Foundations and New Paradigms of Brain Computing: Past, Present, and Future

Stephen Grossberg

Center for Adaptive Systems, Boston University,
677 Beacon Street, Boston, MA 02215
steve@cns.bu.edu

1 Introduction

A paradigm shift: autonomous adaptation to a changing world. There has been rapid progress over the past fifty years in modeling how brains control behavior; that is, in developing increasingly sophisticated and comprehensive computational solutions of the classical mind/body problem. Not surprisingly, such progress embodies a major paradigm shift, but one that is taking a long time to fully take hold because it requires a synthesis of knowledge from multiple disciplines, including psychology, neuroscience, mathematics, and computer science.

Linking brain to mind clarifies both brain *mechanisms* and behavioral *functions*. Such a linkage is needed to develop applications to computer science, engineering, and technology, since mechanisms tell us how things work, whereas functions tell us what they are for. Knowing how things work and what they are for are both essential in any application. Such models represent a paradigm shift because the brain is unrivaled in its ability to autonomously adapt in real time to complex and changing environments. Models that embody adaptive autonomous intelligent responses to unexpected contingencies are just the sorts of models that can fully realize the dream of artificial intelligence.

A method to link brain to mind. By what method can such models be discovered? A successful method has been elaborated over the past fifty years. The key is to begin with behavioral data, typically scores or even hundreds of parametrically structured behavioral experiments in a particular problem domain. One begins with behavioral data because brain evolution needs to achieve behavioral success. Any theory that hopes to link *brain* to behavior thus must discover the computational level on which brain dynamics control behavioral success. This level has proved to be the network and system level. That is why the name *neural networks* is appropriate for these models.

Behavioral data provide a theorist with invaluable clues about the functional problems that the brain is trying to solve. One starts with large amounts of data because otherwise too many seemingly plausible hypotheses cannot be ruled out. A crucial meta-theoretical constraint is to insist upon understanding the behavioral datawhich comes to us as static numbers or curves on a pageas the emergent

R. Pirrone and F. Sorbello (Eds.): AI*IA 2011, LNAI 6934, pp. 1–7, 2011.

properties of a dynamical process which is taking place moment-by-moment in an individual mind. One also needs to respect the fact that our minds can adapt on their own to changing environmental conditions without being told that these conditions have changed. One thus needs to frontally attack the problem of how an intelligent being can *autonomously adapt to a changing world*. Knowing how to do this, as with many other theoretical endeavors in science, is presently an art form. There are no known algorithms with which to point the way.

Whenever I have applied this method in the past, I have never used homunculi, or else the crucial constraint on *autonomous* adaptation would be violated. The result has regularly been the discovery of new organizational principles and mechanisms, which are then realized as a minimal model operating according to only locally defined laws that are capable of operating on their own in real time. The remarkable fact is that, when such a behaviorally-derived model has been written down, it has always been interpretable as a neural network. These neural networks have always included known brain mechanisms. The functional interpretation of these mechanisms has, however, often been novel because of their derivation from a behavioral analysis. The networks have also often predicted the existence of unknown neural mechanisms, and many of these predictions have been supported by subsequent neurophysiological, anatomical, and even biochemical experiments over the years.

Once this neural connection has been established by a top-down analysis from behavior, one can work both top-down from behavior and bottom-up from brain to exert a tremendous amount of conceptual pressure with which to better understand the current model and to discover design principles that have not yet been satisfied in it. Then the new design principles help to derive the next model stage. This Method of Minimal Anatomies acknowledges that one cannot derive "the brain" in one theoretical step. But one can do it incrementally in stages by carrying out a form of conceptual evolution. Applying this method, a sequence of self-consistent but evolving models can be derived, with each subsequent model capable of explaining and predicting more data than its ancestors.

A fundamental empirical conclusion can be drawn from many experiences of this type; namely, the brain as we know it can be successfully understood as an organ that is designed to achieve autonomous adaptation to a changing world. Although I am known as one of the founders of the field of neural networks, I have never tried to derive a neural network. Neural networks arise from a real-time behavioral analysis because they provide natural computational realizations with which to control autonomous behavioral adaptation to a changing world.

New paradigms: Complementary computing, laminar computing, and nano chips. How does the brain carry out autonomous adaptation to a changing world? What new computational paradigms are needed to accomplish this goal?

Complementary Computing clarifies the nature of brain specialization. It provides an alternative to the previous computer-inspired paradigm of independent modules. If there were independent modules in the brain, properties such as visual lightness, depth, and motion would be computed independently, which is not the case. Complementary Computing explains why the brain is specialized into parallel processing streams, and how these streams interact in specific ways to overcome their complementary deficiencies [8].

Laminar Computing clarifies how the ubiquitous organization of cerebral cortex into layered circuits can support, through variations of the same laminar architecture, such different aspects of biological intelligence as vision, speech, and cognition [1] [9] [11] [12].

These new computational paradigms promise to have a major impact on the design of future computers that increasingly embody aspects of human intelligence. For example, it is widely acknowledged that Moores Law will break down within ten years. Current Von Neumann chip designs cannot continue to become increasingly dense without becoming highly noisy and generating too much heat. The DARPA SyNAPSE program, among others, has responded to this challenge by supporting research to design new nano-scale VLSI chips that better embody properties of biological intelligence. The idea is for future computers to contain the fastest traditional chips, which can carry out many functions that human brains cannot, as well as brain-inspired chips whose successive generations can carry out increasingly complex types of characteristically human intelligence, notably autonomous adaptation to a changing world.

Nano-scale chips tend to be noisy chips, unlike the perfect chips in Von Neumann computers on which current AI builds. In order to generate less heat, the new nano-scale chips need to use discrete spikes in time to communicate between processing elements. In order to pack in the necessary processing, they may also need to be organized in processing layers. DARPA turned to the brain for design inspiration because the cerebral cortex, which supports all higher aspects of biological intelligence, provides a paradigmatic example of a noisy, layered, spiking intelligent device. That is why Laminar Computing is starting to change the way in which future chips are being designed. [2] have described, using the example of the 3D LAMINART model of 3D vision, a general method for converting fifty years of neural networks based on continuous rate-based signals into spiking neural networks that are amenable to being embodied in SyNAPSE-style chips.

Research progress: towards autonomous adaptive agents.Using this method, my colleagues and I have developed increasingly detailed and comprehensive neural models of vision and visual object recognition; audition, speech, and language; development; attentive learning and memory; cognitive information processing; reinforcement learning and motivation; cognitive-emotional interactions; navigation; sensory-motor control and robotics; and mental disorders. These models involve many parts of the brain, ranging from perception to action, and multiple levels of brain organization, ranging from individual spikes and their synchronization to cognition. My web page `http://cns.bu.edu/~steve` contains many downloadable articles that illustrate this progress. In my talk at AI*IA, I will summarize some recent theoretical progress towards designing autonomous mobile adaptive agents. One of these developments is summarized below.

2 What Is an Object? Learning Object Categories under Free Viewing Conditions

ARTSCAN model: What-Where stream coordination supports invariant object learning. What is an object? How can we learn what an object is without any external supervision? In particular, how does the brain learn to recognize a complex object from multiple viewpoints, and when it is seen at multiple positions and distances? Such a competence is essential in any mobile autonomous adaptive agent. Consider what happens when we first look at an object that is not instantly recognizable. We make scanning eye movements, directing our foveas to a variety of points of interest, or views, on the object. The objects retinal representations of these views are greatly distorted by cortical magnification in cortical area V1. The brain somehow combines several such distorted views into an object recognition category that is invariant to where we happen to be gazing at the moment. Future encounters with the same object can therefore lead to fast recognition no matter what view we happen to see.

How does the brain know that the views that are foveated on successive saccades belong to the same object? How does the brain avoid the problem of erroneously learning to classify parts of different objects together? Only views of the same object should be linked through learning to the same view-invariant object category. How does the brain know which views belong to the same object, even before it has learned a view-invariant category that can represent the object as a whole? How does the brain do this without an external teacher; that is, under the unsupervised learning conditions that are the norm during many object learning experiences *in vivo*?

My colleagues and I [3] [6] [10] have been developing a neural model to explain how spatial and object attention to coordinate the brains ability to learn representations of object categories that are seen at multiple positions, sizes, and viewpoints. Such invariant object category learning and recognition can be achieved using interactions between a hierarchy of processing stages in the visual brain. These stages include retina, lateral geniculate nucleus, and cortical areas V1, V2, V4, and IT in the brain's What cortical stream, as they interact with spatial attention processes within the parietal cortex of the Where cortical stream.

The model first was developed to explain view-invariant object category learning and recognition [6] [7] [10]. This version of the model is called ARTSCAN. ARTSCAN has been generalized to the *positional* ARTSCAN, or pARTSCAN, model which explains how view-, position-, and size-invariant object categories may be learned [3].

I predict that view-invariant object learning and recognition is achieved by the brain under free viewing conditions through the coordinated use of spatial and object attention. Many studies of spatial attention have focused on its spatial distribution and how it influences visual perception. I predict that spatial attention plays a crucial role in controlling view-invariant object category learning. In particular, several authors have reported that the distribution of spatial attention can configure itself to fit an objects form. Form-fitting spatial attention is sometimes called an *attentional shroud* [13]. ARTSCAN predicts how an objects pre-attentively formed surface representation can induce such a form-fitting attentional shroud. Moreover, while this attentional shroud remains active, I predict that it accomplishes two things.

First, it ensures that eye movements tend to end at locations on the objects surface, thereby enabling views of the same object to be sequentially explored. Second, it keeps the emerging view-invariant object category active while different views of the object are learned and associated with it. Thus, the brain avoids what would otherwise seem to be an intractable infinite regress: If the brain does not already know what the object is, then how can it, without external guidance, prevent views from several objects from being associated? My proposal is that the *pre-attentively formed surface representation of the object* provides the object-sensitive substrate that prevents this from happening, even before the brain has learned knowledge about the object. This hypothesis is consistent with a burgeoning psychophysical literature showing that 3D boundaries and surfaces are the units of pre-attentive visual perception, and that attention selects these units for recognition.

This proposed solution can be stated more formally as a temporally-coordinated cooperation between the brains What and Where cortical processing streams: The Where stream maintains an attentional shroud whose spatial coordinates mark the surface locations of a current "object of interest", whose identity has yet to be determined in the What stream. As each view-specific category is learned by the What stream, it focuses object attention via a learned top-down expectation on the critical features that will be used to recognize that view and its variations in the future. When the first such view-specific category is learned, it also activates a cell population at a higher cortical level that will become the view-invariant object category.

Suppose that the eyes or the object move sufficiently to expose a new view whose critical features are significantly different from the critical features that are used to recognize the first view. Then the first view category is reset, or inhibited. This happens due to the mismatch of its learned top-down expectation, or prototype of attended critical features, with the newly incoming view information [4] [5]. This top-down prototype focuses object attention on the incoming visual information. Object attention hereby helps to control which view-specific categories are learned by determining when the currently active view-specific category should be reset, and a new view-specific category should be activated. However, the view-invariant object category should *not* be reset every time a view-specific category is reset, or else it can never become view-invariant. This is what the attentional shroud accomplishes: It inhibits a tonically-active reset signal that would otherwise shut off the view-invariant category when each view-based category is reset. As the eyes foveate a sequence of object views through time, they trigger learning of a sequence of view-specific categories, and each of them is associatively linked through learning with the still-active view-invariant category.

When the eyes move off an object, its attentional shroud collapses in the Where stream, thereby disinhibiting the reset mechanism that shuts off the view-invariant category in the What stream. When the eyes look at a different object, its shroud can form in the Where stream and a new view category can be learned that can, in turn, activate the cells that will become the view-invariant category in the What stream.

The original archival articles show how these concepts can explain many psychological and neurobiological data about object category learning and recognition. In particular, the model mechanistically clarifies basic properties of attention shifts (engage, move, disengage) and inhibition of return. It simulates human reaction time data about object-based spatial attention shifts, and learns with 98.1% accuracy and a compression of 430 on a letter database whose letters vary in size, position, and orientation.

References

1. Cao, Y., Grossberg, S.: A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. Spatial Vision 18, 515–578 (2005)
2. Cao, Y., Grossberg, S.: Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: A method for converting neural rate models into spiking models. Neural Networks (in press, 2011)
3. Cao, Y., Grossberg, S., Markowitz, J.: How does the brain rapidly learn and reorganize view- and positionally-invariant object representations in inferior temporal cortex? Neural Networks (in press, 2011)
4. Carpenter, G.A., Grossberg, S.: A massively parallel architecture for a self-organizing neural pattern-recognition machine. Computer Vision Graphics and Image Processing 37, 54–115 (1987)
5. Carpenter, G.A., Grossberg, S.: Pattern recognition by self-organizing neural networks. MIT Press, Cambridge (1991)
6. Fazl, A., Grossberg, S., Mingolla, E.: View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. Cognitive Psychology 58(1), 1–48 (2009)
7. Foley, N., Grossberg, S., Mingolla, E.: Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-vew, and crowding (submitted for publication, 2011)
8. Grossberg, S.: The complementary brain: Unifying brain dynamics and modularity. Trends in Cognitive Sciences 4, 233–246 (2000)
9. Grossberg, S.: How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. Behavioral and Cognitive Neuroscience Reviews 2, 47–76 (2003)
10. Grossberg, S.: Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion, and action. Philosophical Transactions of the Royal Society of London, Special Issue Predictions in the Brain: Using our Past to Generate a Future 364, 1223–1234 (2009)
11. Grossberg, S., Kazerounian, S.: Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise. Journal of the Acoustical Society of America (in press, 2011)

12. Grossberg, S., Pearson, L.: Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. Psychological Review 115, 677–732 (2008)
13. Tyler, C.W., Kontsevich, L.L.: Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. Perception 24, 127–153 (1995)