

Reverse Engineering the Brain With a Circuit Diagram Based on a Segmented Connectome and System Dynamics

Walter Schneider, Michael Cole, Sudhir Pathak

Learning Research & Development Center, Center Neural Basis of Cognition, & Psychology Dept. , Univ. of Pittsburgh

Email to Walter Schneider at wws@pitt.edu

Abstract

New connection and activity imaging methods can map most cortical areas, their specialization, control structures and functional links. This will provide a “target circuit diagram” of the human brain. This technology can map the likely 200+ brain areas and 30,000+ fiber tracks, identifying their location, boundaries, volume of tissue, connective topology, causal links, and core functions. The qualitative pattern of the connection structure suggests three critical functions with qualitatively different connective topologies.

The brain appears to be organized into three functional subsystems involving specific representation areas (RAs; e.g., vision, audition, tactile, motor), a domain general Cognitive Control Network (CCN; e.g., attention, decision, task sequencing, affect/arousal), and priority field maps (PFM; e.g., salience, activity, affect). Each system has qualitatively different anatomical connective topology. The synergy of a serial CCN and parallel representation system provides computational advantage over current symbolic or parallel association based computation.

Introduction

The Biologically Inspired Cognitive Architectures (BICA) challenge is to derive principles of operations from neuroscience that usefully guide the development of artificial systems to create artificial learning agents with cognitive capacity at the level of humans. Theorists have used biological inspiration for centuries (e.g., James 1890) to understand cognition. There are however few cases where the biology has produced hard operational constraints that have affected the development of artificial agents. For example, connectionism has been inspired by the idea of computation by populations of simple computational elements but they are rarely seriously specified by the biology. This paper takes the perspective that recent breakthroughs in imaging techniques allow activity and connectivity analysis at sufficient resolution to guide system structure of a Biologically Inspired Cognitive Architecture.

Data sets for reverse engineering human brain

systems. Reverse engineering of a system requires a high precision and richness of the data of internal states to be useful beyond what can be derived from input/output analysis of the system. We believe that the human brain system data of the twentieth century was insufficient for substantial reverse engineering. We had little human activity data and an absence of human connectivity data.

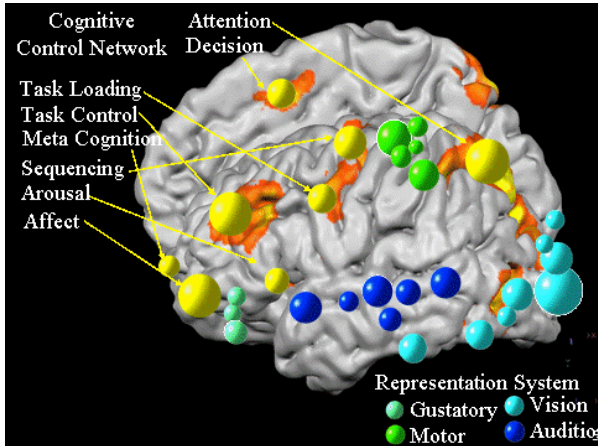
By 2010 there will be a dramatic increase in resolution of both activity and connectivity data.

Think of an analogy of reverse engineering modern computers with the instrumentation of 1940s. Assume you are a 1940 researcher (before the invention of transistors) that got an advanced 2008 laptop and was assigned the task of reverse engineering this “cognitive architecture”. How would one go about creating a Laptop Inspired Cognitive Architecture? Input-output analyses would reveal great complexity and systematicity in the response. It would take a substantial effort to derive useful design information about how the mother board and chips operated given the limited ability of the test instruments of the time to track the high speed processing and small size of the components. Two pivotal analysis methods would include mapping connections and identifying the functional units (chips) and their specialization. Modern circuit boards involve many layers (e.g., 14). Having a circuit diagram of just the first layer would be nearly useless (connections begin but are lost at each crossing so only a small subset of the paths could be followed from source to destination). Only with the ability to follow traces through crossings could one map the topology of the processing system. The second analysis method would be monitoring operations/power use of functional units, the chips of the computer. If one followed the component power over many tasks one could relate processes to functional events (e.g., more energy use (heat generated) in tasks requiring more images to be retrieved from memory chips). Reverse engineering would be tough, and occur at many levels (how are signals transformed in silicon, what is the system organization, how is power utilized, what is an operating system, etc).

The human brain is an evolutionary advanced technology. Our tools are just now allowing mapping of the system dynamics and anatomical structure of human cognition. In 1992 functional Magnetic Resonance Imaging (fMRI) was discovered (Kwong et al., 1992) and it enables non-invasive brain mapping of many condition contrasts. Since then over 20,000 scientific papers have reported brain activation patterns.

In the last three years breakthroughs in brain imaging and analysis provide two pivotal types of analyses of human anatomical brain data. Modern fiber tracking techniques based on multi-shell diffusion weighted imaging (see Hagmann et al. 2006 V J Weeden 2000 2005) allow following fibers across crossings, enabling tracing fibers from source to destination without

1A. Anatomical Space



getting lost when fibers cross¹. This provides the potential of mapping the system circuit diagram of the brain, showing what is connected to what. Recent work by Haggmann and colleagues (2007, 2008) illustrates the ability to track connections showing patterns of local neighborhood and small world connectivity.

There is a second critical challenge: resolving the borders between areas. The borders of brain areas have been unresolvable by twentieth century non-invasive anatomical means (e.g., V2 versus V4 boarder can only be reliably found in primate studies by functionally mapping topologies). Connectivity data without knowing the edges of the functional components greatly limits the conclusions. In the above computer reverse engineering example, assume you could not tell the boundaries of the chips. Connectivity would be hard to interpret where there is a misalignment of presumed and actual area boundaries.

In the last year there have been reports of techniques that allow single subject segmentation of the cortex into regions (V1, V2, V4...) based on and anatomical mapping (Cohen et al. 2007, V J Wedeen 2008, Pathak, et al., 2008). With within-subject segmentation of the brain, detailed connectivity topologies can be derived allowing interpretation of what region is connected to what region and the strength and directionally of that connection (based on techniques such as Granger Causality).

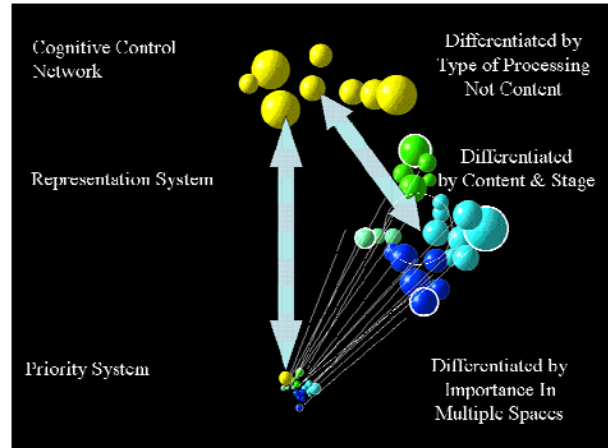
By combining advanced fiber tracking and anatomical segmentation techniques our group is elucidating qualitatively different patterns of connectivity that support different computing functions. This brain connectivity shows an architecture incorporating features of different types of computing (connectionist, symbolic, salience).

Brain System Architecture

The brain can be productively examined at many levels of detail from the molecular to the systems level.

¹ Typically long distance fiber pathways cross multiple times as they combine into major fiber pathways and then exit the fiber pathways.

1B. Process Space



In this paper we examine the systems level with the goal to identify the major patterns of interaction between brain areas and likely computational function. The mammalian cortex is composed of similar cells, cortical layers, and regions across brain areas, evolution, and species (from our earliest evolutionary decedents of the tree shrew to man, from vision to high level decision making).

The following text characterizes the three types of connectivity seen in cortex and suggests the connectivity involves qualitatively different types of processing structures that perform different computational operations. These operations seem similar to major approaches appearing in the computational literature. Figure 1 provides a summary of the structural connectivity topologies and potential functions of the control network. Figure 1A shows the regions in anatomical space. Figure 1B shows them in network interconnectivity space highlighting function. Figure 2 shows a sample of anatomical neural connections that provided the data for related to Figure 1.

1. Domain specific representation systems with hierarchical connectivity.

The representation systems of the brain include sensory (vision, audition, somatopic, gustatory) and motor systems and higher level representation (e.g., social Allison et al., 2000) that are organized in a weakly hierarchical

arrangement. These areas show strong coupling to extrinsic activity (see Hasson 2004). There are detail maps of these systems with the visual system characterized into an estimated 32 areas (Felleman & Van Essen 1992). Similar detail maps exist for the

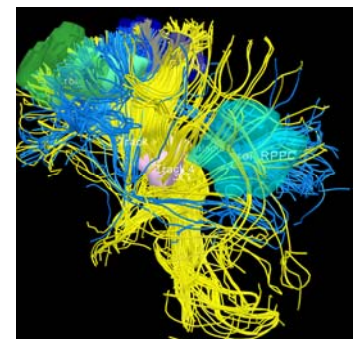


Figure 2 Anatomical connection traces from thalamus (Yellow) and between CCN regions (Blue).

auditory system (Formisano et al. 2003) and somatosensory system (Overduin & Servos, 2004). Recent fMRI methods have demonstrated connective hierarchies in humans in close agreement with primate maps ((Serenio et al., 1995; Wandell et al., 2007) Information comes in from subcortical relay nuclei to an early visual stage (e.g., lateral geniculate to V1) and then branch out to multiple areas (e.g. V1/2/3, hV4, VO-1/2, LO-1/2, IPS-0/1/2/3/4 see Wandell et al. 2007) with sensory systems breaking into 'what' and 'where' systems. Young () has studied connectivity in animal neuroanatomy studies and identifies a pattern of connectivity of each of the areas connecting to a subset (typical 7) of other areas within the sensory area. The key features of the representation areas are tight local organized in a loose hierarchy with early simpler coding of properties. As processing rises in the hierarchy there is often specialization into types of operations such as a 'what' and 'where'/'how' division of the visual system. Learning new representations often requires many experiences with the stimuli.

The processing functions of these areas can be interpreted as representing information in multiple spaces across the processing hierarchy (e.g. visual features, objects, locations, functions). This type of multi-level representation is typical of connectionist multi-layer nets (e.g. PDP books). Loss of these areas can cause selective loss of sensory function (e.g., loss color, motion detection).

2. Domain general Cognitive Control Net (CCN) with wide connectivity through cortical and subcortical nodes. In contrast to the representation regions there are domain general control regions that process stimuli from multiple modalities. The Cognitive Control Net (CCN) has multiple anatomical locations that are distinguished by the type of operation rather than the nature of the content. These areas have been shown to be active in hundreds of tasks (see Cabeza & Nyberg 2000). For example whether doing an auditory tone search or a complex visual motor task such as air traffic control these regions are active. Although not content specific, the regions appear to be process specific. For example Posterior Parietal Cortex (PPC) is more active in attention switching and posterior medial frontal cortex in decision making. This network of areas shows tight functional coupling (correlations typically $r > 0.8$) with less correlation with representation areas (Cole & Schneider, 2007). Early data suggests that the CCN has strong anatomical connectivity as well (Hagmann, et al., 2008, Pathak et al 2008). One aspect of the control system is that it is active during early acquisition of a task and appears to drop out as practice continues and the task becomes automatic (Schneider & Chein, 2003). The control system can rapidly acquire complex rules to compare representations and execute sequential steps as a result of those comparisons.

These cortical areas perform operations on information (switching attention, comparison, decision, response release, sequential step execution, association).

This CCN appears to perform operations similar to productions performing sequential If-Then rules (e.g., SOAR, ACT-R). Anatomically these areas have tight coupling between them and to subcortical areas and connectivity to upper-level representation areas. Loss of individual components of the CCN causes processing failures such as attentional neglect from PPC damage and difficulty in task switching from Dorsolateral Prefrontal Cortex (DLPFC) damage.

3. Priority tracking of scalar signals from the representation areas to control areas. The third class of connections is characterized by structures that are widely connected to many cortical areas with relatively thin cross connections (few fibers per connection). Animal studies show the amygdala and thalamus are connected widely to most areas of the cortex (Young 1993, 2000). In our anatomical connection tracings with DSI we replicate in humans that the amygdala and thalamus are connected to much of cortex. The size of fiber pathways is much larger in connections between cortical areas than from cortex to the priority system. (e.g., 74x between cortical areas fusiform face area {FFA} and other visual areas relative to FFA to thalamus). The many thin link connections is compatible that cortical areas projecting scalar priority signals to subcortical structures that then prioritize requests from large cortical areas to determine what areas need additional processing (see Schneider & Chein, 2003).

An importance function is the evaluation of the importance of information to allow the limited CCN to attend to critical information. A common feature of machine vision systems and attention theories are the presence of priority maps (e.g., Koch, C., & Ullman, 1985; Wolf 1994) These go by various names: salience, pertinence, importance, priority maps. The priority function system performs a role similar to priority interrupts in modern operating systems. The brain has an estimated 10,000 cortical macro columns that are organized into some 200+ cortical areas. This massive parallel architecture is controlled by a mostly serial CCN. A key computational problem is deciding which area and what columns in an area must be attended to by the CCN domain general operations. Attention models typical assume there are planar priority/salience maps which identify peaks sizes of areas of importance for further processing. Anatomically there are several such fields coding different classes of importance (salience, affect, visceral disgust). The loss of such fields can occur by drug induced deactivation or anatomical loss of areas routing the priority information (e.g., Desimone et al., 1989).

Powerful robust computing through synergy of computational architectures. The human cognitive architecture synergistically combines the domain specific representation systems, the domain general CCN, and the priority fields to produce powerful perception and decision making abilities. A detailed description of how the three types of computing interact is provided elsewhere (see Schneider & Chein, 2003). Biological

cognition has combined the strengths of these computational methods. Modern computer applications (e.g., Google) can beat the limited symbolic operations of the CCN. However symbolic systems have difficulty of dealing with similarity and relatedness which severely limits the capacity of purely symbolic approaches. The CCN operates on the representation network that codes the meaning and transformations the sensory systems developed to deal with a world of immense variation. The CCN comparisons use that architecture to allow operations on the meaning representations. The priority maps are also attached to the meaning representations allowing coding across transformations. The presence of multiple maps allows rapid shifting of the CCN depending on context (e.g., a stimulus can trigger a fear response, shifting which importance field is attended to).

Using Biological Data to Inspire & Guide Cognitive Architectures

High resolution data on human brain connectivity and activity provide rich data sets to examine biological computational architectures. Within the next decade we will likely have a detailed segmented connectome map detailing the 200+ cortical areas, the topology and strength of their connectivity, and basic functional specialization. Having a detailed diagram of the human cognitive architecture will elucidate the principles for a biologically specified cognitive architecture. There are a number of key questions that we think need to be resolved, including: A) How are computations by a serial CCN doing quasi symbolic operations integrated with the subsymbolic representation areas? B) How does the CCN effectively control the representation system? C) How does the representation system deal effectively with transfer and similarity? Why is the representation system domain specific whereas cognitive control and episodic memory are domain general? D) How does deep learning occur through experience in multilevel representation systems fast enough in a living/surviving agent (e.g Hinton et al., 2006)?; F) How does the CCN dramatically speed learning relative to typical connectionist learning?; G) How does learning enable the representation system to process well-learned patterns automatically (without the need for the limited CCN)?

References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in cognitive sciences*, 4(7), 267-278

Cohen, A. L., Fair, D. A., Miezin, F. M., Dosenbach, N. U. F., Schlaggar, B. L., & Petersen, S. E. (2007). Defining functional areas in individual human brains using resting functional connectivity fMRI. Washington University, St. Louis, MO.

Cole, M. W. & Schneider, W. (2007). The Cognitive Control Network: Integrated cortical regions with dissociable functions. *NeuroImage*, 37, 343-360.

Corbetta, M. & Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention of the brain. *Nature reviews: Neuroscience*, 3, 201-215.

Desimone, R., Wessinger, M., Thomas, L. & Schneider, W. (1989). Effects of deactivation of lateral pulvinar or

superior colliculus on the ability to selectively attend to a visual stimulus. *Society for Neuroscience Abstracts*, 15, 162.

Fellman, D. J. & Van Essen, D. C. (2003). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1, 1-47.

James, W (1890) *Psychology Vol 1 & 2*. New York: Henry Hold and Company.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303, 1634-1640.

Hinton, G. E., Osindero, S. & Teh, Y. (2006) A fast learning algorithm for deep belief nets. *Neural Computation*. 18, 1527-1554

Kwong K K et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (12), 5675-5679.

Miller, E.K. & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.

Overduin, S. A. and P. Servo (2004). "Distributed digit somatotopy in primary somatosensory cortex." *Neuroimage* 23(2): 462-472.

Pathak S., Martins B., Cole M.W., & Schneider W. (2008). *Anatomical and Functional Segmentation of the Cognitive Control Network: Supporting a preliminary cognitive control network connectome*. Poster presented at Cognitive Neuroscience Society, San Francisco, CA

Ragland, J. D., J. Yoon, et al. (2007). Neuroimaging of cognitive disability in schizophrenia: Search for a pathophysiological mechanism. *International Review of Psychiatry*, 19(4), 419-429.

Schneider, W. & Chein, J.M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science*, 27, 525-559.

Sereno, M. I., A. M. Dale, et al. (1995). "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging." *Science* 268(5212): 889-93.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron Review*, 56, 366-383.

Wedeen, V.J. et al., (2000). Mapping fiber orientation spectra in cerebral white matter with fourier-transform diffusion MR, *Paper Presented at: Proc. Intl. Soc. Mag. Res. Med. (Denver)*

Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G. & Weisskoff, R.M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging, *Magn. Reson. Med.* 54, 1377-1386.

Wedeen, V.J. et al., (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *NeuroImage*, 41,(4) 1267-1277.

Young, M.P. (1993). The organization of neural systems in the primate cerebral cortex. *Biological science*, 252(1333), 13-18.

Young, M. P. & J. W. Scannell (2000). Brain structure-function relationships: advances from neuroinformatics - Introduction. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 355, 3-6.