts in our understanding of the brain. The last hurdle of a comprehensive theory of brain function must be grounded in specifically determining and explaining the function of the neuroanatomy.

1.3 Summary: putting the cognitive puzzle pieces together

If we are going to ever understand our brain, we must principally reverse engineer it. To know WHAT happens is a matter of experiment and observation. To understand HOW it happens is a matter of reverse engineering the system to the detailed point of understanding.

In the last hundred years of exploring the brain, experiments have produced many fragments of information about the brain. Each new experiment and reported observation adds to the number of puzzle pieces that must be put together. In essence, experimental neuroscience (in the context of studying the brain in general) has provided an enormous list of WHAT happens to a subjects observable responses as a result of stimuli and/or physical brain alterations. However, there is little precise discussion of the fundamental neural code of cognition, the underlying

principles of brain organization and function, the answering of why we observe what we do.

The dissertation chapters in some way each illuminate and put together certain pieces of the cognition puzzle in an effort to get to explaining HOW neuroanatomically the brain implements cognitive information processing.

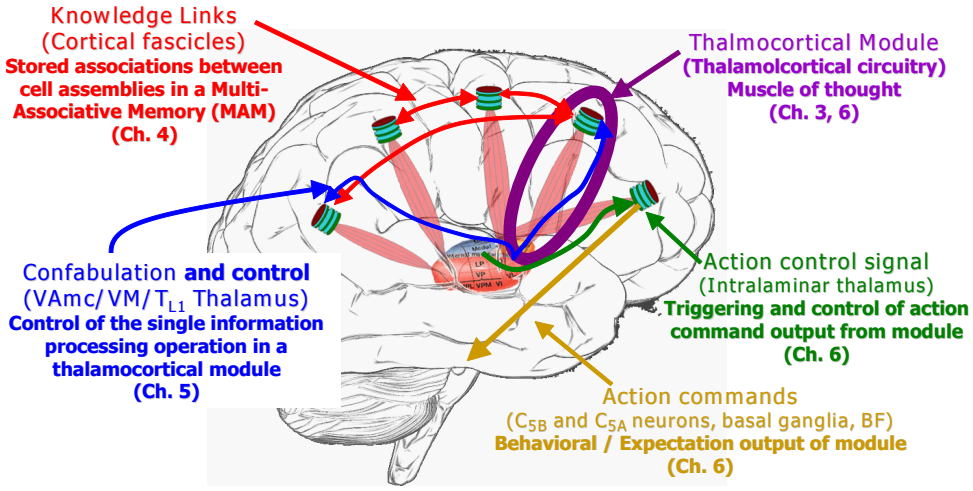

Figure 1.1: Main dissertation overview. Labels in the picture describe the confabulation theory functional hypothesis, followed by the anatomical structure implementing the hypothesized function in parenthesis, followed by a brief description of its role in cognitive information processing in bold. Chapter 2 presents confabulation theory, which provides a working hypothesis of mechanisms underlying cognition. The four basic principles are 1) Thalamocortical modules and symbols; 2) Knowledge links; 3) Confabulation; and 4) Action commands. Chapter 3 discusses the detailed neuroanatomy of the primate non-primary cerebral cortex in order to establish the functional blueprint upon which functions must be mapped. Chapter 4 provides a mathematical model of the formation and utilization of multi-modal associations (knowledge links) in the cerebral cortex. Chapter 5 introduces a working model of a biologically plausible thalamocortical attractor network capable of implementing the necessary and sufficient conditions related to a controlled confabulation operation. Chapter 6 presents a unified explanatory model of the major psychological cognitive processes and their neuronal implementation in the mammalian brain.

Figure-1.1 provides provides an overview of the topics that are covered by the dissertation chapters.

Chapter 2 presents confabulation theory, which provides a working hypothesis of mechanisms underlying cognition. The four basic principles are 1) Thalamocortical modules and symbols; 2) Knowledge links; 3) Confabulation; and 4) Action commands.

Chapter 3 discusses the detailed neuroanatomy of the primate non-primary cerebral cortex in order to establish the functional blueprint upon which functions must be mapped.

Chapter 4 provides a mathematical model of the formation and utilization of multi-modal associations (knowledge links) in the cerebral cortex.

Chapter 5 introduces a working model of a biologically plausible thalamocortical attractor network capable of implementing the necessary and sufficient conditions related to a controlled confabulation operation.

Chapter 6 presents a unified explanatory model of the major psychological cognitive processes and their neuronal implementation in the mammalian brain.

Chapter 7 presents the novel finding, discovered through the rarely used unbiased analytic method of exact sequence matching in entire genomes, regarding the symmetry and genomic diversity of exactly repeated DNA reverse complementary (reco) codes.

1.4 Contributions

The principle contribution of this dissertation is the creation of a comprehensive model of human and mammalian brain function, which provides a consistent and explanatory framework spanning neuroanatomy to psychology, with sufficient detail to correlate to neuroscience experiments at both the neuron and behavior level.

In general the contributions of the dissertation include both theoretical work and computational work and are summarized as follows by chapter:

1. Several modifications and improvements to confabulation theory
    - The modified confabulation theory hypothesis of the direct implementation of knowledge links as opposed to an indirect two-stage synfire chain and accompanying mathematical model.
    - An explicit hypothesis regarding the anatomical implementation of the control of confabulations via layer 1 thalamic projections.
    - The introduction to confabulation theory of the need to control the triggering of action commands and the hypothesized anatomical mechanism.
2. The synthesis of a large body of neuroanatomical studies into a single unified framework of functional cognitive circuits including:
    - A poster displaying the significant body of neuronanatomical facts in a graphical format used to derive the cognitive circuit blueprint
    - A novel hypothesis proposing three distinct functional thalamocortical projections
    - The identification of the 8 major neuroanatomically identifiable populations of cortical pyramidal neurons and hypothesis on their individual function.

3. The introduction of a new mathematical model that generalizes classic associative memories into a many to one, one to many, multi-associative memory (MAM). We derive an explicit mathematical formulation for associative signal and associative interference based on the underlying parameters of the system.

4. We present a new model of thalamocortical information processing

    - The model introduces an explicit form for balanced inhibition in associative memory connectivity utilizing associative signal and interference.
    - The model introduces the mechanisms by which non-symmetric neural fields can be connected in an associative memory and still function equivalent to symmetric neural fields.

5. We present a unified model of cognitive information processing and present detailed hypothesis explaining the neural mechanisms underlying the following cognitive phenomena

    - Cortically represented perceptions
    - Cortically represented behaviors
    - Working memory
    - Hippocampal short term memory
    - Cortically consolidated long term memory
    - Procedural memory
    - Attention
    - Cognitive information processing

2

Confabulation theory: a hypothesis on mammalian cognition

2.1 Introduction

The formal academic study of human and animal cognition has been underway for over two millennia. Yet, even today, about all that can be stated with certainty is that there is strong evidence suggesting that the storage and processing of information involved in all aspects of human cognition (seeing, hearing, planning, language, reasoning, control of movement and thought, etc.) is carried out by the cerebral cortex and its related subcortical nuclei. Beyond general statements of this sort (primarily based on deficits after cortical lesions)(Catani and ffytche, 2005; Penfield and Rasmussen, 1968), nothing is definitively known about how cognition (also referred to here as thinking) works, or about what cognitive knowledge is.

The present chapter serves to introduce the updated hypotheses underlying confabulation theory, which serves as our working hypothesis on the basic structure and function of mammalian cognition. Section 2.2 provides a conceptual framework for the key elements of confabulation theory. Sections 2.3.1, 2.3.2, 2.3.3, and 2.3.4 detail the four key elements of confabulation theory. Finally, section 2.4 briefly surveys some natural language processing experiments using computer implementations of the theory.

2.2 A conceptual framework for confabulation theory

Cognition and/or thought processes are a result of the coordinated contraction of many "muscles of thought", just as movement is the the result of the coordinated contraction of many physical muscles. The structure (i.e. connectivity) of the musculo-skeletal system ultimately determines what movements an organism is capable of performing, and any organism is forced to learn what muscles to contract in what sequence to achieve a movement goal. The structure (i.e. connectivity) of the cognitive nuclei in the mammalian brain (cerebral cortex, thalamus, and basal ganglia) ultimately determines what cognitive processes an organism is capable of performing, and similarly any organism is forced to learn what "muscles of thought" to contract in what sequence to achieve a cognitive or thought process goal.

We hypothesize that cognitive information processing is a direct evolutionary re- application of the early neural circuits evolved to control movement, and thus functions just like movement. Brains (nervous systems) seem to have developed to process sensory inputs and adaptively coordinate individual muscles (Lieber, 2002; Squire, 2004). The survival of certain organisms increased the evolutionary fitness of the neural circuits proficient in coordinating muscle contractions. Since neural circuitry now existed to contract individual muscles, we hypothesize the 6-layered cerebral cortex (Northcutt and Kaas, 1995) and specific thalamus (Jones, 2007) evolved to perform cognitive information processing operations utilizing the same control mechanisms implemented for muscles (i.e. basal ganglia, brainstem, cerebellum, etc.).

Conceptually, brains are composed of many "muscles of thought" (termed *thalamocortical modules* in mammals), each of which represents a single *cognitive dimension* in the individual's *mental universe*. A module is composed of *symbols*. Each symbol is a sparse population of neurons and represents a particular *perception* within the cognitive dimension of a module. For example, if the cognitive dimension of a module encompass the visual representation of faces, then a single symbol may represent a particular face (a single face perception)(Tsao et al., 2006). If the cognitive dimension of the module encompasses words (in any language) then a single symbol represents a particular word (a single word perception). We use the term perception to universally define any symbol in any module without constraint to representing external world information. Note that a perception in the motor cortex could represent a physical movement, or a perception in the cingular cortex might represent a particular emotion. We would call both a symbol (i.e. perception), even though either might be better conceptualized as an internal behavior or state.

Each thalamocortical module is connected to many other modules through the cortical

white matter in the brain. When two symbols are active at the same time, usually in different modules, they are said to co-occur. Co-occurrence in time creates the opportunity to physically associate the two symbols. For instance, after seeing a face and hearing a name together, the symbols representing each may be associated. These learned associations are implemented by synaptic connections between symbols. A single strengthened association between two symbols is termed a *knowledge link*. Collectively knowledge links comprise all cognitive knowledge.

Each thalamocortical module performs the same single information processing operation, which can be thought of as a "contraction of symbols", termed a *confabulation*. Throughout a confabulation, input excitation is delivered to symbols in the module through knowledge links. When a thalamocortical module contracts, there is no physical movement in the brain, rather symbols compete through excitation and inhibition for exclusive activation (a so called "winner-take-all" competition) within that module. As a result of the competition, the number of active symbols is reduced. The contraction of symbols in each thalamcortical module is externally controlled, in the same way that the contraction of a muscle is externally controlled.

Physical muscle contractions are controlled by graded analog inputs provided by alpha motor neurons(Lieber, 2002). Similarly, a confabulation in a thalamocortical module is controlled by a graded analog control input, the thought control signal, which determines how much overall symbol activity there can be in the module. The thought control signal determines how many symbols are in the competition, but has no effect on selecting which symbols are in the competition. Which symbols are in the competition is determined by the knowledge link input from active symbols in other modules. Ultimately, the thought control signal is used to contract the number of active symbols in a module from many active symbols to a single active symbol. The resulting single active symbol is termed the *confabulation conclusion*.

Each time a module reaches a conclusion, the module is triggered (through different subcortical input than the thought control signal) to activate action commands (a separate population of neurons from symbol neurons). Action command outputs are ultimately associated with the conclusion symbol. Depending on the module, action commands may cause direct physical behavior, or be used to control confabulations in other modules. The learned associations between symbols and action commands comprises all skill knowledge.

In summary, the brain is composed of many thalamocortical modules ("muscles of thought"), which, through controlled input, expand and contract the list of active symbols in the module. The list of active symbols is determined by input from active symbols in other modules via knowledge links, thus all the modules interact dynamically, "comparing notes", while a thought control input contracts the number of active symbols in each module to a single conclusion. When a conclusion is reached in a module, those action commands which have a learned

association to that conclusion symbol may be launched. These action commands are proposed to be the source of all cortically launched behaviors.

Thalamocortical modules performing confabulations, communicating cognitive knowledge through knowledge links, and applying skill knowledge through action commands constitute the complete foundation of all mammalian cognition.

2.3 Four key elements of confabulation theory

Confabulation theory is organized into four key elements that we propose form the fundamental underpinnings of all cognition. Although confabulation theory in its most general form likely applies to cognitive information processing in all nervous systems, we focus on describing confabulation theory from the perspective of mammalian neuroanatomy. The dominant neuronal structures (gray matter) and gross anatomical projections (white matter) in all mammals have a virtually identical organization(Striedter, 2005), therefore the four key elements presented here apply equivalently to all mammals, including humans. Although the foundations of cognition seem to be fully covered by the four key elements, many details are still waiting to be elucidated. In the present dissertation, we elaborate on many of the specifics within chapters 3, 4, 5, and 6.

2.3.1 Thalamocortical symbols and modules: representing perceptions within an individual's many cognitive dimensions

The cerebral cortex is a thin (∼3mm) six-layered sheet of neurons surrounding the entire brain that is by far the most developed in humans compared with other mammals(Striedter, 2005). For over a hundred years now, the cerebral cortex has been known to have localized functionality, for example, vision, language, and movement are each processed in separate cortical areas (Finger, 1994; Penfield and Rasmussen, 1968). Even though each area of the cerebral cortex carries out seemingly different types of information processing, every area of the cortex has the same 6-layered structure and equivalent reciprocal axonal connections with some part of the thalamus(Brodmann, 1909; Jones, 2007). The similarity across all regions of cortex and thalamus strongly suggest that how information is stored and processed in each cortical area is the same even though what is stored and processed is different. Surprisingly, given the detailed knowledge of cortical organization, very little is known about exactly how or exactly what is stored and processed in any part of cortex. Confabulation theory proposes that the cerebral cortex is divided into thousands of thalamocortical modules, each on the order of 10's mm^2 (see Figures-2.1

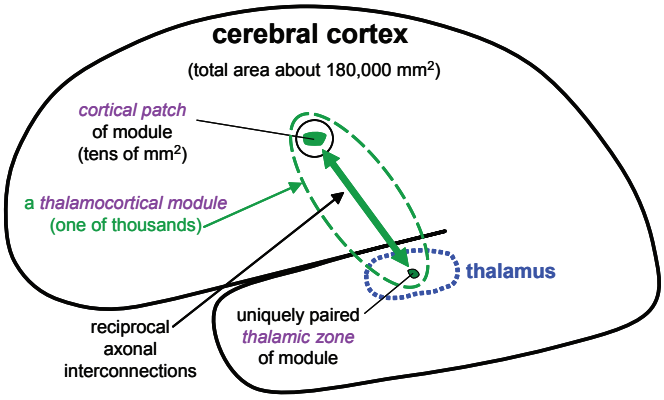

Figure 2.1: A human thalamocortical module (one of thousands in human cerebral cortex). Each thalamocortical module is composed of a localized patch (having an area of a few tens of square millimeters) of the six-layer cortical sheet along with a uniquely paired, reciprocal, small zone of specific thalamus. The cortical patch of each module is reciprocally axonally connected with the thalamic zone of the module. Although cortical patches (and thalamic zones) of different modules are largely disjoint, partial overlaps do likely occur.

and 2.2). Each thalamocortical module performs an identical information processing operation (confabulation) on different symbolic neuronal representations. A single thalamocortical module can be thought of as a "muscle of thought" whose sole information processing operation is a "contraction of symbols".

Each thalamocortical module is responsible for storing and processing information encompassed by one cognitive dimension within the individual's mental universe (e.g. visual perceptions, auditory perceptions, language objects, thought processes, plans, movements, etc.). Each module develops and permanently stores a finite number of symbols, each representing a single perception within the cognitive dimension. For example, if a particular module stores and processes words (the modules cognitive dimension encompasses "words"), then it might contain over 100,000 symbols, each representing a specific English word or phrase ("tree", "jet fighter", "Winston Churchill", etc.). Similarly, examples of visual symbols exist in visual modules(Tanaka, 2003). In humans, each module typically possesses 1,000's to 100,000's of symbols. New symbols are created in childhood and throughout adult life likely through a neuronal self organizing process due to environmental exposure. Once formed, symbols can be effectively permanent, unless damaged. Each cortical module contains 100,000's to 1,000,000's of neurons. However, each symbol is represented by a small population of 100's of those neurons. Because of this sparsity, a large number of symbols can exist within a module without significant interference due to a few shared neurons.

A module is said to be describing a single perception when a single symbol is active in the

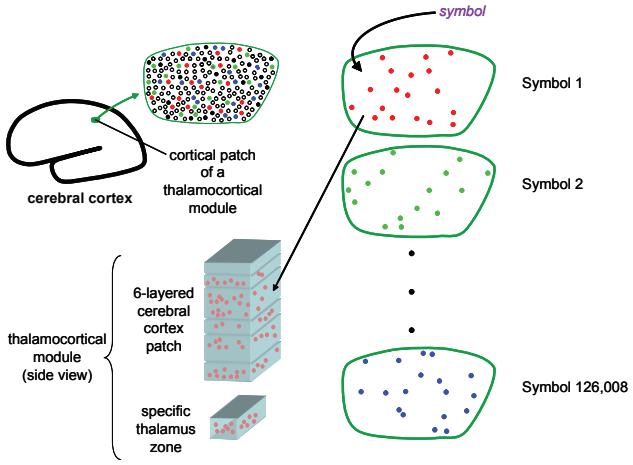

Figure 2.2: A primary function of each thalamocortical module is to store and process information encompassed by one cognitive dimension within the individual's mental universe (e.g. visual perceptions, auditory perceptions, language objects, thought processes, plans, movements, etc.). To carry out this function, Each module develops and permanently stores a finite number of symbols, each representing a single perception within the cognitive dimension (e.g. a symbol may represent a particular word in a module that processes words, or a particular face in a module that processes faces). Each symbol representation is composed of 100's of neurons (shown as colored dots within the enlarged depiction of the module's cortical patch). Here, a module with 126,008 symbols is depicted. See chapter 3 and 6 for many more details.

module. The activation of a symbol is represented by the elevated rate and possibly synchronous firing of its neurons. Only when a symbol is active is it sending excitation through axons to other associated symbols. Because of the sparse coding of symbols, one or several symbols can be fully or partially active at the same time in a module (or none may be active). Consider that the cerebral cortex has one module whose cognitive dimension encompass the visual representation of faces(Tsao et al., 2006), and another whose cognitive dimension encompass words. Suppose a symbol representing "Bob's" face in one module and a symbol representing "Bob's" name in another module have been associated. Hearing the word "Bob" will activate the symbol (i.e. perception) representing the name "Bob" in the word module and consequently, excitation will be delivered to all the symbols (i.e. perceptions) in the face module representing faces. Only those faces that have been associated with word "Bob" will receive excitation. If a name, like "Bob", has been associated with many faces, then additional context (like a last name) would be necessary to disambiguate the activation of a particular "Bob's" face symbol/perception. This begins to illustrate both how symbols in different modules interact and the need for many modules to interact simultaneously, described further in section 2.3.3.

Anatomically, thalamocortical modules have the exact same structure in all mammals(Northcutt and Kaas, 1995), providing strong evidence that cognition functions identically in

all mammals. Understanding the structure of a thalamocortical module, therefore, is essential to understanding how cognition functions. Within every thalamocortical module, the same distinct populations of neurons exist (roughly aligned with the six layers of the cortex and thalamus) each having homotypical projections to different regions of the brain(Lorente de No, 1943). We call each anatomically distinct neuron population a *neural field*. Together, the afferent(input) and efferent(output) connectivity of each neural field defines its function. Further complexity arises because each symbol has a separate sparse neuronal representation within each neural field. Therefore, even within a single module, a symbol exists in multiple neural fields and therefore has multiple functions. Although describing further details is beyond the scope of this introduction, our computational models suggest this detailed anatomical organization is central to performing the controlled winner-take-all competition of confabulation, in addition to communicating knowledge links, with mere neurons and synapses (see section 2.3.3).

2.3.2 Knowledge links: basis of all cognitive knowledge

In 1949 Donald Hebb postulated that learning in brains was the strengthening of synapses linking two groups of neurons (which he called "cell assemblies") with axonal connections between them(Hebb, 1949). He postulated that this occurred whenever the first cell assembly helped cause the second cell assembly to become active (the involved synapses are then strengthened). Ample neurological evidence supports Hebb's postulate; however, no comprehensive examination of the role of cell assemblies in learning has yet occurred. Figure-2.3 illustrates the role of Hebb's idea in confabulation theory. When two symbols are co-active they may become associated by strengthening the synapses linking them. The unidirectional association between two symbols is termed a knowledge link; a reciprocal pair of knowledge links may exist between symbols. Each knowledge link is considered a single item of knowledge. An active source symbol delivers input excitation to all target symbols to which it is connected through knowledge links, where the strength of a knowledge link determines the amount of input excitation that a target symbol receives. Therefore, a knowledge link is an association between two cell assemblies, as Hebb postulated, albeit with a bit more complexity.

In particular, knowledge links are formed over two time-scales (Squire, 2004). Instantaneous/temporary knowledge links are formed indirectly by linking the two co-active symbols via the hippocampus, entorhinal cortex, and related perirhinal/ parahippocampal cortices. Over many sleep periods this indirect knowledge link is consolidated into a direct cortico-cortical knowledge link from one symbol to another symbol (i.e. no longer through the hippocampus). This unidirectional consolidated cortico-cortical knowledge link between two symbols will typically last for decades; even if it is not used. Knowledge links that are used last for life. The collection

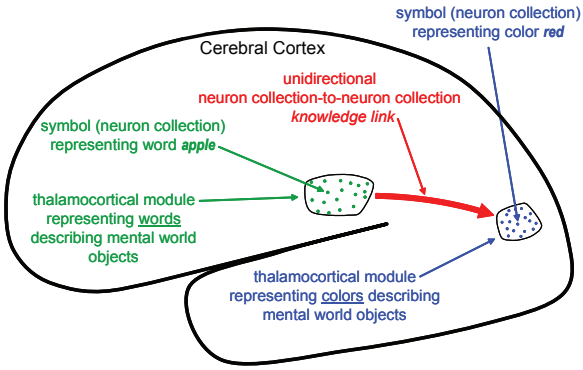

Figure 2.3: A cognitive knowledge link. Here, a human subject is viewing and considering a red apple. A visual thalamocortical module contains an active symbol for the color of the apple. At the same time, a language thalamocortical module contains an active symbol for the English name of the apple. Pairs of symbols which meaningfully co-occur in this manner have unidirectional axonal links, termed knowledge links (each considered a single item of knowledge), established between them via synaptic strengthening. The entire axonal bundle of all unidirectional knowledge links between two modules is termed a knowledge base. Knowledge bases compose the vast majority of cortical white matter. Confabulation theory predicts that knowledge links must be implemented in vast quantities for cognition to be useful, which is consistent with known neuroscience; white matter is the largest structure in the human brain.

of all unidirectional knowledge links connecting a particular source module to a particular target module is termed a knowledge base.

Confabulation theory proposes that the mathematics of cognition relies on the formation, strength, and use of these knowledge links. The strength of a single knowledge link is logarithmically related to the conditional probability $p(\beta|\epsilon)$, where β represents the occurrence of source symbol β, and ϵ the occurrence of the target symbol ϵ (see Figure-2.5). Importantly the quantity $p(\beta|\epsilon)$ is estimated by dividing the number of times that β and ϵ co-occur by the number of total occurrences of the target symbol ϵ . Biologically, this implies that a target symbol (composed of neurons) has a relatively fixed total strength of incoming knowledge links (synapses) that it can physically support, and that the total strength of all incoming knowledge links to a single target symbol is limited. The brain, therefore, cannot form an arbitrary number of strengthened knowledge links, which explains the need to use temporary knowledge link formation and a entire dedicated brain region (the hippocampus) to determine which knowledge links should be consolidated and become permanent. Amazingly the simple biological constraint on neurons and the support of synapses may have enabled the exploitation of the underlying mathematics necessary for cognition.

A major question arises as to whether co-occurrence knowledge of this sort is sufficient to account for human and animal "intelligence". Below, in section 4, we will see that they are.

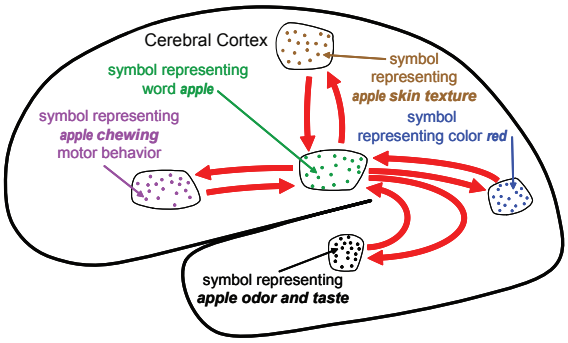

Figure 2.4: Billions of pairs of symbols are connected via knowledge links. The set of all knowledge links joining symbols belonging to one specific source module to symbols belonging to one specific target module is termed a knowledge base. In the human brain, knowledge bases take the form of huge bundles of axons termed fascicles, which together make up a large portion of each cerebral hemisphere's ipsilateral white matter. Each module also typically has a knowledge base to its contralateral 'twin' module (and perhaps to a few others near its twin) - which together constitute the corpus callosum fascicle linking the two cerebral hemispheres. Here, reciprocal knowledge links (red arrows), only some of which are shown, connect various symbols representing different attributes of an apple pairwise with each other. Fig. 4 illustrates an example of knowledge links that may have been formed by experiencing a red apple. Here, five modules are each expressing a symbol describing one attribute of the apple (i.e. one symbol is active in each module). In the center, the symbol representing the English name of the apple is active. Above that, the symbol representing the apple's skin texture is active. To the right, the apple's visual color is active. And to the left and at the bottom the motor chewing process for an apple and the gustatory sensation of the apple are active. When an apple is currently present in the mental world, it is its collection of knowledge-link-connected symbols which are currently active in many modules. There is no "binding problem" [23], because all of these symbols are mutually 'bound' by their previously established pairwise knowledge links. In consonance with the pairwise associationist doctrine established by Aristotle and built up further by a series of leading thinkers on human cognition over the past 500 years confabulation theory contends that such knowledge links - formed on the basis of symbol pair co-occurrence - are the only type of knowledge used (or needed) in cognition.

2.3.3 Confabulation: universal basic operation of thought

The vague notion that cognition employs some sort of "information-processing" has been around for millennia. Today, the understanding of the exact nature of this "cognitive information-processing" is roughly the same as it was in 350 B.C. - the time of Aristotle (arguably the first recorded neuroscientist). Confabulation theory states explicitly that cognition involves only one information-processing operation - confabulation: a simple controlled winner-take-all competition between symbols on the basis of their total input excitation received from knowledge links.

The input excitations arriving at symbol k from different knowledge links are summed to yield the total input excitation for symbol k, I(k) (this summation is noted by the plus signs between the knowledge links in the enlarged illustration of module five). This additive knowledge combination property of thought is what enables the vast information-processing power and

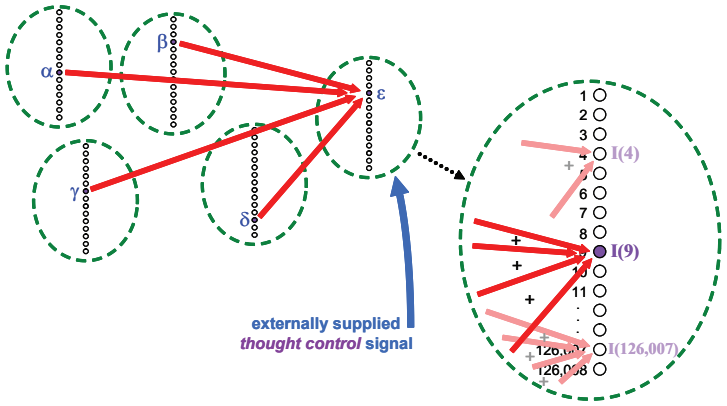

Figure 2.5: Confabulation - the only information-processing operation used in cognition. Here, a concrete example involving five thalamocortical modules is shown (for simplicity, each module is illustrated as a dashed green oval with a list of that module's symbols inside it). During a confabulation, active symbols $(\alpha, \beta, \gamma, \delta)$ in four source modules shown on the left send excitation through knowledge links to symbols in a fifth target module (shown on the right). Each confabulation on every module is controlled by a graded analog thought control signal. The conclusion of a confabulation operation will ultimately be the symbol receiving the most input excitation I (symbol 9 shown on the right). See text for more details. Figure-2.5 illustrates a confabulation. The four source modules on the left each have a single active symbol in them: α, β, γ, and δ. Each active source symbol delivers input excitation to many symbols in the target module through knowledge links. Note that in a brain all modules are typically a source and target module at the same time. The state of the fifth target module, which is about to undergo confabulation, is shown enlarged on the right (red arrows depict individual knowledge links). For illustration, symbol 4 of this module is receiving two active knowledge links, whereas symbols 9 and 126,007 are receiving knowledge links from all four symbols α, β, γ, and δ. Each knowledge link is delivering a certain quantity of input excitation to the neurons of its target symbols. The input from the thought control signal (blue arrow) causes the module to contract, as a result the number of active target symbols decrease. If this control signal allowed only two symbols to be active, then symbols 9 and 126,007 would be active (since they have the most input excitation) and symbol 4 would be inhibited through competition and thus inactive. If the control signal allowed only a single symbol to be active, then symbol 9, having the most input excitation would remain active and symbol 126,007 would be inhibited, resulting in a single conclusion symbol.

flexibility of human cerebral cortex. Note that knowledge links are not neuron to neuron connections, but rather symbol to symbol connections (i.e. many neurons to many neurons); therefore, many 100's to 1,000's of synapses may transmit input excitation to a single target symbol, enabling accurate additive knowledge combination even in the presence of large background noise or individual synaptic failure.

We emphasize that a thalamocortical module does not undergo a confabulation operation unless commanded to do so, in the same way a muscle contracts only when commanded to do so by its motor-neuron input. Upon being commanded to contract (by a deliberately supplied thought control signal, illustrated by a blue arrow in Figure-2.5), each symbol of the fifth module

competes with all others for exclusive activity. During this competition the number of active symbols decreases in proportion to the thought control signal strength (thus a confabulation is a "contraction of symbols"). Since the timing of this contraction is controlled, coordinating the parallel convergence of many modules to a final state may itself involve a significant amount of learning. This learned coordinated control of convergence is termed a thought process. Upon converging to a final conclusion, the neurons representing the symbol with the largest input intensity I (in the example of Figure-2.5, symbol 9) are highly activate and all other symbol-representing neurons are not. This "winner-take-all" competition is called confabulation, and the winning symbol is termed its conclusion.

It may seem mysterious that mere neurons can implement controlled, winner-take-all symbol competition. Within a module, connections between the neural fields in the cortex and the paired thalamic region constitute a neuronal attractor network (see chapter 5), the state of which evolves through cortex-thalamus-cortex oscillations and is modulated by the thought control signal. Each collection of neurons representing a symbol is a stable state of the attractor. A symbol (perception) or multiple symbols (perceptions) can be held active in "working memory" by means of this cortex-thalamus-cortex oscillation. During the oscillation, additional context can be applied through knowledge links to influence the competition. In this way, modules can be made to converge slowly or quickly, and the number of active symbols at any one time can be made to grow or contract to the symbol with the greatest input excitation. In behavioral experiments, subjects can temporarily retain a finite set of sensory domain specific information, which has been termed working memory(Monsell, 1984). We propose that the underlying neural mechanisms of working memory is a controlled continuous thalamocortical oscillation of a single (or possibly several) symbols in a single module. Each module can implement working memory of perceptions within its cognitive domain, thereby distributing working memory throughout the cortex.

Confabulation is hypothesized to be the only information-processing operation of thinking. In the Figure-2.5 example, there is only one confabulation taking place. Ordinarily, confabulations on multiple modules take place together (modules acting as source and target simultaneously), with convergence to the winning symbols slowed somewhat to allow mutual knowledge-link-mediated interaction ("comparing notes" in order to arrive at a mutually consistent confabulation consensus of final conclusions). In this so-called multiconfabulation, millions of relevant items of knowledge (i.e., knowledge links), each emanating from a viable candidate conclusion, are employed in parallel in a "swirling" convergence process. Multiconfabulation is another mechanism enabling the enormous information-processing power and flexibility of thought. As an analogy between movement and thought, a biceps contraction is to a single confabulation, as

the elegant movements of a ballerina are to multiconfabulation.

Confabulation seems quite alien in comparison to existing concepts in neuroscience, computational intelligence, neural networks, computer science, traditional AI, and philosophy. For example, computer CPUs all follow the Turing paradigm: when commanded via a specific, digital, instruction code they execute a pre-defined logical or arithmetic instruction on specified variables. Thalamocortical modules, on the other hand, have only one information-processing "instruction" - confabulation. Further, the thought control signal delivered to the confabulating module from outside the cerebral cortex, is not digital, but analog. Yet the result of each completed confabulation is digital: a single symbol.

A natural question arises as to where the thought control signal originates (see chapter 3). In brief, The most likely source of the thought control signal is a small area of the thalamus (VAmc) close to the mammothalamic tract, which projects diffusely to layer 1 of virtually the entire cerebral cortex(Herkenham, 1980). Early electrophysiology experiments in the cat (before knowledge that these layer 1 projections existed) actually showed that single shock stimulation of this thalamic area caused an immediate activation of almost the entire cerebral cortex as would be expected from a central thought control signal(Hanbery and Jasper, 1953). Although the intralaminar nuclei had for decades largely been the focus of the layer 1 nonspecific projection (Jones, 2007), we now know that the intralaminar nuclei predominantly target layers 5 and 6 of the cortex (Jones, 2007; Herkenham, 1980). From the confabulation theory perspective, the intralaminar nuclei are quite likely involved with the behavioral triggering of action commands discussed in the next section. In addition to layer 1 projections, the VAmc nucleus of the thalamus also receives projections from both the basal ganglia and cerebellum (both highly involved in movement) giving this small thalamic area all the necessary axonal connections to function as the "alpha motor neurons of thought". We encourage the neuroscience community to test this hypothesis, and hope that it serves as an illustration of the predictive utility of confabulation theory.

2.3.4 Action Commands: skill knowledge and the origin of behavior

One of the most obvious aspects of brain function (and therefore one of the most consistently ignored) is that animals typically launch many behaviors every second they are awake. Most of these are microbehaviors (small corrective modifications or addenda to ongoing behaviors), but typically, major new behaviors are launched many times per hour, predicated on newly emerged events. Beyond simple reflexes (e.g., knee jerk) and autonomic reactions (e.g., digestion),

no understanding of how and why behaviors originate currently exists.

Confabulation theory proposes the conclusion action principle (Figure-2.6): every time a confabulation operation on any thalamocortical module reaches a conclusion, an associated set of action commands may be launched from a specific set of neurons. Action commands arise from neural fields that send axons towards subcortical structures (layer V pyramidal neurons). These action commands either launch behaviors immediately (when originating from layer 5b subcortical projetions) or suggest behaviors for further evaluation (when originating from layer 5a projections to the basal ganglia). Confabulation theory postulates that all non-reflexive and non-autonomic behaviors arise in this manner.

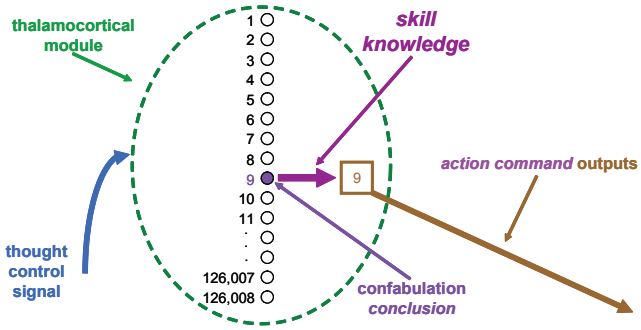

Figure 2.6: The conclusion action principle: hypothesized to be the origin of all non-reflexive and non-autonomic behavior. Here, a thalamocortical module (illustrated abstractly, in consonance with Figure-2.5, as an oval containing a list of the module's symbols) has successfully completed a confabulation operation (under control of its externally supplied thought control signal) and reached a conclusion (symbol number 9 as in Figure-2.5). Whenever a module completes a confabulation and reaches a conclusion it immediately causes a set of action command outputs to be launched (these outputs proceed to subcortical nuclei). The action command outputs that are launched are those which have been previously associated with the conclusion symbol via a subcortically managed skill-learning process (distinct from cortical knowledge link learning). All behaviors are caused by these action commands. The conclusion action principle is the fourth and last of the key elements of confabulation theory. Reproduced from (Hecht-Nielsen, 2007).

The mapping between symbols and action commands composes a different form of memory, termed skill knowledge (or procedural memory), that requires rehearsal and practice. As opposed to cognitive knowledge of facts and events (stored by knowledge links), skill knowledge is not directly consciously accessible(Squire and Zola, 1996). Skill knowledge is a learned association from the conclusion symbol neural field to the action command neural field within a thalamocortical module.

The neuroanatomical location and physiological properties of skill knowledge is very different from cognitive knowledge. First, as opposed to the module-to-module (symbol-to-symbol)

nature of knowledge links, the learned mapping from symbols to action commands lies entirely within a thalamocortical module. Second, unlike a cognitive knowledge link, which may be extremely robust if consolidated over many nights of sleep, skill knowledge is often fragile and short-lived. The impermanence of skill knowledge is required for rehearsal learning of skills (like playing a musical instrument), where gradually more competent skill knowledge needs to supplant earlier, less perfected, skill knowledge. Finally, there are separate learning mechanisms for each type of knowledge. Whereas the learning of cognitive knowledge requires the hippocampus and its related medial temporal lobe, the learning of skill knowledge requires other subcortical structures such as the basal ganglia, intralaminar thalamus, and basal forebrain.

The application of skill knowledge to the launching of action commands is not part of cognitive information processing per se (it comes into play only after each thalamocortical information processing operation has completed its job of reaching a conclusion). However, thought processes are dependent upon the thought control sequences coordinating confabulations in many thalamocortical modules. In the same way that movement sequences (actually, postural goal sequences) are learned, stored, and recalled, so are thought control sequences. These thought control sequences are controlled directly by action commands launched by thalamocortical modules. Therefore, thought (confabulation) begets action (action commands) and action begets thought in an endless cycle during wakefulness. The homunculus hiding behind a curtain pulling the control levers of the brain and body is thus exorcised.

2.4 Confabulation theory experiments: natural language processing

To glimpse the potential of confabulation theory to describe human level capabilities, consider the capabilities of the simple confabulation architecture shown in Figure-2.7. This particular architecture allows sets of three consecutive sentences from the same paragraph of a well-written newspaper story to be represented in terms of symbols. At the bottom level, each module has 63,000 symbols, representing the most common words and punctuation of English. When a sentence is entered, the symbol representing the corresponding word of the sentence is activated in each module. Words are entered in order from left to right and each module only has one active symbol at a time. Modules to the right of each sentence's ending period have no active symbols. The modules of the second and third levels of the architecture have symbols representing words, word phrases, and punctuation.

As tens of millions of such well-written sentence triples from 1990's-vintage newspaper

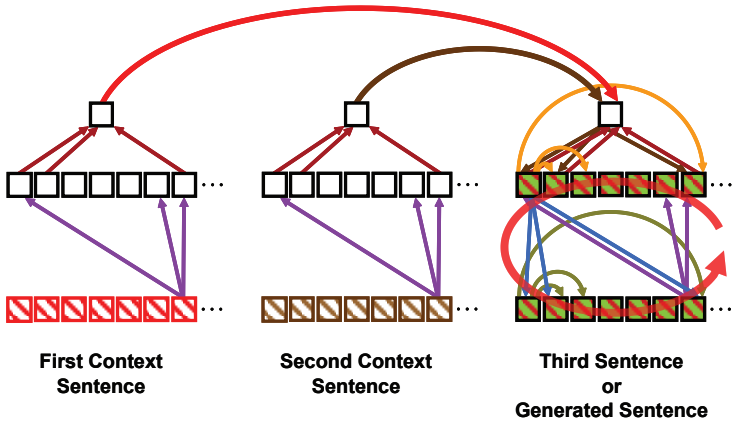

Figure 2.7: For (a)-(f) see figures below. This confabulation architecture (implemented on a computer) consists of hundreds of modules (each indicated by a square - only a few of which are shown) and thousands of knowledge bases (each illustrated by an arrow connecting one module to another - again, only a few of which are shown). This particular architecture likely captures elements of thalamocortical module connectivity in the human brain, but should not be viewed as a reproduction of known connectivity. With hard work, neuroscience research will reveal that connectivity in the near future.

stories are entered and symbols co-occur on the various connected modules, billions of knowledge links arise. Although this architecture is implemented on a computer, it is important to note that the formation of these knowledge links is consistent with the known anatomy in the human brain (see chapter 3).

Once this architecture has completed this "reading" (exposure to a huge amount of text), its "intelligence" can be explored. Consecutive pairs of novel sentences (ones not seen during learning) are read into the modules of the system's first and second sentences (the "context sentences"). The modules of the third sentence are then commanded to confabulate. The multi-confabulation swirling of that thought process, illustrated in Figure-2.7 by a red arrow, represents coordinated confabulations in many modules, which are interacting and mutually converging to single symbols. As each module converges to a single symbol the result is a plausible, although entirely made up (i.e. confabulated), sequence of words in place of a third sentence.

We emphasize that the storage of the knowledge links are consistent with anatomy, the convergent confabulation operation functions identically in each module and can be biologically implemented by a thalamocortical module, and the coordination of the confabulations in the multiple modules requires no more neural circuitry than is used to control the coordination of muscles. Therefore, the simulation of this architecture is extremely biologically consistent and should be viewed as a basic simulation of a human thought process.

As an example of one of these simulated thought processes, we enter two novel consecutive context sentences (obtained from the Detroit Free Press and never before seen by the architecture). The first sentence entered is shown in red. The second sentence is shown in brown. The word string produced by the confabulation process is shown in green.

"Several other centenarians at Maria Manor had talked about trying to live until 2000, but only Wegner made it." (first sentence shown in red - to match the color scheme of Figure-2.7) "Her niece said that Wegner had always been a character - former glove model , buyer for Macy's, owner of Lydia's Smart Gifts downtown during the 1950s and '60s - and that she was determined to see 2000." (second sentence shown in brown), "She was born in the Bronx Borough of New York City." (confabulated third sentence shown in green).
Figure 2.7(a)

Using the same color scheme, the bullet list below presents more examples of the operation of the confabulation architecture of Figure-2.7 (a good fraction of the outputs from randomly chosen fresh consecutive sentence pairs are of this high quality). Sentences are shown as Figures-2.7(b-g). These results were produced by Hecht-Nielsen's research group at Fair Isaac Corporation and were first presented at the Cognitive Computing 2007 conference.

He started his goodbyes with a morning audience with Queen Elizabeth II at Buckingham Palace, sharing coffee, tea, cookies and his desire for a golf rematch with her son, Prince Andrew. The visit came after Clinton made the rounds through Ireland and Northern Ireland to offer support for the flagging peace process there. The two leaders also discussed bilateral cooperation in various fields.
Figure 2.7(b)

Seeing us in a desperate situation, the Lahore airport authorities switched on the runway lights and allowed us to land with barely one to two minutes of fuel left in the aircraft, he said. At Lahore, Pakistani authorities denied Saran's request to accept wounded passengers and women and children, but they refueled the plane. Airport authorities said they were not consulted beforehand.
Figure 2.7(c)

Michelle strengthened from a Category 2 to a Category 4 storm Saturday, with winds reaching 140 mph, but it was expected to weaken before it reached Florida. The storm or its effects could

strike the Keys and South Florida tonight or early Monday, said Krissy Williams, a meteorologist at the National Hurricane Center in Miami. Forecasters warned residents to evacuate their homes as a precaution.
Figure 2.7(d)

But the constant air and artillery attacks that precede the advance of Russian troops have left civilians trapped in southern mountain villages, afraid to venture under the bombs and shells raining on the roads, Chechen officials and civilians said. Residents of the capital Grozny who had fled the city in hopes of escaping to Georgia, which borders Chechnya to the south, have been stuck in the villages of Itum-Kale, 50 miles south of Grozny, and Shatoi, 35 miles south of Grozny. Russian forces pounded the strongholds in the breakaway republic.
Figure 2.7(e)

A total of 22 defendants were convicted after the five-month trial of possessing explosives and plotting terrorist acts, but all were acquitted on charges that they were linked to the Al Qaeda terrorist network. Jordanian authorities now have a second chance on the Hijazi case. The defendants are accused of conspiring with the outlawed rebel group.
Figure 2.7(f)

Now, I must admit that I'm not so sure the Palestinians really wanted to reach a framework agreement, Eran said Tuesday. Eran wondered aloud whether the Palestinian strategy might be to negotiate as much land as possible in the remaining transfers, then declare statehood unilaterally - as the Palestinians have threatened to do before when talks bog down. Netanyahu said the Palestinians would be barred from jobs in Israel.
Figure 2.7(g)

The incident threatens relations between the Americans and Kosovo civilians, whom the peacekeepers were sent to protect after the 78-day NATO bombing campaign. We don't want them here to give us security if they are going to do this, said Muharram Samakova, a neighbor of the girl's family. NATO has struck a military airfield near Pale.
Figure 2.7(h)

These results suggest that the computer simulation must somehow be applying a deep knowledge and understanding of the general functioning of the world. The architecture is capable of linking context across two previous sentences and applying that context to generate a

cogent third sentence. Additionally, the third sentence produced is a grammatically correct, well structured third sentence, yet there are no rules for language in the system. In fact, the identical architecture, will produce cogent third sentences in any language given training data from that language. Interestingly, when born, all humans have the same grossly fixed brain architecture, yet each can learn any language provided that they are continually exposed to it. The emergence of grammatically correct and cogent language production in a biologically consistent architecture provides significant evidence that confabulation theory is in fact describing, in a fairly complete way, fundamental mechanisms of human (mammalian) language function.

2.5 Conclusion

One striking feature about these confabulation architectures is the extremely large quantity of knowledge (one knowledge link is a single item of knowledge) they employ and the effectiveness with which confabulation architectures exploit this knowledge in demonstrating astonishing intelligence. Amazingly, this performance is achieved in a biologically plausible way and lacks traditional "rules" or "algorithms". Since language, speech recognition, and even visual processing systems have been implemented to varying degrees with nothing more than modules, symbols, knowledge links, and thought control signals, we know a wide variety of cognitive tasks can be carried out by confabulation architectures. More sophisticated processes involving interactions between many sensory and behavioral modalities is possible with confabulation architectures. Such tasks (along with movements) can dynamically interact and be selectively activated by hierarchies of modules. Thus, enormously powerful ensembles of thought processes and movement processes can be rapidly selected and integrated.

The most important aspect to this dissertation is the enormous consistency and explanatory framework that confabulation theory provides for understanding brain. The cerebral cortex seems to perform all human capabilities without significant variation in its fundamental structure between cortical areas. This structure also appears to be consistent across all mammalian species. With a definitive hypothesis on the general organization and functional processes underlying cognition, we are ready to test those hypothesis against the experimental details of neuroscience.

Chapter 2, in full, is a modified reprint of the material as it appears in Physics of Life Reviews, Solari, S., Smith, A., Minnett, R., Hecht-Nielsen, R., Elsevier, 2008. The dissertation author was the primary author of this material.

3

Fundamental mammalian non-primary cortical (including thalamic and basal ganglia) anatomical circuits

3.1 Introduction

The past 150 years has produced almost all our detailed neuroanatomical knowledge of the brain (Clarke and Dewhurst, 1972; Finger, 1994); therefore, neuroanatomy is surprisingly one of the youngest, yet most fundamental, scientific fields today.

In nervous systems, anatomical structure determines function. Information is communicated from neuron to neuron through axons as action potentials. In order to understand and determine the function of a brain, we must understand its structure and therefore its neuroanatomy. The classic neuroanatomists in the late 19th and early 20th centuries understood this and were capable of determining many functions of the brain simply through anatomical analysis, well before more advanced experimental techniques existed to verify such hypothesis. For example, the anatomical projections from the eye to the lateral geniculate nucleus of the thalamus to calcarine sulcus of the occipital lobe was clearly used to predict the visual function of primary visual cortex by Meynert. However, most of these original anatomical studies were entirely based on golgi and nissl stains in addition to the ability to trace axon fascicles through

dissection. These techniques are all incapable of determining precise axonal projections between cortical layers and/or other nuclei in the brain. That limitation and the more widespread use of electrophysiology and EEG in the 1930's lead to a great shift in focus in neuroscience, away from studying neuroanatomy and toward experimenting with stimulus response paradigms in nervous systems. To illustrate this point, the last comprehensive histological analysis of the human cerebral cortex was performed in 1925 by Constantin von Economo and Koskinas(von Economo, 1929). 1925! Even though many of the modern day anatomical tools had not even been invented, von Economo's work remains the pre-eminent single body of work on the basic neuronal structure of the human cerebral cortex. Even more surprisingly, von Economo's work is hardly known to non-neuroanatomists in neuroscience, because only a few of his books still exist in libraries. Only recently, in 2007, were the slides he developed of the human cerebral cortex republished(von Economo and Koskinas, 2007). However, the new publication, coming in at 22lbs, ∼2ft x 2ft, and costing over $1000 will hardly become a common sight on neuroscientists shelves. After 80 years, with all that is unknown, it is indefensible that neuroscience has not encouraged an even more comprehensive study of the basic anatomy of the human cerebral cortex; the dominant structure in our brain that we as a human species are attempting to understand.

A re-emergence within neuroanatomy to understand cortical afferent and efferent projections occurred in the late 1960's(Graham and Karnovsky, 1966). Retrograde and anterograde tracing techniques were developed that enabled the ability to inject a localized tracer into one region of the brain, and to visualize the neurons that projected to, or synapses receiving projections from that particular injected region. Although a great deal of new anatomical work was done with tracers, the introduction of even newer electronic investigative tools, such as functional magnetic resonance imaging (FMRI), once again lead to a shift away from the focused analysis of neuroanatomical connectivity and structure. The unfortunate result was that many individual tracing studies were performed, but no study was single handedly comprehensive, nor were the individual studies ever comprehensively integrated into a single unified picture of the entire cerebral cortex, although some consolidations of information have occurred(Schmahmann and Pandya, 2006).

The purpose of this chapter is to present a detailed analysis of non-primary, primate, cortical and associated nuclei connectivity based on a comprehensive survey of the neuroanatomical data that exists today. More than a review, we attempt to integrate anatomical studies across many primate brain regions into a single unified viewpoint of the basic cortical "blueprint" that appears to form the underpinning of all cortical connectivity. We are not suggesting that our viewpoint should be considered the end all model. That model will eventually emerge once a single comprehensive project maps the brain. Our viewpoint is simply one of the best neuroanatomical

model approximations existing today based on available data.

The focus on non-primary cortical areas is critical. Non-primary cortices appear to have the same homotypical organization everywhere in the brain(von Economo, 1929; Brodmann, 1909). In contrast, the primary cortical areas are specialized versions of the basic non-primary blueprint(Northcutt and Kaas, 1995; Brodmann, 1909). In terms of cortical organization, primary cortical areas contain features that do not exist in non-primary cortices; such as, a sub-divided layer 4, layer 4 spiny stellate cells, a lack of striatally projecting layer 5 neurons(visual and auditory only), dominant layer 4 thalamocortical projections. Due to the extensive literature presenting primary sensory cortices as the basic cortical model, many large scale cortical models contain these erroneous features(Izhikevich and Edelman, 2008). If we are to understand and correctly model the entire primate(human) cerebral cortex, it must be based on the correct anatomy.

In order to appropriately understand the basic blueprint of cortical organization, two additional sets of nuclei must be included; the thalamus and basal ganglia. Together, the cerebral cortex, thalamus and basal ganglia form the most critical components of cognitive information processing(Purves et al., 2004). Because these structures are so interconnected, understanding the organization of one requires understanding the basic organization and inter-connectivity of the others. The combined connectivity between the three form several distinct closed information processing loops; which we will later hypothesize to have very specific function (see chapter 6).

The typical approach to a review of cortical organization is to discuss as many studies as possible layer by layer, neuron by neuron without significant mention of the thalamus and basal ganglia(Cajal, 2002; Lorente de No, 1943; Nieuwenhuys, 1994; Bannister, 2005; Douglas and Martin, 2004; Thomson and Bannister, 2003). This makes it difficult to create a unified picture. Additionally, almost all reviews focus on, or utilize extensive results from primary sensory cortical organization (biased toward primary visual cortex), from many species. We adopt a slightly different approach and present an anatomical description of the cerebral cortex and associated structures divided into the major functional projections that are clearly repeated across the entire non-primary primate cerebral cortex. We have integrated across the many studies and present the basic classes of neurons that likely have a stereotyped function. We propose that these basic cognitive circuits form the blueprint of functional networks that are repeated throughout the entire cerebral cortex. We hypothesize that understanding the basic computational operations of these cognitive circuits is all that is needed to understand the fundamental computational operations that are repeatedly performed across/through the entire cerebral cortex, thalamus and basal ganglia.

A synthesized framework, as we are presenting, is only as good as the data that is

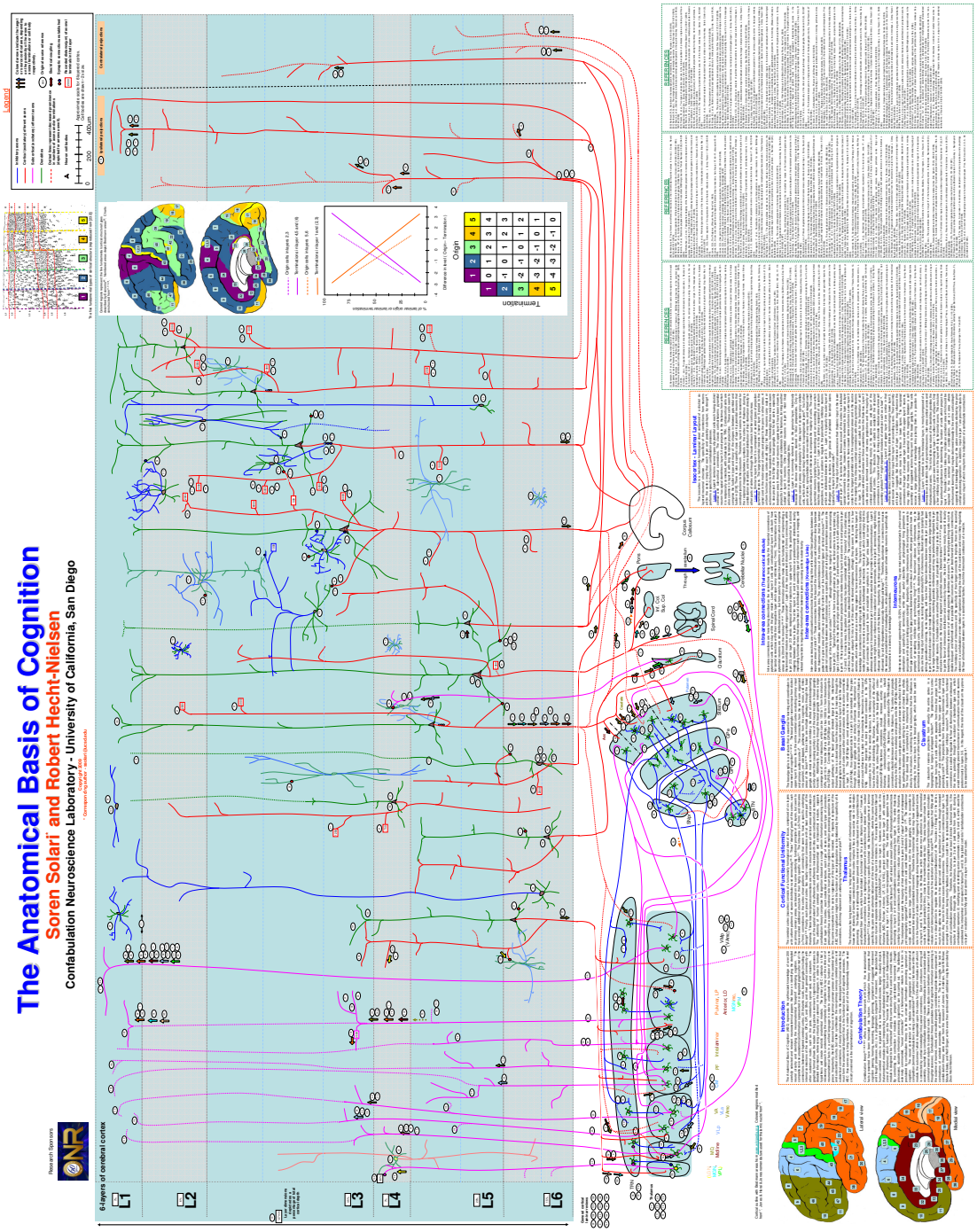

Figure 3.1: Anatomical Basis of Cognition (ABC) poster

synthesized and a picture is worth a thousand words. Therefore, we first developed a comprehensive visual poster, cataloging the primate anatomical connectivity from hundreds of high quality original anatomical studies(see Figure-3.1). This large poster displays the basic anatomical neuronal types for the cerebral cortex, thalamus, basal ganglia and their known projections, where each neuron and projection is labeled by original anatomical studies that were used to determine their existence and properties. The large poster contains significant extra detail, used to form our viewpoint, and is likely best viewed as a large (5ft x 4ft) printed poster or in the online pdf format where zooming is possible. The poster provides a tool for the inquisitive reader who would might want to locate additional information and/or question certain viewpoints or certain diagrammatic depictions. We attempt to follow the graphical reference format in our other figures by listing some of the references we consider most illuminating to particular neurons and projections.

3.2 Generalizations about cognitive circuitry

All brains follow general patterns of connectivity; however, for any "rule" or general principle of connectivity, there can be found an exception to the rule. This phenomena should not be used to discredit the ideas of generality. Exceptions to the rule seem to be a natural outcome of the development and evolution of the brain. Molecular gradients (see Figure 3.2) in the brain seem to exist as specific rules for connectivity(Striedter, 2005), however by the very nature of gradients, no connections will be fully precise, nor will neuron locations be exact. It is in this light that details are presented and inferences made about neuroanatomical principles. Our primary position is that the dominant projections (by number and/or density) and cortical neuron locations represent the general circuitry of the brain, and that this general circuitry should be the basis for models. Exceptions to the underlying basis of each cognitive circuit, similar to exceptions to 6-layered cortex (e.g. primary visual cortex) are viewed here as evolutionary specializations that are not necessary in understanding general brain function, but may be crucial in understanding details of specialized brain function.

Our deductive premise and approach adopts a certain perspective. Each neuron integrates and transforms synaptic input information into output action potentials; hence, neurons can be viewed as performing a functional transformation of synaptic input into action potential output. Assuming that molecular gradients generally determine source to target projections, the nuclei/layer location of a projection is the best indicator of which target neurons (and/or location, apical/basal dendrite, on the target neuron) a source neuron may be genetically programmed to influence and/or connect to. The properties of each synapse (location

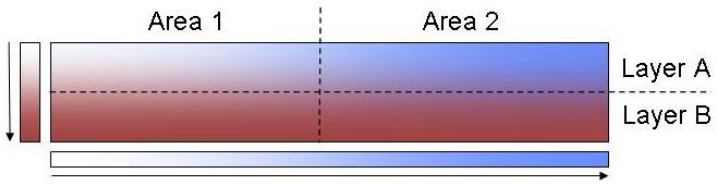

Figure 3.2: Depiction of orthogonal molecular gradients that explain both cortical layering development and cortical area differentiation(Striedter, 2005). Connectivity between layers and areas is then specified in a general fashion, but local connections are random and potentially overlapping.

on neuron/neurotransmitters/types of receptors/etc.) cause different effects in a target neuron. Therefore, in addition to viewing a single neuron as having a particular transformative function, we are also apt to view the location of the target synapses as providing a certain type of functional input to the target neuron. In this regard, if a source neuron projects axons to different cortical layers, the source neuron (or rather the information the neuron sends via action potentials) could play multiple different functional roles depending on which layer the information is going to. We are adopting the viewpoint that a single action potential may have multiple functional roles depending on where it goes. Hence, we argue that a type of projection likely has prescribed function in addition to a function prescribed to a type of neuron.

Finally, it is important to note that very little is actual known about neuron to neuron connectivity. Tracing studies alone can not be used to determine precise connectivity between neurons. They can only be used to determine where axons project. We use the term projection to imply that a neuron sends axons to a particular location and use the names of certain cell types as the recipient of the projection to emphasize certain neurons in those target locations. However, do to the limitations of the experimental techniques we can not conclude definitively that a particular neuron "connects" to another neuron. This subtlety is often lost (even in good reviews) when talking about neuroanatomy. In certain cases actual connections between neuron types have been established, in which case we use the term connection. But a projection and a connection are not the same.

3.2.1 Cortical functional uniformity and modularity

The cerebral cortex (isocortex) consists of a functionally homotypical sheet of tissue, comprised of six layers with uniform patterns of neuron distribution and projections(Brodmann, 1909; Lorente de No, 1943). Those "deviating" parts of the cortex, such as primary sensory areas, are based on the same underlying functional principle of six layers, but seem to have

evolved additional structure for their highly specific roles(Northcutt and Kaas, 1995). The intra-area (<5-7mm) and inter-area (>7mm) projections in the cortex show patterns that appear to rely on an underlying blueprint(Felleman and Van Essen, 1991; Barbas, 1986). Additionally, each piece of cortex has locally reciprocal projections with at least some part of the thalamus. These local reciprocal projections are similar in structure across all cortical areas and correlate (in terms of the spatial extent of the afferents and efferents involved) with intra-area cortico-cortical projections. Even so, all experimental evidence in the cerebral cortex, from lesion studies to electrophysiology to FMRI point to localized cortical modules on the order of a few mm^2 that are functionally disjoint(Catani and ffytche, 2005; Szentagothai, 1975; Tsao et al., 2006; Tanaka, 2003). Each module appears to process a distinct type of information reflecting the external and internal perceptions/behaviors of the individual, such as visual objects, language, executive plans, or movements (Penfield and Rasmussen, 1968; Tanaka, 2003; Goldman-Rakic, 1996; Grafton et al., 1996). Yet, the uniformity of the cerebral cortex, thalamocortical, and cortico-cortical structure supports the conjecture that cognition is based on a single uniform information processing scheme. Each layer in the cortex, each subcortical structure, each type of neuron, and each type of connection presumably has a specific functional role in carrying out the cognitive information processing operations that are the same in each cortical module, regardless of the type of data involved. Our viewpoint is that the cerebral cortex, thalamus, and basal ganglia, only perform a limited few cognitive information processing functions, and that significant insight into the function of each cognitive circuit can be deduced by the specificity of its projections(preferably connections), which represent an underlying functional blueprint or plan(Watakabe et al., 2006).

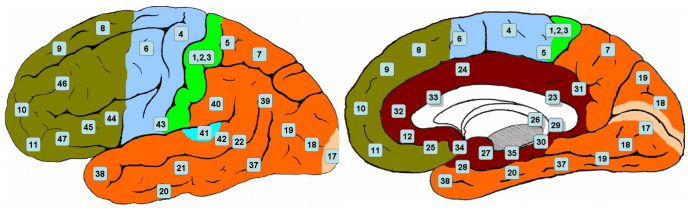

Figure 3.3: Brodmann areal numbering depicted on a colored lateral view (left) and medial view (right) cortical template. Colors were filled in to highlight significant cortical divisions when mapped onto thalamic nuclei and the striatum. The colors (dark green→frontal cortex, blue→pre-motor and motor cortex, orange→parietal and temporal cortices) will be used throughout the paper to designate projections that arise and/or terminate in these cortical regions. Primary sensory, auditory, and visual cortices are separately depicted in light green, cyan, and peach respectively. The uncolored cortical outline with brodmann numbering was reproduced from http://en.wikipedia.org/wiki/Brodmann_area.

3.3 Basic cortical layer organization

Based on morphology and efferent/afferent projections, we have divided the pyramidal neurons in the cerebral cortex into 9 basic functional types. We present an introduction to each cortical layer and the summarized cortical types for that layer. Interneurons are discussed independently. In addition, we include hypothesized roles in italics for each neuron as they might relate to confabulation theory.

3.3.1 Layer 6

We divide the neurons in layer 6 into three distinct functional groups depicted in Figures-3.5 and 3.6:

1. C_{6T} - Layer 6 thalamic projecting pyramidal neurons. *Cortical symbol candidates in confabulation.*

2. C_{6C} - layer 6 claustrum projecting pyramidal neurons. *May be involved in knowledge link communication or regulation between modules through the claustrum.*

3. C_{56} - Layer 5 and 6 cortically projecting pyramidal neurons, includes layer 6 spindle and fusiform neurons. *Participate in the communication of knowledge links, especially of the feedback type to other thalamocortical modules.*

Layer 6 contains several types of pyramidal cells each having distinct cortical and subcortical projections. The classic corticothalamic pyramidal neuron, C_{6T}, sends apical dendrites and large axonal collaterals to layer 3b. It also projects to the thalamus with collaterals to the TRN. Layer 6 corticocortical cells, C_{56}, are also common in layer 6 and layer 5 with long horizontal projections. These cells seem to comprise the fusiform type and many of the shapes in between pyramidal and fusiform (Soloway et al., 2002). One important feature and one that is used to group C_{56} neurons in both layer 5 and 6 is that their apical dendrite rarely extends past the upper portion of layer 5. The third class are layer 6 pyramidal neurons, C_{6C}, that sends distinct projections to the claustrum(Kowianski et al., 1998; Mettler, 1935).

3.3.2 Layer 5

We divide the neurons in layer 5 into three distinct functional groups, one of which is the same as in layer 6 depicted in Figures-3.4 and 3.6:

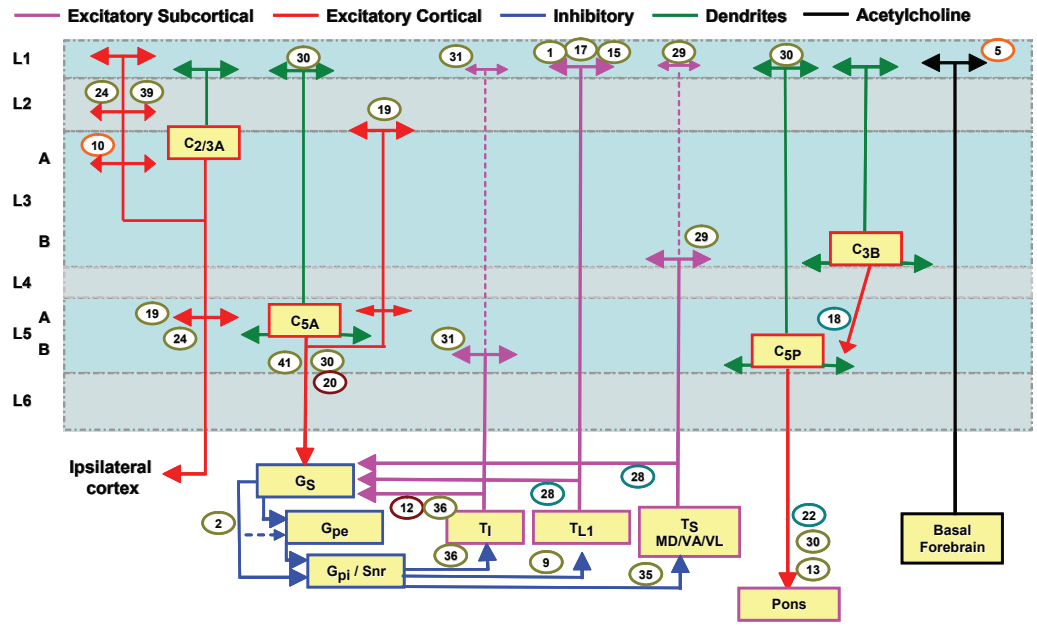

Figure 3.4: Cortico-basal ganglia-thalmo-cortical circuit. See Appendix 3.13 for numbered references. Colored reference labels designate the general cortical area (see Figure-3.3) that the reference corresponds to.

1. C_{5P} - Layer 5 pons projecting pyramidal neurons. *Transmission of direct actions/behaviors and/or behavioral predictions depending on the location of cortex. These compose the final output of action commands.*

2. C_{5S} - Layer 5 striatally projecting pyramidal neurons. *Suggested action commands that a confabulation is considering and provide confabulation state information to the basal ganglia for future control over confabulations.*

3. C_{56} - Layer 5 and 6 cortically projecting neurons, includes layer 6 spindle and fusiform cells. *Participate in the communication of knowledge links, especially of the feedback type to other thalamocortical modules.*

Similar to layer 6, layer 5 is composed of several types of projection pyramidal neurons. The large pyramidal neurons in lower layer 5, C_{5P}, project to the spinal cord in the motor areas of cortex and project to the pons in other areas. A uniform function across cortex would imply that these neurons carry output or action commands to their respective locations. Interestingly these neurons seem to also project to the SNr of the basal ganglia from frontal cortex, especially brodmann area 9(Mettler, 1935; Levesque et al., 1996), again suggesting a role in directly causing actions to happen, or inhibiting them by exciting the SNr. The second class of neurons in upper

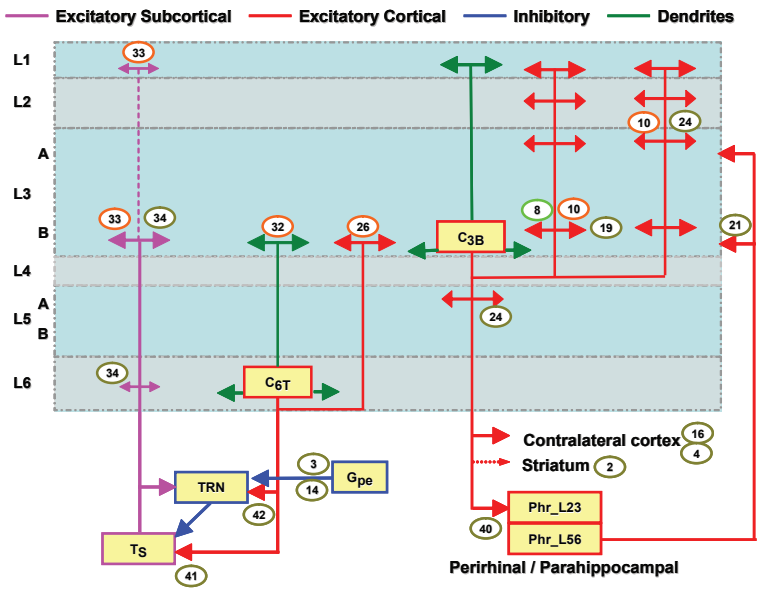

Figure 3.5: Thalamocortical circuit. See Appendix 3.13 for numbered references. Colored reference labels designate the general cortical area (see Figure-3.3) that the reference corresponds to.

layer 5, C_{5S}, project to the striatum with occasional collaterals to the thalamus. They form distinct projections with other C_{5S}, and do not seem to receive input from the C_{5P} cells(Morishima and Kawaguchi, 2006; Molnar and Cheung, 2006). Other pyramidal neurons in layer 5 participate in cortico-cortical projections and are considered as a single group of layer 5/6 cortical projection neurons, C_{56}.

3.3.3 Layer 4

Layer 4 neurons appear to be a single uniform type of small pyramidal neurons depicted in Figure-3.6:

1. C_4 - Layer 4 pyramidal neurons, note this does not include stellate cells since they do not exist outside primary sensory areas. *Feature detector symbols which are used to map feedforward input into symbols in layer 3 and possibly 6. Used to develop/self-organize symbols in layer 3.*

Layer 4 is commonly referred to as the granular layer, historically because of the dense small neurons stained in Nissl preparations. The term granular, historically refers only to the size of cells and NOT to any particular type of cell. Only in primary sensory areas, and especially in V1 does layer 4 contain spiny stellate cells, which are thought to relay information to supragranular layers(Callaway, 2005). In all other parts of cortex, spiny stellate cells are non-existent or very

rare, and instead small pyramidal cells along with interneurons compose the majority of cells. The small pyramidal neurons typically have a descending and an ascending axon which reaches upwards of layer 2(Meyer et al., 1989; Lund et al., 1981). We consider these small pyramidal neurons to form the blueprint for layer 4. In the human primary auditory cortex, small pyramidal cells form the majority of cells, where spiny stellate cells seem to form a minority(Meyer et al., 1989). The focus on of research on primary visual cortex has mis-represented the role of spiny stellate cells in information processing in the brain. In primary visual cortex (V1) they may comprise up to 95% of neurons in layer 4c, however their role in brain function appears to stop there. Spiny stellate cells are rare or non-existent in the rest of the brain (especially outside primary sensory cortices), and this fact is not well known or taught in neuroscience. We conclude the point with a quotation by Lund:

> There are no spiny stellate neurons in V2 in contrast to area V1 where they are the main neuron types of lamina 4. (Lund et al., 1981)

Layer 4 is the primary recipient of claustrum projections, and is in a position to regulate the parvalbumin inhibitory neurons which are distributed between layer 3-5. Layer 4 is typically the outer layer of Baillarger and thus contains a large plexus of myelinated horizontal axons typically from layer 3 pyramidal cells(Braak, 1980). Finally, layer 4 receives the majority of cortical feedforward input from layers 2/3, suggesting that layer 4 both in primary cortices and in other cortices is responsible for the input stages of developing any cell assembly perceptions in thalamocortical modules.

3.3.4 Layer 3

We divide the neurons in layer 3 into two distinct types depicted in Figures-3.4 and 3.5:

1. C_{3b} - Lower layer 3 pyramidal cells projecting to contralateral cortex and possibly hippocampal areas. *Conclusion symbols. These get mapped to C_{5P} cells to launch action commands.*

2. C_{3a} - Upper layer 3 pyramidal neurons long-range ipsilateral cortico-cortical projections. *Involved in the confabulation attractor network and form the major source of knowledge link communication between modules.*

The majority of cortical areas receive their thalamic input in the lower part of layer 3(Jones, 2007). In Lorente de No's famous account of the cortex he attributed special comment to this layer, however it seems he referred to this layer as layer 4a, which in the literature

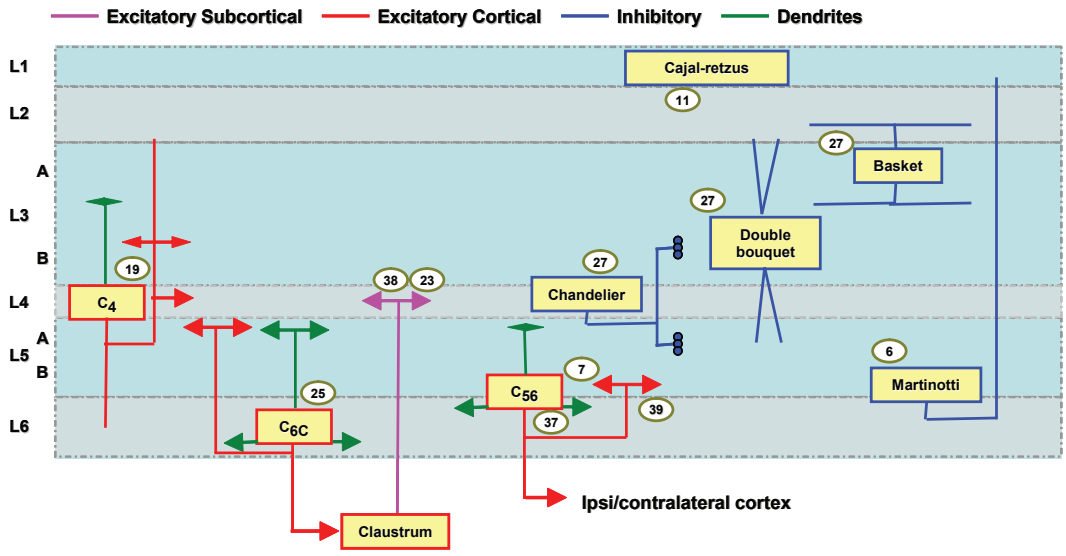

Figure 3.6: Cortical interneuron and claustrum circuit. See Appendix 3.13 for numbered references. Colored reference labels designate the general cortical area (see Figure-3.3) that the reference corresponds to.

seems to have caused some confusion. Lower layer 3 contains the largest pyramidal cells in the supra-granular areas and is the source of the majority of the intra-area cortical projections with other pyramidal neurons in layer 3. Layer 3 cells are also a dominant source of inter-area projections and the majority of colossal projections, terminating mostly in the same area and layer of the contralateral cortex. Finally, the neurons in layer 3 seem to comprise the majority of cortical projections to the perirhinal/perihippocampal cortices, implicating them as the store of perceptions used in hippocampally mediated short-term memory.

3.3.5 Layer 2

Layer 2 neurons appear to be a single uniform type depicted in Figure-3.4:

1. C_2 - Layer 2 short range cortico-cortical pyramidal neurons. *Feature detector symbols used to map feedback input into symbols in layer 3. Used to map top down influences back to symbols.*

C_2 and C_{3a} neurons are similar in that neither receive the direct projections from the specific thalamus that terminate in lower layer 3. Instead both receive much of their input from layer 1 or ascending cortical axons. C_2 neuron proximity with layer 1 suggests that layer 1 excitation may influence the activation of these cells more than those in lower layers. The reciprocal

projections that layer 2 and upper layer 3 have with the upper layer 5 neurons, *implies that these neurons are essential in the mapping between a symbol and suggested actions in C_{5a} being sent to the basal ganglia.* These cells receive a large proportion of cortico-cortical projections placing them in a position to establish significant associations (knowledge links) between thalamocortical modules. The fact that both C_2 and C_4 neurons have similar small granular morphology is likely not a coincidence and probably represents analogous function. Since C_2 neurons receive a larger proportion of cortical "feedback" projections, *they might likely be feedback feature detectors that then map to symbols in layer 3.*

3.3.6 Layer 1

Layer 1, referred to as the molecular layer, is mostly composed of a dense plexus of dendritic tufts of pyramidal neurons, inter-area cortical axons and thalamocortical axons. The rich and dense nature of the dendrites provides the opportunity for a given axon terminating in layer 1 to effect cells in almost all layers. Only a few inhibitory cells exist in the layer with long horizontal axons. This structure has profound significance for determining the function of cells which project to layer 1. *Possible functional roles may include utilization for external control of thalamocortical modules by layer 1 thalamic projections. Utilization by acetylcholine reinforcement learning signals delivered by the basal forebrain to the entire cortical depth supporting skill knowledge learning elicited by the prior cortical processing operation.*

3.3.7 Interneurons

Interneurons represent approximately 15-25% of cortical neurons. They are most easily characterized by their axonal arborization and specificity of projections, while other useful classifiers are physiological firing properties in conjunction with calcium binding protein staining(Alonso-Nanclares et al., 2005). Importantly, inhibitory interneurons are the only known neurons to form gap junctions and only form gap junctions between the same class of interneuron, where gap junctions have the property of spreading inhibition and synchronizing firing(Hestrin and Galarreta, 2005; Amitai et al., 2002; Gibson et al., 1999). In general, inhibitory gabaergic neurons are biased toward the upper layers of cortex(Gabbott and Bacon, 1996).

The dominant classes (although more diversity exists) of interneurons depicted in Figure - 3.6 and may be summarized as:

1. Basket - Most between layers 3 and 5. Smaller versions in layer 2. Named for the basket like shape of synapses they form around the soma of pyramidal neurons suggesting very powerful inhibition capabilities.

2. Chandelier - Mostly between layers 3 and 5. Named for the chandelier looking synaptic boutons that are exclusively formed on the initial axon segment of pyramidal neurons.

3. Martinotti - Typically found in lower layers with an axon that rises to the superficial layers.

4. Double bouquet - Vertically projecting dendrites and axons that span across layers.

5. Cajal-Retzius - exclusively found in layer 1

Basket cells are the majority of interneurons, consisting of approximately 50% of interneurons. Basket cells are typically fast spiking, parvalbumin staining, soma targeting, and have their highest densities between middle layer 3 and upper layer 5. A distinct division can be made between large basket cells and smaller basket cells, where large basket cells have large horizontal axonal arborization stretching possibly millimeters and may be preferentially targeted by layer 5 pyramidal neurons.

Chandelier cells are another important class of parvalbumin inhibitory neurons which provide exclusive terminations on the initial axon segment of pyramidal neurons.

The vertically projecting cells such as double bouquet and Martinotti cells (we include bi-tufted in these) are possible direct sources of inter-layer feed-forward or feed-back projections. Double bouquet cells typically have dendrites that stretch across multiple layers in the vertical direction and a similar projection of axons. Martinotti cells are unique in that they are typically found in the lower layers with dendrites in lower layers and send a vertically projecting axon, which often reaches layer 1.

Cajal-Retzius cells exist exclusively in layer 1 and are the only cells found in layer 1.

Detailed reviews(Gupta et al., 2000; Letinic et al., 2002; Markram et al., 2004; Gabbott and Bacon, 1996; Zaitsev et al., 2005; Defelipe et al., 1999) are essential to get the flavor of interneuron distribution and projections/connectivity. *Interneurons appear to be critical in the attractor network functions that a thalamocortical module must perform (such as synchronization and competition), especially in the balanced feedforward inhibitory projections and overall control over module excitation (see chapter 5).*

3.4 Intra-cortical module circuit

Both Figure-3.5 and 3.4 show pieces of the intra-cortical circuit. The intra-cortical circuit is here defined by the horizontal projections of pyramidal cells that stay within the cortical gray matter. The prominent source of intra-cortical projections are C_2, C_{3a}, C_{3b}, C_{5S}, C_{56} neurons. C_{3b} and C_{56} neurons both have horizontal projections within their respective layer

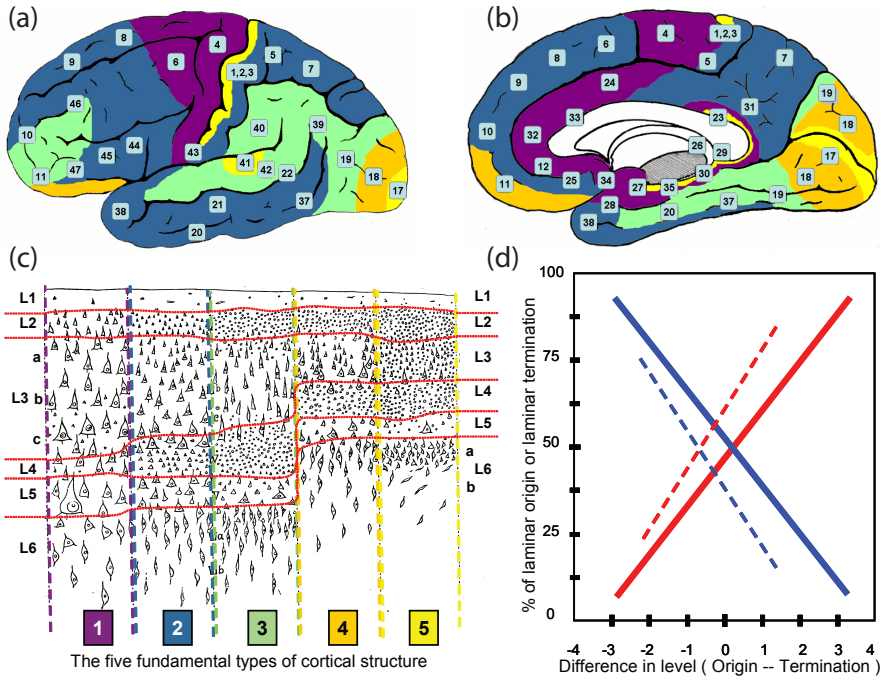

Figure 3.7: Synthesis of von Economo cortical types and laminar cortical projections. (a) Lateral view and (b) medial view human cortical regions colored to correspond to five fundamental cortical types depicted in (c). (c) von Economo five fundamental human cortical types(von Economo, 1929). 1=purple. 2=dark blue. 3=green. 4=orange. 5=yellow. The laminar distribution in the human cerebral cortex can be identified along a smooth numerical gradient, where 5 corresponds to primary-sensory cortices and 1 corresponds to primary motor cortex. Horizontal red lines distinguish layers across types. (d) Laminar (origin/termination) projection percentages predicted by numerical difference of cortical types in (c). von Economo cortical types are correlated with monkey tracing studies from (Barbas and Rempel-Clower, 1997; Barbas and Hilgetag, 2002; Medalla and Barbas, 2006) to produce graph. Dotted red = % neurons originating in layers 2,3. Dotted blue = % neurons originating in layers 4,5,6. Solid red = % synaptic terminations in layers 4,5 and lesser 6. Solid blue = % synaptic terminations in layers 1 and lesser 2,3. For example, a cortical type 2 projecting to cortical type 3 would have a difference of -1, and predict (looking at red) roughly 48% of the projections from the type 2 cortical area would originate from neurons in upper layers 2,3; and roughly 30% of synaptic terminations in the type 3 cortical area would terminate in middle/lower layers 4,5 and to a lesser degree 6.

which span many millimeters (∼7mm). The terminations of C_{3a} and C_{3b} pyramidal neurons are not distributed uniformly, but form patchy or stripe-like patterns of termination which comprise areas up to 20mm^2 in the monkey mostly in layers 2 and 3(Fujita and Fujita, 1996; Pucak et al., 1996; Levitt et al., 1993; de Lima et al., 1990). These termination patterns likely make up the fundamental cortical circuits resulting in a modular organization. C_{5a} neurons also have distant (∼2-3mm) reciprocal projections with C_2 and C_{3a} neurons forming the potential for a distinct mapping between the two layers(Kritzer and Goldman-Rakic, 1995; Levitt et al., 1993). The projections from other neuron types are predominantly distributed quite locally ($< 1mm$, implicating functional roles that require less of a distributed cell assembly type of interaction, and instead may relate to regulating information flow between layers or mapping of actions subcortically.

3.5 Inter-module (cortico-cortical) circuit

The inter-module cortico-cortical circuit is summarized in Figure-3.7. Inter-area projections are cortico-cortical projections that travel through the white matter fiber pathways between two separate areas of cortex(Schmahmann and Pandya, 2006). These projections form the basis of associations and information sharing between two thalamocortical modules. There is a great deal of specificity in cortico-cortical projections. Different populations of pyramidal neurons tend to project contralaterally as opposed to ipsilaterally(Soloway et al., 2002). The cytoarchitectonics of the cortex (as determined by von Econnomo) shows that the laminar pattern of a given area of cortex can generally fit within one of five fundamental types of cortical structure Figure-3.7(a,b,c)(von Economo, 1929; Walker, 1940). The pattern of projections between two cortical areas (as determined by Barbas in the monkey) shows a specific pattern of neuron layer origin and layer termination based on the cytoarchitectonic relationship of the two cortices as shown in Figure-3.7(d)(Barbas and Rempel-Clower, 1997; Rempel-Clower and Barbas, 2000; Barbas, 1986; Barbas et al., 2005b; Rockland, 1992; Medalla and Barbas, 2006; Van Essen, 2005).

As a note, Barbas never cites nor mentions von Economo in her papers and therefore does not appear to have made the connection, and was unaware, that the five types of cortical structure she studied in the monkey relate so significantly to the human, and in fact originated in the human(Barbas, 1986). Von Bonin adopted/translated von Economo's five cortical type description from the human into the monkey in 1947, which appears to be the source of the research branch in monkey literature(von Bonin and Bailey, 1947). Our figure is designed to make the correlation between the original human study and the monkey experiments.

Overall, the majority (i.e. more common type) of projections emanate from layer 2/3.

Typically if the projection is from a more granular (e.g. type 4) cortical area to a less granular (e.g. type 3) cortical area then the cells of origin are dominantly in layer 3 and terminate in layers 4,5 with collaterals in layer 6. If the projection is reversed then projection neurons reside mostly in layer 5 and some in 6 and project to layer 1,2 and 3. In visual areas, this pattern of projections has been correlated with the "hierarchy" of the cortical area, and thus the pattern of projections has been described as feedforward or feedback(Felleman and Van Essen, 1991). The cortico-cortical projections do not always uniquely fit into one of the above mentioned categories, although they do seem to come in a few varieties, which when looked at as a whole may appear to have projections to all layers. Identifying the layers of projections should imply subtle differences in their function. Layer 4 projections typically studied in parietal cortex are associated with a feedforward projections from layer 3, which could imply that this type of projections is being used to form or establish new cell assemblies in a higher cortical area based on lower level cell assemblies, whereas direct layer 3 to contralateral layer 3 projections would be direct cell assembly to cell assembly communication. Projections to layer 5 might have an influence on the selection of actions, whereas layer 6 projections might directly influence cell assembly interaction with the thalamus. Importantly, the laminar specificity and development of cortico-cortical projections is activity dependent, especially in feedback pathways(Price et al., 2006).

The structural areas of the primate cerebral cortex are also clearly defined by differences in gene expressions and these differences likely play a large role in cortico-cortical projections(Yamamori and Rockland, 2006).

3.6 Three functional thalamocortical projections

After reviewing vast amounts of anatomical and electrophysiology literature, we believe that the best functional description of thalamocortical projections divides them into three categories (see Figure-3.4):

1. T_S - Specific thalamic projections. Primarily targeting cortical lower layer 3. *Action potentials to cortex, represent candidate lists of symbols in a confabulation.*

2. T_I - Intralaminar thalamic projections. Primarily targeting cortical layers 5/6 (with layer 1 collaterals). *Responsible for both triggering action commands and influencing the mapping between C_{3b} to C_{5P} neurons.*

3. T_{L1} VAmc/VM/Layer 1 thalamic projections. Cortical layer 1 projections, with emphasis on the central VAmc/VM nuclei providing projections to virtually entire cerebral cortex.

Confabulation control signals to control activation and confabulations within thalamocortical modules.

The division of thalamocortical projections into three categories is new. Past work has typically emphasized either two types of projections, namely specific/non-specific(Lorente de No, 1943) or core/matrix(Jones, 1998), or four types of projections(Herkenham, 1980). The four projections include our three plus a fourth hybrid projection. We suggest here that the three projection viewpoint provides the best model to understand potential functional roles for the different thalamocortical projections observed anatomically. We hypothesize that there are three basic functional roles for thalamocortical projections. Here we revert to our principle that a single action potential may have multiple functions if it projects to two different layers; hence, we are describing our viewpoint regarding three classes of functional projections.

Before we embark on discussing thalamocortical and corticothalamic projections we would like to generally point out the "bible" of information that exists on the thalamus written by Edward Jones(Jones, 2007). The book and its references is unparalleled in its descriptive depth of the thalamus. We will refrain from constantly citing the book and just say that any information regarding the thalamus can almost surely be found there.

3.6.1 Specific thalamic projections

The specific thalamocortical projections are in general those that project to lower layer 3 in the cerebral cortex and whose projections are "specific", in that, the terminal arborization is localized to a small (<a few millimeters) region of the cerebral cortex. The general rule is that T_S nuclei receive reciprocal projections back from cortical C_{6T} pyramidal neurons from the localized area of cortex that T_S projected (see Figure-3.5).

Strong evidence suggests that T_S projections are topographically organized in accordance with the temporal development of the thalamus and cortex (Brysch et al., 1990; Hohl-Abrahao and Creutzfeldt, 1991). These studies show a continuous gradient of projections between the cortex and thalamus, such that two adjacent parts of cortex seem to have adjacent projections in the thalamus. This developmental hypothesis accounts for those instances where widely separated parts of cortex (frontal/parietal) project to the same thalamic nuclei (pulvinar)(Asanuma et al., 1985).

We should be careful of thinking of the thalamus and cerebral cortex as two distinct and separate functional entities. The most distinct interconnected information processing between the two is represented by the specific thalamocortical projections, and we hypothesize that cortical modules should instead be thought of as thalamocortical modules, which are uniform in their

information processing functions.

A primary sensory cortex disclaimer. Projections in primary sensory cortex target layer 4, but in all other parts of the cortex they target layer 3b(Jones, 2007). The early work by Cajal(Cajal, 2002) and Lorente de No(Cajal, 2002; Lorente de No, 1943), along with the unproportionate amount of research dedicated to primary sensory areas, seems to have ingrained layer 4 as the generally taught location of specific thalamocortical projections. As a historical note, Lorente de No's description of "layer 4a" appears to be in reference to what is now commonly referred to as lower layer 3b(Lorente de No, 1943). We suspect this may underly much of initial confusion that has since been propagated. Since 85% or more of the human brain receives specific projections to layer 3b we feel that should be viewed as the prototypical case, and corrected in introductory neuroscience textbooks(Purves et al., 2004).

3.6.2 Intralaminar thalamic projections

The intralaminar nuclei of the thalamus were originally thought to provide the majority of the "non-specific" layer 1 input in the cerebral cortex. Even today that seems to be the general notion in neuroscience. However, others, such as Herkenham, have demonstrated that all intralaminar nuclei distinctly project to layers 5/6 of the cerebral cortex of the mouse(Herkenham, 1980). See Figure-3.4 for a depiction. The most compelling evidence confirming this fact in primates comes from recent single-axon tracing studies in the monkey that undeniably demonstrate the majority of intralaminar (CM/PF) projections do principally terminate in layers 5/6(Parent and Parent, 2005). In addition, intralaminar thalamic projections (in the same study) are largely segregated into those that project exclusively to the cerebral cortex and those that project to the matrix portion of the striatum in the basal ganglia. Finally, although T_I projections are more diffuse in the cerebral cortex than T_S projections, they are still topographically mapped(Brysch et al., 1984) and are generally considered part of distinct cortio-basal ganglia-thalamo-cortical loops.

The fact that intralaminar nuclei project to the lower layers of cortex and are not the source of the layer 1 "non-specific" projection is significant. In the 1950's, research focused on understanding the cortical "recruiting response" due to intralaminar electrode stimulation(Hanbery and Jasper, 1953; Verzeano et al., 1953; Hanbery and Jasper, 1954). The recruiting response (most studied in cats) required pulsed thalamic stimulation of approximately a few HZ. After 10's of milliseconds, strong surface negative wave potentials would appear across wide spread cortical areas. The recruiting response, although more wide spread than T_S stimulation, was topographically organized (which makes sense because of the intralaminar topographic projection). The recruiting response was consequently attributed to the "non-specific" layer 1 thalamic pro-

jections described by Lorente de No in the 1940's. Today, with the knowledge that intralaminar projections target lower cortical layers, in addition to sending projections to the striatum, we can more accurately hypothesize that the recruiting response involved T_I-C_{5S}-basal ganglia-T_I and T_I-basal ganglia-T_I circuits. The explosion of basal ganglia research due to Parkinson's disease has also lead to many models of the basal ganglia as a pace maker, which would additionally point to the basal ganglia projections as the principle reason for the recruiting response observed.

These facts beg the question, "where do the non-specific layer 1 projections come from and what do they do"?

3.6.3 VAmc/VM/Layer 1 thalamic projections

One of the most perplexing (although unfortunately least studied) thalamic projections has been the non-specific layer 1 projections described by Lorente de No in the 1940's. As discussed, the intralaminar nuclei have long been thought to supply the layer 1 projection, which we now know is false. A simple paper in the 1979 by Herkenham essentially answered the question by stating that the small ventromedial, VM, nuclei in the mouse thalamus provided exclusive projections to the outer portion of layer 1 of the entire cerebral cortex with a decreasing density gradient from rostral(front) to caudal(back) (Herkenham, 1979, 1980). See Figure-3.4.

A second notable fact exists relating back to the "recruiting response" that was discussed above in the cat. This fact appears to have been lost in the literature and is so important that we quote:

> During the course of the present experiments, using the less precise stimulating electrode with tips separated by 1 mm., we have discovered a portion of the diffuse projection system which behaves quite differently from that which has previously been described. In the inferior medial portion of what is usually described as VA, at about stereotaxic planes Frontal 11 and 12, Lateral 1 to 3 and Horizontal -2 to +2, we have obtained diffuse short-latency cortical responses in response to a single shock. These diffuse responses, in addition to requiring no recruitment for nearly maximum voltage response, appeared after a latency of only 5-10 msec. We seem to be stimulating here, with the more widely spaced bipolar electrodes, a shortlatency diffuse projection system, which actually does not give true recruiting responses of the type presumably characteristic of the intralaminar system of the thalamus(Hanbery and Jasper, 1953).

These results were obtained before the knowledge existed of the exclusive layer 1 VAmc/VM nuclei projections. The fact that the response is due to a direct connection and elicits responses due to single shocks across virtually the entire cerebral cortex, implies that this local region could be used to "activate" or "control" different areas of the cerebral cortex. Since the electrodes used were separated by 1mm, the possibility exists that the VAmc/VM projection is still topographically mapped (in a small area) to most of the cerebral cortex, such that

small parts of VM/VAmc could activate/control local regions of the cortex independently. We believe a significant investment to test this hypothesis in the monkey would be very productive for neuroscience.

In the monkey, the results to date are more inconclusive, because no one appears to have specifically looked for a single small region of the thalamus that projects to the entire cerebral cortex. Hence, we use the loose term VAmc/VM/layer 1 to describe the hypothesized projection nuclei that almost surely also exists in the primate(human). As soon as the projection is definitively determined a more appropriate nuclei can be used as the naming convention. The ventral thalamus in the monkey does have projections to layer 1(Nakano et al., 1992). Clearly, layer 1 projections in the monkey are more complicated and may likely be distributed by several thalamic nuclei. This fact lead to the hypothesis of a layer 1 projecting thalamic matrix(Jones, 1998). However, it is likely that a small localized region of the thalamus, presumably part of VM/VAmc and near the mammothalamic tract provides significant distinct layer 1 projections.

The layer 1 projection is significant because of its potential for activation of a local area of cortex via the apical dendrites of cortical pyramidal cells. If thought were controlled, control over the activation and information processing in thalamocortical modules would likely have a centralized controller (much like alpha motor neurons control muscles), which we hypothesize is the role of the VM/VAmc small thalamic region projecting to layer 1. In more complex species such as the monkey and human, local feedback/feedforward control of thalamocortical modules by other thalamocortical modules could be implemented directly by the diffuse matrix of layer 1 connections throughout the thalamus. Yet, we hypothesize that both layer 1 type projections likely serve the same functional purpose.

3.7 Cortico-thalamo-cortical circuit

The general distinguishing features of the cortico-thalamo-cortical circuit is the reciprocal projections from the C_{6T} pyramidal neurons to T_S neurons (see left side of Figure-3.5). In general intralaminar nuclei do not receive heavy projections from the cortex. It is unclear whether the VAmc/VM nucleus with distinct layer 1 projections possibly receives a distinct projection from a local area of cerebral cortex (we hypothesize somewhere close to Brodmann's area 9 is probably the best location to look for that cortico-thalamic projection). As we mentioned the specific thalamic projections primarily target lower layer 3, while avoiding layer 4. The C_{6T} pryamidal cells have two distinguishing features, which provide additional evidence for their integration with specific thalamic projections. The C_{6T} pyramidal cells appear to be the only pyramidal cells with a significant apical dendritic tuft in lower layer 3 (Peters et al., 1997). All

other pyramidal cells with large apical dendrites send them to layer 1. Second, the C_{6T} cells also send large axonal projections to lower layer 3 (Lund et al., 1981). By terminating on apical dendritic tufts, specific thalamic projections could act functionally on the C_{6T} neurons in a similar fashion as the layer 1 thalamic projections act on other pyramidal cells with dendritic tufts in layer 1.

Given that the specific thalamic projections, and both C_{6T} axons and dendrites target layer 3B, we might also consider that C_{3B} neurons are a critical component of the the cortico-thalamo-cortical circuit. The C_{3B} neurons are interesting in that they are typically the largest of all cells above layer 4. The C_{3B} neurons are the significant source of the large callosal cortical projections, thus they are in a place to bind information inter-hemispherically (Jacobson and Trojanowski, 1977; Barbas et al., 2005a). The C_{3B} neurons also appear to be the dominant source of projections to the upper layers of perhirhinal/parahippocampal cortices(Witter et al., 1989). Therefore the cortico-thalamo-cortical circuit consisting of three major cell types; T_S, C_{6T} and C_{3B} neurons, forms a distinct circuit capable of thalamocortical oscillations, inter-hemispheric information binding, and additionally the source of the necessary information to form hippocampal mediated short term memory. We hypothesize that the thalamocortical circuit is the fundamental circuit underlying working memory as a thalamocortical oscillation, which would explain the prominence of layer 3 cells in working memory tasks(Lewis et al., 2002).

Numerous researchers have also pointed out projections from layer 5 in the cortex to the thalamus(Sherman and Guillery, 2006; Fujita and Fujita, 1996; Levitt et al., 1993; Kritzer and Goldman-Rakic, 1995; Kakei et al., 2001). These projections have been reported to come from lower parts of layer 5 when projecting to specific motor thalamic nuclei and from upper portions of layer 5 when projecting to the intralaminar nuclei. The projections are distinct from C_{6T} projections as they typically do not target the same specific thalamic nuclei that send axons to their cortical location. The layer 5 projections have additionally been suggested to be the source of large synaptic boutons from cortical projections to the thalamus, and hence of a "driver" type of information(Sherman and Guillery, 1996). Unfortunately, experiments have not definitively answered which layer 5 neurons (C_{5P} or C_{5S}) neurons send their axons to the thalamus. Although, most evidence points to upper layer 5 projections to the thalamus, therefore as collaterals of C_{5S} neurons(Catsman-Berrevoets and Kuypers, 1978; Rockland, 1996). For now the best approach would be to infer from the general source of cortical layer, upper or lower layer 5. Notwithstanding the sparse evidence, the layer 5 projections do appear to be a general, albeit weak rule of thalamocortical projections. We hypothesize the function of these layer 5 projections is most clearly understood by which of the three thalamocortical projection types is targeted, since each of those thalamocortical projections likely has an independent function. Anatomically

the layer 5 projection's function seems separate from the specific cortico-thalamo-cortical circuit.

3.8 Cortico - perirhinal/parahippocampal - cortical circuit

The cortico-perirhinal/parahippocampal-cortical circuit is the axonal highway upon which all short-term memory is formed and by which short-term memory is consolidated into cortico-cortical long-term memory(Squire and Zola, 1996; Squire, 2004; Eichenbaum, 2000). The Phr cortices are located deep in the temporal lobe surrounding the hippocampal formation and lesions to the Phr cortices result in severe deficits in short term memory(Squire and Zola, 1996). They are essentially the interface between the major association areas of cortex and the hippocampus, and are therefore in a place to form the necessary indexing of associations that must occur to store memories. In that regard, we might consider the circuit important from an anatomical standpoint. Even so, little anatomical research has been invested in understanding the afferent/efferent layer projections between the cerebral cortex and parahippocampal/perirhinal (Phr) cortices in primates. Research on the hippocampus has elucidated quite a bit of specificity of the reciprocated Phr, entorhinal, and hippocampal projections. The general topographic connections between association cortices and the Phr are well mapped, but the actual specificity of projections between associative cortices and Phr cortices remain vague at best, especially when we attempt to decipher the exact cortical layer where the projections terminate or emanate(Burwell, 2000; Lavenex et al., 2002; Witter et al., 1989). As shown on the right side of Figure-3.5 the best evidence suggests that layer 3 cells project to the upper layers of the Phr and receive reciprocal projections back from the lower layers of the Phr to which they projected.

These projections suggest that perceptions that must be stored in short-term memory in the hippocampus likely reside in lower layer 3. The reciprocal projections back to the same cortical layer additionally suggest that those perceptions can be reactivated through the same Phr and hippocampal circuitry. The projections fit well with the idea that the cortico-thalamo-cortical circuit maintains working memory perceptions, then these perceptions are transmitted to the Phr and ultimately hippocampus to be "bound" into a short-term memory. The activation of one perception can then re-activate the bound associations in the hippocampus and the reciprocal projections back to cortex can the recover the memory back into a thalamocortical working memory oscillation.

3.9 Cortico-basal ganglia-thalmo-cortical circuit

Figure-3.4 depicts the main projections from the cortex to the basal ganglia. Every area of the cerebral cortex projects to the basal ganglia, except two. Primary visual cortex and primary auditory cortex do not project to the basal ganglia(Borgmann and Jurgens, 1999). This fact alone is one reason to avoid using primary cortices as a basis for the prototypical blueprint of cortical circuitry. The projections from the cerebral cortex come from the C_{5S} neurons in the upper part of layer 5. (Yeterian and Pandya, 1994; Saint-Cyr et al., 1990; Kemp and Powell, 1970; Kunishio and Haber, 1994). In the mouse, these neurons have clearly been shown to be a separate population from the C_{5P} neurons and often form reciprocal projections with each other(Thomson and Morris, 2002).

Based on the neuroanatomical projections, we can decompose the basal ganglia into the following grouping of homotypical functional nuclei:

1. G_S - Striatum - receives input from the cerebral cortex and dopaminergic "reward" input from the SNc and excitatory input from the same thalamic nuclei that the GPi/SNr project to. Sends inhibitory output to the GPi/SNr with collaterals to the GPe and exclusively to the GPe.

2. G_{pe} - External globus pallidus - receives input from the striatum and STN. Sends inhibitory output to the GPi/SNr, STN, and a region of the thalamic reticular nucleus (TRN) surrounding the ventral thalamus (notably close to VAmc/VM).

3. G_{pi} - Internal globus pallidus/Substantia nigra pars reticulata (GPi/SNr) - receives input from the striatum and STN. Sends tonic inhibition predominantly to the T_I thalamus, and also to the T_{L1} thalamus, and certain ventral T_S thalamic nuclei.

4. STN - Subthalamic nucleus - receives input from the cerebral cortex, (C5P) neurons. Sends excitatory input to the GPe from one population and GPi/SNr from another population of neurons.

5. SNc - Substantia nigra pars compacta - provides dopaminergic input to the striatum.

In general the basal ganglia receives topographic projections from the entire cerebral cortex, which has lead to the notion of separate functional loops through the basal ganglia(Smith et al., 2004, 1998; Haber, 2003). We differ in our assessment of the anatomical facts and hypothesize that the pathway through the basal ganglia has a single uniform function everywhere in the basal ganglia, with the only difference being the cortical source of information that is operated on.

In this light, the three main structures of the striatum (caudate, putamen, and nucleus accumbens) can be considered as the same functional element, namely the striatum. Similarly, GPi and the SNr both receive (from the striatum and GPe) and send (to the thalamus) analogous projections. The neurons in these structures also have analogous physiological properties, one of which is sending tonic inhibition to their targets. Therefore, from a functional standpoint, we group these structures into a single functional element, namely the GPi/SNr.

The striatal inputs from the SNc are significant because the determination of the SNc appears to be the cause of parkinson's disease, and therefore a great deal of additional research not mentioned here has been invested into this pathway.

The pathway through the basal ganglia involves a large convergence of neural interaction, which can be 100 to 1 from the striatum to GPe, and 10 to 1 from the GPe to GPi/SNpr, thus 1000 to 1 from the striatum to the GPi/SNpr.

Considering the GPi/SNpr are the dominant motor/cognitive output pathways of the basal ganglia and directed largely at the intralaminar thalamus, there is a closed loop from C_{5S} neurons, through the basal ganglia to the T_I nuclei and back to layer 5 of both the original projecting area of cortex and other cortical areas. *These projections could be used to trigger action commands in C_{5P} neurons.* The GPi/SNpr also send projections to the ventral thalamus (VL/VA/VAmc/VM), thus suggested actions in C_{5S} can indirectly influence the information (symbols in competition) in the motor(VL) or more cognitive(VA) T_S thalamic nuclei, or could be used for the activation (via disinhibition) of thalamocortical modules via layer 1 in the case of T_{L1} nuclei.

One of the most significant and distinguishing aspects of the main basal ganglia pathway is that it is almost entirely inhibitory. Since there are multiple pathways through the basal ganglia and most connections are inhibitory, the effect of disinhibition(allowing activity rather than causing activity) is one of the key principles in basal ganglia function. We do not focus on the traditional view of "direct" and "indirect" pathways through the basal ganglia(Haber, 2003), and rather focus on the fact that cortical stimulation of the striatum likely "allows something to happen" or "prevents a previous allowance for something to happen". The functional question then simply becomes "what happens?". Considering the tonic inhibition that the output nuclei of the basal ganglia exert on the thalamus, again two basic phenomena are likely: 1) Allowing or triggering something to happen through disinhibition of the thalamus, or 2) Preventing or stopping a triggered action that is going to occur or has occurred. The correlation of these events with the type of dopamine receptors in the striatum is likely the basis for the ability to learn to start and stop actions(Herrero et al., 2002). With these two principles in mind we look at the possible effects that cortical C_{5S} excitation would have through three pathways in the

basal ganglia:

- $C_{5S} \rightarrow G_S \rightarrow G_{pi} \rightarrow T_I/T_{L1}$ connections should increase activity in the thalamus. This pathway is associated with D1 dopamine receptors in the striatum. *The pathway allows or triggers action by disinhibiting the tonic inhibition in the thalamus.*

- $(C_{5S} \rightarrow G_S \rightarrow G_{pe} \rightarrow G_{pi} \rightarrow T_I/T_{L1})$ or $(C_{5S} \rightarrow G_S \rightarrow G_{pe} \rightarrow STN \rightarrow G_{pi} \rightarrow T_I/T_{L1})$ connections should decrease activity in the thalamus. This pathway is associated with D2 dopamine receptors in the striatum. *The pathway prevents or stops a triggered action by reducing the inhibition of (and/or increasing the excitation of) the G_{pi} output nuclei thus increasing inhibition in the thalamus.*

- $C_{5S} \rightarrow G_S \rightarrow G_{pe} \rightarrow TRN \rightarrow T_{L1}$ connections should decrease activity in the thalamus. The projection is focused on the ventral thalamus. *The pathway prevents or stops a triggered action from happening through indirectly increasing inhibition in the thalamus.*

The cortex also projects directly to the STN, which is excitatory and projects onto the GPi/SNpr, therefore excitation in the cortex can bypass the striatum and reverse the effect of the first two pathways described above providing for a diversity of closed loop effects. The final closed loop is completed by the intralaminar thalamic projection directly to the striatum.

Another potentially significant but rarely mentioned projection is the GPe projection to the TRN of the ventral thalamus(Hazrati and Parent, 1991; Asanuma, 1994; Gandia et al., 1993). The GPe sends inhibitory projections to the TRN, thus the $C_{5S} \rightarrow G_S \rightarrow G_{pe} \rightarrow TRN \rightarrow T_{L1}$ projection could be used to gate/control thalamocortical information processing, or control signals emanating from T_{L1} nuclei.

3.10 Other sub-cortical circuitry

3.10.1 Cortico-claustrum

The claustrum receives projections from virtually the entire cortex in a topographic, but largely overlapped fashion(Edelstein and Denaro, 2004). See Figure-3.6 for a depiction. The projections from cortex originate from C_{6C} neurons, which are distinct from C_{6T} neurons(LeVay and Sherk, 1981; Katz, 1987). Occasionally collaterals of C_{5S} neurons are found in the claustrum(Parent and Parent, 2006). The projections from the claustrum terminate in layer 4 and seem to preferentially target inhibitory neurons(LeVay, 1986). The claustrum's functional connections are suited to regulate the flow of information between wide areas of the cortex, potentially

through the excitation of chandelier type cells, which would immediately prevent transmission of communication from active neurons predominantly in layers 3-5. In this regard, the role of the claustrum may prove to be essential for cognitive information processing(Crick and Koch, 2005).

3.10.2 Basal forebrain

Acetylcholine is found in only three basic populations of neurons. 1) Motor neurons, 2) Interneurons in the striatum. and 3) The basal forebrain (including the nucleus of Meynert). Acetylcholinesterase staining typically stains layer 1 of most cortices, therefore the B_F projection appears to primarily target layer 1 of most of the cortex(Bigl et al., 1982). See Figure-3.6 for a depiction. We focus here on the basal forebrain because many pieces of indirect evidence suggest that the basal forebrain acetylcholine projection may form a special type of learning signal that reinforces the mapping between C_{3B} and C_{5P} neurons. When staining in monkey and human cortex for acetylcholinesterase, it appears that C_{3B} and C_{5P} neurons are preferentially stained, suggesting their prominent utilization of acetylcholine(Hackett et al., 2001; Bravo and Karten, 1992). An additional correlation exits between these pyramidal types, because there also appears to be a preference for direct synaptic connections between C_{3B} and C_{5P} neurons(Kaneko et al., 2000; Thomson and Deuchars, 1997; Thomson and Bannister, 1998). Finally, basal forebrain lesions "abolish cortical plasticity associated with motor skill learning"(Conner et al., 2003).

Together, the evidence suggests that the acetylcholine delivered by the basal forebrain is critical for the learning of a mapping between C_{3B} and C_{5P} neurons, and that this mapping is the source of cortically learned behaviors and/or skill learning (i.e. action commands).

3.10.3 Cortico-pons

The projections from C_{5P} neurons in primary motor cortex travel to the spinal chord and synapse with motor neurons. These projections are the most direct cortical pathway for the instantiation of physical movement and thus physical behavior. We can, as a result, hypothesize that these neurons code directly for actions/behaviors. By extension, and through the uniformity of cortical function, we can hypothesize that the C_{5P} projections from other thalamocortical modules also directly code for actions/behaviors. However, the C_{5P} neurons in other thalamocortical modules tend to project to the pons as opposed to the spinal chord. The pons has many interwoven nuclei that have many functions and an enormous variety of projections, therefore the effect of C_{5P} projections is likely related to the cognitive dimension that a thalamocortical module processes.

We hypothesize that the C_{3B} to C_{5P} to pons projections essentially learn the mapping

to predict/execute "good" actions and behaviors. C_{3B} represent the perceptual state of the individual. C_{5P} represent the actions/behaviors predicted by that perceptual state. The information sent to the pons has the ability to effect a wide variety of states in the body and brain. We suggest that many the signals sent to the pons have two basic functions. Some are used to induce or move toward certain intrinsic internal states, similar to providing inputs to muscles to move to a physical state. The second type of information is likely predictive signals that are meant to "cancel out" any other internal/external sensory stimuli that arrive at the pons. In this way, the action/behavior signals in the brain are used to develop a working model of the individual's universe, when the model is correct signals "cancel out" in the pons and there is no need to change, because no error signals exist. When the cortical predictive model is incorrect, "error" signals will be produced in the pons alerting the rest of the system to change or do something different.

3.11 Conclusion: a unified theory and anatomical model

Here we summarize our findings by categorizing the previously described the role that neurons and neuroanatomical circuits in the four major components of confabulation theory. Figure-3.8 presents a unified anatomical model with confabulation theory. The four confabulation theory components are 1) Knowledge links; 2) Confabulation; 3) Action commands; and 4) Confabulation control, which are color coded and correspond to the populations of neurons discussed in the text.

3.11.1 The anatomy of knowledge links

The knowledge link circuits are shown in yellow in Figure-3.8. Knowledge links are communicated through associations between cell assemblies. Our multi-associative memory (chapter 4) explicitly describes the underlying mathematics capable of implementing these associations with very sparse connectivity. Knowledge links are essentially consolidated into direct long term cortico-cortical connections after first being temporarily stored in the hippocampus (chapter 6). Anatomically, the C_{3B} neurons form conclusion symbols and send "winning" symbols to the hippocampus for binding. After being bound in the hippocampus, the C_2, C_{3A}, and C_{56} neurons corresponding to the associated symbols are reactivated during sleep. Associations then form between these neurons in one module and the associated symbols in other modules, through two processes: 1) the strengthening of existing synapses and 2) the slow axonal growth of synapses

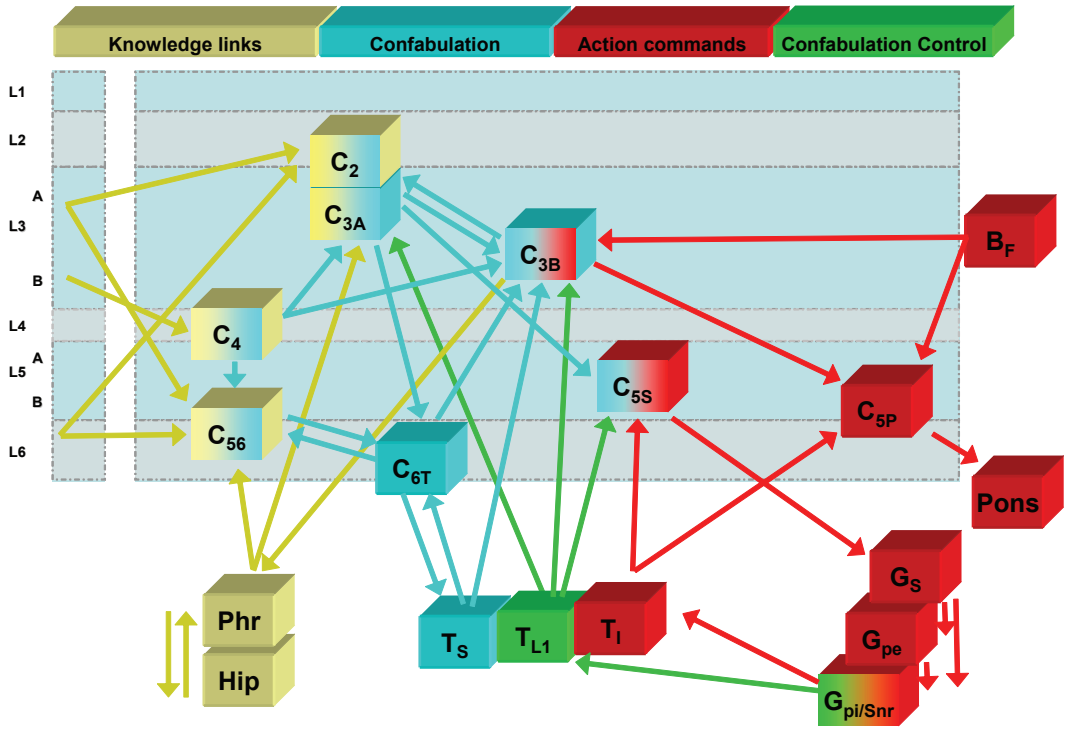

Figure 3.8: Confabulation theory anatomical model. The 4 basic circuits involved in confabulation theory 1) Knowledge links [yellow]; 2) Confabulation [blue]; 3) Action commands [red]; and 4) Confabulation control [green], and the neuronal populations that are used to implement each. Each neuronal population is described in the text.

followed by strengthening. The effect of slow growth likely contributes significantly to the robustness of the association and explains the time delay for consolidation.

3.11.2 The anatomy of a confabulation

The confabulation circuits are shown in blue in Figure-3.8. The confabulation circuitry essentially implements the winner-take-all competition among symbols (see chapter 5 for specific details). All the neurons involved in knowledge links are also involved in a confabulation through recurrent connections; however, the C_{3B}, C_{6T}, and T_S neurons are the main drivers of the attractor network and ultimately are the definitive symbols that represent confabulation conclusions. Even more precisely, the C_{3B} neurons with their projections to the hippocampus and to C_{5P} neurons likely represent the definitive conclusion of a confabulation. A confabulation likely involves T_{L1} excitatory input to control the confabulation process, at which point a thalamocortical oscillation is induced which contains persistent cell assembly excitation. The neurons involved in knowledge links then receive and transmit the current "winning symbols" to similar

neurons in other modules. That information is integrated and context is then applied to the symbols in competition in the thalamocortical oscillation. The T_{L1} input ultimately determines how many symbols should be active at any given moment in time. The hippocampus then records associations between C_{3B} symbols in different modules for later reactivation.

The input from C_{5A} neurons to the striatum is likely used to provide information about the state of confabulations in certain modules. This information is then used by the basal ganglia to control the timing of confabulations and action commands in many modules.

3.11.3 The anatomy of action commands

The action command circuits are shown in red in Figure-3.8. Action commands are ultimately determined by the mapping between C_{3B} and C_{5P} neurons. The firing of C_{5P} neurons would be considered the action commands themselves. The mapping is hypothesized to develop through an acetylcholine reinforcement signal from the basal forebrain. The receipt of the reinforcement signal will tend to help strengthen reinforce the co-activation of winning symbols and action commands that caused the reward signal.

We also hypothesize that action commands must be triggered by the T_I nuclei. In this way, a module can undergo a confabulation through T_{L1} activation, but not launch action commands. Action commands are only launched if sufficient excitation is delivered from the T_I neurons to C_{5P} neurons. Only those C_{5P} neurons receiving both C_{3B} and T_I input will fire. In this way, the basal ganglia is responsible for triggering (or releasing from inhibition) the action commands in certain modules, and can control the timing and launching of action commands in appropriate sequences. The firing frequency of C_{5P} action commands could also be regulated in this way.

Since action commands must be triggered by T_I nuclei, the basal ganglia is essential in learning the timing and control of disinhibition action commands. The input from C_{5A} neurons must give all the necessary state information for the basal ganglia to learn which commands to trigger and which not to.

We note that the reticular activating system can be used to enable or disable action commands in the T_I nuclei; but, thought processes (confabulations) could still occur. We propose this mechanism as the explanation for the underlying mechanisms of dreaming.

3.11.4 The anatomy of confabulation control

The circuitry involved in confabulation control is shown in green in Figure-3.8. Confabulations are controlled in each thalamocortical module through T_{L1} input. These inputs are

analogous to motor neuron inputs to a muscle, therefore, they are dynamic, graded and ultimately determine the state of a confabulation. The control of confabulation in thalamocortical modules must be learned, therefore the basal ganglia is critical in learning the sequence and timing of these control signals. In almost every way, the triggering of action commands is identical to the triggering and control of confabulations other than the target thalamic nuclei.

3.12 Discussion and future work

As we can see, the anatomical knowledge of the major cognitive circuits interlinking the cerebral cortex, thalamus, and basal ganglia is less than ideal. Especially when we attempt to dig down to the exact layer and/or cortical pyramidal type that is sending or receiving projections. Our basic failure in neuroscience over the last 100 years to fundamentally decipher the basic neuroanatomy underlying cortical circuits should not be understated. The failure does not lie on the shoulders of the brave and excellent neuroanatomists who have collected the data which is often painstakingly difficult to obtain. The failure lies on the shoulders of the community as a whole for not providing the impetus through an explicit call for need, and as a consequence providing the necessary monetary funds. Our fundamental conclusion is that our ability to fundamentally understand the brain critically depends on the need to:

1. Completely understand the basic neuroanatomy of the primate(human) cerebral cortex in much more detail.

Other researchers have suggested creating "the connectome" (which at this point is a hypothetical idea)(Lichtman et al., 2008). Instead, we suggest a practical approach that will give dramatic results within 5 years.

We propose that several volunteer human subjects who are terminally ill (with diseases that do not fundamentally effect cortical organization) should have their thought processes recorded with every possible FMRI and non-invasive scanning technique performed on a wide variety of tasks. In post-mortem analysis, their brains should be thinly sectioned and every section should be pictured in 3-D with a confocal microscope. Alternating sections should be stained with nissl and possibly other stains to be determined. The brain should then be reconstructed visually in 3-D and fully mapped back onto all experiments. Although the cost will be huge, the experiment will provide necessary data that neuroanatomists such as von Economo could only have dreamed of, and with which integrative neuroscientists such as myself will use to finally reverse engineer the brain.

3.13 Appendix: references for figures

[1] Arbuthnott, G.W. et al., Distribution and synaptic contacts of the cortical terminals arising from neurons in the rat ventromedial thalamic nucleus. Neuroscience 38 (1), 47-60 (1990).

[2] Arikuni, T. et al., The organization of prefrontocaudate projections and their laminar origin in the macaque monkey: a retrograde study using HRP-gel. J. Comp. Neurol. 244 (4), 492-510 (1986).

[3] Asanuma, C., GABAergic and pallidal terminals in the thalamic reticular nucleus of squirrel monkeys. Exp. Brain Res. 101 (3), 439-451 (1994).

[4] Barbas, H. et al., Parallel organization of contralateral and ipsilateral prefrontal cortical projections in the rhesus monkey. BMC Neurosci 6 (1), 32 (2005).

[5] Bigl, V. et al., Cholinergic projections from the basal forebrain to frontal, parietal, temporal, occipital, and cingulate cortices: a combined fluorescent tracer and acetylcholinesterase analysis. Brain Res. Bull. 8 (6), 727-749 (1982).

[6] Conde, F. et al., Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: distribution and morphology. J. Comp. Neurol. 341 (1), 95-116 (1994).

[7] de Lima, A.D. et al., Morphology of the cells within the inferior temporal gyrus that project to the prefrontal cortex in the macaque monkey. J. Comp. Neurol. 296 (1), 159-172 (1990).

[8] DeFelipe, J. et al., Long-range focal collateralization of axons arising from cortico-cortical cells in monkey sensory-motor cortex. J. Neurosci. 6 (12), 3749-3766 (1986).

[9] Francois, C. et al., Distribution and morphology of nigral axons projecting to the thalamus in primates. J. Comp. Neurol. 447 (3), 249-260 (2002).

[10] Fujita, I. et al., Intrinsic Connections in the macaque inferior temporal cortex. J. Comp. Neurol. 368 (4), 467-486 (1996).

[11] Gabbott, P.L. et al., Local circuit neurons in the medial prefrontal cortex (areas 24a,b,c, 25 and 32) in the monkey: I. Cell morphology and morphometrics. J. Comp. Neurol. 364 (4), 567-608 (1996).

[12] Gimenez-Amaya, J.M. et al., Organization of thalamic projections to the ventral striatum in the primate. J. Comp. Neurol. 354 (1), 127-149 (1995).

[13] Glickstein, M. et al., Corticopontine projection in the macaque: the distribution of labelled cortical cells after large injections of horseradish peroxidase in the pontine nuclei. J. Comp.Neurol. 235 (3), 343-359 (1985).

[14] Hazrati, L.N. et al., Projection from the external pallidum to the reticular thalamic nucleus in the squirrel monkey. Brain Res. 550 (1), 142-146 (1991).

[15] Herkenham, M., Laminar organization of thalamic projections to the rat neocortex. Science 207 (4430), 532-535 (1980).

[16] Jacobson, S. et al., Prefrontal granular cortex of the rhesus monkey. I. Intrahemispheric cortical afferents. Brain Res. 132 (2), 209-233 (1977).

[17] Jarvis, E.D. et al., Avian brains and a new understanding of vertebrate brain evolution. Nat. Rev. Neurosci. 6 (2), 151-159 (2005).

[18] Kaneko, T. et al., Predominant information transfer from layer III pyramidal neurons to corticospinal neurons. J. Comp. Neurol. 423 (1), 52-65 (2000).

[19] Kritzer, M.F. et al., Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 359 (1), 131-143 (1995).

[20] Kunishio, K. et al., Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. J. Comp. Neurol. 350 (3), 337-356 (1994).

[21] Lavenex, P. et al., Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. J. Comp. Neurol. 447 (4), 394-420 (2002).

[22] Lemon, R.N. et al., Comparing the function of the corticospinal system in different species: organizational differences for motor specialization? Muscle Nerve 32 (3), 261-279 (2005).

[23] LeVay, S., Synaptic organization of claustral and geniculate afferents to the visual cortex of the cat. J. Neurosci. 6 (12), 3564-3575 (1986).

[24] Levitt, J.B. et al., Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). J. Comp. Neurol. 338 (3), 360-376 (1993).

[25] Lorente de No, R., Cerebral Cortex: Architecture, Intracortical connections, motor projections in Physiology of the nervous system, edited by J.F. Fulton (Oxford University Press, New York, 1943), pp. 274-301.

[26] Lund, J.S. et al., Anatomical organization of primate visual cortex area VII. J. Comp. Neurol. 202 (1), 19-45 (1981).

[27] Lund, J.S. et al., Local circuit neurons of developing and mature macaque prefrontal cortex: Golgi and immunocytochemical characteristics. J. Comp. Neurol. 328 (2), 282-312 (1993).

[28] McFarland, N.R. et al., Organization of thalamostriatal terminals from the ventral motor nuclei in the macaque. J. Comp. Neurol. 429 (2), 321-336 (2001).

[29] McFarland, N.R. et al., Thalamic relay nuclei of the basal ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. J. Neurosci. 22 (18), 8117-8132 (2002).

[30] Morishima, M. et al., Recurrent connection patterns of corticostriatal pyramidal

cells in frontal cortex. J. Neurosci. 26 (16), 4394-4405 (2006).

[31] Parent, M. et al., Single-axon tracing and three-dimensional reconstruction of centre median-parafascicular thalamic neurons in primates. J. Comp. Neurol. 481 (1), 127-144 (2005).

[32] Peters, A. et al., The organization of pyramidal cells in area 18 of the rhesus monkey. Cereb. Cortex 7 (5), 405-421 (1997).

[33] Rockland, K.S. et al., Single axon analysis of pulvinocortical connections to several visual areas in the macaque. J. Comp. Neurol. 406 (2), 221-250 (1999).

[34] Romanski, L.M. et al., Topographic organization of medial pulvinar connections with the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 379 (3), 313-332 (1997).

[35] Sidibe, M. et al., Efferent connections of the internal globus pallidus in the squirrel monkey: I. Topography and synaptic organization of the pallidothalamic projection. J. Comp. Neurol. 382 (3), 323-347 (1997).

[36] Sidibe, M. et al., Nigral and pallidal inputs to functionally segregated thalamostriatal neurons in the centromedian/parafascicular intralaminar nuclear complex in monkey. J. Comp. Neurol. 447 (3), 286-299 (2002).

[37] Soloway, A.S. et al., Dendritic morphology of callosal and ipsilateral projection neurons in monkey prefrontal cortex. Neuroscience 109 (3), 461-471 (2002).

[38] Tanne-Gariepy, J. et al., Projections of the claustrum to the primary motor, premotor, and prefrontal cortices in the macaque monkey. J. Comp. Neurol. 454 (2), 140-157 (2002).

[39] Tardif, E. et al., Laminar specificity of intrinsic connections in Broca's area. Cereb. Cortex 17 (12), 2949-2960 (2007).

[40] Witter, M.P. et al., Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. Prog. Neurobiol. 33, 161-253 (1989).

[41] Yeterian, E.H. et al., Laminar origin of striatal and thalamic projections of the prefrontal cortex in rhesus monkeys. Exp. Brain Res. 99 (3), 383-398 (1994).

[42] Zikopoulos, B. et al., Prefrontal projections to the thalamic reticular nucleus form a unique circuit for attentional mechanisms. J. Neurosci. 26 (28), 7348-7361 (2006).

Chapter 3, in part, is currently being prepared for submission for publication of the material. Solari, Soren. The author was the primary investigator and author of this material.

4

Multi-associative memory: the basis of knowledge link associations in the cerebral cortex

4.1 Introduction

Our exploration into the storage and utilization of associations (knowledge links) in confabulation theory lead to a new general mathematical framework for understanding the possible function and utilization of associations in nervous systems. We termed that general framework a multi-associative memory (MAM) and the present chapter is dedicated to its formulation and explanation.

Storing and recalling associations between multi-modal perceptions has long been a basic premise underlying cognitive function. Although mathematical models of associative memories exist, traditional associative memories are incapable of storing and recalling general multi-modal associations. Here we extend this past work and present a general mathematical model, termed a multi-associative memory, demonstrating that associations formed between any combination of cell assemblies can be utilized in biologically realistic symbolic information processing. We simulate the storage and recall of natural language with a 10,000 word vocabulary at the neural level. Using natural language as an example we show that the symbolic information processing outcome of a neuronal simulation can be predicted independent of the actual neuronal connectivity. Since the formation of multi-associative memories statistically relies on a few parameters, we propose that DNA codes these variables to guarantee symbolic information processing capabilities in

nervous systems.

All nervous systems are composed of synaptic connections between neurons (Striedter, 2005). Except for the most basic nervous systems, most neuronal connectivity appears to be initially randomly wired between large populations of neurons. The mammalian cerebral cortex is a prime example (Braitenberg and Shuz, 1998). But how can randomly wired brains basically all function the same? Donald Hebb puts forth several significant insights into the formation of associations in nervous systems(Hebb, 1949). Hebb's main hypothesis states that populations of neurons will group into *cell assemblies* that each represent a functional unit or perception. Evidence is growing that these cell assemblies do in fact form in the cerebral cortex(Yoshimura et al., 2005). Secondly, Hebb hypothesizes the rule of synaptic plasticity, infamously paraphrased as neurons that fire together wire together, that in general has proven to be true(Lisman, 1989). However, one of Hebb's main arguments, which is often overlooked, suggests associations are formed between cell assemblies rather than neurons. As he mentions, "an association between two perceptions is likely to be possible only after each one has independently been organized or integrated"(Hebb, 1949). The idea put forth by Hebb is that cell assemblies first form through some developmental self organization and are then associated through co-occurrence learning by strengthening individual synapses. If we adopt the notion that a cell assembly is a discrete symbolic perception, then associations between discrete symbols can be viewed as associations between cell assemblies in nervous systems.

A great deal of research has been done on associative memories; however, the use of traditional associative memories has essentially been limited to pattern completion(Willshaw et al., 1969; Palm, 1980; Amari, 1989; Buckingham and Willshaw, 1992; Graham and Willshaw, 1995; Haines and Hecht-Nielsen, 1988). In a traditional associative memory, a cell assembly can only be associated with one other cell assembly. Therefore, given the partial activation of one source cell assembly the associative memory might be used to recover the associated target cell assembly, thereby completing the pattern. This basic structure has limitations that do not seem to be consistent with the diversity of associations we experience everyday. For example, if individual faces and words are represented by cell assemblies somewhere in the cerebral cortex, we must be able to associate one word, e.g. a first name, with many faces. Conversely, one face may be associated with many words, e.g. both a first name and a last name. These are unconstrained multi-modal associations that can not be formed in traditional associative memories and therefore require the new mathematical formulation that we present here.

We use human language as a concrete implementation of multi-associative memories. Consider a single word to be a symbolic perception. Humans know tens of thousands of words; hence, we must be capable of storing tens of thousands of these symbolic perceptions in our

cerebral cortex. When generating a next word, we must utilize learned associations to ensure that we select a cogent next word. Naturally, humans are not born speaking any given language; therefore, the ability to learn to speak any language, at a minimum requires that two conditions must be met. First, the ability to learn and store tens of thousands of arbitrary words. Second, the ability to form associations between any pairs of words. If a person knows 10,000 words, the cerebral cortex must be capable of both representing the 10,000 words and implementing any one of the 100 million possible associations.

When we consider that mammalian cortical connectivity between individual neurons has been estimated as low as 1-2%(Braitenberg and Shuz, 1998), it may seem impossible that any specific association can be implemented in the cerebral cortex. If words are represented by cell assemblies, and cell assemblies are composed of groups of neurons, then how can any two arbitrary cell assemblies be associated, exclusively through the strengthening of synapses, if there is no guarantee that the neurons in the two cell assemblies are even connected? More so, synaptic excitation is also non-deterministic(Malenka and Siegelbaum, 2001); therefore, how can we be guaranteed of recovering information if there is no guarantee that a certain level of excitation will be delivered by any particular synapse that is strengthened? The answer to these questions is that statistical principles dominate when large populations of neurons are randomly organized, randomly connected, and even when synapses deliver random excitation in multi-associative memories. The function of a multi-associative memory relies exclusively on the specification of only a few parameters; hence, we hypothesize for nervous system architectures that DNA specifies the values of a few biologically controlled parameters, statistically guaranteeing the same function for all similar brains.

The basic principles underlying multi-associative memories, although explained here in terms of language and the cerebral cortex, are in fact general principles and can therefore be applied to all varieties of neuronal connectivity involving associations between large populations of neurons.

Here, we introduce a multi-associative memory model providing a general mathematical and theoretical tool for neuroscientists to understand information processing in populations of neurons (Figure-4.1 and Figure-4.2 and Methods). We simulate a more detailed neuronal implementation of past work in natural language generation, demonstrating the capability to carry out complicated symbolic information processing in a biologically realistic architecture (Figure-4.3). We also derive analytic estimates of the associative signal and associative interference in any multi-associative memory, demonstrating that the information processing performed between neurons in multi-associative memories can be accurately abstracted as symbolic information processing between cell assemblies(Figure-4.4) . The Figure-4.5 summary proposes the hypothesized

role multi-associative memory models play in the biological development (DNA to consolidated cortical connections) of associations in the cerebral cortex.

4.2 Results

4.2.1 MAM symbolic information processing

The introduction of confabulation theory provides a concrete hypothesis describing the fundamental mechanisms underlying mammalian cognition (see chapter 2). One fundamental hypothesis of confabulation theory proposes that all cognitive information processing relies upon the additive combination of graded associations between symbolic perceptions. Confabulation theory explicitly describes the mathematical conditions for the formation and strength of these associations (knowledge links) using real world data, and provides strong evidence that cognitive information processing simply utilizes the additive combination of knowledge links. However, a biologically realistic neuronal mathematical model for the storage and recall of this type of information has not yet been suggested.

In our effort to explore the possible neuronal mechanisms capable of storing and recalling arbitrary associations between symbolic perceptions in the cerebral cortex, we discovered that a basic extension of past work on associative memories is capable of implementing robust symbolic information processing under biologically realistic conditions. Since our extension of past work is rooted in the generalization of classic associative memory matrices to include multiple arbitrary associations (one cell assembly to many, many cell assemblies to one) in addition to multiple convergent association matrices between many neural fields, we term the new formulation a *multi-associative memory*(see section 4.4)

4.2.2 Modeling natural language processing with MAMs

In order to test the multi-associative memories capability to carry out symbolic information processing, we simulate a multi-associative memory neuronal implementation of past published work on symbolic natural language processing.

In past work, Robert Hecht-Nielsen introduces the mathematical concept of *maximizing cogency*(Hecht-Nielsen, 2005). Maximizing cogency can be thought of as selecting, as the conclusion of an information processing operation, the symbol with the largest additive sum of active associations. Hecht-Nielsen proposes that the maximization of cogency is the basic mathematics underlying all cognitive information processing and his paper presents experimental results demonstrating the ability to add a cogent fourth word after three consecutive words from a nat-

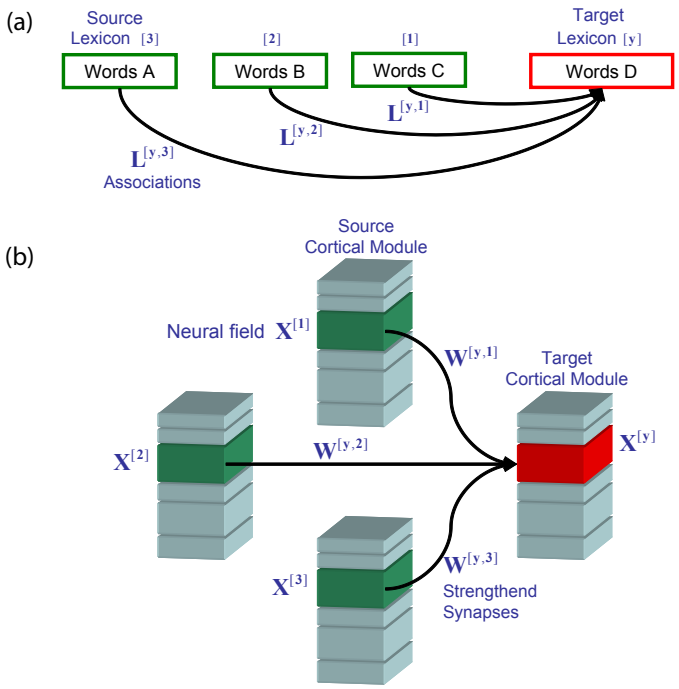

Figure 4.1: Multi-associative memory sentence completion organization. (a) Each lexicon contains 10,000 symbols, which represent the 10,000 most commonly encountered words. Association matrices, $\mathbf{L}^{[y,k]}$, associate individual words between source(green) and target(red) lexicons based on the frequency of their co-occurrences(Hecht-Nielsen, 2005). (b) A multi-associative memory "cortical" implementation of the symbolic word associations is shown as connections between cell assemblies in cortical layer 3 neural fields. Each neural field $\mathbf{X}^{[k]}$ contains 10,000 cell assemblies that each represent one word. Associations between words in (a) are translated into associations between cell assemblies in (b) through the strengthening of existing unstrengthened synaptic connections between neurons. The strengthened neuronal connectivity matrix $\mathbf{W}^{[y,k]}$ is determined by the associations in matrix $\mathbf{L}^{[y,k]}$, which associate the cell assemblies given in the matrices $\mathbf{X}^{[k]}$ and $\mathbf{X}^{[y]}$. See Figure 4.2 and Methods.

ural sentence. The basic idea of the experiment is shown in Figure-4.1(a). Each colored box in Figure-4.1(a) represents a lexicon containing the same 10,000 most commonly encountered words in a training corpus. During training, a 1.4 x 10^9 word English news article text corpus is read serially into the four-contiguous-word window. Three individual 10,000 by 10,000 count matrices, record co-occurence counts between the word in each k^{th} source(green) lexicon and the word in the y target(red) lexicon. After training each association matrix, $\mathbf{L}^{[y,k]}$, had the following number of non-zero(nnz) elements, which each correspond to a single association: nnz($\mathbf{L}^{[y,1]}$)=870,090; nnz($\mathbf{L}^{[y,2]}$)=1,293,503; nnz($\mathbf{L}^{[y,3]}$)=1,520,274 (see Appendix-4.5.1). Even in this simple experiment, we immediately see that a multi-associative memory must be capable of storing and recalling millions of associations. We must emphasize that much more extensive and impressive results regarding language have been published which demonstrate the ability,

using an identical implementation of co-occurrence learning, to generate a contextually accurate, complete, and grammatically correct third sentence that might follow two preceding sentences in a news story (see chapter 2). Therefore, given more powerful computers our neuronal simulation could by extension be used to produce the same full grammatically and contextually accurate sentences.

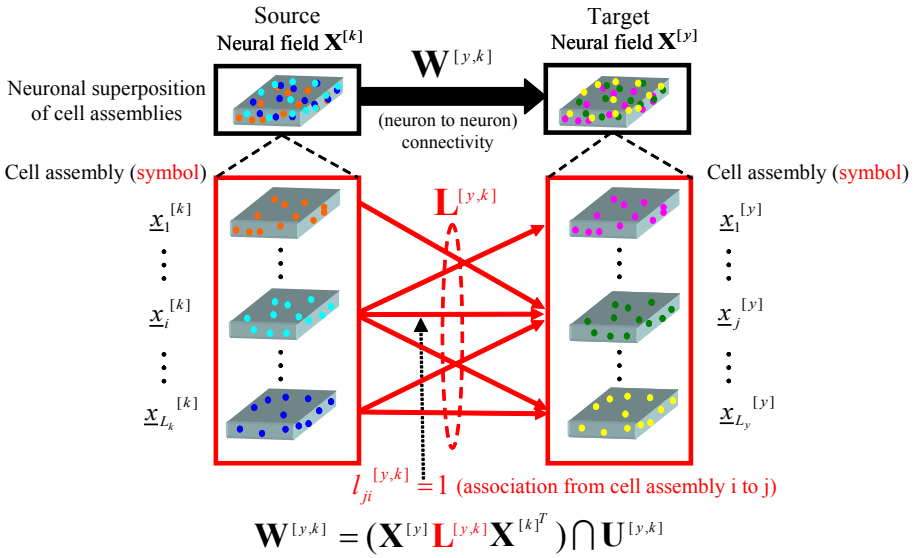

Figure 4.2: Multi-associative memory connectivity. The connectivity between two neural fields, k and y, are shown. $\mathbf{W}^{[y,k]}$ is a binary matrix (defining the connectivity in a multi-associative memory) containing the strengthened neuron to neuron connections from neurons in k to neurons in y. In general there may be many (and possibly reciprocally connected) neural fields, therefore multiple connectivity matrices exist (e.g. $W^{[y,1]}, W^{[y,2]}, W^{[2,y]}$,etc), as in Figure 4.1. Red lines correspond to symbolic organization and associations, where as black lines correspond to neuronal organization and connectivity. The matrix $\mathbf{X}^{[k]}$ is a binary matrix defining the random cell assembly organization in neural field k. Each row of $\mathbf{X}^{[k]}$ represents a single neuron and each column a cell assembly (i.e. the column vector $\underline{x}_i^{[k]}$ identifies all neurons in cell assembly i). One neuron may be part of multiple cell assemblies. The binary matrix $\mathbf{L}^{[y,k]}$ defines the associations between source and target cell assemblies. As shown in red, one source or target cell assembly may be associated with many target or source cell assemblies respectively. The number of source cell assemblies L_k need not be the same as the number of target cell assemblies L_y. The binary connectivity matrix $\mathbf{U}^{[y,k]}$ is the initial sparse random unstrengthened connectivity matrix between neurons. $\mathbf{U}^{[y,k]}$ and the intersection symbol indicates that synapses are only strengthened if they exist. See Methods for more details.

Figure-4.1(b) shows the multi-associative memory simulation of the word experiment in a biologically realistic context. We emphasize that our simulation represents information processing in an adult mammalian brain in which memories have been consolidated into cortico-cortical connections(Squire, 2004). Three source (green) neural fields $\mathbf{X}^{[k]}$ and one target neural field (red) $\mathbf{X}^{[y]}$, each consisting of $N = 1,000,000$ neurons are shown. The biologically realistic

neural fields represent the excitatory pyramidal neurons in layer 3 of a six layered human cerebral cortex with approximate area 30mm^2(von Economo, 1929)(see Appendix-4.5.2). We depict the biological implementation in this form because cortio-cortical connections are known to have specific organization and selectivity between layers(Barbas and Rempel-Clower, 1997; Barbas, 1986). Each word in a lexicon in Figure-4.1(a) is represented by a cell assembly, a sub-population of neurons in a neural field, in Figure-4.1(b). Neurons are randomly assigned to cell assemblies and cell assemblies randomly assigned to represent words. Each neuron has the same probability of being in any one of the $L = 10,000$ cell assemblies in the neural field, such that the average number of neurons in a cell assembly is $M = 300$. Based on clinical studies in humans, a sparse encoding of a few hundred neurons per cell assembly appears to be reasonable(Quiroga et al., 2007). Prior to any learning, each source neuron has a probability ($p = 0.03$) of of forming an unstrengthened "silent" synapse with any target neuron(Craig and Lichtman, 2001). The association matrices, $\mathbf{L}^{[y,k]}$, created in the previous experiment are assumed to have been learned and temporarily stored in the hippocampus, which is not directly modeled. Next, only cell assemblies which contained learned associations in the $\mathbf{L}^{[y,k]}$ matrix are co-activated in paired neural fields. The reactivation of cell assemblies in the cerebral cortex is here hypothesized to occur during sleep cycles, via hippocampal associative playback(Sutherland and McNaughton, 2000). As cell assemblies are co-active all existing unstrengthened synapses between co-active neurons are strengthened, and the learned associations are consolidated into cortico-cortical synaptic connections between the neural fields.

Figure 4.2 shows the structure underlying a single multi-associative memory connectivity matrix(see Methods). Notice one source cell assembly may be associated with many target cell assemblies and vice versa. At the completion of training, the multi-associative memory in Figure-4.1(b) consists of three connectivity matrices describing strengthened synaptic connections between neurons (see Methods). All symbolic associations in $\mathbf{L}^{[y,k]}$ are ultimately stored in the multi-associative memory neuronal connectivity matrix $\mathbf{W}^{[y,k]}$.

In the original symbolic natural language processing experiment three starter words are activated in the source lexicons and the word which maximizes cogency is selected as the result in the target lexicon. The cogent third word is that one with the largest sum of active symbolic associations from the source words. For example, in the original experiment the three words "college students learn" are activated in each source lexicon. Based on the the learned matrices $\mathbf{L}^{[y,k]}$, only one word, "math", receives associations from all three source words, hence, "math" is the cogent next answer.

The biological simulation activates the ∼300 neurons in each source cell assembly that represent a given word. An additional 1000 neurons are randomly activated in each source neu-

ral field as noise. All active source neurons then deliver stochastic synaptic excitation, normally distributed with mean 1 and standard deviation 0.3, to all their strengthened synaptic connections. The A_y most highly excited target neurons are activated (discussed below) and all other target neurons are inhibited. Figure-4.3 shows the percentage of neurons active in each target cell assembly after a single simulation run. The target cell assembly with the largest fraction of its neurons activated is considered to be the cogent next word selected in the biological simulation. In Figure-4.3(a) for the same sentence initialization "college students learn", the cell assembly with the largest activation percentage is "math". In fact, Figure-4.3(a,b,c) result in the identical fourth word selections as in the original paper(Hecht-Nielsen, 2005). The four different initializations highlight the effect of the overall number and distribution of active associations in multi-associative memories. Initialization with "the the the" (Figure 4.3d) produces so much associative interference (see below) that all words have low activation and either no word is selected or a random word is selected. Effectively, the system produces an answer that can be read as "I don't know".

Here, we point out a couple important points. The simulation basically involves only one time step, and as a result, the plot in Figure-4.3 only represents an initialization of the target neural field. Any further processing (i.e. converging to the highest activated cell assemblies) undoubtedly continues on in time and may involve dynamic processes such as an attractor network(Hopfield, 1982; Abeles, 1991; Durstewitz et al., 2000). If we make an assumption that a local neuronal attractor network exists in the target cortical module, which is capable of converging over time to the cell assembly with the largest initial activation, then we can definitively say that a multi-associative memory can be used to maximize cogency.

4.2.3 Control of information processing in MAMs

A primary function, suggested by our experiments, of a multi-associative memory is to carry-out dynamic symbolic information processing operations between many cell assemblies in multiple neural fields simultaneously. In nervous systems, the firing of action potentials by neurons is the primary source of information transfer between neurons. Therefore, controlling which neurons fire action potentials, and which do not, will ultimately effect information processing in nervous systems. As a result, we hypothesized that the ability to roughly control the overall number of neurons which fire action potentials, independent of any underlying neuronal dynamics, is sufficient to control information processing in multi-associative memories.

Two basic neuroscience facts suggest that this type of control could be reasonably implemented in the cerebral cortex. First, although the firing of an action potential is a complex dynamic process(Izhikevich, 2007), in general, for a homogenous population of neurons, more

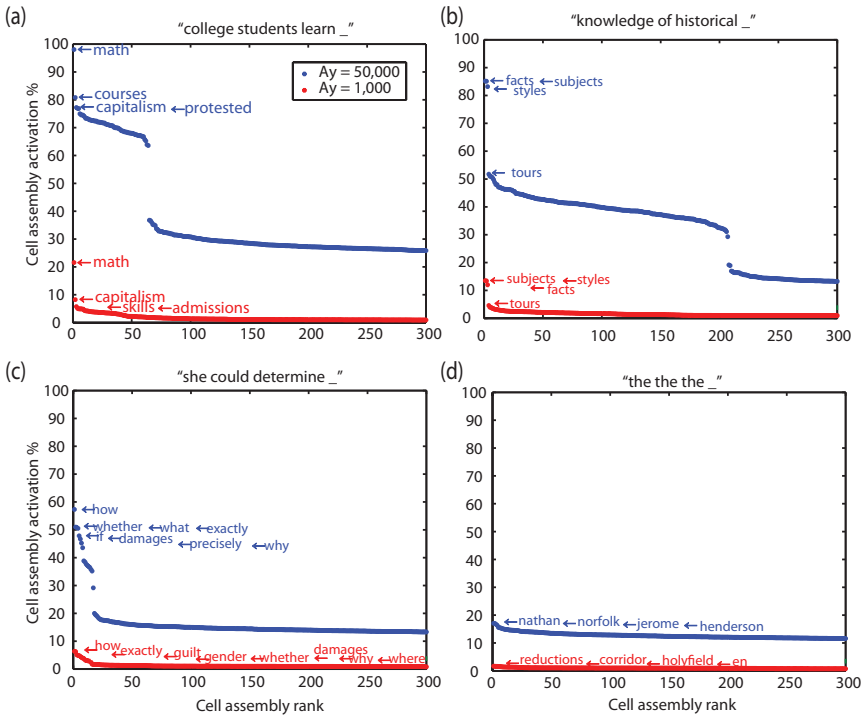

Figure 4.3: Multi-associative memory sentence completion results. Each source neural field $X^{[k]}$ is initialized by activating all neurons in a single cell assembly representing a particular word, along with 1000 randomly selected other neurons as additional noise. The words representing the cell assembly initialized in each source neural field are shown as "$X^{[3]}$ $X^{[2]}$ $X^{[1]}$ ". Each active synapse delivers some amount of random excitation selected from a gaussian distribution with mean=1 and variance=0.3. The A_y most highly excited neurons in the target neural field $X^{[y]}$ are activated and all others are inactivated. The resulting target cell assembly activations (fraction of active neurons in the cell assembly) are computed and the top 300 of 10,000 are plotted in order of their activation rank (highest to lowest x-axis). The words represented by the four highest ranked cell assemblies are labeled and the highest ranked cell assembly is considered the cogent 4^{th} word selection. Two different scenarios are plotted on each graph for different A_y values ("blue" A_y=50,000(50k), "red" A_y=1k). Notice the highest ranked cell assemblies roughly stay the same even with 50 times difference in the number of active target neurons. Four different word initializations are used to demonstrate the effect of the associative signal and associative interference (see Figure-4.4) due to the number of active afferent associations in the target neural field $X^{[y]}$. (a) Initialization "college students learn" results in 1(3), 63(2), and 581(1) [read: 1 target cell assembly receiving 3 associations, 63 target cell assemblies receiving 2 associations and 581 target cell assemblies receiving 1 association from the initialized words.] (b) Initialization "knowledge of historical" results in 3(3), 204(2), and 9639(1). (c) Initialization "she could determine" results in 8(3), 1930(2), and 4383(1). (d) Initialization "the the the" results in 8690(3), 1253(2), and 40(1).

highly excited neurons will fire before less highly excited neurons. Second, local inhibition is a prominent feature of the cerebral cortex(Markram et al., 2004). Local feedback inhibition caused by highly excited pyramidal neurons firing action potentials will tend to prevent other less highly excited neurons from firing. Therefore, in our experiments we choose to avoid simulating a particular neuronal model to test the hypothesized effect of simply allowing the A_y most highly excited neurons to fire and inhibiting the rest.

Figure-4.3 demonstrates the effect of this type of information processing control. Each plot shows two curves (red and blue). The blue curve shows the cell assembly activation when $A_y = 50,000$ target neurons are activated (i.e. "allowed to fire action potentials"). The red curve shows the effect when $A_y = 1,000$ neurons are activated. As can be seen, the overall order of cell assemblies activation percentage is mostly unchanged, even though the absolute activation decreases. Assuming an attractor network can converge to the most highly activated cell assembly, the selection of the most highly activated cell assembly is very robust to the overall number of neurons that are activated. We conclude that multi-associative memory information processing is robust against orders of magnitude differences in the control of overall neuronal activation. The resulting prediction is that the control of information processing in the cerebral cortex only requires rough control of overall neuronal activation within a local cortical module.

4.2.4 MAM associative signal and associative interference

Multi-associative memories, by construction, store associations between cell assemblies (symbolic perceptions) embedded in synaptic connections between neurons. The results presented in Figure 4.3 demonstrate that multi-associative memories can in fact accurately reproduce symbolic natural language information processing. Our result leads to a natural question; namely, to what degree and under what conditions do multi-associative memories perform symbolic information processing? We answered this question by running parameter sweeps across many multi-associative memory initializations (see Appendix 4.5.5) and by deriving an analytic expression, the cell assembly excitation (see Theorem 1 in Methods). The cell assembly excitation is only dependent on the symbolic information state and basic underlying parameters, which defines the associative signal and associative interference into each target cell assembly. Figure 4.4 displays the estimated fraction of cell assembly excitation, $\frac{S_j}{I_j}$, that is due to associative signal for each simulation run in Figure 4.3. The symbolic estimation for all 10,000 cell assemblies is shown and the corresponding words are displayed for the first few cell assemblies. Notice the general agreement between the order of the most highly ranked cell assemblies between Figure 4.3 and Figure 4.4. The large gaps between some cell assemblies in both plots highlight the fact that some cell assemblies are associated with all 3 of the initialized words, some with 2, some

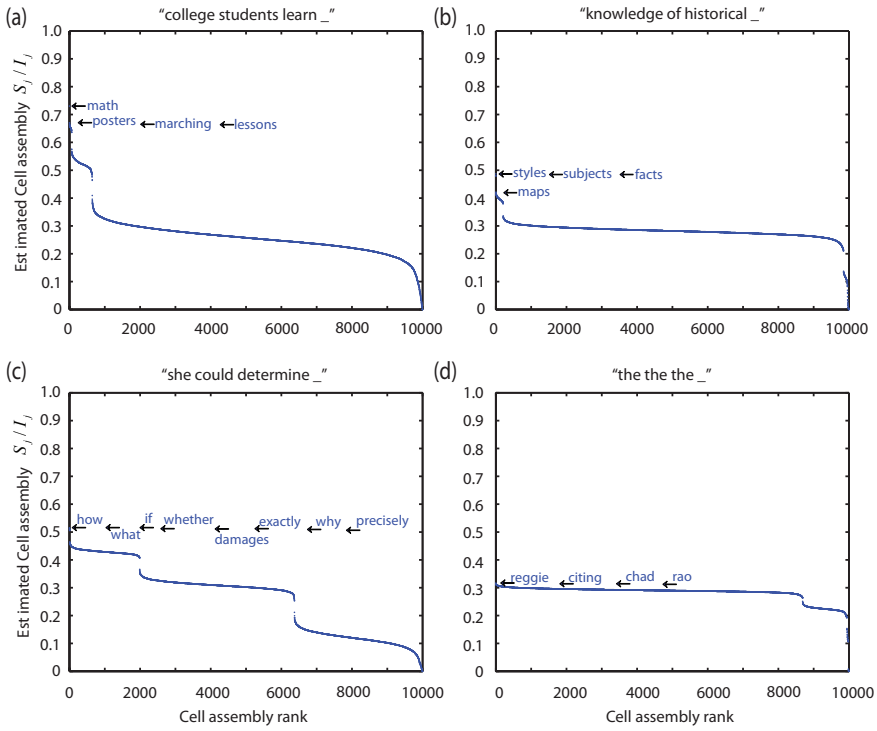

Figure 4.4: Multi-associative memory associative signal and associative interference. The fraction of each cell assemblies estimated symbolic cell assembly excitation due to associative signal is plotted for each target cell assembly in each of the word completion experiments shown in Figure 4.3. The fractional values are plotted in order of cell assembly rank (those with the highest fraction to those with the lowest). The words representing the top four(a,b,d) or eight(c) cell assemblies are shown for comparison to the resulting cell assembly activations obtained through simulation, where arrows indicate values for the corresponding words. Notice the predictive value of the estimated symbolic fractional input values in determining the final cell assembly activations from the neuronal simulation.

with 1, and some with 0.

The strength of an association is determined exclusively by three terms, the average number of neurons in a source cell assembly, M, the probability of connectivity between neurons, p, and the mean strength of a strengthened synapse, s. In the experiment, the probability of connectivity between neurons is %3, and yet any one of the 100 million possible associations can be formed. The basic reason is that for any pair of source, i, and target, j, cell assemblies, a target neuron in cell assembly j will receive connections from on average pM neurons in the source cell assembly i. Therefore, a neuron in target cell assembly j is statistically guaranteed to receive on average $pM = (0.03)300 = 9$ strengthened synapses from source cell assembly i. When these synapses are activated some synapses will deliver more excitation and some less, but the synapses on average will deliver some mean excitation value s. The number of strengthened connections

and the mean value of synaptic excitation defines the mean strength of an association. What if there were 20,000 cell assemblies? Could any one of the now 400 million possible associations still be formed? Yes, because no matter how many cell assemblies or possible associations there are, the value pM does not change. The problem occurs when attempting to recover information stored in the associations, because more cell assemblies and more associations results in more associative interference.

Associative interference is exclusively a result of cell assembly overlap, stated otherwise, a single neuron being part of multiple cell assemblies. If a single cell assembly is activated, cell assembly overlap will deliver synaptic excitation to neurons in target cell assemblies that were never associated with the source cell assembly. Since associations form between entire cell assemblies not individual neurons, a critical realization is that associative interference is only calculable in reference to a specific target cell assembly (see Appendix 4.5.3).

In general under most cell assembly overlap conditions ($r_k > 2$), and when large numbers of associations exist (in our simulation millions), the associative interference acts very much like additive gaussian noise. Therefore, plotting the fraction of target cell assembly excitation which is associative signal (Figure 4.4), provides an estimate of the likelihood that a target cell assembly will have more of its neurons among the A_y most highly excited neurons over other cell assemblies. In essence, it relates to the entropy of the neural field information state. We mention this point, because there is an array of additional research that might be pursued by experts in information theory.

Comparing Figure-4.3 and Figure-4.4 demonstrates that the symbolic information processing capability of a multi-associative memory is clearly predicted by the estimated symbolic target cell assembly associative signal and associative interference. The ability to accurately predict information processing in neuronal simulation with estimates of the symbolic excitation onto cell assemblies, suggests that the information processing that populations of neurons perform can in fact be abstracted accurately as symbolic information processing. We have demonstrated that the precision with which symbolic information processing can be performed is determined by the symbolic estimation of information processing, in turn based on the initialized information state of source neural fields and the underlying parameters in the multi-associative memory.

4.3 Discussion

Our multi-associative memory model provides an explanation to many questions, such as: "How is it that brains can be initially wired randomly and yet all function identically in the sense of performing identical information processing?", and "What architecture might allow

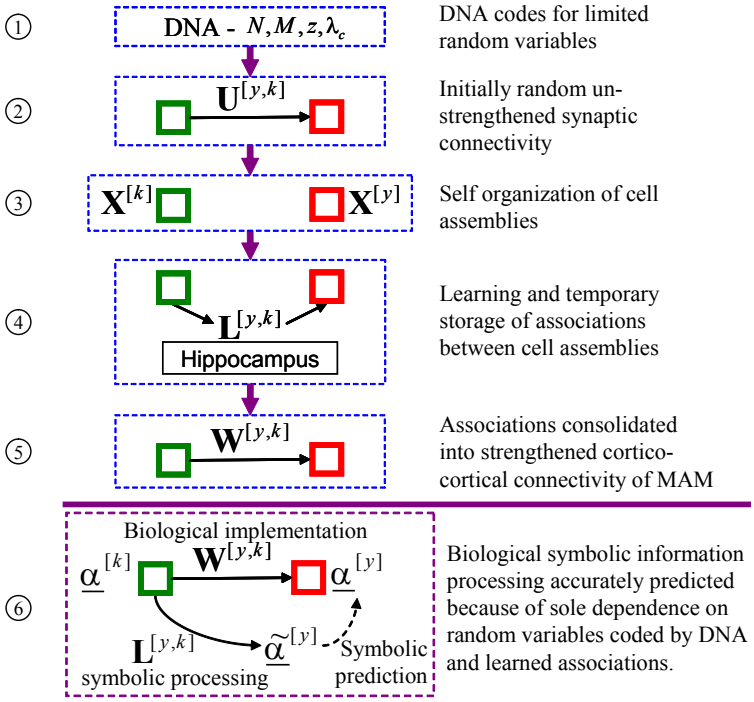

Figure 4.5: Hypothesized evolution of a cortical multi-associative memory. A depiction of the hypothesized formation of MAM (multi-associative memory) cortico-cortical connections. (1) Parameters that control overall MAM development are stored in DNA, N, M, and λ_c. (2) The neurons in the source neural field (green box) send axonal fascicles to the target neural field (red box) and randomly form unstrengthened synapses with target neurons, defined by $\mathbf{U}^{[y,k]}$. (3) Environmental exposure causes self-organization of cell assemblies, approximated by randomly assigning neurons to cell assemblies, defined by $\mathbf{X}^{[k]}$. The random assignment of neurons to cell assemblies demonstrates that even after random initial connectivity, the inclusion of a specific neuron in a specific cell assembly is not needed to form arbitrary associations. (4) During a learning period, the hippocampus records co-occurrences between cell assemblies and ultimately determines which cell assemblies to associate, defined by $\mathbf{L}^{[y,k]}$. (5) Temporary hippocampally mediated associations are consolidated into direct cortico-cortical synaptic connections, $\mathbf{W}^{[y,k]}$, by cell assembly reactivation during sleep. (6) The basic biological implementation in fact carries out symbolic information processing, as is demonstrated by the prediction of target cell assembly activations by the MAM symbolic cell assembly excitation estimates. Importantly, the symbolic estimates are completely determined by the parameters coded by DNA, learned associations and the source symbolic excitation; therefore, symbolic information processing in MAMs can be guaranteed through DNA coding and control of overall neuronal activation.

different mammalian species' cerebral cortices to scale across orders of magnitude in size and still form and recall associations identically?" We demonstrate that providing the statistics of a few random variables during initialization is sufficient to create and recall arbitrary universal associations between symbolic perceptions under biologically realistic cerebral cortical connectivity. The statistics of the underlying random variables could be controlled by genetic coding in DNA and provide dramatic scalability and robust information processing with the multi-associative

memory architecture(see also Appendix-4.5.5). Figure-4.5 summarizes both the development of the model in our simulation and the hypothesized development of multi-associative memories in the mammalian cerebral cortex.

Several features of the multi-associative memory are especially relevant to cognitive information processing in mammals.

The multi-associative memory is scalable. Information processing is not dependent on the specific size of parameters in the multi-associative memory, but rather the relationship between parameters which define the associative signal and associative interference. The multi-associative memory associative signal is proportional to the probability of connectivity between neurons combined with cell assembly size. If the product of both parameters are held roughly constant, then the basic input strength for any association will be constant across any multi-associative memory size range. This fact provides a possible explanation for why the brain of a mouse and a human have the same basic structure, have such different sizes, and yet could still function identically.

The multi-associative memory provides a substrate upon which arbitrary knowledge can be stored and recalled. Even with a very low probability of connectivity between individual neurons, all cell assemblies are still statistically connected to all cell assemblies. Therefore, associations can be formed between any two arbitrary cell assemblies instantaneously simply by strengthening existing synapses. The consolidation of knowledge in the multi-associative memory does not require the addition of new synapses, nor does it depend on the existence of any individual synapse. In real nervous systems dynamic growth of synapses and additional self-organization could be used to regulate subtle differences in association strengths more precisely.

Only gross statistical random variables are needed to randomly configure a multi-associative memory to function in a robust manner. Our multi-associative memory model therefore provides a basic explanation for how all mammalian brains can be initially connected completely randomly but function identically. The parameters that need to be initially determined, such as, average number of synapses per neuron, average number of neurons per neural field, average size of a cell assembly or the number of cell assemblies per neuron, could be reasonably assumed to be directly or indirectly coded by DNA. In fact, the multi-associative memory predicts that the basic statistical parameters are all that should be coded by DNA, therefore any brain configured randomly with the same underlying parameters will function identically with high probability.

The multi-associative memory predicts a concrete mathematical role for the hippocampus. The ability to recall information accurately in multi-associative memories requires control over the formation of associations. For instance, interference is significantly reduced by keeping

the average number of efferent associations per source cell assembly, λ_c, low. As a result, our model predicts the role of the hippocampus in mammalian memory development is to learn and store "temporary" association matrices $\mathbf{L}^{[y,k]}$ between cortical areas, that over time should be consolidated into direct cortico-cortical connections between cortical neural fields (see Appendix-4.5.2). Our hippocampal hypothesis is consistent with prior models of the hippocampus and its established clinical function(Teyler and Rudy, 2007; Squire, 2004).

Finally, an indirect result of the multi-associative memory model is a reconsideration of confabulation theory's original biological hypothesis on symbol to symbol association implementation. The original hypothesis suggests an indirect two stage synfire chain connectivity for association implementation. That hypothesis was predicated on the thought that direct associations between cell assemblies could not be implemented robustly due to the sparse (low probability) connectivity between neurons seen in biological tissue. With the multi-associative memory, associations can be realistically implemented directly and robustly between neural fields with sparse connectivity between neurons and then recalled appropriately.

At the present time in neuroscience, we have an unprecedented amount of work being performed at the molecular/neuron level and at the behavioral level. However, due to the limitations in experimental techniques, there is an invisible gap of knowledge at the systems neurobiology level. The multi-associative memory provides visibility into the basic mathematical principles that may be exploited at the system level and provides several concrete hypothesis on the specific implementation of associations in the cerebral cortex. We hope the multi-associative memory model will be found useful as a concrete theoretical tool for the developing field of integrative neuroscience.

4.4 Methods

The mathematical formulation of the multi-associative memory ultimately has roots in Willshaw's 1969 mathematical formulation of the non-holographic associative memory(Willshaw et al., 1969); however, the multi-associative memory is a significant generalization of past work on associative memories, and extends to a much larger class of problems.

In our use of terminology, we generally refer to cell assemblies as the discrete population of neurons that represents a symbol. A full table of terminology is provided in the symbol list of the dissertation.

4.4.1 MAM connectivity

Let a single *neural field* k contain N_k neurons. A *cell assembly* i in neural field k is defined as a sub-population of the N_k neurons and represented by a N_k by 1 column vector $\underline{x}_i^{[k]}$, where each element $x_{ai}^{[k]} \in \{0,1\} \forall\ a, i$. The superscript $[k]$ will be used to denote a neural field k. If $x_{ai}^{[k]} = 1$, then the a^{th} neuron in neural field k is part of cell assembly i, otherwise if $x_{ai}^{[k]} = 0$, it is not. Source and target neural fields will typically be referred to by the indices k and y respectively, source and target cell assemblies will typically be refereed to by the indices i and j, and source and target neurons will typically be referred to by the indices a and b.

If there are L_k such cell assemblies, then we define the N_k by L_k neural field matrix as, $\mathbf{X}^{[k]} = \begin{bmatrix} \underline{x}_1^{[k]} & \underline{x}_2^{[k]} & \cdots & \underline{x}_{L_k}^{[k]} \end{bmatrix}$, where the i^{th} cell assembly in neural field k is defined by $\underline{x}_i^{[k]}$. If we create the neural field matrix randomly, such that, each neuron is part of on average, r_k cell assemblies, then the average number of neurons per cell assembly is given by M_k, where $p(x_{ai}^{[k]} = 1) = \frac{r_k}{L_k} = \frac{M_k}{N_k}$.

Associations are created uni-directionally between cell assemblies in a source neural field and cell assemblies in a target neural field. If source neural field k has L_k cell assemblies and target neural field y has L_y cell assemblies, the L_y by L_k association matrix is given by $\mathbf{L}^{[y,k]}$, where each element $l_{ji}^{[y,k]} \in \{0,1\} \forall j, i$. Any superscript, such as $[y,k]$, denotes that k is the source neural field and y is the target neural field. If element $l_{ji}^{[y,k]} = 1$ then cell assembly j in target neural field y receives a knowledge link from cell assembly i in source neural field k. For practical application, the $\mathbf{L}^{[y,k]}$ matrix is created through co-occurrence learning. Note that any element may be set to one therefore the matrix is capable of implementing arbitrary associations. We will define the average number of efferent knowledge links per source symbol by the variable λ_c, and the average number of afferent knowledge links per target symbol by the variable λ_r. λ_c and λ_r correspond to the average number of 1's in a column and row respectively of the knowledge base matrix $\mathbf{L}^{[y,k]}$. In order to incorporate the probabilistic nature of the connectivity between neurons, we define a N_y by N_k matrix, $\mathbf{U}^{[y,k]}$, such that $p(u_{ba}^{[y,k]} = 1) = \frac{z^{[y,k]}}{N_y}\ \forall\ b, a$. The value $z^{[y,k]}$ is the average number of unstrengthened "silent" synapses made by each source neuron in source field k with neurons in target field y. The matrix $\mathbf{U}^{[y,k]}$ can be viewed as an unstrengthened probabilistic synaptic connectivity matrix between neurons in the source field k and target field y.

Consider a hebbian learning rule, where all the unstrengthened "silent" synapses existing between neurons in all associated cell assemblies are strengthened or "unsilenced".

Definition 1 (Multi-associative memory connectivity)**.** *The connectivity matrix,* $\mathbf{W}^{[y,k]}$*, defines strengthened synaptic connections between source neural field* k *and target neural field* y*, such*

that

$$\mathbf{W}^{[y,k]} = \left(\mathbf{X}^{[y]}\mathbf{L}^{[y,k]}{\mathbf{X}^{[k]}}^{T}\right)\bigcap\mathbf{U}^{[y,k]}. \tag{4.1}$$

The $\bigcap$ operator is the element by element "AND" operation, such that for the equation $\mathbf{W} = \mathbf{X}\bigcap\mathbf{U}$, if $x_{ba} \geq 1$ and $u_{ba} \geq 1$ then $w_{ba} = 1$, otherwise $w_{ba} = 0$. Notice that after this operation all elements in the connectivity matrix, $\mathbf{W}^{[y,k]}$, are zero or one.

The term $\mathbf{X}^{[y]}\mathbf{L}^{[y,k]}{\mathbf{X}^{[k]}}^{T}$ in (4.1) represents all possible strengthened synapses between neurons in cell assemblies that have been associated. The matrix $\mathbf{U}^{[y,k]}$ is used to determine which of those strengthened connections exist, or in other words were capable of being strengthened.

4.4.2 MAM similarity to past models

We should note here the similarity of (4.1) to past work. First, all traditional associative memories can be viewed as a special case of equation 4.1, where the association matrix is the identity matrix, $\mathbf{L}^{[y,k]} = \mathbf{I}^{[y,k]}$. If $\mathbf{L}^{[y,k]}$ equals the identity matrix, and $p(u_{ba}^{[y,k]} = 1) = 1$(i.e. full initial connectivity), then the matrix $\mathbf{W}^{[y,k]}$ reduces to the original Willshaw connectivity matrix for single associative memories (Willshaw et al., 1969). If $y = k$ and $\mathbf{L}^{[y,k]} = \mathbf{I}^{[y,k]}$, then (4.1) takes the form of an autoassociative memory, such as a Hopfield network(Hopfield, 1982). Finally, if $y = k$, the main diagonal terms in $\mathbf{L}^{[y,k]}$ equal zero, and the first diagonal terms are all ones, then (4.1) could be viewed as modeling a synfire chain(Abeles, 1991). The past work mentioned is limited to the case of a single source neural field, k, resulting in a single connectivity matrix. The multi-associative memory may have multiple source neural fields resulting in multiple $\mathbf{W}^{[y,k]}$ matrices. In addition a single multi-associative memory may combine combinations of the past work mentioned. The present model is therefore a generalization of many related formulations of associative memories, and as a result all present analysis can be readily applied to this past work.

4.4.3 Neuronal activity

We first define activity of a neuron, followed by the activity of a cell assembly. Neurons communicate information via action potentials (Hodgkin and Huxley, 1952), and at time steps of approximately 1ms, action potentials can be modeled as discrete events without affecting the phenomenological behavior of neurons (Izhikevich, 2007).

Definition 2 (Neural field activity)**.** *The activity of all neurons in a neural field k at time t, represented by a N_k by 1 column vector $\underline{\eta}^{[k]}(t) \in \{0,1\}^{N_k}$.*

If, $\eta_i^{[k]}(t) = 1$, then neuron i is said to have fired an action potential at time t. If, $\eta_i^{[k]}(t) = 0$, then it did not fire an action potential. We will refer to a neuron that has fired an action potential at time t as active, otherwise it is inactive.

4.4.4 Neuronal excitation

Due to the discrete nature of action potentials, we assume that a source neuron communicates to all its target neurons through synapses only when the source neuron is active. In our model, if a neuron a is active then all of its target synapses b are activated, meaning they deliver some amount of synaptic excitation to all connected neurons at time t given by the synaptic variable $s_{ba}^{[y,k]}(t)$. The synaptic variable $s_{ba}^{[y,k]}(t)$ may be constant, a random variable, or could be a function describing the dynamics of each synapse. All synaptic dynamics between two neural fields is described by the synaptic matrix $\mathbf{S}^{[y,k]}(t)$. Let $s^{[y,k]}$ represent the average synaptic excitation delivered by a strengthened synapse. For our simulations every individual synaptic excitation is given a random value selected from a normal distribution with mean 1 and standard deviation 0.3.

Definition 3 (Neural field excitation)**.** *The total excitation delivered to all neurons in a neural field y at time t, represented by a N_y by 1 column vector $\underline{\widetilde{\eta}}^{[y]}(t) \in \Re^{N_y}$, where*

$$\underline{\widetilde{\eta}}^{[y]}(t+1) = \sum_{\forall k} \left(\mathbf{S}^{[y,k]}(t) \bullet \mathbf{W}^{[y,k]}\right) \underline{\eta}^{[k]}(t). \tag{4.2}$$

Note that the tilde,~, is used to denote excitation vs. activity. The Hadamard product, $\bullet$, signifies element by element matrix multiplication. The term inside the parentheses represents a new randomly weighted synapse matrix for every time t. Notice that the neuronal excitation $\underline{\widetilde{\eta}}^{[y]}(t+1)$ in (4.2) depends on the activity, $\underline{\eta}^{[k]}(t)$, and connectivity, $\mathbf{W}^{[yk]}$, from all source neural fields k. Axonal delays are not explicitly included in the present model. However, the neuronal activity vector could easily be replaced with a different vector of neuronal activity that represents the source neural field activity with appropriate delay. In this case, the results regarding cell assembly information processing in the target neural field can be viewed without loss of generality.

The transformation from neuron excitation to neuron activity, $\widetilde{\eta}_a^{[y]} \Rightarrow \eta_a^{[y]}(t)$, depends on the choice of neuron model. However, given the neural field activity at time t, the synaptic excitation at time $t+1$ is independent of the choice of a neuron model. Therefore, when we consider given prior neural field activity we do not need to specify a neuron model to analyze the distribution of synaptic excitation in multi-associative memories.

4.4.5 Cell assembly activity and neural field information state

Given the neural field activity we can define the neural field information state.

Definition 4 (Neural field information state (cell assembly activity))**.** *The neural field "k" information state is the fraction of each cell assemblies' neurons that are activated at a given time t, represented by a L_k by 1 column vector, $\underline{\alpha}^{[k]}(t)$, such that*

$$\alpha_i^{[k]}(t) = \frac{\sum_{\forall a} \left(\eta_a^{[k]}(t) \bigcap x_{ai}^{[k]}\right)}{\sum_{\forall a} x_{ai}^{[k]}} \text{ for } i = 1, 2, ..., L_k. \tag{4.3}$$

The numerator in (4.3) is the number of active neurons in cell assembly i at time t. The denominator is the total number of neurons in cell assembly i, thus $\alpha_i^{[k]}(t)$ is the fraction of the neurons in cell assembly i that are active at time t.

Proposition 1. *In a multi-associative memory the neural field information state is a measure of the instantaneous information content of a neural field.*

4.4.6 Symbol excitation

Definition 5 (Symbol excitation)**.** *The symbol excitation delivered to a target symbol j in neural field y, $\tilde{\alpha}_j^{[y]}$, is the sum of all source symbol activations with which the target symbol has been associated, where the vector, $\underline{\tilde{\alpha}}^{[y]}$, containing all target symbol excitations is given by*

$$\underline{\tilde{\alpha}}^{[y]} = \sum_{\forall k} \mathbf{L}^{[y,k]} \underline{\alpha}^{[k]}. \tag{4.4}$$

The symbol excitation is the mathematical equivalent of the symbolic input excitation utilized in confabulation theory. A target symbol (cell assembly) excitation value, $\tilde{\alpha}_j^{[y]}$, can also be thought of as a "linear sum of fractionally active associations", and thus represents the total sum of active associations. For example, if $\tilde{\alpha}_j^{[y]} = 0.3$, then the target symbol (cell assembly) j is receiving a third of an active association. If $\tilde{\alpha}_j^{[y]} = 2.3$ then the target symbol j is receiving two and one-third active associations, and so forth.

4.4.7 Information processing control

The control of a multi-associative memory selects the $A_y(t)$ most highly excited neurons to be activated in neural field y at a given time t, while all other less excited neurons are inhibited.

The use of the variable A_y in our analysis, allows us to ignore neuronal dynamics, yet still come to robust conclusions on the information processing that is implementable in multi-associative memories and nervous systems.

4.4.8 MAM associative signal and associative interference

The purpose of developing an analytic expression for the associative signal and associative interference is to relate the purely symbolic information processing (having no interference) determined by the symbolic excitation in (4.4), to multi-associative memory information processing defined by the neuronal excitation in (4.2).

We present theorem 1 on the strength (associative signal plus associative interference) of symbolic cell assembly excitation. Additional details and insights on the symbolic cell assembly excitation can be found in Appendix-4.5.3. Details on the theorem derivation and proofs can be found in Appendix-4.5.4.

Theorem 1 (Symbolic cell assembly excitation)**.** *If a random number of neurons are uniformly activated in $\mathcal{K}$ source neural fields, given the symbol excitation vector, $\tilde{\underline{\alpha}}^{[y]}$, where the multi-associative memory is defined by the connectivity matrices, $\mathbf{W}^{[y,k]}$, identically parameterized by the underlying variables $M_k, r_k, r_y, N_k, N_y, L_k, L_y, \lambda_c, \lambda_r, z^{[y,k]}$, and synaptic excitation variable, $s^{[y,k]}$, the expected value of the synaptic input onto a neuron b in target cell assembly j, $I_j = E\left[\tilde{n}_{bj}^{[y]} | \tilde{\underline{\alpha}}^{[y]}\right]$, can be viewed as the sum of two independent random variables, namely the target cell assembly associative signal signal ($\mathbf{S_j}^{[y]}$) and associative interference ($\mathbf{N_j}^{[y]}$),*

$$I_j = E\left[\tilde{n}_{bj}^{[y]} | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\mathbf{S_j}^{[y]} | \tilde{\underline{\alpha}}^{[y]}\right] + E\left[\mathbf{N_j}^{[y]} | \tilde{\underline{\alpha}}^{[y]}\right] = S_j s^{[y,k]} + (A_{\mathcal{K}} - S_j)\, p_w s^{[y,k]}, \tag{4.5}$$

where, the expected number of associative signal source neurons that are active is given by

$$S_j = \frac{\mathcal{K} N_k z^{[y,k]}}{N_y} \left(1 - \left(1 - \frac{M_k \tilde{\alpha}_j^{[y]}}{\mathcal{K} N_k r_k}\right)^{r_k}\right),$$

the expected number of total source neurons that are active is given by

$$A_{\mathcal{K}} = N_k \mathcal{K} \left(1 - \left(1 - \frac{L_k M_k \sum_{\forall i} \tilde{\alpha}_i^{[y]}}{L_y \mathcal{K} N_k \lambda_r r_k r_k}\right)^{r_k}\right),$$

and the probability of strengthened connectivity between any two neurons is given by

$$p_w = \left(1 - \left(\left(1 - \frac{r_y \lambda_c}{L_y^2}\right)^{L_y} + \left(1 - \left(1 - \frac{r_y \lambda_c}{L_y^2}\right)^{L_y}\right)\left(1 - \frac{r_k}{L_k}\right)\right)^{L_k}\right) \frac{z^{[y,k]}}{N_y},$$

where, for the simplifying assumption $L_y = L_k$ and biologically reasonable values, the expectations can be reduced to the more intuitive approximations

$$S_j \approx \tilde{\alpha}_j^{[y]} M_k \frac{z^{[y,k]}}{N_y},$$

the sum of symbolic excitation on cell assembly j times the number of neurons per source cell assembly times the probability of connectivity between neurons approximates the number of active synapses on a neuron in cell assembly j,

$$A_{\mathcal{K}} \approx \left(\sum_{\forall i} \tilde{\alpha}_i^{[y]} \right) \frac{M_k}{\lambda_r r_k},$$

the sum of all symbolic excitations times the number of neurons per source cell assembly divided by the product of the average number of associations per source cell assembly and the average number of neurons per cell assembly approximates the total number of active source neurons,

$$p_w \approx \left(1 - e^{-\frac{r_k r_y \lambda_c}{L_k}} \right) \frac{z^{[y,k]}}{N_y}.$$

Although the theorem enforces that all neural fields have the same parameters, in general if many different parameters exist, each connectivity between neural fields could be treated independently by the same theorem by setting $\mathcal{K} = 1$. Because of the independence of neural field connectivity, the symbolic cell assembly excitation calculated from theorem 1 can be added together for each independent neural field connectivity. In this case, each $\tilde{\alpha}^{[y]}$ would have to be calculated independently from each source neural field initialization and should be written $\tilde{\alpha}^{[y,k]}$. Therefore, the treatment of multiple source neural fields with different parameters is easily handled with the same theorem.

Proof. See section 4.5.4 for the proof and derivation of the probability distributions for each variable. □

4.5 Appendix

4.5.1 Association matrix $\mathbf{L}^{[yk]}$ experimental structure

The association matrix $\mathbf{L}^{[y,k]}$ is not arbitrary. Each element of the matrix represents a learned association between two symbols. In the real world, associations are made between certain symbols (such as words) and not between others. In order to address the non-random structure of the $\mathbf{L}^{[y,k]}$ matrices, we rely on an established mathematical neuroscience theory termed confabulation theory that discusses the creation of associations between symbols in the cerebral cortex(Hecht-Nielsen, 2007)(Solari et al., 2008).

In brief, we will relate confabulation theory to the present model. Confabulation theory proposes that all cognition is the result of the maximization of cogency, argued as biologically

equivalent to the maximization of the confabulation product $p(\underline{x}_a^{[k]}|\underline{x}_j^{[y]})p(\underline{x}_b^{[k+1]}|\underline{x}_j^{[y]})...$, where $\underline{x}_a^{[k]}$ and $\underline{x}_b^{[k+1]}$ are source cell assemblies in neural fields k and $(k+1)$ respectively, and $\underline{x}_j^{[y]}$ is a target cell assembly in neural field y. Given the number of co-occurrences of active source cell assemblies with active target cell assemblies, each conditional probability can be calculated. From these probabilities, confabulation theory provides the mathematical framework in order to create the association matrices $\mathbf{L}^{[y,k]}$.

Each $\mathbf{L}^{[y,k]}$ matrix represents associations between symbols. In our experiment, we simulate a plausible natural formation of the matrix by creating associations between individual words (each word is one symbol) in well written text. For comparison with prior theoretical models, the association matrices used in the present experiment were identical to those created in a previous paper(Hecht-Nielsen, 2005). For implementation in the present model, all matrices were converted to binary matrices by the heavyside operation $\mathbf{L}^{[y,k]} = H(\mathbf{L}^{[y,k]})$, which converts all non-zero elements to one. The consequence, as compared with the past experiment, is that the present experiments only deal with the existence of associations without implementing directly the strength of associations. For continuity we summarize the creation of the three matrices $\mathbf{L}^{[y,1]}$, $\mathbf{L}^{[y,2]}$, $\mathbf{L}^{[y,3]}$.

Figure 4.1 shows the organization of the association matrices $\mathbf{L}^{[y,k]}$ based on the use of real text data. Each colored box in Figure 4.1 represents a lexicon containing the same 10,000 most commonly encountered words in the training corpus. In the neural simulation, we let each neural field represent a lexicon and each cell assembly in a neural field represent a word. Therefore each $\mathbf{X}^{[k]}$ neural field contains 10,000 cell assemblies, and associations between words are translated into associations between cell assemblies.

During training, a 1.4 x 10^9 proper English training text corpus is read serially into the four-contiguous-word window. Three individual 10,000 by 10,000 count matrices, $\mathbf{C}^{[y,k]}$, record the co-occurence counts between the word in each source(green) position and the word in the target(red) position. After all counts were recorded, any $c_{ij}^{[y,k]}$ element less than 4 counts, is set to zero. The conditional probability, $p(\psi|\lambda)$, between each source(ψ) and target(λ) word is then approximated as $c_{\lambda\psi}^{[y,k]} / \sum_{\forall j} c_{\lambda j}^{[y,k]}$. If the conditional probability is greater than 0.001 the association matrix element $l_{\lambda\psi}^{[y,k]} = 1$ otherwise $l_{\lambda\psi}^{[y,k]} = 0$.

After training each association matrix has the following number of non-zero(nnz) elements: nnz($\mathbf{L}^{[y,1]}$)=870,090; nnz($\mathbf{L}^{[y,2]}$)=1,293,503; nnz($\mathbf{L}^{[y,3]}$)=1,520,274.

Here we would like to emphasize that all the association matrices created by confabulation theory principles have an identifiably similar structure. As discussed in the signal to noise analysis, the structure of $\mathbf{L}^{[y,k]}$ has important consequences in the performance of the system.

The row and column sum histograms of $\mathbf{L}^{[y,1]}$ are shown in Figure 4.6. Figure 4.6(a)

shows a histogram of the row sum of the matrix $\mathbf{L}^{[y,1]}$ related to λ_r. Figure 4.6(b) shows a histogram of the column sum of the matrix $\mathbf{L}^{[y,1]}$ related to λ_c. The first observation from (a) is that the number of source cell assemblies associated with a single target cell assembly can be approximated by a binomial distribution with a median≈87. In contrast, the number of target cell assemblies associated with a single source cell assembly, shown in (b), can be approximated by a log-normal distribution with median≈18. The key point is that for any randomly activated source cell assembly, the total number of active associations will typically be very small relative to the total number of cell assemblies, thus in some sense λ_c for active cell assemblies is small for most active source cell assemblies.

The effect is really demonstrated when multiple neural fields are simultaneously transmitting information to a target neural field. The larger the number of active associations the larger the noise, therefore the log-normal distribution statistically ensures that for any randomly activated cell assemblies the number of active associations will be low, thus improving the signal to noise ratio of each combination of associations. As a result, we hypothesize that the structure of associations in other real world data, and furthermore, associations in the brain should have a similar structure.

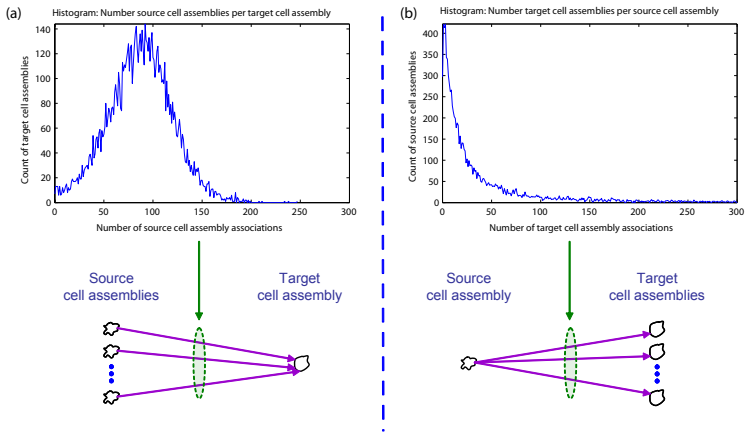

Figure 4.6: Association matrix $\mathbf{L}^{[yk]}$ row/column sum histograms. Histograms shown for row(a) and column(b) sums of the association matrix $\mathbf{L}^{[yk]}$ for $k = 1$. (a) shows a normal distribution of many to one type associations. (b) shows a log-normal distribution of one to many type associations. The log-normal structure of one to many type associations is likely an important aspect of information processing in generalized associative memories.

4.5.2 Biological realism of cortical depiction and development

Figure-4.1 depicts a 6-layered cortical module with connections between layer 3 neural fields. Our depiction is not random and relies on a great deal of neuroscience facts to appropriately depict cortical organization (see chapter 3). Figure-4.5 additionally relies on many neuroscience facts and hypothesis discussed here.

Structurally, the cerebral cortex is a thin continuous sheet of neurons, which evolved as a 6 layered structure(Northcutt and Kaas, 1995). Like layers in a cake, each layer is visually distinguishable by an alternating density of neuronal cell bodies when stained. Connections between cortical areas in the brain can be predicted by the source and target layers from which the connections arise(Barbas, 1986; Thomson and Bannister, 2003; Yeterian and Pandya, 1994), since the large bundles of axons(fascicles) connecting different areas of the brain are established genetically(Schmahmann and Pandya, 2006). The projections from one area of cortex to another are relevant for associating different information, because, spatially localized cortical areas are responsible for processing specific information objects, such as visual objects, visual motion, language production, or language comprehension(Broca, 2006; Desimone et al., 1984; Tsao et al., 2006). The axonal projections connecting large populations of neurons in one cortical area to another provide a genetically programmed axonal highway capable of associating different types of information, yet individual synaptic connections are probabilistic(Braitenberg and Shuz, 1998). Additionally, in vivo electrophysiological recordings demonstrate that individual neurons respond maximally to specific attributes of information objects represented by each cortical area, for example a single neuron may respond to a specific face in a cortical area that responds to faces in general(Quiroga et al., 2005). Since it seems unlikely that a single neuron by itself is exclusively necessary for information storage or processing, populations of neurons(cell assemblies) most likely encode information(Young and Yamane, 1992; Hebb, 1949).

Within a 6-layered cortical volume in the mammalian cerebral cortex there exist many different populations of neurons. Certain populations may reside in different cortical layers, or others in the same layer may utilize different neurotransmitters(Gupta et al., 2000; Thomson and Bannister, 2003). The axonal projections between different cortical areas depend on the source and target populations of neurons in each cortical area(Barbas and Hilgetag, 2002).

The utilization of 1,000,000 neurons in a simulation is selected based on known anatomical human data. All parameters were based on von Economo's cytoarchitectonic data on the human brain(von Economo, 1929). Numbers are used from area TA, corresponding to much of Wernicke's area in the brain, known for processing language. In von Economo's estimates, a

unit of measure is given as $0.001mm^3$. There are 40 cells per unit in layer III. Assuming 85% of neurons are pyramidal, the number of layer III pyramidal neurons per unit in area TA is estimated as $0.85 * 40 = 34$. The thickness of layer III in TA is given as $0.95mm$, therefore there are approximately $0.95 * 34 * 1000 = 32,300$ neurons in $1mm^2$ of human layer III TA cerebral cortex. In $30mm^2$ of TA cortex there are approximately 969,000 total neurons, which we round to 1,000,000.

Regarding the implementation of random synaptic activity. Cortical synapses in the brain are far from static structures. They can form between an axon and dendrite that are spatially close, or they can disappear after they have formed(Craig and Lichtman, 2001). Once a synapse has formed, the synaptic effect on the target neuron is not a fixed quantity. Many synapses can be classified as "silent" synapses(Malenka and Siegelbaum, 2001), which might be considered unstrengthened synapses, in that they have little to no effect on the post-synaptic target. Unstrengthened synapses remain silent until a change takes place which transforms the synapse into a strengthened synapse(Malenka and Siegelbaum, 2001).

In strengthened synapses synaptic efficacy involves the probabilistic quantal release of neurotransmitter(Malenka and Siegelbaum, 2001). Therefore, for every pre-synaptic action potential activating a synapse to release neurotransmitter, the effect on the target neuron is a probabilistic quantity.

Finally, we provide a biological description of the formation of the association matrix $\mathbf{L}^{[y,k]}$ to demonstrate the consistencies with known neuroanatomy and neuroscience fact.

Imagine the two neural fields k and y reside in layer 3 of cortical areas that develop information representations independently, such as auditory words and visual objects respectively. During development, a sub-population of neurons in the neural field, group to respond exclusively to one item of information represented by that cortical area. That sub-population is a cell assembly, and may represent one word among all words stored in a cortical area. During everyday experiences active cell assemblies will co-occur in each neural field. The layer 3 neural fields are both directly connected with each other and with the hippocampus. If two active cell assemblies co-occur, the hippocampus and related structures determine if the co-occurrence is "worthy" of forming an association. If "worthy", the hippocampus temporarily stores the association between the two cell assemblies as an index(Teyler and Rudy, 2007). Over time, presumably during sleep, the hippocampus would activate the cell assemblies in each neural field in order to strengthen the unstrengthened synapses that exist directly between the cell assemblies.

Anatomical evidence exists showing reciprocal connections between all cortical association areas and the hippocampus necessary to implement such an architecture(Witter et al., 1989).

A direct prediction of this model consistent with present neuroscience evidences is that the removal of the hippocampus would result in the inability to form new associations, and should have a temporally graded loss of associations. The associations being discussed would be related to declarative rather than non-declarative memories(Squire, 2004). Using relationships between words as training data is consistent with declarative memory organization.

In our model, the hippocampus would thus be responsible for creating and maintaining the $\mathbf{L}^{[yk]}$ matrices. Here, the hippocampus initially stores and implements associations between neural fields, and over time, the direct synapses would be strengthened for associations that were deemed by some biological mechanism as important.

We propose that the strengthening of associations involves two processes. First, the existing unstrengthened synapses are strengthened. Second, not implemented in simulations, further strengthening of an association involves an effective increase in the probability of connectivity between the two associated cell assemblies. This process would involve the formation of new synapses and structural plasticity between the neurons in each cell assembly. The second process is likely to contribute more significantly to the temporally graded nature of associations.

4.5.3 Associative signal and associative interference comments

First and most importantly, associative signal signal and associative interference in multi-associative memories must be referenced to a specific target cell assembly! We cannot talk about the signal and interference strength of a multi-associative memory "channel" or the signal and interference of the multi-associative memory connectivity as a whole. Since the multi-associative memory is defined by the learned associations between cell assemblies (symbols), the associative signal onto a target cell assembly j is the excitatory input delivered to j from source cell assemblies i with which it has been associated (i.e. $l_{ij}^{[y,k]} = 1$). The associative interference onto a target cell assembly is that excitatory input delivered to cell assembly j from source cell assemblies i with which it has NOT been associated (i.e. $l_{ij}^{[y,k]} = 0$).

Discussing the synaptic input onto a cell assembly is initially not a well formed problem, since individual synaptic input is by definition from neuron to neuron. However, for any neuron in a specific target cell assembly j, we can partition each active synaptic input into one of two non-overlapping populations of source neurons. The two populations are called the "associative signal" and "associative interference" with reference to target cell assembly j. Figure 4.7 demonstrates sources of associative signal (solid black lines) and associative interference (dashed black lines) in a multi-associative memory. Associative interference is exclusively a result of target cell

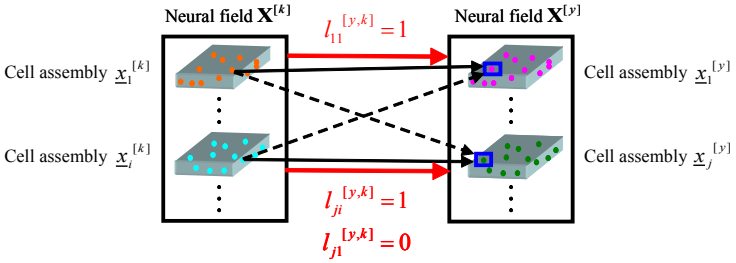

Figure 4.7: Multi-associative memory associative signal and interference neuronal depiction. The blue box on the right highlights a single neuron that is part of two target cell assemblies. In put into the target neuron is defined in reference to a specific target cell assembly, not the neuron itself. Associative signal (solid lines) and associative interference (dashed lines) are shown for the two target cell assemblies. Since target cell assembly $\underline{x}_1^{[y]}$ has been associated with $\underline{x}_1^{[k]}$, synaptic input into neuron b, when viewed as part of $\underline{x}_1^{[y]}$, from neurons in $\underline{x}_1^{[k]}$ or other associated cell assemblies are considered associative signals. Any other active neurons in non-associated cell assemblies, such as from $\underline{x}_i^{[k]}$ are considered associative interference. Notice that synaptic inputs into the same neuron b, when viewed as part of $\underline{x}_j^{[y]}$, from source cell assembly $\underline{x}_i^{[k]}$, are now considered associative signal while all other synaptic inputs are considered associative interference.

assembly overlap. In figure 4.7 the same neuron b (blue box on right) is part of two different target cell assemblies $\underline{x}_1^{[y]}$ and $\underline{x}_j^{[y]}$. Since target cell assembly $\underline{x}_1^{[y]}$ has been associated with $\underline{x}_1^{[k]}$, synaptic input into neuron b, when viewed as part of $\underline{x}_1^{[y]}$, from neurons in $\underline{x}_1^{[k]}$ or other associated cell assemblies are considered associative signals. Any other active neurons in non-associated cell assemblies, such as from $\underline{x}_i^{[k]}$ are considered associative interference. Notice that synaptic inputs into the same neuron b, when viewed as part of $\underline{x}_j^{[y]}$, from source cell assembly $\underline{x}_i^{[k]}$, are now considered associative signal while all other synaptic inputs are considered associative interference.

As a result, we immediately see that the same synaptic connection can be considered associative signal in reference to one cell assembly and associative interference in reference to another. Therefore, it is crucial to realize that associative signal and interference in multi-associative memories only has meaning in reference to a specific target cell assembly. Again the neurons considered as associative signal are those that are active and in a source cell assembly i that has been associated with target cell assembly j (i.e. $l_{ij}^{[y,k]}=1$). The neurons considered as associative interference are all other active source neurons which are not in any of the source cell assemblies that are associated with the target cell assembly j, yet still provide synaptic input onto neurons in target cell assembly j due cell assembly overlap.

Figure 4.8 shows the sources of associative signal and associative interference from the perspective of the mathematical model. Since the connectivity of all neurons is determined by equation 4.1, we can determine which matrix configurations account for associative signal and which for associative interference. Between Figure 4.7 and Figure 4.8 we attempt to graphically

depict the sources of associative signal and associative interference in multi-associative memories. Again, input and interference are in reference to a target cell assembly, NOT to an individual neuron or synapse. The same synapse may be associative signal in reference to one target cell assembly and associative interference to another.

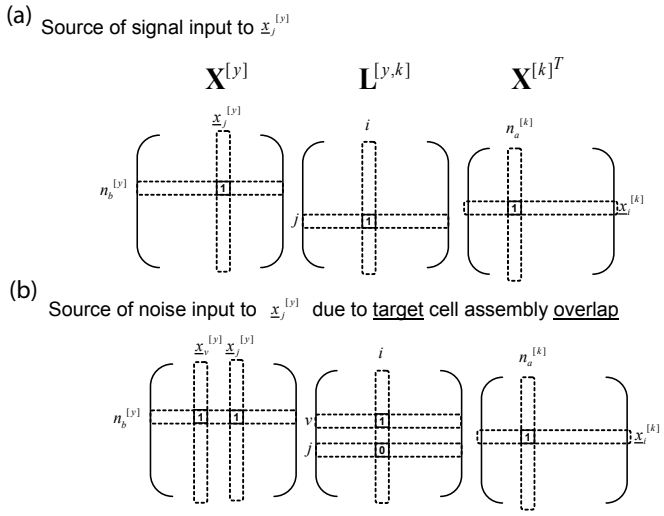

Figure 4.8: Sources of associative signal and interference as viewed from equation 4.1. In (a) source cell assembly i is associated with target cell assembly j, therefore any synaptic input to neurons in j from neurons in source cell assembly i are considered associative signal. In (b) source cell assembly i is NOT associated with target cell assembly j, but target neuron b is part of target cell assembly v which has been associated with i. Source cell assembly i will therefore provide some amount of input into target cell assembly j even though they haven't been associated, which is considered associative interference.

One immediate benefit of the formulation of associative signal and interference in a multi-associative associative memory enables insight into parameters in multi-associative memories that might be optimized to improve the performance of symbolic information processing. Having spent significant time analyzing multi-associative memories, we will attempt to mention some of the critical variables regarding the associative signal and interference that are clearly important, hopefully providing the reader with additional insight.

The product $r_y r_k \lambda_c$ is a critical term useful for ratio of associative signal to associative interference. The terms r_y and r_k determine the number of cell assemblies per neuron. Therefore, both r_y and r_k describe the amount of overlap between cell assemblies. The less overlap between cell assemblies, the lower the cross talk, the lower the interference, and the higher the associative signal to interference ratio. If $r_y = 1$ with zero variance, then there will be zero noise. The term λ_c represents the number of associations per source cell assembly. Importantly, this parameter can be controlled during the learning of associations in the $\mathbf{L}^{[y,k]}$ matrix and suggests that minimizing

the number of efferent associations per source cell assembly is critical to the performance.

We predict certain biological mechanisms that might exist to improve the associative signal to interference ratio for more accurate symbolic information processing. First, r_y and r_k must be greater than 1 in order to store more cell assemblies. However, keeping the overall variance in the number of cell assemblies per neuron low will decrease the variance on the noise, therefore, we predict that neurons are in general part of the same number of cell assemblies with low variance. Second, a critical feature of the $\mathbf{L}^{[y,k]}$ matrix will improve the SNR. During actual information processing, a limited number of source cell assemblies will be fully active, therefore, keeping the column sums of the matrix small (i.e. the number of efferent associations per cell assembly) will decrease the noise substantially. We predict the need to minimize SNR through establishing few efferent source cell assembly associations is one possible role for the hippocampus and the slow consolidation of cortically based long term memory.

4.5.4 Associative signal and interference proofs

Note, in the present proofs section we interchangeably refer to associative signal as "signal" and associative interference as "noise".

Associative signal derivation

Definition 6 (Multi-associative memory signal)**.** *The number of active synaptic inputs onto a neuron in a target cell assembly j which originate from neurons in source cell assemblies i that have been associated with the target cell assembly j, such that $l_{j,i}^{[y,k]} = 1$.*

Lemma 1. *If a random number of neurons are uniformly activated in $\mathcal{K}$ source neural fields, then given the knowledge link excitation value for target cell assembly j, $\tilde{\alpha}_j^{[y]}$, where the multi-associative memory connectivity matrices, $\mathbf{W}^{[y,k]}$, are identically parameterized by M_k, N_k, N_y, r_k, $z^{[y,k]}$, $s^{[y,k]}$, the distribution of the number of associated strengthened synaptic inputs, $p\left(\mathcal{S}_j = x|\tilde{\alpha_j}^{[y]}\right)$, onto any target cell assembly j is given by*

$$p\left(\mathcal{S}_j = x|\tilde{\alpha_j}^{[y]}\right) = B\left(x; \mathcal{K}N_k; \frac{z^{[y,k]}}{N_y}\left(1 - \left(1 - \frac{M_k\tilde{\alpha_j}^{[y]}}{\mathcal{K}N_k r_k}\right)^{r_k}\right)\right), \tag{4.6}$$

where,

$$S_j = E\left[\mathcal{S}_j|\tilde{\alpha_j}^{[y]}\right] = \frac{\mathcal{K}N_k z^{[y,k]}}{N_y}\left(1 - \left(1 - \frac{M_k\tilde{\alpha_j}^{[y]}}{\mathcal{K}N_k r_k}\right)^{r_k}\right), \tag{4.7}$$

which, for $\frac{M_k\tilde{\alpha}_j^{[y]}}{\mathcal{K}N_k} << 1$ and large $\mathcal{K}N_k$ can be reasonably approximated as

$$S_j \approx \tilde{\alpha}_j^{[y]} M_k \frac{z^{[y,k]}}{N_y}. \tag{4.8}$$

Proof. We need to find the expected value of the number of active neurons in associated cell assemblies. Since $\tilde{\alpha}_j^{[y]}$ is the sum of the fraction of neurons that are active in each associated cell assembly in each neural field, then $\frac{M_k \tilde{\alpha}_j^{[y]}}{\mathcal{K}}$ represents the expected sum of all neurons in associated cell assemblies in a single neural field IF there is no overlap of neurons in separate cell assemblies. Note that if a neuron is active and in multiple associated cell assemblies then it will contribute multiple times to the value of $\tilde{\alpha}_j^{[y]}$. The expected value of the total number of 1's in any matrix $\mathbf{X}^{[k]}$ is $N_k r_k$. From (4.3), if a neuron is active then we can view a single row in the matrix $\mathbf{X}^{[k]}$ as being multiplied by 1. If a neuron is inactive then the row in matrix $\mathbf{X}^{[k]}$ is multiplied by 0. For a single neural field, we can state $p(\eta_a^{[k]} x_{ai}^{[k]} = 1 | \tilde{\alpha}_j^{[y]}) = \frac{M_k \tilde{\alpha}_j^{[y]}}{\mathcal{K} N_k r_k}$, therefore $p(\eta_a^{[k]} x_{ai}^{[k]} = 0 | \tilde{\alpha}_j^{[y]}) = \left(1 - \frac{M_k \tilde{\alpha}_j^{[y]}}{\mathcal{K} N_k r_k}\right)$. As noted, if $\eta_a^{[k]} = 1$ then all 1's in a row of $\mathbf{X}^{[k]}$ must contribute to the sum $\tilde{\alpha}_j^{[y]}$. For a neuron to be inactive, it is equivalent to $p(\eta_a^{[k]} x_{ai}^{[k]} = 0)$ over r_k trials, one for each cell assembly the neuron is a part of, or $p(\eta_a^{[k]} = 0) = \left(1 - \frac{M_k \tilde{\alpha}_j^{[y]}}{\mathcal{K} N_k r_k}\right)^{r_k}$. The number of active and associated neurons is therefore given by a binomial distribution with probability of success equal to $\frac{z^{[y,k]}}{N_y}\left(1 - \left(1 - \frac{M_k \tilde{\alpha_j}^{[y]}}{\mathcal{K} N_k r_k}\right)^{r_k}\right)$, which is the probability of $\eta_a^{[k]} = 1$ AND the propagability of the neuron being connected $\frac{z^{[y,k]}}{N_y}$. Since there are N_k trials in $\mathcal{K}$ neural fields, the expected value of the number of synaptic inputs is simply $N_k \mathcal{K}$ times the last probability discussed. The total excitation includes the contribution of the strength of each synaptic input. Since the synaptic contribution is independent of the number of synaptic inputs we can just multiply by the expected value of a single synaptic excitation resulting in equation (4.7).

The second half of the proof follows from $\lim_{n\to\infty}(1-\frac{x}{n})^n = e^{-x}$ and from $\lim_{n\to\infty} n(1 - e^{-\frac{x}{n}}) = x$, therefore using the limits as approximations for large n, completing the proof. □

Intuitively, the signal is basically the sum of the average synaptic input delivered by all active neurons in associated cell assemblies, times the initial unstrengthened probability of connectivity between any two neurons.

Associative interference derivation

As will become evident in our discussion, the noise of communication between symbols in neural fields is most efficiently characterized by the probability that two neurons are randomly connected.

Lemma 2. *If a random number of neurons are uniformly activated in $\mathcal{K}$ source neural fields, then given the knowledge link excitation vector $\underline{\tilde{\alpha}}^{[y]}$, where the multi-associative memory connectivity*

matrices, $\mathbf{W}^{[y,k]}$*, are identically parameterized by* $M_k, N_k, N_y, r_k, z^{[y,k]}$*, the total number of active source neurons,* $\mathcal{A}$*, has a probability distribution given by*

$$p(\mathcal{A}_{\mathcal{K}} = x|\underline{\tilde{\alpha}}^{[y]}) = B\left(x; N_k\mathcal{K}; 1 - \left(1 - \frac{L_k M_k \sum_{\forall i} \tilde{\alpha}_i^{[y]}}{L_y \mathcal{K} N_k \lambda_r r_k r_k}\right)^{r_k}\right), \tag{4.9}$$

where,

$$A_{\mathcal{K}} = E\left[\mathcal{A}_{\mathcal{K}}|\underline{\tilde{\alpha}}^{[y]}\right] = N_k\mathcal{K}\left(1 - \left(1 - \frac{L_k M_k \sum_{\forall i} \tilde{\alpha}_i^{[y]}}{L_y \mathcal{K} N_k \lambda_r r_k r_k}\right)^{r_k}\right), \tag{4.10}$$

which for biologically reasonable values, and $L_y = L_k$ *can be approximated and simplified as*

$$A_{\mathcal{K}} \approx \left(\sum_{\forall i} \tilde{\alpha}_i^{[y]}\right) \frac{M_k}{\lambda_r r_k} \tag{4.11}$$

Proof. For a single source neural field, if we are given the cell assembly activation vector $\underline{\alpha}$ for that neural field, we can calculate the expected number of active neurons in that neural field. If we count all active neurons in cell assemblies independently then the total number of active neurons is given by the product $M_k \sum_{\forall i} \alpha_i$. The total number of possible independent cell assembly neuron activations is given by the product $N_k r_k$, which is the number of 1's in the matrix $\mathbf{X}^{[k]}$. If we view a neuron $\eta_a^{[k]}$ as multiplying its corresponding row a in the matrix $\mathbf{X}^{[k]}$ by its activity, then the probability that $p(\eta_a^{[k]} x_{ia}^{[k]} = 1) = \frac{M_k \sum_{\forall i} \alpha_i}{N_k r_k}$. In this case, since there are r_k cell assemblies per neuron, the probability of a 1 in a randomly selected cell assembly i in row a is $\frac{M_k \sum_{\forall i} \alpha_i^{[k]}}{N_k r_k r_k}$. In order for the neuron $\eta_a^{[k]}$ to be inactive, its corresponding "activity" in cell assembly i in row a must be zero for all r_k cell assemblies that it is a part of. Therefore the probability that a single neuron $\eta_a^{[k]} = 0$ in a single neural field is $\left(1 - \frac{M_k \sum_{\forall i} \alpha_i}{N_k r_k r_k}\right)^{r_k}$. The total number of active neurons from $\mathcal{K}$ neural fields with average equal activation, $\sum_{\forall i} \alpha_i$, is then given by a binomial distribution $p(\mathcal{A}_{\mathcal{K}} = x) = B\left(x; N_k\mathcal{K}; 1 - \left(1 - \frac{M_k \sum_{\forall i} \alpha_i}{N_k r_k r_k}\right)^{r_k}\right)$. Since we are only given $\underline{\tilde{\alpha}}^{[y]}$ we must estimate $\sum_{\forall i} \alpha_i$. From equation 4.4 and the definition of $\mathbf{W}^{[y,k]}$ in equation 4.1 and noting that all neurons are randomly activated with equal probability in all identically parameterized source neural fields,

$$E\left[\sum_{\forall j} \tilde{\alpha}_j^{[y]}\right] = E\left[\sum_{\forall j}\left(\sum_{\forall k} L^{[y,k]} \underline{\alpha}^{[k]}\right)_j\right] = \lambda_r L_y \mathcal{K} \frac{E\left[\sum_{\forall i} \alpha_i^{[k]}\right]}{L_k}$$

for any k, therefore,

$$E\left[\sum_{\forall i} \alpha_i\right] = \frac{L_k \sum_{\forall j} \tilde{\alpha}^{[y]}}{\lambda_r L_y \mathcal{K}}.$$

Substituting the expected value of average source cell assembly activation into the binomial distribution completes the proof for the probability distribution.

The two biologically reasonable conditions that must hold are $\frac{\sum_{\forall i}\tilde{\alpha}_i^{[y]}}{\mathcal{K}\lambda_r L_y r_k} << r_k$ and The simplification follows from $\lim_{n\to\infty}(1-\frac{x}{n})^n = e^{-x}$ and from $\lim_{n\to\infty} n(1-e^{-\frac{x}{n}}) = x$, therefore using the limits as approximations for large n and substituting completes the proof. □

Lemma 3. *Given a multi-associative memory connectivity matrix,* $\mathbf{W}^{[y,k]}$*, parameterized by the variables* $r_k, L_k, r_y, L_y, N_y, \lambda_r, \lambda_c, z^{[y,k]}$*, the probability,* $p_w = p(w_{b,a}^{[y,k]} = 1)$*, that a randomly selected source neuron* a *in neural field* k *is connected to a randomly selected target neuron* b *in neural field* y *is given by*

$$p_w = \left(1 - \left(\left(1 - \frac{r_y\lambda_c}{L_y^2}\right)^{L_y} + \left(1 - \left(1 - \frac{r_y\lambda_c}{L_y^2}\right)^{L_y}\right)\left(1 - \frac{r_k}{L_k}\right)\right)^{L_k}\right)\frac{z^{[y,k]}}{N_y}. \tag{4.12}$$

Proof. First, from the left term in equation 4.1, the product $x_{bj}^{[y]} l_{j,i}^{[y,k]} x_{ai}^{[k]} = 0$ must hold $\forall i, j$ in order for neuron a <u>not</u> to be connected to neuron b. By definition, $p(x_{b,j}^{[y]} = 1) = \frac{r_y}{L_y}$ and $p(l_{j,i}^{[y,k]} = 1) = \frac{\lambda_c}{L_y}$, therefore the probability that the product of the two equals zero is $p(x_{b,j}^{[y]} l_{j,i}^{[y,k]} = 0) = p_{bj} = \left(1 - \frac{r_y\lambda_c}{L_y^2}\right)$. If we index this probability over j then the product must hold for L_y trials, and the probability of $x_{b,j}^{[y]} l_{j,i}^{[y,k]} = 0$ for all L_y trials is $p_{bj}^{L_y}$. Note that $x_{b,j}^{[y]} l_{j,i}^{[y,k]} = 0$ is a sufficient condition for $x_{bj}^{[y]} l_{j,i}^{[y,k]} x_{ai}^{[k]} = 0$. However, we must also consider the condition that $x_{b,j}^{[y]} l_{j,i}^{[y,k]} = 1$ and $x_{ai}^{[k]} = 0$, which has a probability equal to $\left(1 - p_{bj}^{L_y}\right)\left(1 - \frac{r_k}{L_y}\right)$. The probability that $x_{bj}^{[y]} l_{j,i}^{[y,k]} = 0$ or [$x_{b,j}^{[y]} l_{j,i}^{[y,k]} = 1$ and $x_{ai}^{[k]} = 0$] must be indexed over i for L_k trials, which gives the sum of the two probabilities to the L_k power. Finally, the probability of a connection is one minus that probability, giving the form of term in the first parentheses in (4.12).

Second from the right term in (4.1), $u_{b,a}^{[y,k]} = 1$ must also hold for a connection to exist. By definition the probability $p(u_{b,a}^{[y,k]} = 1) = \frac{z^{[y,k]}}{N_y}$. The final probability is the product of the probabilities for the first and second necessary conditions, thereby completing the proof. □

Lemma 4. *Under the biologically plausible conditions that,* $\lambda_c << L_y$*,* $r_y << L_y$*, and* $r_k << L_k$*,* $p(w_{b,a}^{[y,k]} = 1)$ *can be accurately approximated as*

$$p(w_{b,a}^{[y,k]} = 1) \approx \frac{z^{[y,k]}}{N_y}\left(1 - e^{-r_k\left(1 - e^{-\frac{r_y\lambda_c}{L_y}}\right)}\right), \tag{4.13}$$

which if $L_y = L_k$ *and* L_k *is large, reduces to*

$$p(w_{b,a}^{[y,k]} = 1) \approx \frac{z^{[y,k]}}{N_y}\left(1 - e^{-\frac{r_k r_y \lambda_c}{L_k}}\right). \tag{4.14}$$

Proof. Since, $\lim_{n\to\infty}(1-\frac{x}{n})^n = e^{-x}$, then under the condition that $x << n$, $(1-\frac{x}{n})^n \approx e^{-x}$. The proof is simply three substitutions in lemma 3 using this approximation and some basic re-arrangement of the equation. The biological plausibility requires that the average number of

afferent associations per source cell assembly, λ_c, is kept small relative to the total number of target cell assemblies. Additionally, the condition must hold that the number of cell assemblies r_k or r_y that a target or source neuron is a part of is a small fraction of the total number of cell assemblies. Regarding the simplification. First we make the two substitution, $r_k = \frac{r_k L_k}{L_k}$ and $L_y = L_k$, giving $\frac{z^{[y,k]}}{N_y} e^{-\frac{r_k L_k}{L_k}\left(1-e^{-\frac{r_y \lambda_c}{L_k}}\right)}$. Next we note that $\lim_{n\to\infty} n(1-e^{-\frac{x}{n}}) = x$, therefore for large L_k the equation can be reasonably approximated as stated. □

Synaptic input proof

Given the instantaneous symbol excitation vector $\tilde{\underline{\alpha}}^{[y]}$, and the multi-associative memory parameterizing variables, all active synaptic input onto neurons in a target cell assembly j, defined by (4.2), can be viewed as the sum of two independent random variables, $\mathbf{S}_j$ and $\mathbf{N}_j$. The signal $\mathbf{S}_j$ is the total synaptic excitation contributed by neurons that are active in source cell assemblies, which are associated with the target cell assembly j. The noise, $\mathbf{N}_j$, is the total synaptic excitation contributed by all other active neurons. The synaptic excitation that each synapse delivers is a random variable itself, call it $\mathcal{X}$, where as defined $E[\mathcal{X}] = s^{[y,k]}$. The total amount of synaptic excitation is merely the sum of each individual synaptic contribution, where $\mathbf{S_j} = \sum_{i=1:\mathcal{S}_j} \mathcal{X}_i$ and $\mathbf{N_j} = \sum_{i=1:\mathcal{N}_j} \mathcal{X}_i$. Now, $\mathcal{S}_j$ is a random variable representing the number of signal inputs and $\mathcal{N}_j$ is a random variable representing the number of noise inputs. Note that the random variables representing the number of synaptic inputs in each case are independent from the value of the synaptic excitation for each synapse, therefore $E[\mathbf{S_j}] = E[\mathcal{S}_j] s^{[y,k]}$ and $E[\mathbf{N_j}] = E[\mathcal{N}_j] s^{[y,k]}$. Additionally, because the two populations of neurons have zero overlap and are thus independent, we can write $E\left[\tilde{n}_{b,j}^{[y]} | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\mathbf{S}_j + \mathbf{N}_j | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\mathcal{S}_j | \tilde{\underline{\alpha}}^{[y]}\right] s^{[y,k]} + E\left[\mathcal{N}_j | \tilde{\underline{\alpha}}^{[y]}\right] s^{[y,k]}$. For a cell assembly j the number of signal synaptic inputs is entirely determined by the symbol excitation, $\tilde{\alpha}_j^{[y]}$ and we can write $E\left[\mathcal{S}_j | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\mathcal{S}_j | \tilde{\alpha}_j^{[y]}\right]$. This fact is easily proved since the value $\tilde{\alpha}_j^{[y]}$ is the exclusive sum of the activations of all associated cell assemblies, which contain all active signal neurons. The number of noise neurons that are connected to a neuron in cell assembly j, $\mathcal{N}_j$, is given by the number of neurons that are NOT signal neurons AND are connected, therefore we can write $\mathcal{N}_j = (\mathcal{A}_\mathcal{K} - \mathcal{S}_j) p_w$, where $\mathcal{A}_\mathcal{K}$ is a random variable representing the total number of active neurons, and p_w is the probability that a source neuron is randomly connected to a target neuron. Noting that the signal and noise neurons are independent, we can now write $E\left[\mathcal{N}_j | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\left(\mathcal{A}_\mathcal{K} - \mathcal{S}_j\right) p_w | \tilde{\underline{\alpha}}^{[y]}\right] = E\left[\mathcal{A}_\mathcal{K} | \tilde{\underline{\alpha}}^{[y]}\right] p_w - E\left[\mathcal{S}_j | \tilde{\underline{\alpha}}^{[y]}\right] p_w = E\left[\mathcal{A}_\mathcal{K} | \tilde{\underline{\alpha}}^{[y]}\right] p_w - E\left[\mathcal{S}_j | \tilde{\alpha}_j^{[y]}\right] p_w$. Thus our task is reduced to finding $S_j = E\left[\mathcal{S}_j | \tilde{\alpha}_j^{[y]}\right]$, given by lemma 1, finding $A_\mathcal{K} = E\left[\mathcal{A}_\mathcal{K} | \tilde{\underline{\alpha}}^{[y]}\right]$, given by lemma 2, and finding p_w given by lemma 3, thereby completing the proof.

4.5.5 Robustness analysis

In our robustness analysis, we address the variation of parameters that can be biologically controlled and asses the consequences on information processing performance. We re-ran simulations for the cell assembly initializations with the words "knowledge of historical" corresponding to Figure-4.3(b). Robustness results are summarized in Figures 4.9 and 4.10. We vary the possibly genetically controlled parameters p, $N_k = N_y = N$, and M, in addition to the parameter A_y, which can be viewed as a learned control signal. For the parameter search we use the formal neuron model and only run one time step, therefore we temporarily drop references to time with the assumption that all target neural fields are analyzed one time step after the source neural field initialization. Simulation results do not fundamentally change if using a spiking neuron model which captures the full range of possible neuronal dynamics.

We use an information processing performance measure to asses the relationship between the symbol excitation, $\widetilde{\alpha}_j^{[y]}$, and cell assembly activity, $\alpha_j^{[y]}$, in the target neural field y. Comparing simulation results between symbol excitation and activity gives the ability to determine some properties of the function, $\mathcal{F}(\cdot)$, in $\underline{\alpha}(t)^{[y]} = \mathcal{F}\big(\underline{\widetilde{\alpha}}(t)^{[y]}\big)$. Recall that the symbol excitation is a linear sum of fractionally active associations and the cell assembly activity vector is a representation of the neural field information state. Essentially, as defined above the performance measure is used to address the question: When will a cell assembly with a larger sum of fractionally active associations always have a larger cell assembly information state than cell assemblies with a smaller sum of associations?

We define the performance measure, $D(t, a, b)$, for integer values of a and b as

$$\begin{aligned} D(t,a,b) = \quad & min\big(\alpha_j(t)^{[y]};\ \text{s.t. } a \leq \widetilde{\alpha}_j(t)^{[y]} < (a+1)\big) \\ - \quad & max\big(\alpha_j(t)^{[y]};\ \text{s.t. } b \leq \widetilde{\alpha}_j(t)^{[y]} < (b+1)\big). \end{aligned} \tag{4.15}$$

If $a > b$ and the performance measure is positive then those cell assemblies with cell assembly excitation between a and $a+1$ will always have a larger cell assembly activity than cell assemblies with excitation between b and $b+1$, for the given simulation. When only a few cell assemblies are fully activated in each source neural field, the excitation vector has values that are exclusively between integer values.

Each simulation involves activating at some time $t_{\mathcal{X}}$ all the neurons from a chosen set of cell assemblies in a neural field k. Each neural field may have a different set of cell assemblies that are activated. We will use the notation, $\mathcal{X}^{[k]}$, to describe the set of activated cell assemblies in neural field k. For example, $\mathcal{X}^{[2]} = \{3, 431, 1842\}$, states that the all neurons in each cell assembly, $x_3^{[2]}$, $x_{431}^{[2]}$, $x_{1842}^{[2]}$, will be activated in neural field 2. Generally, the initialization of the neurons in a neural field can be stated as providing the appropriate excitation at some time, $t_{\mathcal{X}}$,

to each neuron such that

$$\underline{\eta}^{[k]}(t_{\mathcal{X}}) = \bigcup_{i \in \mathcal{X}^{[k]}} \underline{x}_i^{[k]}, \tag{4.16}$$

where $\bigcup$ represents the element by element "OR" operation, which states that if any element $x_{ij}^{[k]} = 1$ then the corresponding element $\eta_j^{[k]}$ should equal 1.

The initialization can be described as a superposition of active items of information in a source neural field.

To simplify the simulation, a single cell assembly is activated independently in each source neural field. The three cell assemblies initialized corresponded to "knowledge of historical" as in the main text. With the neurons from only one cell assembly activated, the activity in other cell assemblies in a single neural field is small due to overlap, where for example, $\alpha_j^{[1]} << 1 \; \forall j \neq$ "*knowledge*". This being the case for each source neural field, the symbol excitation

$$\widetilde{\underline{\alpha}}^{[y]} = \mathbf{L}^{[y1]}\underline{\alpha}^{[1]} + \mathbf{L}^{[y2]}\underline{\alpha}^{[2]} + \mathbf{L}^{[y3]}\underline{\alpha}^{[3]}, \tag{4.17}$$

will only have values that are very close to 0, 1, 2, or 3.

In each simulation target cell assembly activities are calculated from (4.3) and displayed, however we group cell assemblies by their excitations from (4.2). The cell assembly excitation vector, $\widetilde{\underline{\alpha}}^{[y]}$, has four groups. The first group, $\widetilde{\alpha}_j^{[y]} \approx 3$, has 3 cell assemblies $j = \{$"facts","styles","subjects"$\}$. The second group, $\widetilde{\alpha}_j^{[y]} \approx 2$, has 204 cell assemblies. The third group, $\widetilde{\alpha}_j^{[y]} \approx 1$, has 9639 cell assemblies. The rest of the cell assemblies have $\widetilde{\alpha}_j^{[y]} \approx 0$.

Cell assembly activities(maximum, mean, minimum) are displayed for each group in Figures 4.9 and 4.10 according to color. The group $\widetilde{\alpha}_j^{[y]} \approx 3$ is displayed in red, $\widetilde{\alpha}_j^{[y]} \approx 2$ is displayed in green, and $\widetilde{\alpha}_j^{[y]} \approx 1$ is displayed in blue. Each x-axis position on every graph in Figures 4.9 and 4.10 represents an independent simulation.

The performance measure $D(1,3,2)$ is plotted as a dashed black line in each figure. $D(1,3,2)$ simply represents the difference between the least highly activated cell assembly with 3 associations and the most highly activated cell assembly with 2 associations. $D(1,3,2)$ represents the degree to which more highly excited cell assemblies can be differentiated from less highly excited cell assemblies. For parameters resulting in $D(1,3,2) > 0$, cell assemblies with 3 associations are always more highly active than cell assemblies with 2 associations. From Figures 4.9 and 4.10 it is immediately clear that for a wide range of parameters, $D(1,3,2) > 0$. We mention that the choice of which cell assembly to initially activate effects the total number and distribution of active associations, which will in turn may effect the results, however the general qualitative results hold. The results are meant to illustrate the qualitative nature of the performance measure through parameter variations.

In Figure 4.9(a-c) we vary the number of activated neurons, A_y, in the target neural field from 0 to 100,000. This simulates a control signal that determines the number of neurons that fire in a neural field. In (a-c) $D(1,3,2)$ is a non-linear function of the number of activated neurons, rising quickly and then saturating or decreasing. In (c), for $p = 0.05$, $D(1,3,2)$ achieves a clear maximum at $A_y = 10,000$. Overall cell assembly activity also rises non-linearly as a function of A_y. Together these results demonstrate that there exists a capability to optimize the $D(1,3,2)$ via control of the overall neural field activity level.

The parameter, p, corresponds to the probability of connectivity between two neurons in connected neural fields. As a lower bound we use $p = 0.01$, which correlates to the lower end of random probability of connection calculated in the mouse cerebral cortex(Braitenberg and Shuz, 1998). We test increasing probabilities of connectivity that may be a result of learning and/or temporary increases in developmental connectivity(Rakic et al., 1986).

Figure 4.9(d-f) show the effect of varying the probability of connectivity, p, from 0 to 1. In (d-f) the result on $D(1,3,2)$ of increasing the probability of connectivity is non-linear, rising quickly and then saturating or decreasing. In (a-b), when the number of neurons per cell assembly, M, is 50 and 250 respectively, $D(1,3,2)$ achieves a maximum value at approximately p=0.2. Because of the non-linearity, we can deduce that small changes in the connectivity will have a large effect on $D(1,3,2)$.

In Figure 4.10(a-c) we see that even for the lowest probability of connectivity, p=0.01, $D(1,3,2)$ is positive for $M \geq 100$. This demonstrates that even with extremely sparse connectivity larger cell assembly excitations result in larger cell assembly activity.

In general, from Figure 4.9 and Figure 4.10(a-c), increasing the probability of connectivity initially improves $D(1,3,2)$. However, from Figure 4.9(a,b) it is clear that for a given cell assembly size, M, and threshold, $\Phi^{[y]}$, the capability exists to optimize $D(1,3,2)$ as a function of the probability of connectivity p. The result implies that synaptic structural plasticity may be used to significantly strengthen associations.

In Figure 4.10(d-f), increasing the number of neurons, N, in a neural field will increase $D(1,3,2)$ up to some saturation point. Increasing N decreases the overlap of neurons in cell assemblies, therefore decreasing crosstalk. If the probability, p, of connectivity between two neural fields were a function of N_y, which is plausible, this result may not hold. Likely, p would decrease with increasing target neuron number N_y, therefore we may expect that some optimal value for $D(1,3,2)$ would be achieved with increases in N_y, providing an upper limit to the benefit of increasing N.

The average number of neurons in a cell assembly is M. We varied the value M from 50 to 800 neurons per cell assembly. Figure 4.10 demonstrates that there exists values of M

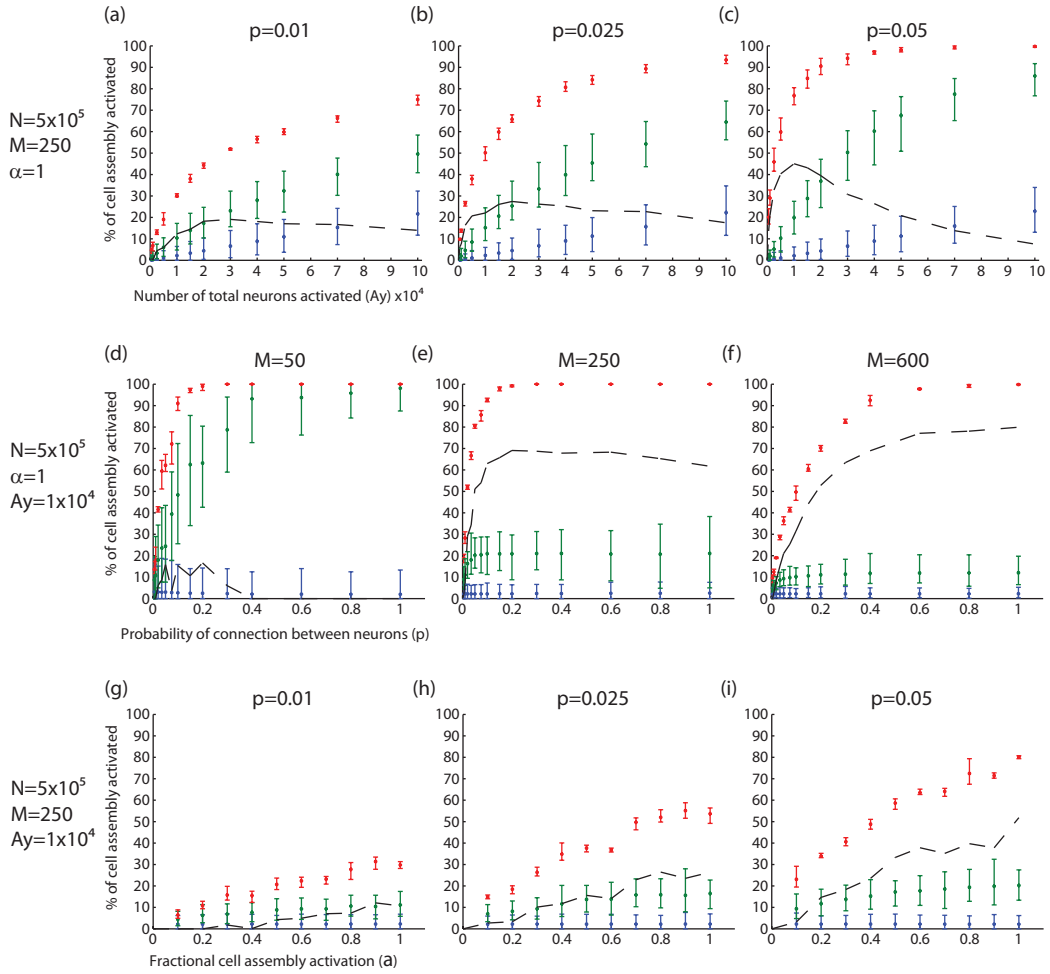

Figure 4.9: Performance over A_y, $\alpha^{[k]}$, and p. The y-axis of (a-i) denotes the percentage of neurons activated in target cell assemblies, $100 * \alpha_i^{[y]}$, grouped by excitation values(number of associations) $\widetilde{\alpha}_i^{[y]} \approx 3$(red), $\widetilde{\alpha}_i^{[y]} \approx 2$(green), $\widetilde{\alpha}_i^{[y]} \approx 1$(blue). Each vertical colored bar displays max(horizontal), mean(star), and min(horizontal) cell assembly activity values grouped by cell assembly excitation values. The performance measure(PM) is shown as a black dashed line representing the difference between the smallest cell assembly activation receiving 3 associations and the largest cell assembly activation receiving 2 associations. Parameters include the total number of neurons in a neural field($N = N_k = N_y$), the total number of neurons activated(A_y), the average number of neurons per cell assembly(M), the probability of connectivity between neurons(p), and the fraction of neurons initially activated in source cell assemblies($\alpha = \alpha^{[k]}$).

which will maximize $D(0, 3, 2)$. The optimal M values are found between 150-350 for a wide range of probabilities $p = 0.01$ to $p = 0.05$ and wide range of neuron number $N = 250,000$ to $N = 1,000,000$.

This predicts that information representation in the cerebral cortex is likely stored in cell assemblies composed of a few hundred neurons.

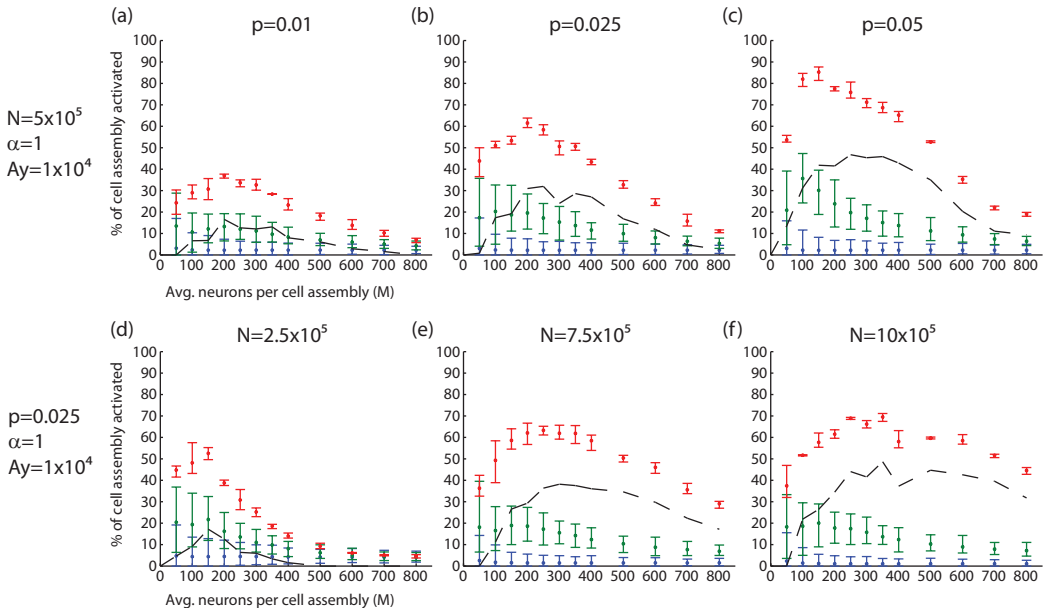

Figure 4.10: Performance over M. Description of parameters given in Figure 4.9. (a) Across a wide range of neuron connectivity probabilities, p=0.01 on left to p=0.05 on right, the performance measure(PM) (black dashed line) reaches a maximum when the average number of neurons per cell assembly, M, is either 200 to 250. (b) The M producing maximum PM tends to increase as N increases, left plot to right plot. Larger N values allow more M values producing higher PM (black dashed line) due to the decrease in cell assembly overlap and interference.

In Figure 4.9(g-i) the initial number of neurons activated in each source cell assembly is varied. In this case (4.16) is modified so that on the left side only a fraction of the neurons specified are activated. The plots (g-i) show that target cell assembly activity increases linearly with source cell assembly activity. Importantly the slopes for the groups $\widetilde{\alpha}_i^{[y]} \approx 3$ and $\widetilde{\alpha}_i^{[y]} \approx 2$ are different. The slope of activation is larger for cell assemblies with larger excitations, showing that larger excitations result in their activity growing faster.

Chapter 4, in full, has been submitted for publication of the material as it may appear in Network, Solari, S. and Hecht-Nielsen, R. 2009. The dissertation author was the primary investigator and author of this material.

5

Controllable contextually dependent thalamocortical attractor network: the basis of confabulations

5.1 Introduction

The purpose of our thalamocortical attractor network model, and the present chapter, is to demonstrate that the necessary and sufficient requirements to implement the single cognitive information processing proposed by confabulation theory, confabulation, can be achieved robustly with a biologically realistic thalamocortical model. Our model and results additionally present novel details on the hypothesized anatomical implementation of cognitive information processing in the mammalian brain.

Proposition 2. *A biologically plausible attractor network model consistent with confabulation theory must achieve the following necessary and sufficient information processing requirements:*

1. *Scalable capacity (as a function of number of neurons) to store 10,000's to 100,000's of symbols.*

2. *Controllable persistent activation of multiple symbols, or in other words the controlled superposition of the number of active symbols in working memory.*

3. *The controlled trimming of the active set of symbols to a subset of the active symbols, based on contextual input and/or control signals.*

4. *The prevention of symbols that are not part of the set of active symbols to re-enter the set of active symbols even if they receive contextual input.*

5. *The controlled convergence a single highly active symbol, which can be held alone in working memory.*

In researching a biologically plausible attractor network model capable of implementing these requirements we discovered several more general principles underlying a class of associative memory attractor networks.

We discovered that the neural fields forming a recurrent attractor network need not be symmetric in size (as is the case with cortical and thalamic neural fields), and demonstrate the basic principles of associative input and associative interference underlying attractor network function.

We also discovered that introducing the concept of balanced inhibition into the attractor network essentially normalizes any recurrent excitation to guarantee stability and creates an architecture where open loop control could be easily learned and closed loop control requires only gross measurements of neuronal activity.

We demonstrate that all requirements above can be accomplished with a simple control law in a formal neuron model. We follow the formal neuron simulation with a more biologically realistic simulation demonstrating similar information processing capabilities with phenomenological spiking neurons and synapses with dynamic conductances.

5.2 Relationship to past work

A great deal of research has been invested into the study of associative memories (Willshaw et al., 1969; Abeles, 1991; Amari, 1989; Kohonen, 1989; Palm, 1980; Buckingham and Willshaw, 1993). The psychological phenomena of working memory has now also been extensively studied in psychology (Baddeley, 1981, 2003; Monsell, 1984) and at the more detailed level of electrophysiology (Goldman-Rakic, 1996). Models of working memory have in the past used recurrent excitation, similar to associative memory connectivity, and neuromodulation as a basis for rough control underlying working memory (Brunel and Wang, 2001). However, these and other past models have analyzed the role of selected parameters regulating the dynamics of individual neurons, and thus the system, such as AMPA or NMDA synaptic conductances (Compte et al., 2000; Durstewitz et al., 2000). In contrast, we approach the problem of working

memory and thalamocortical attractor networks with a concrete hypothesis on the underlying mathematical operations that must be performed at the system level. While the cerebral cortex is clearly a dynamic network, we argue here, that basic working memory and cognitive information processing as a whole, likely relies only on a few basic principles that can be largely explained in terms of recurrent associative memory connectivity, balanced inhibition, and control of overall neural field activity.

Our results fundamentally contradict the modern viewpoint of working memory that robust delay activity on the order of 15-40Hz must be a function of NMDA receptors, because AMPA receptors act on too short a time-scale(Durstewitz et al., 2000). Almost all past work on working memory seems to have focused exclusively on recurrent cortical connectivity. Here we demonstrate that robust delay activity in working memory on the order of 40Hz can be a result of explicit delays in the cortico-thalamo-cortical loop. We propose that these conduction delays instead form the basis of the cortically measured gamma (∼40Hz) frequencies.

5.3 Thalamocortical model

In order to clarify and simplify terminology between confabulation theory concepts and more general dynamical concepts, we will refer to a single information processing attractor state as a symbol. Biologically, as described in chapter 4, a single symbol can be represented by a single neuronal cell assembly and therefore a cell assembly can be described as a symbol.

5.3.1 Thalamocortical anatomy

As we have mentioned, anatomical structure inevitably constrains anatomical function. It makes sense therefore to begin with a brief description of the anatomical basis for our model. Figure-5.1 demonstrates the basic anatomical model and certain model simplifications. The accuracy of the initial model is supported by the detailed review in chapter 3. Even so, the model is far from a biologically perfect reconstruction of cortical anatomy, and significant room is left for additions and improvement. Our goal in discussing the anatomy is to highlight that the basic principles underlying the thalamocortical attractor, map well onto the known anatomical circuitry.

Figure-5.1(a) highlights the thalamocortical circuitry which we feel is most significant in the thalamocortical attractor dynamics.

In terms of the number of neurons in each neural field, we have estimated that for a thalamocortical module, the number of neurons in the specific thalamus are approximately 4-5 times fewer than in the C_{6T} or C_{3B} neural fields respectively. The specific thalamus is therefore

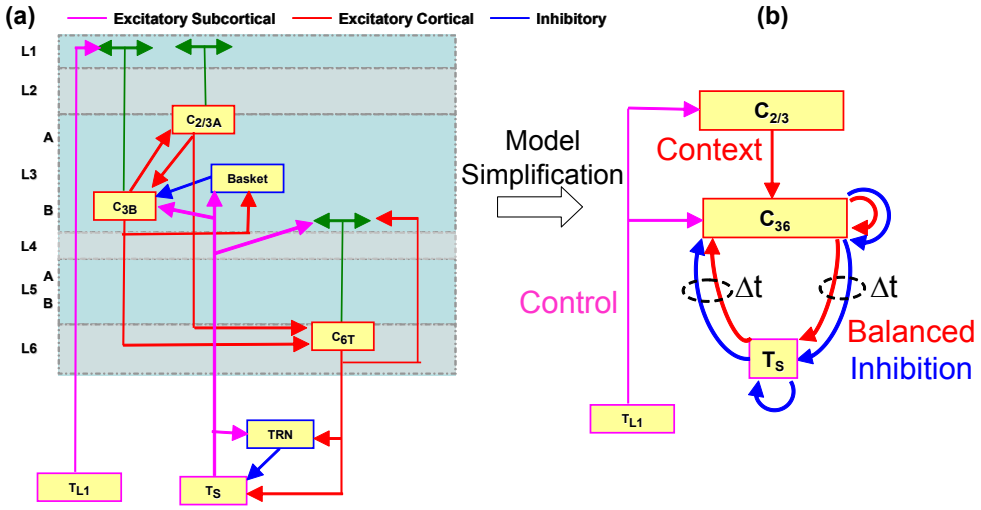

Figure 5.1: Attractor network anatomical model.

shown as a smaller volume compared to other neural fields and the discrepancy in neural field size is modeled directly.

In terms of the fundamental frequencies underlying the thalamocortical attractor, the corticothalamic (layer 6 to specific thalamus) and the thalamocortical projections are of special importance. In the human cerebral cortex the distance between the cortex and thalamus can range between 20-50mm (measurements from (Nolte and Angevine, 2000)). Action potentials traveling at 3-4mm/ms would take approximately 10ms to travel from the cortex to the thalamus and the same time to travel from thalamus to cortex. The result is a 20ms round trip time for reentrant cortico-thalamo-cortical excitation. Given that neurons do not fire instantaneously and intra-cortical delays exist, if we add an additional 5ms for the reentrant action potentials, we are left with about a 40Hz oscillation. The 40Hz oscillation is significant because it represents the base level gamma frequency extra-cranially recorded during cognitive tasks (Buzsaki, 2006). Simply by analyzing the anatomical connectivity the hypothesis arises that the gamma frequency oscillation is a result of thalamocortical reentrance. Even so, most research has focused on the fact that coupled interneuron dynamics can also lead to a gamma 40Hz type of oscillation(Buzsaki, 2006).

Anatomical and electrophysiological measurements give credible evidence to the existence of feedforward inhibition both in the thalamocortical and cortico-thalamic pathway. Layer 6 pyramidal cells projecting to the specific thalamus are known to have many collateral projections in the thalamic reticular nucleus (TRN). The TRN being a significant source of inhibition onto thalamic cells suggests that layer 6 produces significant feedforward inhibition. The spe-

cific thalamus neurons in turn send projections to the TRN on the way to synapsing in lower layer 3. The thalamic-TRN-thalamic projections are a source of feedback inhibition. Finally, the thalamocortical projections in fact synapse directly onto inhibitory cortical interneurons, which supports thalamocortical feedforward inhibition. In addition to the anatomical analysis, electrophysiology measurements suggest that feedforward balanced inhibition is a common feature in nervous systems(Shadlen and Newsome, 1998; van Vreeswijk and Sompolinsky, 1996; Berg et al., 2007). Balanced inhibition in general refers to expected equal amounts (in terms of mean value) of synaptic inhibition and excitation onto a neuron. Balanced inhibition is the critical component that normalizes excitatory interference in our model and enables robust multiplicative control in the dynamic simulation.

In the mouse somatosensory cortex electrophysiological studies have shown that the cortex is driven by synchronous thalamic excitation(Bruno and Sakmann, 2006).

Throughout the cerebral cortex, nearly all pyramidal neurons send their apical dendritic tuft to layer 1. The source of synaptic terminals in layer 1 is widespread, including many varieties of neurotransmitter other than glutamate coming from sub-cortical nuclei, mostly hierarchically "feedback" cortico-cortical projections, and VM/VAmc thalamic plus some intralaminar thalamic projections. But, the specific thalamus does not project to layer 1, therefore the cortico-thalamo-cortico loop described above could reasonably be used to perform specific information processing operations in a reentrant loop, while layer 1 signals could be used to regulate and control that information processing. Much as in a muscle, the interconnection of muscle fibers provides the working substrate for contraction, but external motor neuron input controls the contraction. The layer 1 excitatory input, especially the VM/VAmc thalamic input, is also highlighted because of recent studies that suggest that excitation in the apical dendritic tufts of pyramidal cells could have a multiplicative effect on downstream synaptic excitation. This comprises a central feature of our model.

In relation to our discussion of layer 1 projections, an especially important fact exists with the cortico-thalamo-cortical loop. As discussed in chapter 3, the specific thalamus projects to lower layer 3 in non-primary cortices. Notably, C_{6T} neurons send their apical dendritic tufts to lower layer 3. As mentioned, all other pyramidal neurons with significant apical dendritic tufts have their apical dendrites in layer 1 of the cortex. The synaptic excitation, and its effect, of the specific thalamic projections on layer 6 pyramidal cells might then be hypothesized to be analogous in its action as the layer 1 VM/VAmc thalamic excitation (which does not come from the specific thalamus) onto other pyramidal cells.

Finally, cortico-cortical connections exist in vast amounts. The majority of ipsilateral and contralateral projections arise and terminate in layer 2 and 3, and are virtually all excitatory

synaptic connections. Therefore, if we assume that thalamocortical modules are each processing different information, then the contextual input (thalamocortical attractor states) are communicated to other thalamocortical modules via excitatory synaptic excitation in layer 2 and 3. The multi-associative memory chapter 4, describes the underlying model of these connections in detail.

In summary, our model hypothesizes that attractor states are imbedded in reciprocal thalamocortical connections and within intra-cortical connections. The attractor states are cell-assemblies and map from one neural field to another with associative memory connectivity. We propose that the layer 1 input provides excitatory input that serves as multiplicative input to the internal thalamocortical attractor excitations and the external contextual input. Without the multiplicative input, we hypothesize that the thalamocortical attractor does not have sufficient reentrant excitation to maintain any attractor dynamics. Therefore, the layer 1 input controls the overall activity of the thalamocortical attractor just as a motor neurons control the overall contraction of a muscle. Analogously, we hypothesize that the thalamic projections that connect to the apical dendritic tufts of layer 6 neurons, serves as multiplicative input to layer 6 neurons, thereby activating the layer 6 output. In this way, attractor states can self-excite independent of the layer 1 input. We hypothesize this may be necessary for working memory, although our experimental results are inconclusive on this point.

5.3.2 Thalamocortical attractor model simplification

Figure-5.1(a) shows a biologically accurate depiction of cortical neuroanatomical projections. We have implemented simulations that attempt to include biologically realistic connections, however, several problems emerge that make definitive conclusions difficult without comparison to biological experiments.

The first problem is that no biological experiment exists that is capable of generating the necessary detail for comparison to the multitude of neuronal types and connections. At best, a few neurons can be directly patched and recorded in vivo at any one time, or multi-electrode arrays might be able to record from many tens to a few hundred of neurons simultaneously in a local area. In our experiments a single symbol is composed of 250 neurons which are distributed over a large area. Therefore, we can not yet align our results with biological experiments.

Another problem is the number of free parameters that can be tuned. Some of the parameters that can be tuned include, number of neurons in a neural field, number of neurons in a symbol, probability of connectivity, the dynamics of each neuron, strength of each synapse, and dynamics of each synapse. With ten's of thousands of neurons and millions of synapses the space of parameters is large. Our hypothesis predicts that only a few functional operations on

symbols should be carried out by the anatomy. We therefore chose to simplify the anatomical model in places that we felt did not effect the overall ability of the model to demonstrate the principles underlying the function of the network.

The first major simplification is the removal of inhibitory interneurons. Instead a single source neural field provides both inhibitory and excitatory stimulation onto a target neural field. The purpose of the simplification was to clearly illustrate the functional role of balanced inhibition. Creating balanced inhibition with an intermediary population of inhibitory interneurons is possible, but the parameters are non-unique and the model requires extensive "tweaking". Presenting a model with significant tweaking is hard to justify and it is hard to gain any insight into the principles of its function. We therefore chose to implement a principled approach to the inclusion of balanced inhibition, which provides clear justification and a clear functional outcome that can be tested biologically.

The other major simplification in Figure-5.1(b) is the combination of the C_{3B} and C_{6T} neural fields into a single C_{36} neural field. In this model we are saying that the C_{3B} and C_{6T} neural fields can be abstracted (as a first approximation) into a single functional neural field. The hypothesis is that the T_S input to lower layer 3 activates the C_{6T} neurons via their apical dendrites. The same projection also provides direct input to the neurons in the C_{3B} neural field, which then project back directly or through an intermediary neural field onto C_{6T} neurons. We hypothesize that the T_S stimulation of C_{3B} and co-activation of C_{6T} can reasonably be approximated by a single neural field. As we have proposed earlier, biologically, C_{3B} and C_{6T} neural fields need to be separate, because C_{3B} neurons must trigger action commands in C_{5B} independent of the C_{6T} neural field information being sent down to the thalamus T_S.

5.4 Results

Both simulations contain 25,000 neurons in the C_2 and C_{36} neural fields, and 5,000 neurons in the T_S neural field. The fact that the thalamus has 5 times less neurons is consistent with known anatomy and is implemented to illustrate the fact that the attractor network has no loss of function. There are 400 symbols in each neural field. In C_2 and C_{36} symbols are composed of approximately 250 neurons. In T_S symbols are composed of approximately 50 neurons. The initial probability of unstrengthened connectivity is 0.1 for all connections except T_S to C_{36}, which is 0.5. Notice the five fold increase in connectivity probability. The reduction in symbol size and increase in probability of connectivity is one of the principles that allows the attractor network to function with a 5 fold difference in cortical and thalamic neuron number. The strength of synapses can also be used to compensate for the difference in size, but the key principle is that

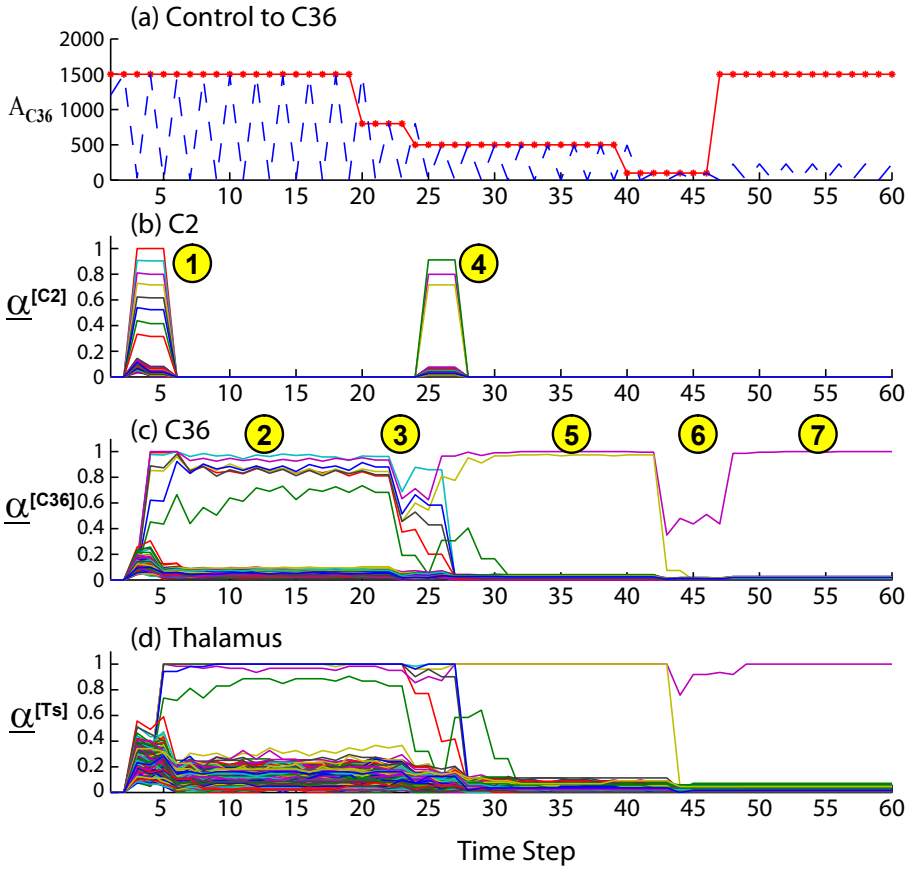

Figure 5.2: Formal neuron attractor simulation. (a) The A_{36} control value activates the A_{36} most highly excited neurons in the C_{36} field that are greater than firing threshold. (b) Symbol activation in neural field C_2. (c) Symbol activation in neural field C_{36}. (d) Symbol activation in neural field T_S. See section 5.4.1 for details.

the strength of the associative signal stays roughly constant (see 4.2.4 and 5.7.2).

Neurons in each neural field were selected at random to be in a symbol, such that each neuron was a part of exactly $r = 4$ symbols. Note that $r = ML/N$, where N is the number of neurons in a neural field, M is the average number of neurons in a symbol, and L is the number of symbols. See also 5.7.1 for a detailed discussion. Maintaining a constant r value with low (zero) variance for each neural field ensures low variance on the distribution of synapses per neuron.

5.4.1 Formal neuron model results

The formal neuron simulation results are shown in Figure-5.2. The simulation utilizes most of the simplified model connectivity shown in Figure-5.1(b). See 5.6.2 for details.

Yellow circles depict seven time points that will be discussed as the results. Figure-5.2(a)

shows the control applied to the C_{36} neural field in red, and the actual number of active neurons in blue. The control was designed to achieve the basic attractor results for discussion. The attractor network is robust to even large ($> 10\%$) perturbations in the control signal, therefore non-unique control signals exist to achieve a particular information processing result. In biological systems the control signal would be learned, just as control input to muscles is learned to perform certain movements.

The control strategy for the formal neuron simulation is simply to activate the A_{C36} most highly excited neurons (above a certain threshold) in the C_{36} neural field at each time step. See also methods section 5.6. The purpose of using our particular choice of control strategy is two-fold: 1) to use a control strategy that could be reasonably implemented in the cerebral cortex and 2) to demonstrate the effectiveness of such simple imprecise control in information processing. The idea behind the control is that T_{L1} thalamic input can regulate the overall level of activity in a thalamocortical module, equivalent to "allowing" a fixed number of neurons to fire. Notice that the T_S neurons do not receive any control signal, therefore any neuron in the T_S neural field with excitation greater than threshold will be activated. The limiting of neuronal activation in the cortex will tend to limit the excitation in the T_S neural field in the next time step.

Figure-5.2(b-d) show the neural field information state (symbol activity) for the C_2, C_{36}, and T_S neural fields respectively. The neural field information state is a vector with elements representing the fraction of the number of neurons active in each symbol (see section 4.4.5). Each line in the plot represents one symbol's activity over time. The symbol activity is calculated within a moving 3 time step window to smooth the plots.

Time point 1 (yellow circle 1) in Figure-5.2(b) shows the initialization of the simulation. The initialization is meant to start off where the results in chapter 4 left off. Eight symbols are initialized with 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3 symbol activities respectively in neural field C_2. At the same time 10% of all other neurons in all neural fields are randomly activated as noise. The control signal is set to 1500, which means that approximately 6 symbols could be fully activated (note $6 * M = 6 * 250 = 1500$). The control immediately enforces that all eight initialized symbols can not all become active in C_{36}, thereby creating competition for activation.

Time point 2 shows that 7 symbols are both fully and partially active within a thalamocortical loop. Notice that all 7 symbols can not be fully active because of the control and so there is some change in the activity of each symbol as their neurons compete for activation. At this point it is clear that multiple partially active symbols can be maintained without interference indefinitely with control. The control signal can also clearly be used to determine how many symbols should be maintained in working memory.

At time point 3, the control is reduced to 750. With only 750 neurons active, only 3 symbols could be fully active, so the activation of all symbols begins to decrease as their neurons compete for activation. Notice the least active symbol in green is the first to loose activation and disappears into the noise floor.

At time point 4, the control is further reduced to 500, but at the same time additional contextual input is activated in neural field C_2. The contextual input activates 3 symbols in C_2, with 0.9, 0.8, and 0.7 activation. All three of these symbols were also initialized at time point 1. Importantly, the symbol with 0.9 activation in C_2 is the "green" symbol that was eliminated from the competition at time point 3, therefore this symbol is not receiving significant reentrant T_S input. However, the other two symbols are receiving T_S input because they are both partially active. The combination of control signal, T_S input, and C_2 contextual input ensures that only the two symbols that are both in working memory AND receive contextual input stay in working memory. Notice even though the green symbol receives more contextual excitation, the symbol has already been trimmed from working memory and is therefore not allowed back into working memory.

At time point 5, only two symbols are in working memory. The two symbols correspond to those that were active in working memory and received contextual input.

At time point 6, the control signal is dropped to $A_{C36} = 100$, forcing the neurons in the last two symbols to compete for excitation. With only 100 neurons active, less than half of one symbol can be active.

At time point 7, the control signal is raised to ensure that the single symbol left in working memory is fully activated. The control is raised back up to 1500 to demonstrate that other symbols can not return into the competition. The explanation, as shown by the blue dotted line in (a) is that all other neurons do not cross threshold, therefore only the neurons in the winning symbol are capable of being excited at this point.

5.4.2 Phenomenological neuron model results

The phenomenological neuron simulation results are shown in Figure-5.3. Figure 5.4 displays the membrane potential and synaptic current for a single neuron randomly selected from the "winning" symbol in the simulation. The phenomenological simulation utilizes all the connectivity shown in the simplified model in Figure-5.1(b), see also methods 5.6. The phenomenological simulation implements a dynamical model of each neuron as described in 5.6.4. Additionally, synaptic conductances have first order dynamics related to two types of excitatory receptors (AMPA, NMDA) and two types of inhibitory receptors ($GABA_A$, $GABA_B$). In general AMPA and $GABA_A$ act on a time-scale of a few milliseconds and provide strong post synaptic

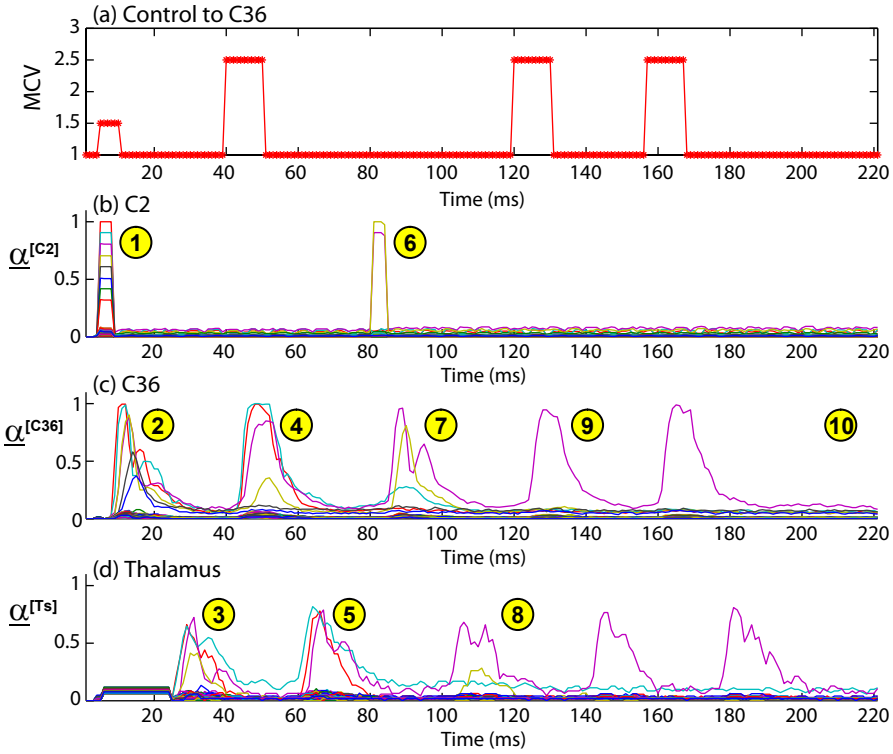

Figure 5.3: Attractor network results. See Figure-5.1 for model configuration. (a) The multiplicative control value (MCV) multiplies all synaptic current into all neurons in the C_{36} field. (b) Symbol activation in neural field C_2. (c) Symbol activation in neural field C_{36}. (d) Symbol activation in neural field T_S. See section 5.4.2 for details.

currents, therefore they are the drivers in the attractor dynamics. NMDA and $GABA_B$ receptors act on a time scale of the order of a hundred milliseconds and provide 1/100 of the synaptic current of AMPA/$GABA_A$, therefore their role is in maintaining overall levels of excitability in neurons. Similar results in other simulations were obtained even when NMDA and $GABA_B$ receptors are removed. A running hypothesis in neuroscience is that NMDA receptors are critical for learning and the acquisition of AMPA receptors, but may not be the driving force in information processing(Malenka and Siegelbaum, 2001). Our results would agree with such hypotheses.

The control signal for the phenomenological model is different from the formal neuron model. Figure-5.3(a) shows the control signal input vs time. The control signal applied to the C_{36} field is a multiplicative control value (MCV), which simply multiplies the sum of all input current into all neurons in C_{36}. Again the neurons in T_S do not receive any control signal. A discussion of the control signal is presented later.

Ten time locations (yellow circles) are indicated in Figure-5.3 in order to facilitate discussion of the results. For comparison, the basic phenomenological simulation initialization

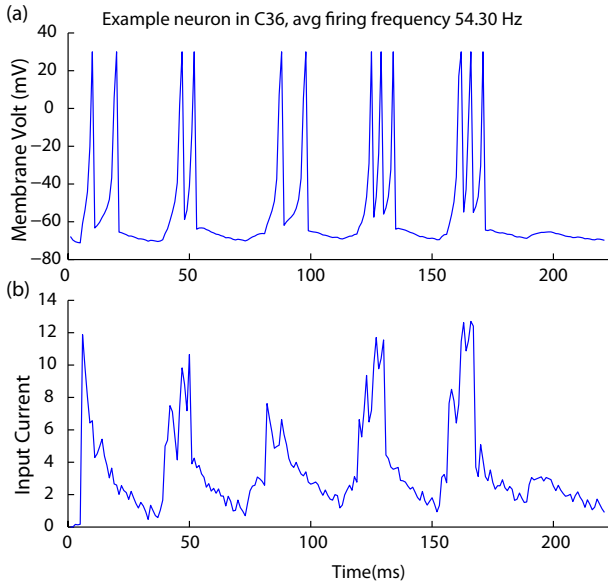

Figure 5.4: Phenomenological attractor network neuron firing. Neuron voltage and input current shown for a single neuron randomly selected from the winning symbol (purple) in neural field C_{36} in Figure-5.3. Input current is in μA. Because of the form of the dynamical neuron model, input current value also corresponds exactly to an excitatory post-synaptic potential (EPSP) equal to the input current value shown, but in mV. As shown the average firing frequency of the neuron is approximately 54Hz, corresponding to the gamma frequency.

and progression is similar to the formal neuron simulation. The symbol activation is again calculated in a 3ms sliding window to smooth the plots. Noise is introduced throughout the phenomenological simulation by adding normally distributed random synaptic current with 0.3 standard deviation to each neuron at every time step.

Time point 1 (yellow circle 1) in Figure-5.3(b) shows the same initialization in neural field C_2 as in the formal neuron simulation. Eight symbols are initialized with 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3 symbol activities respectively in neural field C_2. At the same time the control signal is raised to MCV=1.5. The excitation from active symbols in C_2 is transferred to C_{36} through the associative memory connectivity.

Time point 2 in Figure-5.3(c) shows the activation of symbols in C_{36} as a result of the C_2 initialization and control input. Two important phenomena occur. First the overall activation value of each symbol in C_{36} maintains the rough order of activation values in the initialized C_2 field. Since the control in this case does not explicitly limit the overall number of active neurons, all the symbols are partially active. Second less activated symbols are delayed in time relative to more highly activated symbols. The competitive temporal effect of symbol activation is highlighted by this fact, since symbols that are activated first will send out action potentials

and tend to inhibit neurons in other symbols that have not yet fired, but might have had they not been inhibited. The temporal effect is unique to the phenomenological simulation and is not seen in the formal neuron simulation.

Time point 3 in Figure-5.3(d) shows the resulting activation after action potentials have travelled from C_{36} to T_S. Note that the initial control input has now been reduced back to 1.0, thus no feedback amplification of synaptic input occurs in C_{36}. The conduction delay between C_{36} and T_S is 10ms, therefore at least 10ms of the delay in symbol activation in T_S is a result of conduction delays. Although slightly difficult to distinguish, only the 4 most highly active symbols have significant activation in T_S. This is where the T_S thalamic nuclei "reflect" back up to the C_{36} field any symbols with sufficient C_{36} activation. Critically, the 4 symbols that are active in T_S now represent the set of active symbols in working memory.

At time point 4, in Figure-5.3(b), the control signal has been raised temporarily to 2.5 at approximately the same time as the action potentials from T_S reach C_{36}. Note again that the conduction delay from T_S to C_{36} is also 10ms. The control signal is also reduced back to 1.0 a short time after increasing. The combination of the magnitude and delay of activation in T_S plus the control constrains the activation in C_{36} to 3 symbols, therefore a competitive operation has been implemented trimming the list of active symbols in working memory to a subset of the symbols in working memory based on competition without context.

At time point 5, Figure-5.3(d), three active symbols are clearly seen. With appropriate control values these three symbols could be held indefinitely in working memory.

At time point 6, Figure-5.3(b), two contextual symbols are active in C_2. One of the symbols corresponds to one of the symbols active in working memory, and the other corresponds to the yellow symbol that lost activation at time point 4. The control signal is left equal to 1.0.

At time point 7, Figure-5.3(c), only the two symbols receiving contextual input have any activation, however the symbol in working memory (purple) gets input from T_S as well as contextual input, therefore its activation over time is larger.

At time point 8, Figure-5.3(d), only the (purple) symbol from above has sufficient excitation to get "reflected" back from the T_S neural field.

At time point 9, Figure-5.3(c), the control is raised to 2.0 thereby strongly activating any symbol in working memory. Only one (purple) symbol receives enough T_S input to activate. As can be seen, as long as the control input is pulsed at approximately a 40Hz frequency corresponding to 20ms round trip thalamocortical conduction time plus neuronal synaptic integration delays, the symbol can be held indefinitely in working memory.

At time point 10, the control signal is set back to 1.0 (removed) and activity dies to zero in the thalamocortical module.

5.5 Discussion

We utilized the formal neuron model to demonstrate the enormous practicality of a thalamocortical attractor network with balanced inhibition and a simple control law, namely allowing the A_{C36} most highly excited neuron to fire and inhibiting the rest. The efficiency and accuracy with which computations can be performed is enormous with this control structure. In addition the control can be very imprecise and yet confabulations function perfectly. The large question is how this control might be implemented in real nervous systems.

The inclusion of balanced inhibition is critical to the function of the network. Inhibition enables stability of the network, but balancing the inhibition, such that each neuron has the expected same total sum of inhibition and excitation allows the neuron to remain quiescent in the presence of noise. Since random noise will on average activate the same number of inhibitory and expiatory synapses the neurons overall synaptic input stays close to zero. The threshold nature of neurons ensures that neurons will only robustly fire action potentials in the presence of signal. Multiplicative control can then be used to determine the strength of the signal necessary to robustly cause a neuron to fire.

In both simulations, the control signal used was designed to achieve certain basic attractor state evolutions for discussion. As such, control was applied in an open loop fashion. One question arises as to how the brain might utilize and implement closed loop control. One major structure in the brain with significant projections to the hypothesized control nuclei T_{L1} is the basal ganglia. If the control loop was closed, the measurements that the basal ganglia clearly has access to are action potentials sent from the C_{5S} neural field. At a minimum, the basal ganglia likely has information to the overall level of activity in a thalamocortical module. With that information, a reasonable control law can be designed to integrate the activity of multiple modules in a coordinated fashion. This is one area of research that must be pursued and that is likely to contribute significantly to our understanding of thought processes in the future.

Our second phenomenological simulation illustrates that a dynamical model of neurons and synapses utilizing identical connectivity can implement the same basic functions. The control in this case was changed to a multiplicative control on the neural field. Here, some measure of overall neuronal activity is required to appropriately select a control value, and we suggest that the basal ganglia likely performs this measurement and determines the appropriate control value. However, as we initially mentioned our model implementation is less than ideal. We hypothesize that the organization of intra-cortical connectivity (excitatory and inhibitory) in addition to external neurotransmitter input might in fact be used to implement an analogous control law as the formal neuron model. How exactly this might be directly implemented is a topic for future research, but neuroscience experiments might also be designed to test this hypothesis.

The simulations demonstrate that the necessary and sufficient conditions required by confabulation theory are quite easily implementable with a biologically plausible model of thalamocortical interaction. The simulation presents several additional hypothesis on the implementation of control over thought in the cerebral cortex, the utilization of balanced inhibition, and the role of gamma oscillations in information processing.

5.6 Methods

See Methods in 4.4 for a more detailed description of variables.

5.6.1 Neural field generation

Formal and phenomenological simulations used same neural field generation. L=400 for all neural fields. Cortex parameters $N_{C2} = N_{C36} = 25000, M_{C2} = M_{C36} = 250$. Specific thalamus, T_S, parameters $N_T = 5000, M_T = 50$. All neural fields created with constant r value, in each case $r = \frac{NM}{L} = 4$. The constant r value ensures that every neuron in a neural field is part of exactly r cell assemblies. See Appendix 5.7.1 for a discussion.

5.6.2 Connectivity

The connectivity for both formal and phenomenological simulations is based on the model simplification in Figure-5.1. The connectivity of the thalamocortical attractor network can be summarized as sparse associative memory connectivity for excitatory pathways and random balanced inhibition connectivity for inhibitory pathways.

The sparse associative memory connectivity between source neural field k and target neural field y can be described by the equation (see also Equation 4.1 for details):

$$\mathbf{W}^{[y,k]} = \mathbf{X}^{[y]}{\mathbf{X}^{[k]}}^T \bigcap \mathbf{U}(p^{[y,k]}), \tag{5.1}$$

and balanced inhibition connectivity is described by

$$\mathbf{W}^{[y,k]} = \mathbf{U}(p^{[y,k]}). \tag{5.2}$$

Initial unstrengthened connectivity probability from source field k to target neural field y is given by $p^{[y,k]}$ which is used to randomly create the matrix $\mathbf{U}$ such that each element in the matrix is set to 1 with probability $p^{[y,k]}$.

For associative memory connectivity (corresponding to red arrows in Figure-5.1), the following values apply to the model: $p^{[C_{36},T_S]} = 0.5$, $p^{[C_{36},C_2]} = 0.1$, $p^{[T_S,C_{36}]} = 0.1$, $p^{[C_{36},C_{36}]} = 0.1$, $p^{[T_S,T_S]} = 0.1$.

For balanced inhibitory connectivity (corresponding to blue arrows in Figure-5.1), the value $p^{[y,k]} = 0.5 * p_w^{[y,k]}$, where $p_w^{[y,k]}$ is computed from Theorem-1. The unstrengthened probability of connectivity used in Theorem-1 are those defined above for associative memory connectivity. The purpose of multiplying p_w by 0.5 is to make the average number of inhibitory synapses on each target neuron one-half the number of excitatory synapses. As a consequence, for balanced inhibition, the inhibitory synapses must be twice as strong as the excitatory synapses.

Formal neuron simulation connectivity

The formal neuron model simulation has no time delays between neural fields, therefore Δt=1 for all neural fields. We also removed three connectivity matrices from the formal neuron simulation. The excitatory and balanced inhibitory connections from C_{36} back onto itself were removed, as was the inhibitory connection from T_S back onto itself. The fact that the formal neuron model does not have any temporal dynamics and there are no significant conduction delays means that the feedback necessary to ensure some base level of excitation in C_{36} and inhibitory feedback to improve stability in the T_S case were not needed.

Phenomenological neuron simulation connectivity

The connectivity for the phenomenological neuron simulation contains all arrows in Figure-5.1. Conduction time delays were used. All time delays were set to 1ms, however the conduction delays $\Delta t = 10ms$ from C_{36} to T_S and from T_S to C_{36}.

5.6.3 Formal simulation

All excitatory synaptic excitation were random variables with mean 1 and standard deviation of 0.1. C_2 neurons were either set to 1 or 0 at a given time step. C_{36} neurons had to have excitation greater than 15 to be candidate active neurons, then the top A_{C36} (from the control value) most highly excited candidate neurons were selected and activated, set to 1 for one time step. The T_S neural field had a fixed firing threshold for all neurons. If any T_S neuron excitation was greater than 5, the neuron was activated, set to 1 for one time step. All neuronal activation was computed for each neural field at each time step and the simulation was run one time step at a time.

5.6.4 Phenomenological simulation

Phenomenological neuron

We use the Izhikevich phenomenological neuron model for simulations involving spiking neurons. The model neuron is identical to that previously described(Izhikevich, 2003)(Izhikevich, 2007). The update equations for the membrane voltage, v, of a neuron is given by

$$\dot{v} = v + 0.04v^2 + 5v + 140 - u + I_{syn} \tag{5.3}$$

$$\dot{u} = u + D_a\big(D_b v - u\big), \tag{5.4}$$

with an auxillary after-spike resetting

$$\text{if } v(t) \geq 30mV, \text{ then } \Rightarrow \begin{cases} v(t) = D_c \\ u(t) = u(t) + D_d \end{cases} \tag{5.5}$$

The selection of dimensionless parameters D_a, D_b, D_c, and D_d determine the dynamics of each neuron. In simulations, we model each neuron as an excitatory cortical pyramidal regular spiking(RS) neuron and adopt a similar methodology to calculate a set of random parameters for each RS neuron as found in (Izhikevich, 2003). The parameters are as follows: $D_a = 0.02$, $D_b = 0.2$ for all neurons. We added small variance to two parameters to gain variety in the dynamics of each neuron, such that each neuron was given a value $D_c = -65 + 15x^4$ and $D_d = 8 - 6x^4$, where x is a random variable with uniform distribution on the interval [0,1]. Since $x = 0$ corresponds to a traditional RS neuron, rasing x to the 4^{th} power biases the neurons toward RS neurons.

Phenomenological synaptic conductances

Synaptic excitation was determined by the following equation. Note that membrane voltage v is around $-65mV$ at the resting potential. We drop references to a specific neural field.

$$I_{syn} = -g_{ampa}v - g_{gabaA}(v + 70) - g_{gabaB}(v + 90) - g_{nmda}\frac{(v + 80)/60)^2}{1 + ((v + 80)/60)^2}v \tag{5.6}$$

Each conductance variable g followed first order differential dynamics $\dot{g} = -g/\tau$. Where $\tau_{ampa} = 5$, $\tau_{gabaA} = 7$, $\tau_{nmda} = 150$, $\tau_{gabaB} = 150$. For a synaptic event (action potential arriving at synapse) each conductance was re-initialized to the following values $g_{ampa} = 0.0025$, $g_{gabaA} = 0.005$, $g_{nmda} = 0.000025$, $g_{gabaB} = 0.00005$. One exception was the synaptic connections between the source C_{36} to target T_S. In this one case the four conductances were set to twice the normal value. The purpose for the increase is to improve the probability that thalamic neurons will fire in response to cortical excitation.

5.7 Appendix

5.7.1 Constant cell assemblies per neuron, r

The basic premise and connectivity underlying associative memories was first mathematically presented by David Willshaw in 1969(Willshaw et al., 1969). In essence, two neural fields composed of N neurons are connected via unstrengthened synapses. Each neural field contains L cell assemblies and each cell assembly contains a random collection of M neurons in the neural field. There is an association mapping between the neural fields, such that, exactly one cell assembly in the source field is associated with one cell assembly in the target field. For each source cell assembly, every source neuron has all its existing synapses strengthened to all target neurons in the associated target cell assembly. The result is an associative memory connectivity, where the neuronal connectivity is entirely dependent on the neural field organization specifying which neurons are in which cell assemblies. Neural field organization can be defined by a single matrix,$\mathbf{X}$, where each column represents a cell assembly and each row represents a neuron. If $x_{ij} = 1$ then the ith neuron is part of the jth cell assembly (see section 4.4).

There are three basic methods of randomly constructing a neural field matrix $\mathbf{X}$. Recall that there are N rows corresponding to each neuron and L columns corresponding to each cell assembly. The methods are:

1. Independently assigning each neuron to each cell assembly with probability $p = \frac{M}{N} = \frac{r}{L}$.

2. Assigning to each cell assembly, exactly M randomly selected neurons.

3. Assigning to each neuron, exactly r randomly selected cell assemblies.

In all cases the average number of neurons per cell assembly will be $M = \frac{Nr}{L}$ and the average number of cell assemblies per neuron will be $r = \frac{LM}{N}$. However, the variance of each variable, M and r, will be different in each case. For the first case, both are random variables with a binomial distributions; hence, as random variables, M(column sums of $\mathbf{X}$) and r(row sums of $\mathbf{X}$) have variance equal to $Np(1-p) = M(1-\frac{M}{N})$ and $Lp(1-p) = r(1-\frac{r}{L})$ respectively. In the second case, M has zero variance and r still has a binomial variance. In the third case, r has zero variance and M has the binomial variance.

The significance of these facts is highlighted in the distribution of the number of afferent and efferent synapses in the associative memory connectivity matrix. Recall from 4.1 that a multi-associative memory has the connectivity $\mathbf{W} = \mathbf{X}^{[y]}\mathbf{L}\mathbf{X}^{[k]^T} \bigcap \mathbf{U}$. For this discussion we are only discussing connectivity between two neural fields, source field k and target field y, therefore we drop the $[y, k]$ reference in connectivity matrices. Traditional associative memories are a special

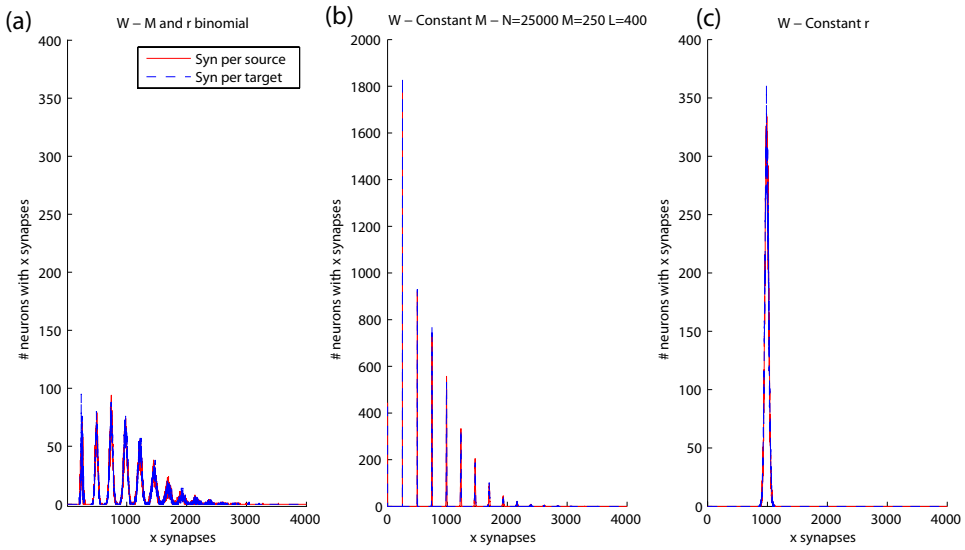

Figure 5.5: Associative memory connectivity distributions. Histograms are plotted of the $\mathbf{W}$ matrix column sums (synapses per source neuron) and row sums (synapses per target neuron) for the three neural field organizations. (a) M and r both binomially distributed. (b) M constant. (c) r constant. Both neural fields were constructed with N=25,000 M=250 L=400 r=4.

case when the symbol to symbol association matrix equals the identity matrix, or $\mathbf{L} = \mathbf{I}$. We also temporarily consider the case of full connectivity where every element of $\mathbf{U}$ is equal to 1. As a result the connectivity simplifies to $\mathbf{W} = \mathbf{X}^{[y]}\mathbf{X}^{[k]^T}$. To illustrate the point, we create neural fields according to the three cases and asses the distribution of synapses in the matrix $\mathbf{W}$. Figure-5.5 displays row and column sum histograms of $\mathbf{W}$ for each of the three cases. A clear difference is seen in the distribution of the number of synapses per neuron in Figure-5.5(c), the case of constant r.

The explanation is rooted in the general principle underlying associative connectivity. Namely, if two cell assemblies are associated then every neuron in one cell assembly is connected to every neuron in the other. As a result, a neuron will send or receive approximately M synapses per cell assembly that it is a part of. The visible peaks in Figure-5.5(a-b) are a result of the variance in r. Some neurons are part of one cell assembly ($\sim$250 synapses), some two($\sim$500 synapses), some three($\sim$750 synapses), etc. When r is constant, in this case $r = 4$, each neuron sends and receives on average 1000 synapses with variance proportional to the variance in M. If the connectivity is initially sparse, the fundamental nature of the results don't change. The number of connections merely decrease with some probability, which has the effect of shifting the plots along the x-axis to the left without changing their basic shape.

The traditional approach to associative memory recall has been to focus on the choice

of target neuron threshold. If the target neuron receives more than the threshold number of synapses then the neuron is active, otherwise it is inactive. Effective target cell assembly recall utilizing a fixed threshold is dependent on the number of neurons in a source cell assembly. If every cell assembly has a fixed number of neurons then the threshold is easy to set and can be set close to M if desired, without fear of some source cell assemblies having fewer neurons than the threshold. In practice, associative memories are quite robust to small variations in M. As a consequence of the bias toward threshold selection, past work on associative memories seems to exclusively have focused on either case 1 or case 2. The benefits of maintaining a constant r is explored here.

In the brain, synapses are dynamic structures(Craig and Lichtman, 2001). New synapses form and unused synapses disappear. There is no tangible evidence to suggest that the number of synapses per neuron varies widely over orders of magnitude. On the contrary, most neurons appear to have a relatively uniform distribution of synapses. There is also a variety of indirect evidence, such as the neurotrophin hypothesis (Schinder and Poo, 2000), that argue for self-regulation of synapses by each neuron, which would likely drive similar neurons to having approximately the same number of synapses. We can now state with reasonable confidence that in an associative model of cortical neuronal connectivity, the third case described above is the most biologically likely.

A major computational benefit of fixing r arises as a result of adding balanced inhibitory connections between the two neural fields.

5.7.2 Balanced inhibition

For clarity, we correlate our approach with chapter 4. The nature of associative memories and "why they function" as they do is easily described by the relationship between associative signal and associative interference. Although others have analyzed the connectivity and excitation in associative memories, their formulation did not include associative signal and interference(Buckingham and Willshaw, 1992; Graham and Willshaw, 1995; Buckingham and Willshaw, 1993). The framing in this context allows us to easily introduce balanced inhibition as interference normalization/cancellation in our model. Although described in some detail in **??**, it is useful to provide a concrete example for the present model.

The number of afferent target synapses in Figure-5.5 can be viewed as target synaptic excitation if all the neurons in the source field were activated. The histogram plotted is equivalent to plotting the histogram of the excitation vector $\underline{\tilde{\eta}}^{[y]} = \mathbf{W}\underline{\eta}^{[k]}$, if $\underline{\eta}$ is a vector of all ones, meaning all source neurons are active. In contrast, if we activate neurons from a single source cell assembly j, then we can similarly plot the histogram of synaptic excitation. The excitation due to a single

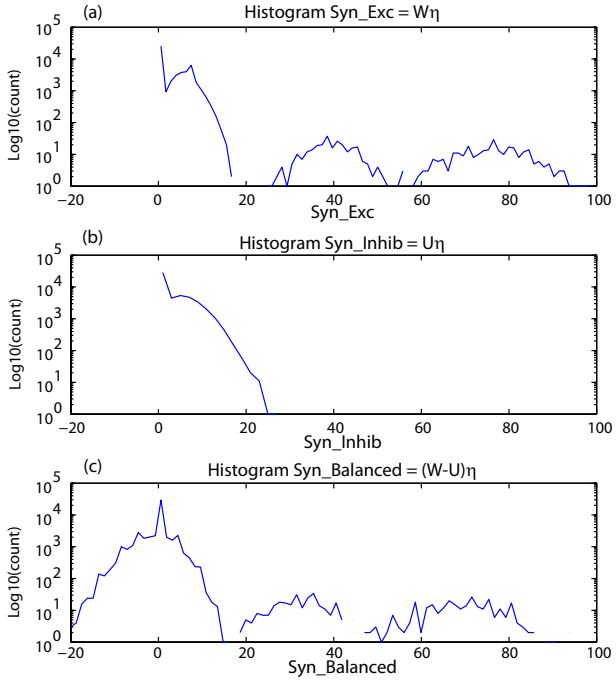

Figure 5.6: Associative memory balanced inhibition example. Each plot displays a histogram of excitation, where excitation value bins are given on the x-axis and the log of the histogram count is displayed on the y-axis. The activity of neurons η was determined by activating all the neurons in a single cell assembly, half the neurons in another cell assembly and activating another 250 neurons randomly. (a) Excitatory synapses only histogram. Notice the two peaks on the right correspond to the excitations due to the fully active and half active cell assemblies. Interference and noise is seen in the peak on the left. (b) Inhibitory synapses only histogram. The inhibition matrix was created randomly so that each target neuron would receive the same sum of inhibition as excitation. Because of the random construction, only interference is present. (c) The excitatory minus inhibitory excitations demonstrates the normalization of interference with balanced inhibition. Notice the left most peak is now centered at zero and the two signal peaks are still clearly separable but shifted left.

active source cell assembly j is given by $\underline{\tilde{\eta}}^{[y]} = \mathbf{W}\underline{\mathbf{x}}_j^{[k]}$, where $\underline{\mathbf{x}}_j^{[k]}$ is the j^{th} column vector of $\mathbf{X}^{[k]}$. According to theorem 1, the neurons in target cell assembly j will only receive associative signal input, and as a first approximation, we can assume that the neurons in all other cell assemblies receive associative interference input.

Figure-5.6 shows a slightly more complicated version of the aforementioned condition. The initial probability of connectivity is set to 30%, such that each element in the matrix $\mathbf{U}$ is set to one with probability $p = 0.3$. In Figure-5.6(a), the synaptic excitation is displayed for a source neural field activation with all the neurons active from one source cell assembly j, half the neurons active from cell assembly h, and another $M = 250$ random neurons activated. The total number of activated neurons is approximately $A = 675$. Three clear peaks are seen in the figure corresponding to the synaptic excitation on neurons in target cell assembly j, h and other cell

assemblies. The parameter p_w from 4.4.8 is the probability that two randomly selected neuron will receive an excitatory connection. Therefore, to create balanced inhibition, we merely need to use this probability for the inhibitory connectivity. In our case and the case of the simulations, the inhibitory synapse strength were twice those of excitatory synapses, therefore the probability of connectivity for inhibition is $2p_w$.

As is predicted by theorem 1, cell assembly j should receive approximately $M * p = 250 * 0.3 = 75$ associative signal inputs and $(A - Mp) * p_w * p = 600 * 0.0117 * 0.3 = 2.1$ interference inputs. Cell assembly h should receive $\frac{M}{2}p$ signal inputs and $642 * p_w * p = 2.25$ interference inputs, and all other cell assemblies should receive approximately $675 * p_w * p = 2.37$ interference inputs. As demonstrated by this example the parameter p_w is the critical parameter for determining interference. The parameter p_w represents the probability that a synapse will be strengthened between two randomly selected neurons if the synapse exists. The parameter p determines if the synapse exists.

As seen in Figure-5.6(c) the role of balanced inhibition is simply to normalize the interference about a zero mean.

Chapter 5, in part, is currently being prepared for submission for publication of the material. Solari, S. and Hecht-Nielsen, R. The dissertation author was the primary investigator and author of this material.

6

The neural code of cognition: unifying cognitive information processing in the mammalian brain

6.1 Introduction

The unified model in this chapter attempts to tie together several puzzle pieces surrounding cognitive information processing. We utilize and integrate concepts from confabulation theory (chapter 2), the detailed neuroanatomy underlying cognitive information processing (chapter 3), the storage and utilization of associations in the cerebral cortex at the neuron level (chapter 4), the proposed universal thalamocortical information processing operation at the neuron level(chapter 5), and we add psychologcial phenomena in this chapter. The chapter attempts to stitch all of the above into a single neural code of cognition, by explaining the array of psychological phenomena related to cognitive information processing in terms of the actual function of individual neuronal populations.

In order to discuss any neural code of cognition, we must attempt to define what we are talking about. One approach might start through community consensus and utilize that of wikipedia, stating that cognition is generally taken to mean "the process of thought". For more specificity we might utilize Markl's definition of cognition as "the ability to relate different

unconnected pieces of information in new ways and to apply the results in an adaptive manner" (Markl,H. 1985 from (Striedter, 2005)).

Although both fairly elegant, we can immediately see that an attempt to unify cognitive information processing with the abstract definitions above, is a lot like scooping water with a strainer. The model and explanation will not hold water. Instead, we attempt to simplify the problem by defining the function of each of the major structures in the mammalian brain. By defining the function of each structure we are then able to (and will) provide explanations for the widespread psychological phenomena/observations regarding memory and cognitive information processing. By defining the basic function of each major neuroanatomical structure, there is little room for interpretation. If we were able to know the function of each major anatomical structure, then cognition can simply be defined by the information processing capabilities that are physically available to us. A benefit of this approach is the inability to invent or add additional mechanisms of thought processes. The structures in the brain are physical, once their functions and interactions are explained then we are done in our quest to explain cognition.

In any modeling approach, neuroscience experimental fact should not merely be restated as a theory, the facts must be integrated as part of a larger whole. Any viable theory of cognition must not only explain all of our everyday experiences as humans and must more importantly be consistent with the existing experimental facts in a variety of disciplines from neuroanatomy, to electrophysiology, to psychology.

Although we are ambitious in our scope, we are cautious in our approach. We recognize that the full intricacies and details of cognition are far from fully understood. Many details still need to be worked out and many years of difficult experiments will be required to validate any model and theory. However, we produce a model consistent across a wide variety of experimental disciplines, upon which past experiments can be fitted and future experiments can be developed and tested. We hope the rigorously anatomical focus will facilitate the development of future experimental correlations, because ultimately any model must be tested by experiment. Our goal here is to provide what we believe to be, the first comprehensive unified theoretical, anatomical, computational, and psychological model of cognitive information processing in the mammalian brain.

6.2 Psychological background information

In order to tie together psychological phenomena and our model of cognition, we must define the psychological phenomena. Psychology and the terms that are used has a long distinguished history itself with many nuances and disagreements. Our goal is not to present a

detailed review of psychological phenomena, but rather to select, based on our judgement, respected psychological descriptions that maintain consistency with our models neuronal descriptions. Regardless, we do attempt to be comprehensive in psychological descriptions, because, we hypothesize that all major psychological phenomena can ultimately be explained in terms of those neuronal mechanisms that we discuss here.

Experimental studies in humans and primates have elucidated two general types of memory, declarative and nondeclarative (procedural) memories(Squire, 2004). We will utilize these two types of memory to explain explicitly the storage, recall, and processing of information.

6.2.1 Declarative memory

To describe declarative memory, we will utilize Larry Squire's definition. He writes,

> ...declarative memory is the kind of memory that is meant when the term 'memory' is used in everyday language. It refers to the capacity for conscious recollection about facts and events(Squire, 2004).

Although unsettled in the literature, our model, in conjunction with what has been experimentally verified, suggests that declarative memory is composed of three independent yet interacting anatomical systems operating on separate time-scales. Namely working memory, short-term memory, and consolidated long-term memory.

Working memory operates on the time scale at which attention can be maintained, seconds to minutes(Baddeley, 1981; Monsell, 1984). Based on psychological studies we adopt Monsell's viewpoint and description of working memory as

> Working memory is no more (or less) than a heterogeneous array of independent temporary storage capacities intrinsic to various subsystems specialized for processing in specific domains.(Monsell, 1984)

Working memory appears to exist fundamentally in the cerebral cortex. One of the principle properties underlying working memory is the notion that he storage and processing of information both occur in the same subsystems which underly the working memory itself.

Short term memory is mediated by, and requires, the perirhinal/parahippocampal (Phr) cortices and hippocampal formation to function properly. Short term memory operates on the order of seconds to years. The classic example of the underlying structures involved in short term memory is the patient H.M. who had large bilateral portions of his medial temporal lobe resected to control epileptic seizures(Milner, 2005). H.M. was unable to form any new declarative memories after the operation, but otherwise had little other changes to his persona and memory.

Consolidated long term memory is declarative memory that was initially stored in the hippocampus via short term memory, but has been consolidated into the cerebral cortex and no

longer requires the medial temporal lobe for use. The form of consolidated long-term memory appears to be

6.2.2 Procedural (Non-declarative) memory

We also adopt Squire's definition for procedural memory.

> [Procedural memory] is expressed through performance rather than recollection ... the memories are revealed through reactivatation of the systems within which the learning originally occurred(Squire, 2004).

The distinguishing feature of procedural memory is that through repetition and practice, behavioral memories can be learned without declarative recall of how the memory was learned.

6.3 The neural code of cognition

The entire explanatory model that follows is a hypothesis, therefore for readability, we simply state the models hypothesis of function as fact even though we are extremely aware of the contentious nature of some concepts. The anatomical description terminology corresponds to chapter 3. Viewing Figure-3.8 in conjunction with the following discussion should help clarify the direct relationship of our discussion to Confabulation Theory.

6.3.1 Representing perceptions in the individual's universe

The first step in developing the neural code of cognition is to establish how perceptions are learned, stored and represented in the cerebral cortex. We argue that the cerebral cortex is divided into many functionally discrete modules varying on the order of a few $10mm^2$ that include all six layers of the cortex and a proportional volume of reciprocally connected specific thalamic nuclei. Each module is programmed genetically to receive certain afferent projections from other modules and subcortical nuclei. Since each module does not receive information from all neurons in the brain, the information that is communicated to the module is constrained. As a result, the incoming projections (and the information they communicate) define an information hyperplane within which perceptions must form. We refer to the information hyperplane that defines each module as a single *cognitive dimension*. Perceptions are simply frequently encountered combinations of afferent information patterns. Because each perception is constrained within a cognitive dimension, perceptions within a single cognitive dimension are anatomically all grouped and stored within a single module. Confabulation theory terms the functional representation of each perception a *symbol*. The functional segregation of each module anatomically is likely a

result of the fact that genetically programmed axons will have a "mean" location that they initially target. The anatomical borders surrounding this mean location will receive less coherent input than the center. As adjacent modules are activated asynchronously, either by control signals, or incoming stimulation, lateral inhibition will likely create a functional boundary between the two. In the mammalian brain, numerous functional boundaries in the cerebral cortex are clearly represented by distinct changes in cortical structure as is the case at the border of V1 and V2, parahippocampal and perirhinal cortices, or primary motor and primary sensory cortices(Brodmann, 1909). We propose that an unbiased 3-D detailed analysis (much more detailed than the last comprehensive study done by von Economo in 1925) of human cytoarchitectonics may even discover all the functional modules in a single brain. Although, the development of modules is an ongoing research topic, simple preliminary models of this functional segregation look promising(unpublished observations), but significant work needs to be done.

Each perception within a module develops (see Figure-6.1(a)), because frequent information patterns will have the tendency to form interconnected cell assemblies just as Hebb postulated(Hebb, 1949). However, our research suggests that even within a single module a perception has multiple representations depending on the layer/class of pyramidal cell (a.k.a neural field) within which it resides. One way to think about this is that each neural field has a function within the module (defined by its efferent projections) and each perception has a cell assembly representation within each neural field. Although each cell assembly self-organizes, our visualizations and computational models depict a cell assembly in a neural field as a randomly selected subset of neurons in the neural field. A confusing aspect arises because the cell assemblies in the various neural fields that each represent a single perception form associative connectivity with each other. Hence, one might argue that a perception can be viewed as a single cell assembly within the module. But, the distinction of multiple functional representations is not semantic. Because neural fields have different efferent projections and thus different functions, the activation of a single perception within one neural field vs another will have varying functional consequences. These functions will be discussed in terms of the formation and utilization of the various types of memory. We should also mention the stark contrast of our modular, localized view of cell assembly perceptions to commonly encountered cortically distributed views of cell assembly perceptions(Fuster, 2003).

We can state some specifics regarding the development of perceptions in a module. As seen in Figure-6.1(a) feed-forward developmental input typically terminates in C_4 and feedback developmental input in C_2. The common feature is that both C_4 and C_2 are composed of numerous small granular cells. Therefore, due to their size, the C_4 and C_2 projections are relatively restricted to intra-module connections. These neural fields act as feature attractors to

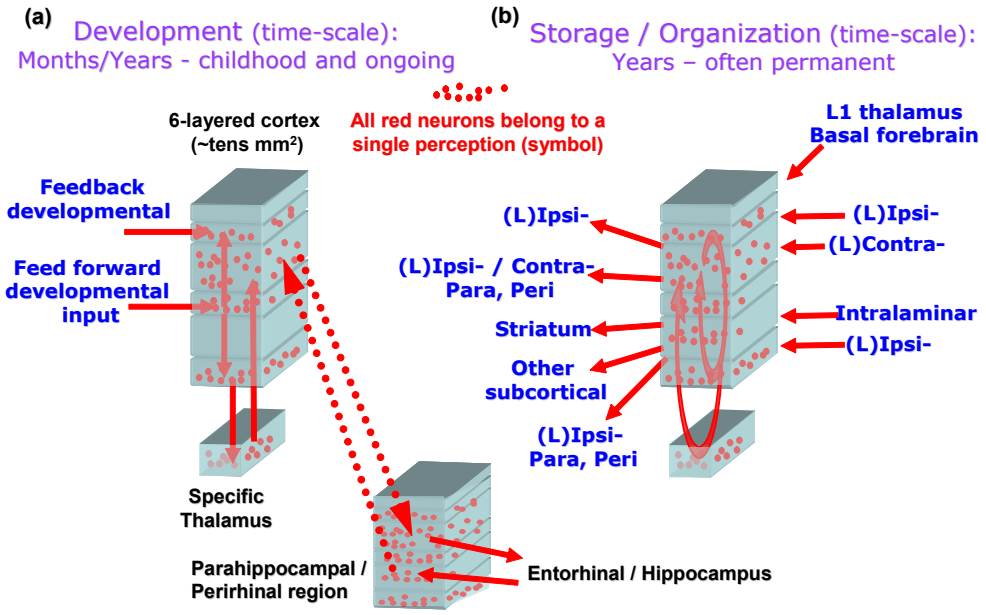

Figure 6.1: The development and organization of a thalamocortical module.

perceptions in the module. The feature attractor cell assemblies are a commonly seen patterns of activation from the incoming projections. The feature attractors in C_4 are like pieces or subsets of the full perceptions that will develop in the module. Feature attractors develop in these neural fields to basically translate incoming information from other source modules into cell assembly percpetions in the target module. During development the C_4 feature attractors are used to develop perceptions in other layers within the module. Once developed the C_4 cell assemblies map source module input into perceptions in the target module. Fukushima's neocognitron is an excellent example of a working computational model utilizing a two stage concept of cell assembly formation(Fukushima, 1980). Additionally, once the perceptions have formed within a module the need for feedback (C_2) or feedforward (C_4) feature attractors may diminish. This may explain why in many instances the C_2 field is very clear in the juvenial brain but may become indistinguishable from C_{3A} in the adult brain(Brodmann, 1909).

The need for feature attractors (a C_2 or C_4 neural field) is dependent on the cognitive dimension of the module. Sensory regions of the brain require the development of more abstract perceptions by combining lower level (in hierarchical terms) perceptual information, hence the need for C_4. In motor regions of the brain, feature attractors are not needed because motor perceptions are not developed from combinations of internal and/or external stimuli. Therefore, C_4 neural field is not needed in regions such as primary motor cortex (see Figure-3.7).

After feature attractors develop in C_4, more general and invariant cell assembly per-

ceptions begin to develop in the C_{3A}, C_{56}, C_{3B}, and C_{6T} neural fields. We propose that the perceptions which develop in the module undergo an integrated parallel development. First, re-entrant cortico-thalamo-cortical (C_{6T}-T_S-C_{6T}/C_{3B}) connectivity is likely utilized to form the cell assembly representations and connections used in thalamocortical information processing. For many modules, perceptions are simply formed and processed at this stage with no further development.

For associative cortices throughout the brain, a concurrent perception must form in the Phr cortices. As seen in Figure-6.1(a) representations are forming in the reciprocally connected Phr cortices in conjunction with the modules perceptions, and we propose that these are not separable. We must emphasize that in order for associations to form between perceptions, each perception must have a cell assembly representation in C_{3B} and two cell assembly representations in the Phr cortices. The first cell assembly resides in the upper layers of the Phr cortices and receives input from the C_{3B} perception. The second cell assembly resides in the lower layers of the Phr cortices and provides input to the C_{3B} module perception. Therefore a perception in a thalamocortical module must develop in conjunction with Phr cortices, otherwise the perception cannot be associated with other perceptions. The cell assembly perceptions in C_{3B}, ultimately those that project to the Phr cortices, are the "final" form of a perception in a module. These connections form the basis for short term memory. Chapter 4 clearly demonstrates that these reciprocal connections can be made with low probabilities of connectivity. Another confusing aspect may arise because the Phr cortices likely function as modules themselves. We propose that the reason allocortices, such as the perirhinal and entorhinal, in essence do not contain a C_4 neural field is because they do not have and/or require feature attractors, they simply have copies of perceptions from association cortices.

The secondary form of perceptions are cell assemblies that form in C_{3A} and C_{56}. The cell assemblies in these neural fields are essentially mapped directly to the C_{3B} cell assemblies. These cell assemblies are utilized to form the basis of the intra-cortical interactions in the thalamocortical thalamocortical attractor network described in chapter 5. The C_{3A} and C_{56} cell assemblies form the majority of associations with perceptions in other modules and within one module. These are the cell assemblies that form associations (knowledge links) in cortically consolidated long term memory described in chapter 4.

The neural fields and cell assemblies discussed relate exclusively to the perceptions of our universe. The expression/activation of the neurons comprising a perception in a thalamocortical module in itself defines a perception and by extension what it means to perceive. These perceptions merely reflect the the organisms internal state of the world. They do not however directly represent the actions that an organism is able to take in response to those perceptions.

Actions and behaviors are stored in separate neural fields.

6.3.2 The cortical representation of an individual's behaviors

In addition to perceptions, thalamocortical modules also store what we term *behaviors* (action commands in confabulation theory). There are two distinct types of behaviors corresponding to distinct populations of neurons, C_{5P} and C_{5S}. The first and most easily described are the C_{5P} *predictive behaviors*. The best example of these behaviors is their projection from primary motor cortex. In primary motor cortex, C_{5P} neurons project directly to the spinal cord. As a result of their firing action potentials, these neurons directly cause the contraction of muscles. Hence, they are behaviors related to the contraction of muscles in a module whose cognitive dimension is muscle contractions. By extension, we propose that the C_{5P} in other modules carry out the same function, but represent behaviors that correlate to the cognitive dimension of those individual modules. A module that represents a perceived state of the external world should likely have behaviors associated with the cognitive dimension within which the perceptions exist. The firing of C_{5P} neurons should result in the predictive behavior of an active perception in C_{3B}. The C_{5P} neurons in most of the brain project to the pons. We suggest that vast majority of behaviors are simply predictive behaviors corresponding to signals that arrive in the pons, including those from other modules. In this way, when expectations/predictions do not align with other signals arriving at the pons (or anywhere else for that matter) error signals can be generated and the brain has the ability to distribute goal/drive signals that compel changes in the future firing of predictive behaviors. This enables the creation of accurate models of the world (internal/external) in which we live. The example of primary motor cortex is no different. The C_{5P} neurons represent a predictive behavior for a particular perception of movement in that module. If the perception launches a predictive behavior (say a muscle contraction) and the muscle contraction is correct, meaning no internal nervous system error signals are generated, then no change is needed from the mapping of perception to predicted behavior. If on the other hand, the outcome of the muscle contraction generated errors, then a new predictive behavior (a different set of C_{5P} neuronal output) must be selected in the future that will not generate errors. Notice that perceptions cannot change, only the mapping between perceptions and predictive behaviors can change. We likely may be able to acquire new predictive behaviors as well. Note that the psychological consequences of the hypothesis of fixed perceptions are enormous.

Another clue as to the nature of predictive behavior launching we are describing, is the predominant connection from C_{3B} to C_{5P} neurons. For example, in rat primary motor

cortex layer 3 predominantly connects with layer 5 neurons monosynaptically(Kaneko et al., 2000). The mapping between C_{3B} to C_{5P} as we discussed in chapter 3 is likely reinforced by the cholinergic basal forebrain projections. Therefore, the way in which we reinforce and learn predictive behaviors is through changes in the mapping from C_{3B} to C_{5P} as a result of error derived cholinergic stimulation from the basal forebrain. The intralaminar nuclei of the thalamus, T_I, projections are utilized to alter the C_{5P} output in order to change predictive behaviors more quickly.

The second type of behavior is stored in C_{5A}, which we term *control behaviors*. These are control behaviors because they represent the information that the basal ganglia receives from the cortex, and are ultimately used to control two things. 1) the triggering of predictive behaviors via the intralaminar thalamus and 2) the control of thalamocortical module confabulations via T_{L1}. Our research did not address computational models of the basal ganglia, therefore, the least is known or hypothesized about the specific form of the C_{5A} input or the type of information transfer that occurs through the basal ganglia. The C_{5A} neural field has large reciprocal projections with the $C_{2/3A}$ neural fields. The C_{5A} neural field is likely involved in representing states of perceptions as they are undergoing selection. We did to a degree also model the control signal emanating from T_I and T_{L1} thalamic nuclei, which receive large projections from the basal ganglia. Both are simply graded control signals of a similar nature targeting a specific module. Therefore, the basal ganglia likely learns transformations from C_{5A} input, that represents the state and/or number of perceptions involved in confabulations, to the simple timing of two independent control signals per module, namely predictive behavior triggering and confabulation control. Undoubtedly, the total composition of behaviors in an individual includes various learned mappings in the basal ganglia.

6.3.3 Perception

What is perceiving? Perceiving is simply the elevated firing of an existing perception, represented by cell assemblies in various neural fields, within a module. Perceptions are capable of being associated only when the perception is highly active in the C_{3B} neural field and the perception is transferred to the Phr cortices for binding with other perceptions in the hippocampus. A perception can only be recalled when a perception in a source module is used to activate it, via direct cortico-cortical projections or indirectly through the Phr. Perceiving is consequently graded, dependent upon two basic factors. First, the existence of a perception in a module. Second, the number of modules (or cognitive dimensions) being utilized to perceive. The accuracy, as one might measure in a psychological test, of a perception combines the degree to which the perception in any given module is an accurate reflection of the external stimuli, and

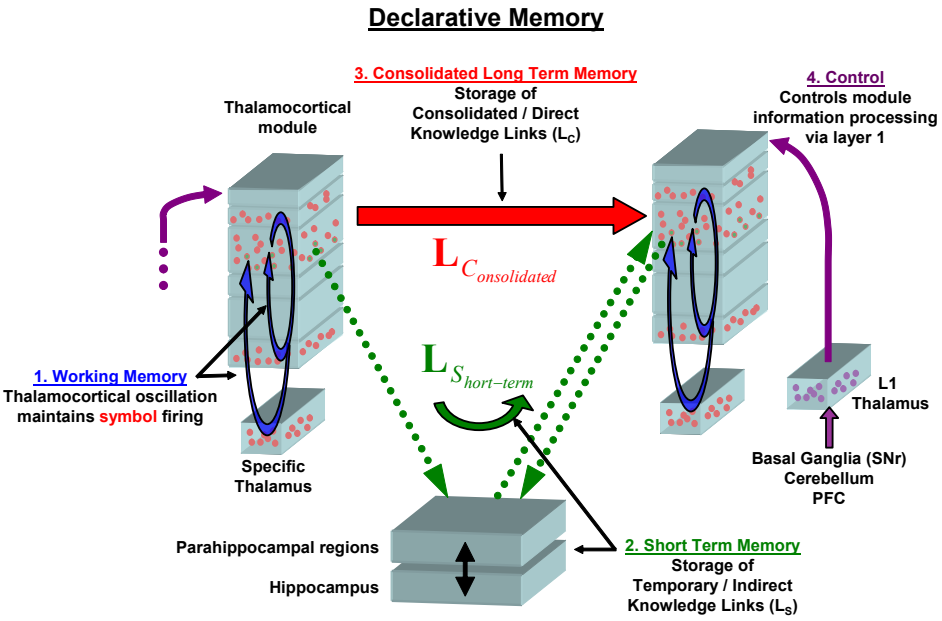

Figure 6.2: Declarative memory.

how many modules are simultaneously utilized to perceive the external stimuli. The degree to which perceptions are accurate reflections of the external/internal world are simply a result of exposure. The more exposure the more accurate the perception in a module. Here the notion of critical periods in cortical development are essential. If one does not develop certain perceptions in a module during a critical time period that exclusively enables the development of those perceptions, then one can never experience those perceptions. And if perceptions must be developed in a hierarchy, then higher level perceptions can not be composed of perceptions that do not exist. The organism is categorically limited to representing the external world with perceptions that did develop and exist in a module.

Perceiving the world, and any tests of perceiving the world, is influenced by the predictive behaviors (C_{5P}) that are mapped to perceptions (C_{3B}) in any module. Testing a subject requires a response, and their response is dependent on the predictive behaviors that are mapped to perceptions. If an "incorrect" predictive behavior has been mapped to a perception, then the response will not be a reflection of the accuracy of the internal perception but rather the particular C_{3B} to C_{5P} mapping.

6.3.4 Thalamocortical working memory

Working memory is the persistent firing of one or more perceptions in a thalamocortical oscillation within a single module as seen in Figure-6.2. Chapter 5 describes several computational mechanisms underlying the thalamocortical oscillation. Since working memory can be implemented by maintaining perceptions in any module, working memory is by nature distributed. Hence, our choice regarding the accuracy of Monsell's description. Surprisingly Monsell's work is not given much mention in the literature regarding working memory. Working memory must be controlled in some way. We propose that the layer 1 thalamocortical projections are in fact the control signals underlying working memory. The major involvement of areas of pre-frontal cortex, which relates to Badley and Hitch's descriptions, in working memory are likely two fold. First, by the nature of the information that is stored in those modules, things like behavioral plans and goals, working memory is easy to set up in these modules within an animal experiment. Second, there is likely an area of frontal cortex, probably the equivalent of Brodmann's area 9, that acts as "primary thought cortex". One might be tempted to think we are discussing the equivalent of the "central executive" in classical literature; however, the layer 1 projections from the thalamus implement the control of working memory. Although the thalamic nuclei itself is fairly central, the inputs to those thalamic projections are wide ranging, hence the central executive controlling thought is distributed. The corticothalamic projections from "primary thought cortex" likely have direct projections to the VAmc/VM nuclei of the thalamus that produce the wide spread layer 1 thalamic projections. In this way activity in the primary thought cortex could be used to control confabulations in other thalamocortical modules almost directly. This would be the analog to primary motor cortices' control over muscles.

6.3.5 Hippocampal short term memory

As seen in Figure-6.2, when a thalamocortical module is maintaining symbol firing in working memory, the C_{3B} neurons are highly active and send information to the Phr cortices, which in turn send information to the hippocampus. The co-activation in time (or sequentially in time) of perceptions in two distinct modules enable the hippocampus to bind the two perceptions together based on emotional signals from the limbic system and/or frequent co-occurrence. The present hypothesis is very similar to past ideas related to indexing theory(Teyler and DiScenna, 1986; Teyler and Rudy, 2007). And in a sense that is exactly what we suggest occurs. As we stated, each perception in associative cortices has two representations (afferent and efferent) in the Phr. Once two perceptions are bound in the hippocampus (one perception could be bound to many others simultaneously, but we refer to a simpler case), the re-activation of one will

re-activate its corresponding perception in the upper layers of Phr, which in turn will unbind or reactivate all related bound perceptions in the hippocampus. As perceptions are unbound, the perceptions in the lower layers of the Phr will be reactivated and send excitation back to the perception in the source thalamocortical module. Here is the interesting aspect of the Phr acting as a thalamocortical module. The upper layer perceptions (C_3B in the Phr) are perceptions in the traditional sense defined above. The lower layer representations (C_5P in the Phr) are in fact predicted behaviors as defined above. Therefore, the Phr cortices can utilize procedural memory and skill learning to create mappings from upper layers to lower layers. These are the basis of semi-permanent knowledge links that will be consolidated. As is illustrated, we propose the temporary knowledge links involved in short term memory are primarily stored through these two independent mechanisms

It is also interesting to note that the perirhinal cortex is by far the largest and most developed in humans compared with other primates(Burwell, 2000). It may not be surprising that the perirhinal cortex is interconnected with auditory association cortices and other polymodal association cortices. In contrast, the parahippocampal cortex is only slightly larger in the human than monkey and is interconnected with visual association cortices. The increased ability to address and store knowledge links via the perirhinal cortex is likely one of the significant factors in human evolution and our use of language.

In summary, short term memory is when a perception active in working memory in one module, activates a perception in another module through the stored indirect hippocampal associations. Since the there is only a mapping needed between source perceptions and target perceptions, the equivalent of a single weighted matrix $L_{S_{hort-term}}$ must be stored by the hippocampus and related structures. The matrix simply stores the weighted values of associations between perceptions (i.e. knowledge links), which can be expressed as a single number for each perception-perception association. In fact, the matrix likely has a third dimension which evolves over time to store episodic memories. The key point is that the form of the matrix is no different than the form of the matrix used in cortically consolidated long-term memory.

6.3.6 Cortically consolidated long term memory

Consolidated long term memory requires the equivalent of transferring the single short term matrix, which is stored in the hippocampus, into a direct cortico-cortical $L_{C_{onsolidated}}$ matrix of associations between perceptions. Chapter 4 explicitly describes the conditions under which these new associations form. In fact, they are formed as a multi-associative memory. Note that these matrices, **L** as shown in Figure-6.2, store identical forms of information, therefore identical information processing in thalamocortical modules can be applied to both short-term

and consolidated long-term memory. The individual associations that exist in the separate short term and consolidated long term systems have no problem jointly integrating information because the associations simply add. Once an association is consolidated into long term memory, the firing of a perception in a source cortical module is sufficient to reactivate associated perceptions in other modules without going through the hippocampus.

The transfer of memory from hippocampal mediated short term memory to consolidated long term memory, simply requires the reactivation of associations in the $L_{S_{hort-term}}$ matrix stored in the hippocampus. As a result, perceptions in the cortex will be reactivated and the equivalent direct cortico-cortical association can slowly form and strengthen. This slowly transfers the elements of the $L_{S_{hort-term}}$ matrix to the $L_{C_{long-term}}$. The process is depicted in Figure-6.4.

6.3.7 Attention

The effects of attention have been described from psychology to electrophysiology. In general, we can say that attention is the ability to focus on certain stimuli (increasing the ability to detect and perceive the attended stimuli) while potentially ignoring distracting stimuli. Studies that measure neuronal firing rates of visual stimuli in monkeys show the basic property of attention is increased sensitivity to the attended stimuli or stimuli location(Reynolds and Chelazzi, 2004). In other words, the ability to increase the signal level of an attended stimuli.

In our model, attention is defined by maintaining or accentuating the control of the activation of certain thalamocortical modules while preventing or inhibiting the controlled activation of others. Attention does not directly focus any input onto perceptions (C_{3B}) in a module. However, there are actually two independent effects of attention. First, increasing the control signal to a particular thalamocortical module (assuming the multiplicative control of chapter 5) will tend to increase the sensitivity of any and all perceptions in that particular module. The prediction of this type of sensitivity increase in all perceptions within a module is exactly that measured in experiments(Reynolds and Chelazzi, 2004). If measuring the firing rate of stimuli in an attended module, the contrast of the external stimuli required to elicit activity in the stored perception will be lower.

The second type of attentional effect will arise when one module is receiving attention control, as described above, but its perceptions in working memory are transmitting information to other modules. The module receiving designated control input will undergo the first type of attentional effect. The target modules receiving input from those active perceptions in source modules should experience a second type of attentional effect. Perceptions, in $C_{3A/3B}$, in the target module will receive associative input from active perceptions in other modules. Therefore, those target perceptions will be biased to become active in the module. This might also be

measured as an increased sensitivity or "gain" on the target perceptions.

Experimentally testing the hypothesis of two independent effects of attention should be feasible. The first type of attentional effect should cause an increase the sensitivity of all perceptions within a single module, whether they are the desired stimuli or not. The second type of attentional effect should only increase the sensitivity of those perceptions in a module that have been associated with perceptions that are active in another module that is receiving attention.

6.3.8 Procedural (Non-declarative) memory

Basal Forebrain

The full mechanisms underlying all describable procedural learning would likely have to include virtually all subcortical nuclei. Here we focus on describing only the cortically stored procedural memories. As depicted in 6.3, these procedural memories are defined as the mapping from C_{3B} to C_{5P}. As discussed in the storage of predicted behaviors in C_{5P} this mapping is likely reinforced by acetylcholine.

In the baboon, there are only two major sources of cholinergic neurons outside the motor neuron system. The two locations are the substantia innominate (including the basal nucleus of meynert) and a subpopulation of neurons in the striatum(Satoh and Fibiger, 1985). One common feature of both the striatum and basal ganglia is in their role in learning and more importantly skill learning. This suggest that the acetylcholine projections may possibly have analogous function in each. Despite the widespread cortical innervation, the cortical innervation of local areas of the basal forebrain has been shown to be relatively discrete implying that each could be individually targeting a cortical module(Bigl et al., 1982). Another line of evidence shows that when the basal forebrain is lesioned, cortical plasticity associated with motor skill learning (procedural memory) stops(Conner et al., 2003).

We hypothesize then that the cholinergic projections to the cortex serve to possibly reinforce certain synaptic connections between neurons that were recently co-active. First, in awake active states in the cortex, the cholinergic projection could be used to strengthen the mapping from C_{3B} to C_{5P}. A second use to the cholinergic projections could occur during sleep, which would allow memories to be consolidated from short-term memory to long term memory.

Basal ganglia skill learning

The basal ganglia is the least developed model regarding the nuclei we studied. Its likely function is measuring the state of confabulations in all modules in the brain and determining

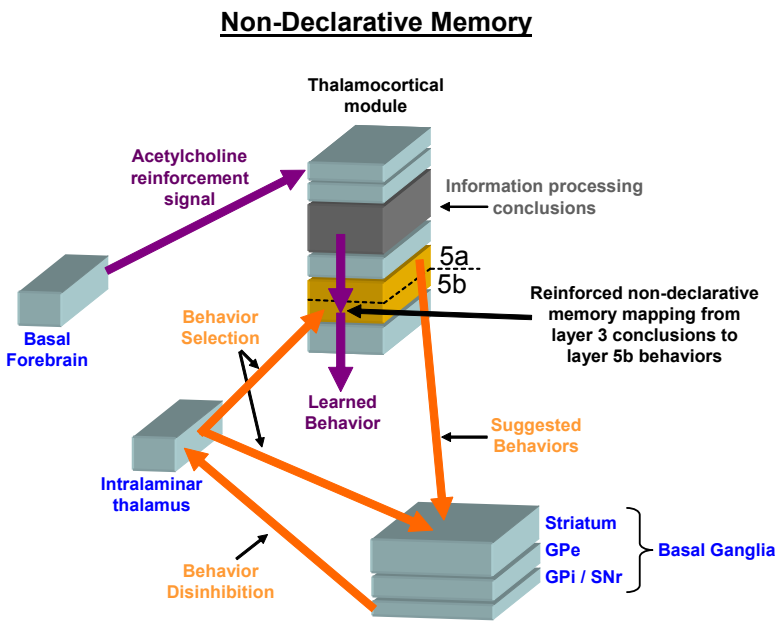

Figure 6.3: Non-Declarative memory.

the timing and sequence of control over modules in the brain, much like controlling the timing of muscle contractions in the body. The basal ganglia must learn two independent types of control for each module in the brain: 1) Confabulation commands via T_{L1}, and 2) The triggering of predictive behaviors via T_I. The feedback from T_I to the striatum can be used to shut down behaviors that have been triggered or could also be used to launch the next behavior in a sequence.

Cerebellar skill learning

Our models did not explicitly address any involvement of the cerebellum. However, its significant role in both movement and thought deserves, if but briefly, to be placed in the context of our discussions(Ramnani, 2006). The cerebellum (Purkinje cells) in the cerebellar cortex project to the deep cerebellar nuclei, which in turn project to the ventral thalamus(Voogd, 2003). The Purkinje cells are tonically inhibitory; the ultimate action of the cerebellum implementing an effect on the ventral thalamus is therefore disinhibitory. Inhibition of Purkinje cells will tend to allow deep cerebellar nuclei to fire thus delivering excitation to the ventral thalamus. We make special notice that the special effect of disinhibition is utilized by the cerebellum and the basal ganglia. It should be interesting that the two structures most involved in learning to deliver patterns of input to the thalamus are both disinhibitory pathways.

The deep cerebellar nuclei project to both ventral T_S and T_{L1} thalamic nuclei, but not to intralaminar T_I nuclei. The unique projections suggest a certain functional role for the

cerebellum. First, it appears that the cerebellum does have projections to the ventromedial T_{L1} nuclei. These projections will have the effect of controlling confabulations in thalamocortical modules; hence the cerebellum can learn to directly control thought processes through regulating which modules undergo confabulations. Second, the projections to the specific VA/VL T_S nuclei will ultimately provide excitation to C_{3B} neural fields in motor and pre-frontal cortices. Since the C_{3B} neural field defines perceptions in a module, the cerebellum can strongly effect perceptions in thalamocortical working memory in the modules to which it projects. Hence, the cerebellum can likely insert appropriate perceptions into thalamocortical information processing conclusions, which will result in learned predictive behaviors being launched in C_{5P}. It appears curious that the cerebellum does not target the intralaminar nuclei T_I of the thalamus, since these projections are directed at C_{5P} outputs. Researching the computational role of influencing motor perceptions directly rather than the predicted behaviors which directly cause movement in primary motor cortex may be a fruitful endeavor.

6.4 Speciation and variations of thought

The differences in the capabilities within and between species can be explained in the existence, size, structural connectivity, and ability to control muscles. Movement capability is reflected by the differences between physical muscles and the structural connectivity of bones. Cognitive capability is reflected by differences between muscles of thought (thalamocortical modules) and the structural connectivity of axons and fascicles.

Clearly evolution took millions of years to establish the genetic code for the correct connectivity capable of endowing mammals, primates, then humans with such significant intelligence capabilities.

But just as the alteration of a few joints and muscles enabled upright walking and the use of opposable thumbs, so did a few alterations in cortico-cortical connectivity and the introduction of a few new thalamocortical modules enable abstract language. In language we gain all the human capabilities that other species don't have.

6.5 Cognition: the ballet of thought

Cognition then, is then the result of developing, strengthening and learning to control muscles of thought. As opposed to regular muscles these thalamocortical modules likely have two control inputs. The first regulates the state of information processing (confabulations) in the module via T_L1 input, the second regulates the timing/triggering of predicted behavioral output

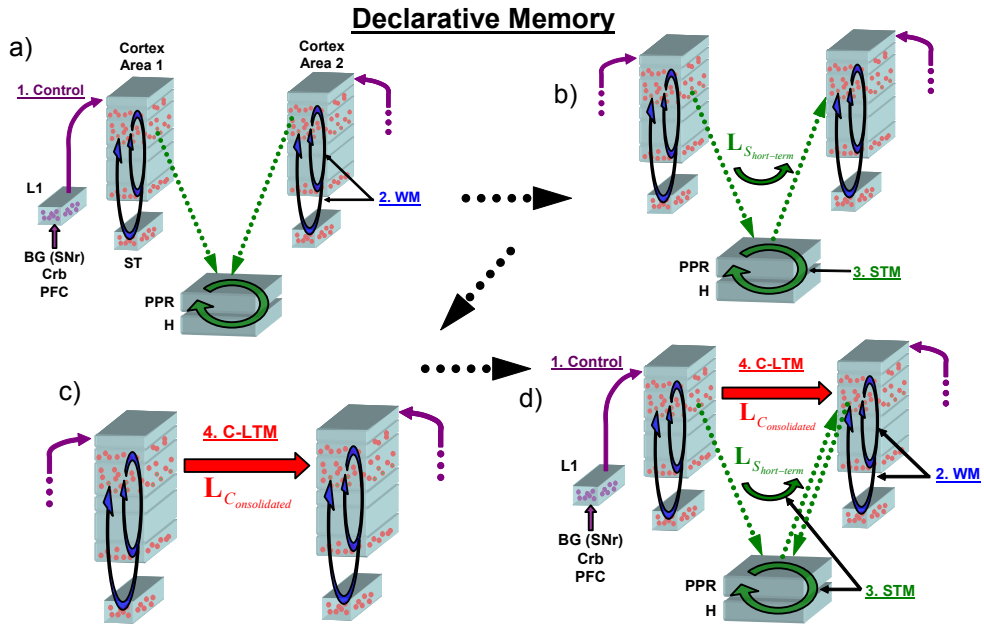

Figure 6.4: Declarative memory. The figure shows the anatomical implementation of declarative memory on the three time scales upon which it operates. a) Working memory (WM) - A control signal from the ventromedial thalamus(VM) provides input to layer 1 of a thalamocortical module. The control signal is analogous to the alpha motor neuron input to a muscle and is controlled by the same circuitry involved in movement, basal ganglia(BG), cerebellum(Crb), and pre-frontal cortex(PFC). The control signal instigates a competitive information processing operation and holds one symbol (shown by red neurons) active in a thalamocortical oscillation in each cortical area. While the symbol is active in each cortical area the parahippocampal region (PPR) and hippocampus(H) link the two symbols together. b) Short-term memory (STM) is a reactivation of the link between two symbols given that one symbol is active. Because of the discrete nature of symbol storage in modules the hippocampus merely has to store a single matrix $L_{Short-term}$ of associations between symbols in modules. c) Consolidated long-term memory (C-LTM) is the transfer of the same associations stored by the hippocampus into direct cortico-cortical associations. The matrix $L_{Consolidated}$ is essentially a temporally graded copy of the $L_{Short-term}$ matrix, thus the implementation of knowledge across many time scales is easily unified. d) Memory at the three time scales work seamlessly together during everyday thought and are easily unified in the present framework.

through T_I input. With these two control mechanisms, thoughts and thought processes can be learned and controlled. Certain muscles of thought will undergo confabulations which then are used to determine (through the basal ganglia, and ultimately through T_{L1} and T_I) which modules should confabulate and which modules should send out their predicted behavioral conclusions at what time. Just like physical muscle control both of these control signals are analog and can therefore control the fine state of both information processes relatively independently. Why can we dream without acting? The simple explanation would be because the T_I behavior triggering is simply shut down, but each muscle of thought could still confabulate and be coordinated through T_L1 input.

The beauty of thought arises because of the virtually infinite combinations of sequences in which muscles of thought can be confabulated and predicted behaviors launched from each module. Each module must develop (form perceptions that will be used in thought processes) and then based on those perceptions, learn sequences of thought processes (confabulations and behavioral triggering) that generate desired information processing thought outcomes. The elegance of thought need not require a complicated explanation, just as the elegance of a ballerina's dance need not require a complicated explanation. The ballerina has simply worked hard to strengthen her muscles, and learned to control them in tightly controlled sequences to obtain the elegance she demonstrates in her ballet of movement. The brain and cognition are no different. One must work hard to strengthen (form perceptions) in each of their muscles of thought, then one must learn to control their muscles of thought in tightly controlled sequences of confabulations and behavior triggering to produce elegance in the ballet of thought. The devil as they say, is in the details.

6.6 Conclusion

A significant amount of research needs to still be performed. Numerous neuroscience experiments need to be performed to validate our model. We must also expand and detail many aspects of the model, especially in the realm of the internal function of the basal ganglia. But overall we believe that sufficient detail now exists to begin designing experiments to test the model. Even if many aspects are proven wrong, our model has served a nobel purpose, for it gives a precise hypothesis at the neuron level for experimentalists to test, and that is one of the major contributions I hope to have given to the world, and to those interested in understanding the brain, as a scientist.

Chapter 6, is currently being prepared for submission for publication of the material. Solari, Soren. The dissertation author was the primary investigator and author of this material.

7

Symmetry and genomic diversity of exactly repeated DNA reverse complimentary (reco) codes

7.1 Introduction

For over half a century it has been widely assumed that the 3-base ($4^3 = 64$ element) DNA protein code is the only Earth lifeform genetic information expression mechanism(Gerstein et al., 2007). Here, we show that DNA also seems to utilize a second, very different, type of code, which we call DNA reco (reverse complementary, pronounced 'r-ee-co') codes, and that DNA reco codes are abundantly present throughout each genome. DNA reco codes are a stricter class of what has previously been termed "DNA repeat sequences"(Jurka et al., 2007; Smit, 1999). Repeat sequences are sections of DNA that appear, albeit with modifications (reco codes have no modifications), within genomes many times(Batzer and Deininger, 2002). Some basic evolutionary mechanisms potentially explaining how repeat sequences are copied seem to be largely understood(Kriegs et al., 2007; Jurka, 2004). Many DNA sequences, including repeats, were thought to be non-gene coding and potentially non-functional, and for many years researchers have viewed them as "junk DNA"(Ohno, 1972; Doolittle and Sapienza, 1980). Here we describe properties of exactly repeated DNA base sequences, and their reverse complimentary pairs, in eleven different complete genomes, which significantly strengthen the evidence that this "junk DNA" interpretation is incorrect.

The existence of repeated sequences of DNA nucleotide bases in genomes has been known

since the 1960's(Britten and Kohne, 1968). The evolutionary origin of repetitive sequences lies in a variety of proposed mechanisms, such as via reverse transcriptase, by which sections of DNA can be copied and re-inserted into the genome at a different location(Jurka et al., 2007). Certain basic sections of DNA, typified by the Alu sequence in primates(Schmid and Jelinek, 1982), are said to be more prone to copying and have been copied up to a million times in the Human genome(Batzer and Deininger, 2002). These repeated DNA sequences have traditionally been dismissed as evolutionary artifacts with no functional relevance in an organism, although this interpretation is changing in the field(Eller et al., 2007).

Research into DNA repeat sequences, in and across genomes, utilizes programs such as RepeatMasker to calculate gross sequence similarity to various DNA template sequences , as opposed to exactly matching and counting the number of identical repeats of a DNA sequence(Jurka et al., 2005). For example, several ∼300 base pair template Alu sequences (see Appendix 7.7.3), used in research to identify thousands of the "same" repeat sequence, actually never exactly occur in the Human genome(Jurka, 2000). In the three base pair protein code, the change of a single base pair can result in a different amino acid and thus create a different protein. From this basic fact we might assume that the exact sequence of bases in a DNA sequence should be extensively paid attention to, yet many sections of DNA are simply dismissed as the 'same repeat sequence' when differing by tens or hundreds of base pairs. The structure (exact base pair sequence) of DNA ultimately determines its functional properties. If a reasonably long segment of DNA (e.g. 33-36 base pairs) identically repeats tens of thousands of times in a genome, and if it has been highly conserved through evolution, we should pay more attention to it and explore its possible function. Here, we report several heretofore unknown properties of the basic structure of DNA and of these exactly repeated sequences, reco codes, based upon the application of a straightforward, but computationally demanding, unbiased DNA analytics method (whole-genome assay of exactly repeated sequences of a wide range of lengths). This new evidence supports our interpretation of these sequences as important codes, not "junk".

7.2 Reco code occurrence and critical length

There are approximately 20,000 unique DNA reco codes of length 33 in the Human genome (see Appendix 7.7.2). Figure-7.1(a) lists the six most frequently appearing of these. Note that approximately 50% of each of these reco codes' appearances is within known gene sequences(Hsu et al., 2006). Considering that gene coding regions only comprise ∼7% of the entire human genome(Lander et al., 2001), reco code positioning is strongly biased towards gene coding regions. Again, this suggests that reco codes are not random "junk DNA". Figure-7.1(b)

(a)

length	total # occurrences	DNA characteristic length reco code	# in genes	# in exons	# RC pair in same gene
36	37391	GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC	18774	329	5534
36	37153	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC	18700	326	5534
33	30419	TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT	14957	219	4441
33	30754	ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA	15251	230	4441
33	25784	GGTGAAACCCCGTCTCTACTAAAAATACAAAAA	12374	202	3751
33	25624	TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC	12393	158	3751

(b)

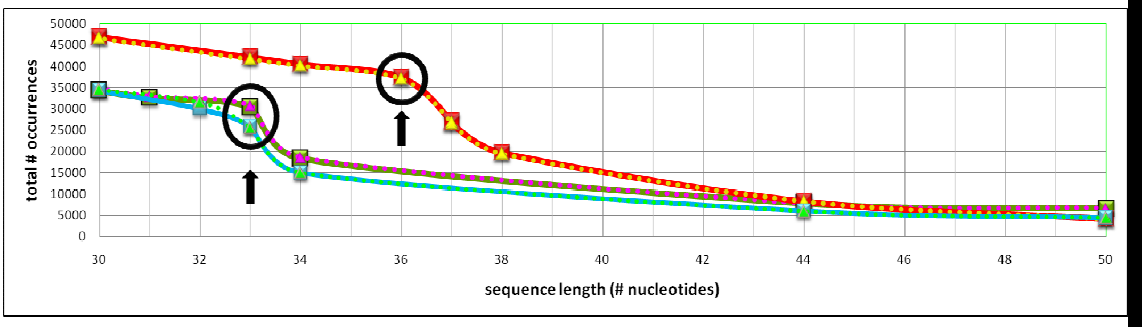

Figure 7.1: Six families of Human DNA reco codes. (a) The six most frequently occurring reco codes in the human genome (ignoring those composed entirely of microsatellites - see below), showing their characteristic length, exact number of occurrences, number of occurrences in known genes, number of occurrences within exons, and the maximum number of paired occurrences, in any single gene, of both that reco code and its reverse compliment (RC) reco code. (b) Each reco code family consists of a series of progressively longer sequences, starting with a 30-base long seed reco code, to a final reco code of length of 50 bases. Note that each family curve seems to have an abrupt knee (the top point of which - indicated by the arrows - defines that family's characteristic length). The sequences of each of these families are the reverse compliments of the sequences of exactly one other family's. Note the astoundingly high similarity of the curves for paired reverse compliment families. These curves are typical of reco code families.

describes the occurrence frequency vs. sequence length curves for six reco code families; where each longer sequence on the curve contains the addition of base pairs that maximize that new sequence's occurrence count. Clearly, as the sequence length is increased in each family, the occurrence count must monotonically decrease. Note that these family curves typically have an obvious "knee" where the occurrence count dramatically falls off with sequence length. The length at the upper point on this knee is what we define as that reco code family's characteristic length. Our preliminary investigation suggests that a having a characteristic length is a common feature of complex reco code families of all species.

We note that the six codes in Figure-7.1 would be traditionally classified as parts of Alu sequences (see Appendix 7.7.3). However, each reco code, as shown in Figure-7.1(b), belongs to a family with a clear characteristic length. The existence of a characteristic length for a sequence under evolutionary pressures would seem to imply a functional importance for that exact

sequence. We would like to point out that the term Alu sequence can be misleading, because the term essentially describes a DNA sequence that shares a common evolutionary origin (in the way it was copied) to other sequences. We should not assume that two different sequences both classified as Alu then have the same function. For example, sequences with Alu origin have already been shown to be involved in alternative splicing, but that does not imply all Alu sequences have this function(Lev-Maor et al., 2003). Additionally, as seen in Figure-7.1(b), we find an astoundingly close match between the curve of each reco code family and that of its reverse compliment family. This suggests that every time a DNA reco code is copied and reinserted into the genome, its exact reverse compliment is also copied into the same strand at a different location, or possibly the DNA reco code is copied identically into opposing strands in different locations. These facts appear to be completely new.

7.3 Reco code evolutionary genomic diversity

Another discovery we have made concerns the evolutionary differences and conservation of DNA reco codes across species. As shown in Figure-7.2, the similarity of reco codes between three Human reference genomes is high (92-96%). The similarity between Humans and Chimpanzee is fairly high (86-88%); as is the similarity between Humans (and Chimpanzee) and the Rhesus monkey (70-74%). However, the similarities between the Human reco code lists and those of: Dog, Horse, Mouse, Rat, Zebrafish, and Drosophila (2.5%, 1.4%, 1.2%, 1.1%, 0.8%, and 0.08%, respectively) are much lower. Based on a brief survey, the small percentage similarities are exclusively a result of simple reco code overlap; therefore, the complex reco code overlap for those cases in general is 0%. Rats and Mice are moderately similar (14%). These relationships between species clearly illustrates the enormous influence of evolution on DNA reco codes. Reco codes vast and persistent presence in all DNA, and divergence with species separation time, in such a wide variety of species, again strongly suggests that DNA reco codes play an essential, conserved, evolution-driven, role in animal cell biology (and probably in plants, fungi, and other kingdoms, as well). Clearly, reco codes may play a crucial role in speciation.

Our method (employing fixed-length sequences and requiring exact matching of repeats) might be criticized. For example, some would note that our lists are filled with sequences that are contained within Alus(Price et al., 2004). However, when comparing species we find high occurrence codes that are unique to each species and are not marked by current tools, such as RepeatMasker, as a known repetitive sequence. For instance, the sequence CGGACTGCG-GACTGCAGTGGCGCAATCTCGGCT occurs 904, 812, and 811 times in each human genome, but never occurs in any other genome. This particular sequence is not recognized by Repeat-

Species	Hum_Ref	Hum_Celera	Hum_AltRef	Chimp	Rhesus	Dog	Horse	Mouse	Rat	Zebrafish	Drosophila
Hum_Ref	x	96.15	96.27	88.09	74.09	2.59	1.42	1.32	1.21	0.82	0.08
Hum_Celera	92.90	x	95.50	87.01	73.23	2.52	1.40	1.24	1.15	0.79	0.08
Hum_AltRef	91.85	94.32	x	86.59	72.93	2.52	1.39	1.22	1.14	0.79	0.08
Chimp	76.84	78.56	79.16	x	70.66	2.20	1.28	1.02	0.95	0.67	0.08
Rhesus	59.27	60.64	61.15	64.80	x	2.74	1.06	1.38	1.23	0.75	0.08
Dog	0.78	0.79	0.79	0.76	1.03	x	0.94	1.22	1.10	0.63	0.08
Horse	0.33	0.34	0.34	0.34	0.31	0.74	x	0.14	0.14	0.14	0.07
Mouse	1.24	1.21	1.20	1.10	1.62	3.79	0.55	x	14.68	1.57	0.09
Rat	1.08	1.05	1.06	0.96	1.36	3.24	0.53	13.88	x	1.48	0.09
Zebrafish	0.68	0.68	0.69	0.64	0.78	1.75	0.50	1.40	1.39	x	0.08
Drosophila	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	x

Figure 7.2: Percentage matches between reco codes of different species. The reco list of 33-base repeated DNA sequences which appear at least three times on each of at least three chromosomes was determined for each species shown (see Supplementary Information). The list of each column's species was then compared with that of each row's species to determine what percentage of the row's species' list entries are listed in that column's list. These are the values shown here. The percentages are not symmetrical because, for example, the lists for different genomes are of different lengths. The three human genomes were obtained from the GenBank database, and all the others are from the UCSC Browser.

Masker(version open-3.2.7, RMLib: 20090120) as a repetitive sequence, even though it occurs hundreds of times uniquely in the Human Genome. Its reverse compliment occurs 951, 870, and 854 times respectively in each human genome. Many other such examples exist between all combinations of species. And that is just the point. Twenty years of study of repeat sequences, such as the Alu, almost entirely from a "junk DNA" perspective, has not yielded the specific facts presented here. We suggest a more rigorous approach to analyzing and quantifying the comparisons between exact sequence matches and their reverse compliments should become an added standard practice in genomics.

7.4 Symmetry of Reco codes in all DNA

While analyzing the occurrences of reco codes and their reverse compliments in single genomes we noticed yet another astonishing fact regarding the overall structure of DNA, which seems to have been previously unnoticed. Namely, as illustrated in Figure-7.3, roughly 81% of the reco codes of each species of Figure-7.2 have their reverse compliments on the list as well. Appendix 7.7.2 presents examples of the astounding reverse complimentary pairing for each species. Surprisingly, the fact that every time a DNA sequence is copied and reinserted into the

genome, it seems it's exact reverse compliment is also copied into the same strand at a different location, could not be found stated anywhere as a biological fact. The average percentage difference in the occurrence counts between paired reverse complementary sequences is about 8% (with no obvious consistent bias towards one sequence or its pair). The percentage is slightly distorted because lower count codes will have a higher percentage difference for the same count difference. For example, from Figure-7.1 we find that the 33-base DNA reco code ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA appears 30754 times in the Human genome, and it's reverse compliment sequence, TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT, appears 30419 times. The average occurrence count deviation for this sequence pair across all Human chromosomes is 3.32% (the pair's occurrence counts are not consistently biased), and the maximum count deviation (on chromosome 17) is 7.84%. The minimum deviation (on chromosome 8) has occurrence counts of 1152 and 1156 respectively. Additional examples can be found in Appendix 7.7.2. Although basic reverse complimentary pairing has not gone unnoticed (Ohno and Yomo, 1991; Wang and Leung, 2006; Stenger et al., 2001), the surprisingly constant percentages of Figure-7.3 strongly suggest that reverse complimentary pairing has a specific function that has been preserved across a good fraction of a billion years of evolution.

Species	% Rev. Comp.	Avg. % count difference
Hum_Ref	81.29	8.11
Hum_Celera	81.28	7.96
Hum_AltRef	81.36	8.02
Chimp	80.96	8.22
Rhesus	80.99	8.38
Dog	81.48	8.39
Horse	83.63	8.73
Mouse	81.80	8.08
Rat	80.30	7.73
Zebrafish	78.82	10.84
Drosophila	85.53	14.07

Figure 7.3: Here, for the genomes of Figure-7.2, are the percentages of the 33-base-length reco codes, of each species' list, which also have their reverse compliment on the list. For the totality of these reverse compliment sequence pairs, the average genome-wide occurrence count difference (expressed as an averaged percentage difference of each pair's smaller count sequence relative to that pair's larger count sequence) is also shown. Note the astounding stability of these numbers across species whose last contacts range from a few million to half a billion years. Again, this strongly suggests that reco codes are of critical functional importance.

Finally, an additional example is presented to illustrate the scope of the conservation

of reverse complementary pairs. Many sequences classified as DNA reco codes are composed of microsatellites (1-6 base pair repeating units). We use the term simple DNA reco code for any reco code comprised entirely of microsatellite sequences, and complex DNA reco codes for all others. Simple DNA reco codes are much more common in lower species than in humans. Only 12 of the top 100 most frequent DNA reco codes in the human are simple as compared to 95 of the top 100 in the mouse. Surprisingly, in all species, even the simple DNA reco codes maintain the same reverse complementary pairing. For example, in the mouse, the length 33 DNA reco code consisting exclusively of AAAG repeats occurs 97,198 times and the reverse complementary reco code consisting of CTTT repeats occurs 96,412 times. Similarly, the DNA reco code consisting of TATC repeats occurs 58,194 times and its paired reverse complementary reco code with GATA repeats occurs 57,617 times. Even though the reverse complimentary pairs may on average be separated by tens of thousands of base pairs (often including many other reco codes), the occurrence counts of reverse complementary pairs are consistently similar across the entire genome and within chromosomes. The symmetry of reverse complimentary DNA sequences described here appears to be a fundamentally conserved and distinct feature in the structure of all DNA.

7.5 Discussion

Since, as with politics, all chemistry is local, the exact conservation of these repeated sequences in and across species argues very strongly that they are not junk. They are likely important functional elements of DNA; which makes our calling them codes a reasonable decision. The additional facts that these codes make up a much higher fraction of DNA than proteins, and that they are conserved across similar species, strongly supports the concept that reco codes are important. The count balance of reco codes and their reverse compliments suggests that these codes are used separately on both strands of DNA, but always with the opposite starting orientation.

The exact function of reco codes is yet to be determined, and will likely require extensive and creative experimental paradigms to unravel. However, we consider four possible functional roles for reco codes. First, the full mechanisms underlying gene transcription are far from understood(Kornberg, 2007). Reco codes could be used extensively in the mediation of transcriptional regulation (Kornberg, 2005; Jurka, 2008). Reco codes are capable of providing extremely accurate and precise locations for interactions with RNA/protein molecules. Their symmetric existence in and out of genes, in conjunction with their reverse complimentary pairs, potentially enable them to act as DNA sequence elements which help regulate the rate of progression of gene expression.

Second, DNA reco codes may act as markers during meiosis for cross-over. Third, the genetic similarity alone between species (e.g. 80% match of genes, 40% nucleotide alignment match between mouse and human)(Waterston et al., 2002), does not seem to fit the large differences between species. This suggests that reco codes may provide a more accurate indication of speciation. For example, 14% reco code similarity between mice and rats might better explain the severe difficulties in combining their DNA (Waksmundzka, 1994). If so, reco codes would suggest that the DNA of humans and other primates would be more easily combined than mice and rats. Finally, DNA chromatin is a complex 3-dimensionally folded structure, the structure of which is important in gene regulation(Wolffe, 1998). In chromatin, when distant sequences are brought into close proximity, the symmetry of repeat sequences and their reverse complimentary pairs may be significant for either/both the folding of chromatin and/or identifications of certain sections to simultaneously unfold for gene expression.

Given the fifty five years which have passed since the Watson and Crick discovery (Watson and Crick, 1953), one would expect that in 2008 every major hospital on Earth could create, in a few weeks, a zero-age replacement heart from a patient skin scraping. Yet today's scientific and medical knowledge of DNA lies vastly short of this. The new facts presented here strongly suggest that DNA reco codes are not "junk", but instead represent critically important, and strongly exactly conserved, DNA structures worthy of greatly intensified investigation.

7.6 Methods

We analyzed the entire genome of three complete human sequences, as well as the entire genome for the chimpanzee, rhesus monkey, dog, mouse, rat, horse, zebrafish, and drosophila (see 7.7.2 for source genomes used). For each entire genome (chromosome Y was ignored for all genomes), we extracted all contiguous DNA sequences of a fixed length L. In the human genome, with approximately 3 billion base pairs, this equates to analyzing approximately 3 billion sequences of length L. We analyzed lengths L ranging from 30 to 400 for all species. Any nucleotide sequence that repeated identically (all L nucleotides 100% the same) more than 3 times in a single chromosome and in at least 3 chromosomes we termed a reco code, and added it to that species' list. The collection of all reco codes of a length L for a particular genome formed the reco code list. For a random DNA sequence, the probability of a DNA sequence, L=30, being classified as a reco code is much less than 10-20. We are therefore guaranteed that all sequences termed reco codes are extremely unlikely to be randomly occurring. For a list of each species see Appendix 7.7.2.

7.7 Appendix

7.7.1 List of Reco Code families displayed in Figure-7.1

Figure-7.4 shows the six families of Human DNA reco codes from which the graph in Figure-7.1 was created. The occurrence counts for each sequence are the number of times that particular exact sequence appears in the entire Human genome.

7.7.2 Estimated unique reco codes for each genome

Viewing the actual reco code lists and the occurrences for each reco code in each species provides enormous insight into the structure of the codes. In addition, researchers may use these lists to formulate practical experiments. We selected length L=33 lists because of the prevalence of four of the top six codes in humans having this characteristic length. We give a modified "unique", as described below, reco list for all species at L=33 for direct comparison. Note the species comparison from Figure-7.2 was performed on the original lists, not the "unique" lists. Our original reco code lists are a reflection of all identically repeating sequences in a given genome. One main issue arises when creating an unbiased reco code list with a fixed length L. The list will contain any and all repeated sequences. As a result, if there exists a DNA sequence with a length longer than L, which repeats, then all sub-sequences of length L from that sequence will be found on the list. Therefore, two reco codes on this list may possibly be overlapped with each other in a longer DNA sequence. In order to provide a rough estimate of the number of unique reco codes that may exist at a given length we perform a crude alignment analysis of the original reco code list. Every reco code was checked against all reco codes which had a higher occurrence. If the reco code matched a contiguous sequence of approximately 50% of its base pairs, then we considered the codes overlapped and dropped the less frequent code from the list. This was done for all reco codes and results in a "unique" reco code list. We only present a selection (10 most frequent) of the unique reco codes for each species for the reader to get a feel for "what is going on". Even with only showing the top 10, the tight pairing of reco codes is clearly demonstrated. The full files have been submitted as supplementary online material for publication. The source of genomic data is listed for each species. The columns are explained as follows:

- column 1: total occurrence number in the genome
- column 2: "unique" reco code corresponding to column 1
- column 3: reverse compliment of reco code in column 2 for easy comparison - shown in red

name	length	count	DNA reco code
GCCT_33	50	6708	GGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGG
	44	7851	GCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGC
	34	18599	ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGAC
	33	30754	ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA
	31	32959	ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGG
	30	34700	ATTCTCCTGCCTCAGCCTCCCGAGTAGCTG
AGGC_33	50	6644	CCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACC
	44	7854	GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGC
	34	18354	GTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
	33	30419	TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
	31	32590	CCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
	30	34325	CAGCTACTCGGGAGGCTGAGGCAGGAGAAT
GGAT_36	50	4194	CTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGC
	44	8274	CCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC
	38	19719	GGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACC
	37	26663	GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACC
	37	27111	GGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC
	36	37391	GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC
	34	40498	CCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
	33	42127	CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
	30	46872	CCTCCCAAAGTGCTGGGATTACAGGCGTGA
ATCC_36	50	4251	GCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAG
	44	8366	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGG
	38	19524	GGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCC
	37	26547	GGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
	37	26950	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCC
	36	37153	GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
	34	40183	TGGCTCACGCCTGTAATCCCAGCACTTTGGGAGG
	33	41726	TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
	44	6085	CCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA
	34	15001	ATGGTGAAACCCCGTCTCTACTAAAAATACAAAA
	33	25784	GGTGAAACCCCGTCTCTACTAAAAATACAAAAA
	32	29991	GTGAAACCCCGTCTCTACTAAAAATACAAAAA
	30	34297	GAAACCCCGTCTCTACTAAAAATACAAAAA
TAGA_33	50	4498	TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGT
	44	6050	TTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCT
	34	14998	TTTTTTGTATTTTTAGTAGAGACGGGGTTTCACC
	33	25624	TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC
	32	31584	TTTTTGTATTTTTAGTAGAGACGGGGTTTCA
	30	34534	TTTTTGTATTTTTAGTAGAGACGGGGTTTC

Figure 7.4: The six families of Human DNA reco codes from which the graph in Figure-7.1 was created. Each code shown represents the sequence which maximizes the occurrence count.

Figure 7.5: Top ten highest occurrence DNA unique reco codes examples for each genome. Subparts of the figure (a)-(k) are continued below.

(Hum_Ref)GenBank Reference Human Genome NCBI Build 36 v3, 24 March 2008.

```
72898 TGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACA
71871 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGT
47321 TATATATATATATATATATATATATATATAT ATATATATATATATATATATATATATATATA
42127 CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
41726 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
30754 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
30419 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA
26519 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
25895 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
25784 GGTGAAACCCCGTCTCTACTAAAAATACAAAAA TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC
```

Figure 7.5(a)

(Hum_altRef)GenBank Human Alternate Reference (shotgun sequence) genome NCBI Build 36 v3, 24 March 2008.

```
70544 TGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACA
69122 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGT
45179 TATATATATATATATATATATATATATATAT ATATATATATATATATATATATATATATATA
39828 CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
39483 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
29090 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
28762 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA
24320 GGTGAAACCCCGTCTCTACTAAAAATACAAAAA TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC
24234 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC GGTGAAACCCCGTCTCTACTAAAAATACAAAAA
19661 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
```

Figure 7.5(b)

(Hum_Celera)GenBank Human Celera (shotgun sequence) genome NCBI Build 36 v3, 24 March 2008.

```
67221 TGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACA
66126 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGT
41129 TATATATATATATATATATATATATATATAT ATATATATATATATATATATATATATATATA
40610 CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
40261 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
29507 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
```

29325 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA

24842 GGTGAAACCCCGTCTCTACTAAAAATACAAAAA TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC

24584 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC GGTGAAACCCCGTCTCTACTAAAAATACAAAAA

18991 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA

Figure 7.5(c)

(Chimp)UCSC Browser Chimpanzee genome (panTro2) Mar. 2006.

42602 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

42408 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

35432 CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG

34955 TGGCTCACGCCTGTAATCCCAGCACTTTGGGAG CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA

27543 ATATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATAT

26543 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT

26069 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA

23436 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

22853 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

20511 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACC GGTGAAACCCCGTCTCTACTAAAAATACAAAAA

Figure 7.5(d)

(Rhesus)UCSC Browser Rhesus genome (rheMac2) Jan. 2006.

57560 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

57068 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

41612 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

40553 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

38854 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT

37837 TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA

37366 ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT

37066 TCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT ATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGA

31726 CTCCCAAAGTGCTGGGATTACAGGCTTGAGCCA TGGCTCAAGCCTGTAATCCCAGCACTTTGGGAG

31473 TGGCTCAAGCCTGTAATCCCAGCACTTTGGGAG CTCCCAAAGTGCTGGGATTACAGGCTTGAGCCA

Figure 7.5(e)

(Dog)UCSC Browser Dog genome (canFam2) May 2005.

93395 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT

90198 TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA

76567 GGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGC GCAGAGACACAGGCAGAGGGAGAAGCAGGCTCC

76545 GCAGAGACACAGGCAGAGGGAGAAGCAGGCTCC GGAGCCTGCTTCTCCCTCTGCCTGTGTCTCTGC

56426 ATAAATAAATAAATAAATAAATAAATAAATAAA TTTATTTATTTATTTATTTATTTATTTATTTAT

55772 TTTATTTATTTATTTATTTATTTATTTATTTAT ATAAATAAATAAATAAATAAATAAATAAATAAA

52987 ATATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATATAT

38256 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

37977 ACACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

28396 GTCTCCAGGATCGCGCCCTGGGCCAAAGGCAGG CCTGCCTTTGGCCCAGGGCGCGATCCTGGAGAC

Figure 7.5(f)

(Mouse)UCSC Browser Mouse genome (mm9) July 2007.

460784 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

459239 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

362303 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

359938 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

147148 TATATATATATATATATATATATATATATATAT ATATATATATATATATATATATATATATATATA

97198 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT

96412 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA

92724 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC

90868 TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA

58194 TCTATCTATCTATCTATCTATCTATCTATCTAT ATAGATAGATAGATAGATAGATAGATAGATAGA

Figure 7.5(g)

(Rat)UCSC Browser Rat genome (rn4) Nov. 2004.

638462 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

635651 ACACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

354705 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

352818 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

155493 ATATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATATAT

49407 TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA

49136 ATAGATAGATAGATAGATAGATAGATAGATAGA TCTATCTATCTATCTATCTATCTATCTATCTAT

48965 GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC

47575 TCTATCTATCTATCTATCTATCTATCTATCTAT ATAGATAGATAGATAGATAGATAGATAGATAGA

26727 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT

Figure 7.5(h)

(Horse)UCSC Browser Horse genome (equCab1) Jan. 2007.

11639 TTTTGCTGAGGAAGACTGGCCCTGAGCTAACAT ATGTTAGCTCAGGGCCAGTCTTCCTCAGCAAAA

11516 ATGTTAGCTCAGGGCCAGTCTTCCTCAGCAAAA TTTTGCTGAGGAAGACTGGCCCTGAGCTAACAT

10637 ATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATAT

8153 TTTTTTTTTGAGGAAGATTAGCCCTGAGCTAAC GTTAGCTCAGGGCTAATCTTCCTCAAAAAAAAA

8031 GTTAGCTCAGGGCTAATCTTCCTCAAAAAAAAA TTTTTTTTTGAGGAAGATTAGCCCTGAGCTAAC

7742 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

7660 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

6084 CTGAGGCAGCGTCCCACATGCCACAACTAGAAG CTTCTAGTTGTGGCATGTGGGACGCTGCCTCAG

6050 CTTCTAGTTGTGGCATGTGGGACGCTGCCTCAG CTGAGGCAGCGTCCCACATGCCACAACTAGAAG

5927 CTGCTGCCAATCCTCCTCTTTTTGCTGAGGAAG CTTCCTCAGCAAAAAGAGGAGGATTGGCAGCAG

Figure 7.5(i)

(Zebrafish)UCSC Browser Zebrafish genome (danRer5) July 2007.

2350663 ATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATAT

349161 ACACACACACACACACACACACACACACACA TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

342805 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT ACACACACACACACACACACACACACACACACA

176497 ATAGATAGATAGATAGATAGATAGATAGATAGA TCTATCTATCTATCTATCTATCTATCTATCTAT

173382 TCTATCTATCTATCTATCTATCTATCTATCTAT ATAGATAGATAGATAGATAGATAGATAGATAGA

51756 TTATTATTATTATTATTATTATTATTATTA TAATAATAATAATAATAATAATAATAATAATAA

51390 TAATAATAATAATAATAATAATAATAATAA TTATTATTATTATTATTATTATTATTATTA

38337 TCTCTCTCTCTCTCTCTCTCTCTCTCTCT AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA

37348 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT

34119 ATGGATGGATGGATGGATGGATGGATGGA TCCATCCATCCATCCATCCATCCATCCATCCAT

Figure 7.5(j)

(Drosophila)UCSC Browser Drosophila Melanogaster genome (dm3) Apr. 2006.

658 AATATAATATAATATAATATAATATAATATAAT ATTATATTATATTATATTATATTATATTATATT

617 ATATATATATATATATATATATATATATATA TATATATATATATATATATATATATATATATAT

438 TATTATTATTATTATTATTATTATTATTAT ATAATAATAATAATAATAATAATAATAATAATA

380 TAATAATAATAATAATAATAATAATAATAA TTATTATTATTATTATTATTATTATTATTA

317 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

258 AAATAATTGTCTGAATATGGAATGTCATACCTC GAGGTATGACATTCCATATTCAGACAATTATTT

252 CACACACACACACACACACACACACACACAC GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG

245 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

236 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG CACACACACACACACACACACACACACACAC

200 ATAGATAGATAGATAGATAGATAGATAGATAGA TCTATCTATCTATCTATCTATCTATCTAT

Figure 7.5(k)

7.7.3 Location of top six reco codes within a template Human Alu sequence.

The following 300 base pair sequence is one example of an Alu template sequence listed in RepBase Update. Although used to identify Alu sequences, this exact 300 base pair sequence is never found anywhere in the Human Genome. The highlighted regions correspond to three of the six most frequent DNA reco code sequences. Notice the other three reco codes in Figure-7.1 are simply the reverse compliments of these sequences and therefore correspond to the reverse compliment of this entire 300 base pair sequence.

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGA
GGCGGGCGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGG
TGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCG
CGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTG
AACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTGCACTCC
AGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAA

Figure 7.6: Example Alu sequence with reco codes.

Chapter 7, in part, has been submitted for publication of the material and is in the editorial review processes. Solari, S., Tsigelny, I., Sharikov, Y., Hecht-Nielsen, R. The dissertation author was the primary investigator and author of this material.

References

Abeles, M. *Corticonics: neural circuits of the cerebral cortex*. Cambridge University Press, Cambridge, 1991.

Alonso-Nanclares, L., Anderson, S., Ascoli, G., Benavides-Piccione, R., Burkhalter, A., Buzsaki, G., Cauli, B., DeFelipe, J., Fairen, A., Feldmeyer, D., Fishel, G., Fregnac, Y., Freund, T. F., Fukuyi, K., Glarreta, M., Goldberg, J., Helmstaedter, M., Hensch, T., Hestrin, S., Kisvarday, Z., Lambolez, B., Lewis, D., McBain, C., Marin, O., Markham, H., Monyer, H., Muoz, A., Petersen, C., Rockland, K., Rossier, H., Ruby, B., Somogyi, P., Staiger, J. F., Tamas, G., Thomason, A., Toledo-Rodriguez, M., Wang, X.-J., Wang, Y., West, D., and Yuste, R. Petilla 2005: Nomenclature of features of gabaergic interneurons of the cerebral cortex, 2005.

Amari, S.-I. Characteristics of sparsely encoded associative memory. *Neural Networks*, **2**, 451–457, 1989.

Amitai, Y., Gibson, J. R., Beierlein, M., Patrick, S. L., Ho, A. M., Connors, B. W., and Golomb, D. The spatial dimensions of electrically coupled networks of interneurons in the neocortex. *J Neurosci*, **22**(10), 4142–52, 2002.

Asanuma, C. Gabaergic and pallidal terminals in the thalamic reticular nucleus of squirrel monkeys. *Exp Brain Res*, **101**(3), 439–51, 1994.

Asanuma, C., Andersen, R. A., and Cowan, W. M. The thalamic relations of the caudal inferior parietal lobule and the lateral prefrontal cortex in monkeys: divergent cortical projections from cell clusters in the medial pulvinar nucleus. *J Comp Neurol*, **241**(3), 357–81, 1985.

Baddeley, A. The concept of working memory: a view of its current state and probable future development. *Cognition*, **10**(1-3), 17–23, 1981.

Baddeley, A. Working memory and language: an overview. *J Commun Disord*, **36**(3), 189–208, 2003.

Bannister, P. A. Inter- and intra-laminar connections of pyramidal cells in the neocortex. *Neuroscience Research*, **53**, 95–103, 2005.

Barbas, H. Pattern in the laminar origin of corticocortical connections. *J Comp Neurol*, **252**(3), 415–22, 1986.

Barbas, H., and Hilgetag, C. C. Rules relating connections to cortical structure in primate prefrontal cortex. *Neurocomputing*, **44-46**, 301–308, 2002.

Barbas, H., Hilgetag, C. C., Saha, S., Dermon, C. R., and Suski, J. L. Parallel organization of contralateral and ipsilateral prefrontal cortical projections in the rhesus monkey. *BMC Neurosci*, **6**(1), 32, 2005a.

Barbas, H., Medalla, M., Alade, O., Suski, J., Zikopoulos, B., and Lera, P. Relationship of prefrontal connections to inhibitory systems in superior temporal areas in the rhesus monkey. *Cereb Cortex*, **15**(9), 1356–70, 2005b.

Barbas, H., and Rempel-Clower, N. Cortical structure predicts the pattern of corticocortical connections. *Cereb Cortex*, **7**(7), 635–46, 1997.

Batzer, M. A., and Deininger, P. L. Alu repeats and human genomic diversity. *Nature Reviews. Genetics*, **3**(5), 370–9, 2002.

Berg, R. W., Alaburda, A., and Hounsgaard, J. Balanced inhibition and excitation drive spike activity in spinal half-centers. *Science*, **315**(5810), 390–3, 2007.

Bigl, V., Woolf, N. J., and Butcher, L. L. Cholinergic projections from the basal forebrain to frontal, parietal, temporal, occipital, and cingulate cortices: a combined fluorescent tracer and acetylcholinesterase analysis. *Brain Research Bulletin*, **8**(6), 727–49, 1982.

Borgmann, S., and Jurgens, U. Lack of cortico-striatal projections from the primary auditory cortex in the squirrel monkey. *Brain Res*, **836**(1-2), 225–8, 1999.

Braak, H. *Architectonics of the Human Telencephalic Cortex*, volume 4 of *Studies of Brain Function*. Springer-Verlag, Berlin, 1980.

Braitenberg, V., and Shuz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, Berlin, Germany, 2nd edition, 1998.

Bravo, H., and Karten, H. J. Pyramidal neurons of the rat cerebral cortex, immunoreactive to nicotinic acetylcholine receptors, project mainly to subcortical targets. *Journal of Comparative Neurology*, **320**(1), 62–8, 1992.

Britten, R. J., and Kohne, D. E. Repeated sequences in dna. hundreds of thousands of copies of dna sequences have been incorporated into the genomes of higher organisms. *Science*, **161**(841), 529–40, 1968.

Broca, P. P. Loss of speech, chronic softening and partial destruction of the left anterior lobe of the brain. *Nepal Journal of Neuroscience*, **3**(2), 96–101, 2006.

Brodmann, K. *Brodmann's 'Localisation in the Cerebral Cortex'*. Smith-Gordon, London, 1909.

Brunel, N., and Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci*, **11**(1), 63–85, 2001.

Bruno, R. M., and Sakmann, B. Cortex is driven by weak but synchronously active thalamocortical synapses. *Science*, **312**(5780), 1622–7, 2006.

Brysch, I., Brysch, W., Creutzfeldt, O., Hayes, N. L., and Schlingensiepen, K. H. The second, intralaminar thalamo-cortical projection system. *Anat Embryol (Berl)*, **169**(2), 111–8, 1984.

Brysch, W., Brysch, I., Creutzfeldt, O. D., Schlingensiepen, R., and Schlingensiepen, K. H. The topology of the thalamo-cortical projections in the marmoset monkey (callithrix jacchus). *Exp Brain Res*, **81**(1), 1–17, 1990.

Buckingham, J., and Willshaw, D. On setting unit thresholds in an incompletely connected associative net. *Network*, **4**(4), 441–459, 1993.

Buckingham, J., and Willshaw, D. J. Performance characteristics of the associative net. *Network*, **3**(4), 407–414, 1992.

Burwell, R. D. The parahippocampal region: corticocortico connectivity. *Annals of the New York Academy of Sciences*, **911**, 25–42, 2000.

Buzsaki, G. *Rhythms of the brain*. Oxford, New york, 2006.

Cajal, S. R. y. *Texture of the nervous system of man and the vertebrates*, volume 3. Springer-Verlag, New Nork, 2002.

Callaway, E. M. Structure and function of parallel pathways in the primate early visual system. *Journal of Physiology*, **566**(1), 13–19, 2005.

Catani, M., and ffytche, D. H. The rises and falls of disconnection syndromes. *Brain*, **128**, 2224–39, 2005.

Catsman-Berrevoets, C. E., and Kuypers, H. G. Differential laminar distribution of corticothalamic neurons projecting to the vl and the center median. an hrp study in the cynomolgus monkey. *Brain Res*, **154**(2), 359–65, 1978.

Clarke, E., and Dewhurst, K. *An Illustrated History of Brain Function*. Stanford Publications, Oxford, 1972.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex*, **10**(9), 910–23, 2000.

Conner, J. M., Culberson, A., Packowski, C., Chiba, A. A., and Tuszynski, M. H. Lesions of the basal forebrain cholinergic system impair task acquisition and abolish cortical plasticity associated with motor skill learning. *Neuron*, **38**(5), 819–29, 2003.

Craig, A. M., and Lichtman, J. W. Synapse formation and maturation. In *Synapses*, editors W. M. Cowan, T. C. Sudhof, and C. F. Stevens. The Johns Hopkins University Press, Baltimore, 2001.

Crick, F. C., and Koch, C. What is the function of the claustrum? *Philos Trans R Soc Lond B Biol Sci*, **360**(1458), 1271–9, 2005.

de Lima, A. D., Voigt, T., and Morrison, J. H. Morphology of the cells within the inferior temporal gyrus that project to the prefrontal cortex in the macaque monkey. *J Comp Neurol*, **296**(1), 159–72, 1990.

Defelipe, J., Gonzalez-Albo, M. C., Del Rio, M. R., and Elston, G. N. Distribution and patterns of connectivity of interneurons containing calbindin, calretinin, and parvalbumin in visual areas of the occipital and temporal lobes of the macaque monkey. *J Comp Neurol*, **412**(3), 515–26, 1999.

Desimone, R., Albright, T., Bruce, C. G., and C. Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, **4**, 2051–2062, 1984.

Doolittle, W. F., and Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**(5757), 601–3, 1980.

Douglas, R. J., and Martin, K. A. Neuronal circuits of the neocortex. *Annu Rev Neurosci*, **27**, 419–51, 2004.

Durstewitz, D., Seamans, J. K., and Sejnowski, T. J. Neurocomputational models of working memory. *Nat Neurosci*, **3 Suppl**, 1184–91, 2000.

Edelstein, L. R., and Denaro, F. J. The claustrum: a historical review of its anatomy, physiology, cytochemistry and functional significance. *Cell Mol Biol (Noisy-le-grand)*, **50**(6), 675–702, 2004.

Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nature Reviews. Neuroscience*, **1**(1), 41–50, 2000.

Eller, C. D., Regelson, M., Merriman, B., Nelson, S., Horvath, S., and Marahrens, Y. Repetitive sequence environment distinguishes housekeeping genes. *Gene*, **390**(1-2), 153–65, 2007.

Felleman, D. J., and Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, **1**(1), 1–47, 1991.

Finger, S. *Origins of Neuroscience: A History of Explorations into Brain Function.* Oxford University Press, New York, 1994.

Fujita, I., and Fujita, T. Intrinsic connections in the macaque inferior temporal cortex. *J Comp Neurol*, **368**(4), 467–86, 1996.

Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**, 193–202, 1980.

Fuster, J. M. *Cortex and Mind.* Oxford University Press, New York, 2003.

Gabbott, P. L., and Bacon, S. J. Local circuit neurons in the medial prefrontal cortex (areas 24a,b,c, 25 and 32) in the monkey: I. cell morphology and morphometrics. *J Comp Neurol*, **364**(4), 567–608, 1996.

Gandia, J. A., De Las Heras, S., Garcia, M., and Gimenez-Amaya, J. M. Afferent projections to the reticular thalamic nucleus from the globus pallidus and the substantia nigra in the rat. *Brain Res Bull*, **32**(4), 351–8, 1993.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S., and Snyder, M. What is a gene, post-encode? history and updated definition. *Genome Research*, **17**(6), 669–81, 2007.

Geschwind, N. Disconnexion syndromes in animals and man. ii. *Brain*, **88**(3), 585–644, 1965.

Gibson, J. R., Beierlein, M., and Connors, B. W. Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, **402**(6757), 75–9, 1999.

Goldman-Rakic, P. S. Regional and cellular fractionation of working memory. *Proc Natl Acad Sci U S A*, **93**(24), 13473–80, 1996.

Grafton, S. T., Arbib, M. A., Fadiga, L., and Rizzolatti, G. Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research*, **112**(1), 1996.

Graham, B., and Willshaw, D. Improving recall from an associative memory. *Biological Cybernetics*, **72**(4), 337–346, 1995.

Graham, J., R. C., and Karnovsky, M. J. The early stages of absorption of injected horseradish peroxidase in the proximal tubules of mouse kidney: ultrastructural cytochemistry by a new technique. *Journal of Histochemistry and Cytochemistry*, **14**(4), 291–302, 1966.

Gupta, A., Wang, Y., and Markram, H. Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science*, **287**, 273–278, 2000.

Haber, S. N. The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat*, **26**(4), 317–30, 2003.

Hackett, T. A., Preuss, T. M., and Kaas, J. H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative Neurology*, **441**(3), 197–222, 2001.

Haines, K., and Hecht-Nielsen, R. A bam with increased information storage capacity. In *Proceedings, 1988 International Conference on Neural Networks*, 181190, Piscataway, NJ, 1988. IEEE Press.

Hanbery, J., and Jasper, H. Independence of diffuse thalamo-cortical projection system shown by specific nuclear destructions. *J Neurophysiol*, **16**(3), 252–71, 1953.

Hanbery, J., and Jasper, H. The non-specific thalamocortical projection system. *J Neurosurg*, **11**(1), 24–5, 1954.

Hazrati, L. N., and Parent, A. Projection from the external pallidum to the reticular thalamic nucleus in the squirrel monkey. *Brain Res*, **550**(1), 142–6, 1991.

Hebb, D. O. *The Organization of Behavior: A Neuropsychological Theory*. Lawrence Erlbaum Associates, New Jersey, 1949.

Hecht-Nielsen, R. Cogent confabulation. *Neural Networks*, **18**(2), 111–5, 2005.

Hecht-Nielsen, R. *Confabulation Theory*. Springer, New York, 2007.

Herkenham, M. The afferent and efferent connections of the ventromedial thalamic nucleus in the rat. *J Comp Neurol*, **183**(3), 487–517, 1979.

Herkenham, M. Laminar organization of thalamic projections to the rat neocortex. *Science*, **207**(4430), 532–5, 1980.

Herrero, M. T., Barcia, C., and Navarro, J. M. Functional anatomy of thalamus and basal ganglia. *Childs Nerv Syst*, **18**(8), 386–404, 2002.

Hestrin, S., and Galarreta, M. Electrical synapses define networks of neocortical gabaergic neurons. *Trends in Neurosciences*, **28**(6), 304–309, 2005.

Hodgkin, A. L., and Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol*, **117**(4), 500–544, 1952.

Hohl-Abrahao, J. C., and Creutzfeldt, O. D. Topographical mapping of the thalamocortical projections in rodents and comparison with that in primates. *Exp Brain Res*, **87**(2), 283–94, 1991.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *PNAS*, **79**(8), 2554–2558, 1982.

Hsu, F., Kent, W. J., Clawson, H., Kuhn, R. M., Diekhans, M., and Haussler, D. The ucsc known genes. *Bioinformatics*, **22**(9), 1036–46, 2006.

Izhikevich, E. M. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, **14**(6), 1569–1572, 2003.

Izhikevich, E. M. *Dynamical Systems in Neuroscience*. MIT Press, Cambridge, 2007.

Izhikevich, E. M., and Edelman, G. M. Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(9), 3593–8, 2008.

Jacobson, S., and Trojanowski, J. Q. Prefrontal granular cortex of the rhesus monkey. i. intrahemispheric cortical afferents. *Brain Res*, **132**(2), 209–33, 1977.

Jones, E. G. Viewpoint: the core and matrix of thalamic organization. *Neuroscience*, **85**(2), 331–45, 1998.

Jones, E. G. *The Thalamus*, volume 1-2. Cambridge University Press, Cambridge, second edition, 2007.

Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics*, **16**(9), 418–20, 2000.

Jurka, J. Evolutionary impact of human alu repetitive elements. *Current Opinion in Genetics and Development*, **14**(6), 603–8, 2004.

Jurka, J. Conserved eukaryotic transposable elements and the evolution of gene regulation. *Cellular and Molecular Life Sciences*, **65**(2), 201–4, 2008.

Jurka, J., Kapitonov, V., and Smit, A. Repetitive elements: Detection, 2005.

Jurka, J., Kapitonov, V. V., Kohany, O., and Jurka, M. V. Repetitive sequences in complex genomes: structure and evolution. *Annual Review of Genomics and Human Genetics*, **8**, 241–59, 2007.

Kakei, S., Na, J., and Shinoda, Y. Thalamic terminal morphology and distribution of single corticothalamic axons originating from layers 5 and 6 of the cat motor cortex. *J Comp Neurol*, **437**(2), 170–85, 2001.

Kaneko, T., Cho, R., Li, Y., Nomura, S., and Mizuno, N. Predominant information transfer from layer iii pyramidal neurons to corticospinal neurons. *J Comp Neurol*, **423**(1), 52–65, 2000.

Katz, L. C. Local circuitry of identified projection neurons in cat visual cortex brain slices. *J Neurosci*, **7**(4), 1223–49, 1987.

Kemp, J. M., and Powell, T. P. The cortico-striate projection in the monkey. *Brain*, **93**(3), 525–46, 1970.

Kohonen, T. *Self-Organization and Associative Memory*. Springer-Verlag, Berlin, third edition, 1989.

Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences*, **30**(5), 235–9, 2005.

Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(32), 12955–61, 2007.

Kowianski, P., Dziewiatkowski, J., Berdel, B., Lipowska, M., and Morys, J. The corticoclaustral connections in the rat studied by means of the fluorescent retrograde axonal transport method. *Folia Morphol (Warsz)*, **57**(2), 85–92, 1998.

Kriegs, J. O., Churakov, G., Jurka, J., Brosius, J., and Schmitz, J. Evolutionary history of 7sl rna-derived sines in supraprimates. *Trends in Genetics*, **23**(4), 158–61, 2007.

Kritzer, M. F., and Goldman-Rakic, P. S. Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J Comp Neurol*, **359**(1), 131–43, 1995.

Kunishio, K., and Haber, S. N. Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. *J Comp Neurol*, **350**(3), 337–56, 1994.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. Initial sequencing and analysis of the human genome. *Nature*, **409**(6822), 860–921, 2001.

Lavenex, P., Suzuki, W. A., and Amaral, D. G. Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *Journal of Comparative Neurology*, **447**(4), 394–420, 2002.

Letinic, K., Zoncu, R., and Rakic, P. Origin of gabaergic neurons in the human neocortex. *Nature*, **417**(6889), 645–9, 2002.

Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. The birth of an alternatively spliced exon: 3' splice-site selection in alu exons. *Science*, **300**(5623), 1288–91, 2003.

LeVay, S. Synaptic organization of claustral and geniculate afferents to the visual cortex of the cat. *J Neurosci*, **6**(12), 3564–75, 1986.

LeVay, S., and Sherk, H. The visual claustrum of the cat. i. structure and connections. *J Neurosci*, **1**(9), 956–80, 1981.

Levesque, M., Gagnon, S., Parent, A., and Deschenes. Axonal arborizations of corticostriatal and corticothalamic fibers arising from the second somatosensory area in the rat. *Cereb Cortex*, **6**(6), 759–70, 1996.

Levitt, J. B., Lewis, D. A., Yoshioka, T., and Lund, J. S. Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J Comp Neurol*, **338**(3), 360–76, 1993.

Lewis, D. A., Melchitzky, D. S., and Burgos, G. G. Specificity in the functional architecture of primate prefrontal cortex. *J Neurocytol*, **31**(3-5), 265–76, 2002.

Lichtman, J. W., Livet, J., and Sanes, J. R. A technicolour approach to the connectome. *Nature Reviews. Neuroscience*, **9**(6), 417–22, 2008.

Lieber, R. L. *Skeletal muscle structure, function, and plasticity.* Lippincott Williams and Wilkins, Philadelphia, second edition, 2002.

Lisman, J. A mechanism for the hebb and the anti-hebb processes underlying learning and memory. *Proc Natl Acad Sci U S A*, **86**(23), 9574–8, 1989.

Lorente de No, R. Cerebral cortex: Architecture, intracortical connections, motor projections. In *Physiology of the nervous system*, editor J. Fulton, Outlines of Physiology Series, 274–301. Oxford University Press, New York, 2nd ed. edition, 1943.

Lund, J. S., Hendrickson, A. E., Ogren, M. P., and Tobin, E. A. Anatomical organization of primate visual cortex area vii. *J Comp Neurol*, **202**(1), 19–45, 1981.

Malenka, R. C., and Siegelbaum, S. A. Synaptic plasticity. In *Synapses*, editors W. M. Cowan, T. C. Sudhof, and C. F. Stevens. The Johns Hopkins University Press, Baltimore, 2001.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci*, **5**(10), 793–807, 2004.

Medalla, M., and Barbas, H. Diversity of laminar connections linking periarcuate and lateral intraparietal areas depends on cortical structure. *Eur J Neurosci*, **23**(1), 161–79, 2006.

Mettler, F. A. Corticofugal fiber connections of the cortex macaca mullatta. the frontal region. *J Comp Neurol*, **61**(3), 509–542, 1935.

Meyer, G., Gonzalez-Hernandez, T. H., and Ferres-Torres, R. The spiny stellate neurons in layer iv of the human auditory cortex. a golgi study. *Neuroscience*, **33**(3), 489–98, 1989.

Milner, B. The medial temporal-lobe amnesic syndrome. *Psychiatr Clin North Am*, **28**(3), 599–611, 609, 2005.

Molnar, Z., and Cheung, A. F. Towards the classification of subpopulations of layer v pyramidal projection neurons. *Neurosci Res*, **55**(2), 105–15, 2006.

Monsell, S. Components of working memory underlying verbal skills: a "distributed capacities" view. In *International Symposium on Attention and Performance X*, editors B. H., and B. D., volume 10, 327–350. Erlbaum, Hillsdale, NJ, 1984.

Morishima, M., and Kawaguchi, Y. Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *J Neurosci*, **26**(16), 4394–405, 2006.

Nakano, K., Tokushige, A., Kohno, M., Hasegawa, Y., Kayahara, T., and Sasaki, K. An autoradiographic study of cortical projections from motor thalamic nuclei in the macaque monkey. *Neuroscience Research*, **13**(2), 119–37, 1992.

Nieuwenhuys, R. The neocortex. an overview of its evolutionary development, structural organization and synaptology. *Anat Embryol (Berl)*, **190**(4), 307–37, 1994.

Nolte, J., and Angevine, J. B. *The human brain in pictures and diagrams.* Mosby, Philadelphia, second edition, 2000.

Northcutt, R. G., and Kaas, J. H. The emergence and evolution of mammalian neocortex. *Trends Neurosci*, **18**(9), 373–9, 1995.

Ohno, S. So much junk dna in our genome. In *Evolution of Genetic Systems: Brookhaven Symposium in Biology*, editor H. Smith, volume 23. Gordon and Breach, New York, 1972.

Ohno, S., and Yomo, T. The grammatical rule for all dna: junk and coding sequences. *Electrophoresis*, **12**(2-3), 103–8, 1991.

Palm, G. On associative memory. *Biological Cybernetics*, **36**(1), 19–31, 1980.

Parent, M., and Parent, A. Single-axon tracing and three-dimensional reconstruction of centre median-parafascicular thalamic neurons in primates. *J Comp Neurol*, **481**(1), 127–44, 2005.

Parent, M., and Parent, A. Single-axon tracing study of corticostriatal projections arising from primary motor cortex in primates. *J Comp Neurol*, **496**(2), 202–13, 2006.

Penfield, W., and Rasmussen, T. *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. Hafner Publishing Company, New York, 1968.

Peters, A., Cifuentes, J. M., and Sethares, C. The organization of pyramidal cells in area 18 of the rhesus monkey. *Cereb Cortex*, **7**(5), 405–21, 1997.

Price, A. L., Eskin, E., and Pevzner, P. A. Whole-genome analysis of alu repeat elements reveals complex evolutionary history. *Genome Research*, **14**(11), 2245–52, 2004.

Price, D. J., Kennedy, H., Dehay, C., Zhou, L., Mercier, M., Jossin, Y., Goffinet, A. M., Tissir, F., Blakey, D., and Molnar, Z. The development of cortical connections. *Eur J Neurosci*, **23**(4), 910–20, 2006.

Pucak, M. L., Levitt, J. B., Lund, J. S., and Lewis, D. A. Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J Comp Neurol*, **376**(4), 614–30, 1996.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Mantia, A.-S. L., McNamara, J. O., and Williams, S. M. *Neuroscience*. Sinauer Associates, Inc., Sunderland, thrid edition edition, 2004.

Quiroga, R. Q., Reddy, L., Koch, C., and Fried, I. Decoding visual inputs from multiple neurons in the human temporal lobe. *Journal of Neurophysiology*, **98**(4), 1997–2007, 2007.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. Invariant visual representation by single neurons in the human brain. *Nature*, **435**(7045), 1102–1107, 2005.

Rakic, P., Bourgeois, J., Eckenhoff, M., Zecevic, N., and Goldman-Rakic, P. Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science*, **232**(4747), 232–235, 1986.

Ramnani, N. The primate cortico-cerebellar system: anatomy and function. *Nat Rev Neurosci*, **7**(7), 511–22, 2006.

Rempel-Clower, N. L., and Barbas, H. The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function. *Cereb Cortex*, **10**(9), 851–65, 2000.

Reynolds, J. H., and Chelazzi, L. Attentional modulation of visual processing. *Annu Rev Neurosci*, **27**, 611–47, 2004.

Rockland, K. S. Configuration, in serial reconstruction, of individual axons projecting from area v2 to v4 in the macaque monkey. *Cereb Cortex*, **2**(5), 353–74, 1992.

Rockland, K. S. Two types of corticopulvinar terminations: round (type 2) and elongate (type 1). *J Comp Neurol*, **368**(1), 57–87, 1996.

Saint-Cyr, J. A., Ungerleider, L. G., and Desimone, R. Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *J Comp Neurol*, **298**(2), 129–56, 1990.

Satoh, K., and Fibiger, H. C. Distribution of central cholinergic neurons in the baboon (papio papio). i. general morphology. *Journal of Comparative Neurology*, **236**(2), 197–214, 1985.

Schinder, A. F., and Poo, M. The neurotrophin hypothesis for synaptic plasticity. *Trends Neurosci*, **23**(12), 639–45, 2000.

Schmahmann, J. D., and Pandya, D. N. *Fiber Pathways of the Brain*. Oxford University Press, New York, 2006.

Schmid, C. W., and Jelinek, W. R. The alu family of dispersed repetitive sequences. *Science*, **216**(4550), 1065–70, 1982.

Shadlen, M. N., and Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci*, **18**(10), 3870–96, 1998.

Sherman, M. S., and Guillery, R. W. Functional organization of thalamocortical relays. *Journal of Neurophysiology*, **76**(3), 1367–1395, 1996.

Sherman, M. S., and Guillery, R. W. *Exploring the Thalamus and its role in cortical function*. Cambridge, Massachusetts, 2nd edition, 2006.

Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics and Development*, **9**(6), 657–63, 1999.

Smith, Y., Bevan, M. D., Shink, E., and Bolam, J. P. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience*, **86**(2), 353–87, 1998.

Smith, Y., Raju, D. V., Pare, J. F., and Sidibe, M. The thalamostriatal system: a highly specific network of the basal ganglia circuitry. *Trends Neurosci*, **27**(9), 520–7, 2004.

Solari, S., Smith, A., Minnett, R., and Hecht-Nielsen, R. Confabulation theory. *Physics of Life Reviews*, **5**, 106–120, 2008.

Soloway, A. S., Pucak, M. L., Melchitzky, D. S., and Lewis, D. A. Dendritic morphology of callosal and ipsilateral projection neurons in monkey prefrontal cortex. *Neuroscience*, **109**(3), 461–71, 2002.

Squire, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, **82**, 171–177, 2004.

Squire, L. R., and Zola, S. M. Structure and function of declarative and nondeclarative memory systems. *Proc Natl Acad Sci U S A*, **93**(24), 13515–22, 1996.

Stenger, J. E., Lobachev, K. S., Gordenin, D., Darden, T. A., Jurka, J., and Resnick, M. A. Biased distribution of inverted and direct alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Research*, **11**(1), 12–27, 2001.

Striedter, G. F. *Principles of Brain Evolution*. Sinauer Associates, Inc., Sunderland, 2005.

Sutherland, G. R., and McNaughton, B. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology*, **10**(2), 180–6, 2000.

Szentagothai, J. The 'module-concept' in cerebral cortex architecture. *Brain Res*, **95**(2-3), 475–96, 1975.

Tanaka, K. Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, **13**(1), 90–99, 2003.

Teyler, T. J., and DiScenna, P. The hippocampal memory indexing theory. *Behav Neurosci*, **100**(2), 147–54, 1986.

Teyler, T. J., and Rudy, J. W. The hippocampal indexing theory and episodic memory: updating the index. *Hippocampus*, **17**(12), 1158–69, 2007.

Thomson, A. M., and Bannister, A. P. Postsynaptic pyramidal target selection by descending layer iii pyramidal axons: dual intracellular recordings and biocytin filling in slices of rat neocortex. *Neuroscience*, **84**(3), 669–83, 1998.

Thomson, A. M., and Bannister, A. P. Interlaminar connections in the neocortex. *Cerebral cortex*, **13**(1), 5–14, 2003.

Thomson, A. M., and Deuchars, J. Synaptic interactions in neocortical local circuits: dual intracellular recordings in vitro. *Cereb Cortex*, **7**(6), 510–22, 1997.

Thomson, A. M., and Morris, O. T. Selectivity in the inter-laminar connections made by neocortical neurones. *J Neurocytol*, **31**(3-5), 239–46, 2002.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science*, **311**(5761), 670–4, 2006.

Van Essen, D. C. Corticocortical and thalamocortical information flow in the primate visual system. *Progress in Brain Research*, **149**, 173–183, 2005.

van Vreeswijk, C., and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, **274**(5293), 1724–6, 1996.

Verzeano, M., Lindsley, D. B., and Magoun, H. W. Nature of recruiting response. *J Neurophysiol*, **16**(2), 183–95, 1953.

von Bonin, G., and Bailey, P. *The Neocortex of Macaca Mulatta*. The University of Illinois Press, Illinois, 1947.

von Economo, C. *The Cytoarchitectonics of the Human Cerebral Cortex*. Oxford University Press, London, 1929.

von Economo, C., and Koskinas, G. N. *Atlas of Cytoarchitectonics of the Adult Human Cerebral Cortex*. Karger, 2007.

Voogd, J. The human cerebellum. *J Chem Neuroanat*, **26**(4), 243–52, 2003.

Waksmundzka, M. Development of rat x mouse hybrid embryos produced by microsurgery. *Journal of Experimental Zoology*, **269**(6), 551–9, 1994.

Walker, A. A cytoarchitectural study of the prefrontal area of the macaque monkey. *J Comp Neurol*, **73**(1), 59–86, 1940.

Wang, Y., and Leung, F. C. Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. *FEBS Letters*, **580**(5), 1277–84, 2006.

Watakabe, A., Ichinohe, N., Ohsawa, S., Hashikawa, T., Komatsu, Y., Rockland, K. S., and Yamamori, T. Comparative analysis of layer-specific genes in mammalian neocortex. *Cereb Cortex*, 2006.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**(6915), 520–62, 2002.

Watson, J. D., and Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**(4356), 737–8, 1953.

Willshaw, D. J., Buneman, O. P., and Longuet-Higgins, H. C. Non-holographic associative memory. *Nature*, **222**, 960–962, 1969.

Witter, M., Groenewegen, H., Silva, F. L. D., and Lohman, A. Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Progress in Neurobiology*, **33**, 161–253, 1989.

Wolffe, A. P. *Chromatin Structure and Function*. Academic Press, San Diego, 1998.

Yamamori, T., and Rockland, K. S. Neocortical areas, layers, connections, and gene expression. *Neurosci Res*, 2006.

Yeterian, E. H., and Pandya, D. N. Laminar origin of striatal and thalamic projections of the prefrontal cortex in rhesus monkeys. *Exp Brain Res*, **99**(3), 383–98, 1994.

Yoshimura, Y., Dantzker, J. L., and Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature*, **433**, 868–873, 2005.

Young, M. P., and Yamane, S. Sparse population coding of faces in the inferotemporal cortex. *Science*, **256**, 1327–1331, 1992.

Zaitsev, A. V., Gonzalez-Burgos, G., Povysheva, N. V., Kroner, S., Lewis, D. A., and Krimer, L. S. Localization of calcium-binding proteins in physiologically and morphologically characterized interneurons of monkey dorsolateral prefrontal cortex. *Cereb Cortex*, **15**(8), 1178–86, 2005.