

**A half century of progress towards a unified neural theory of mind and brain  
with applications to autonomous adaptive agents and mental disorders**

Stephen Grossberg

<sup>1</sup>Center for Adaptive Systems

Graduate Program in Cognitive and Neural Systems

Departments of Mathematics & Statistics, Psychological & Brain Sciences,  
and Biomedical Engineering

Boston University

Boston, MA 02215

Invited article for the book

**Artificial Intelligence in the Age of  
Neural Networks and Brain Computing**

R. Kozma, C. Alippi, Y. Choe, and F. C. Morabito, Eds.

Cambridge, MA: Academic Press

Submitted: November 11, 2017

*All correspondence should be addressed to*

Professor Stephen Grossberg

Center for Adaptive Systems, Room 213

Boston University

677 Beacon Street

Boston, MA 02215

Phone: 617-353-7857

Fax: 617-353-7755

Email: [steve@bu.edu](mailto:steve@bu.edu)

**Running Title:** Mind, brain, autonomous agents, and mental disorders

## **Abstract**

This article surveys some of the main design principles, mechanisms, circuits, and architectures that have been discovered during a half century of systematic research aimed at developing a unified theory that links mind and brain, and shows how psychological functions arise as emergent properties of brain mechanisms. The article describes a theoretical method that has enabled such a theory to be developed in stages by carrying out a kind of conceptual evolution. It also describes revolutionary computational paradigms like Complementary Computing and Laminar Computing that constrain the kind of unified theory that can describe the autonomous adaptive intelligence that emerges from advanced brains. Adaptive Resonance Theory, or ART, is one of the core models that has been discovered in this way. ART proposes how advanced brains learn to attend, recognize, and predict objects and events in a changing world that is filled with unexpected events. ART is not, however, a “theory of everything” if only because, due to Complementary Computing, different matching and learning laws tend to support perception and cognition on the one hand, and spatial representation and action on the other. The article mentions why a theory of this kind may be useful in the design of autonomous adaptive agents in engineering and technology. It also notes how the theory has led to new mechanistic insights about mental disorders such as autism, medial temporal amnesia, Alzheimer’s disease, and schizophrenia, along with mechanistically informed proposals about how their symptoms may be ameliorated.

## **Towards a unified theory of mind and brain**

A major scientific and technological revolution in understanding autonomous adaptive intelligence is currently underway. How the brain works provides a critical example of such intelligence. This revolution has been supported, in part, by publications over the past 50 years of design principles, mechanisms, circuits, and architectures that are part of an emerging unified theory of biological intelligence. This emerging theory explains and predicts how brain mechanisms give rise to mental functions as emergent properties.

This theory has clarified how advanced brains are designed to enable individuals to autonomously adapt in real time in response to complex changing environments that are filled with unexpected events. Its results hereby provide a blueprint for designing increasingly autonomous adaptive agents for future applications in engineering and technology. Many large-scale applications in engineering and technology have already been developed; e.g., <http://techlab.bu.edu/resources/articles/C5>.

As part of the development of the biological theory, the data from thousands of psychological and neurobiological experiments have been explained and predicted in a unified way, including data about perception, cognition, cognitive-emotional dynamics, and action. These results include an emerging unified theory of what happens in an individual brain when it consciously sees, hears, feels, or knows something; how seeing, hearing, feeling, and knowing can be integrated into unified moments of conscious experience; and how unconscious processes can influence a brain's decision-making (Grossberg, 2017b).

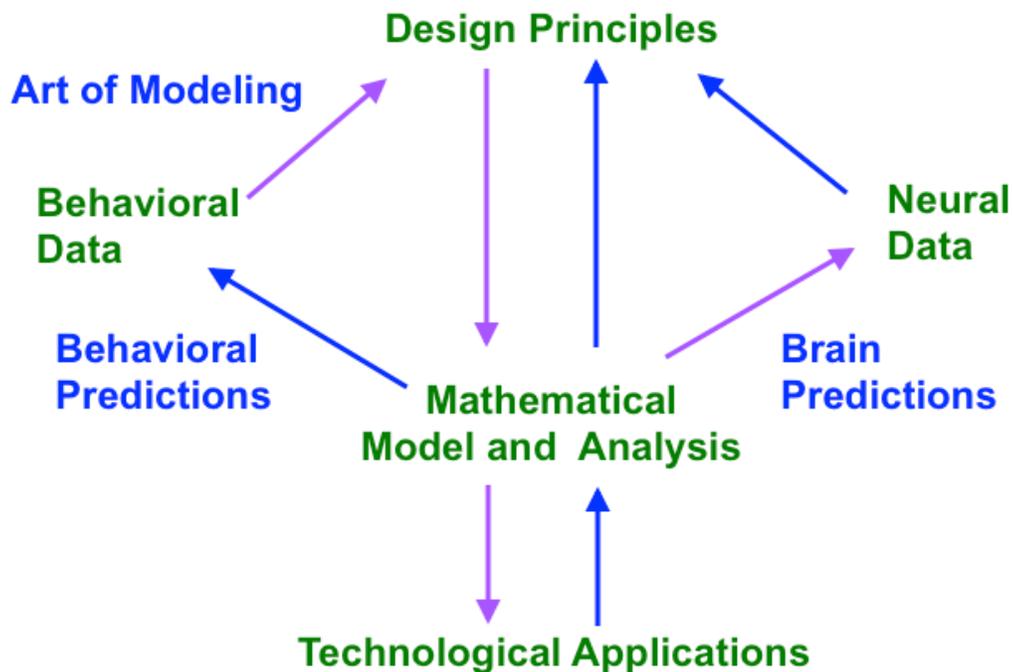
As sufficiently mature models of typical, or normal, behaviors became understood, it also became possible to increasingly explain brain mechanisms and behavioral symptoms of mental disorders. Applications to autism, schizophrenia, and medial temporal amnesia were among the first to be made; e.g., Carpenter and Grossberg (1993), Grossberg (2000b), and Grossberg and Seidman (2006). Additional applications have been recently made towards explaining how the dynamics of learning, memory, and cognition may break down during Alzheimer's disease, why slow wave sleep disorders are often correlated with Alzheimer's disease and other mental disorders, and how symptoms of Fragile X syndrome and autistic repetitive behaviors may arise (Grossberg, 2017a; Grossberg and Kishnan, 2017), and how these insights may help to guide new clinical therapies.

How did a theory that was developed to explain data about the learning and performance of typical, or normal, behaviors lead to explanations of data about mental disorders? This happened when it began to be noticed that, when various model brain mechanisms become imbalanced in prescribed ways, then formal analogs of behavioral symptoms of different mental disorders emerged. In autism, these imbalances include underaroused emotional depression in the drive representations of regions like the amygdala, hypervigilant learning and narrowing of attention in the recognition learning circuits of brain regions like the temporal and prefrontal cortices, and a failure of adaptively-timed learning in brain regions like the hippocampus, basal ganglia, and cerebellum (Grossberg and Seidman, 2006). In this way, one could begin to understand the neural mechanisms and behavioral symptoms of mental disorders on a continuum with neural mechanisms and behavioral properties of typical behaviors.

Said in another way, after one does due diligence in discovering and characterizing the brain mechanisms of normal behaviors, then mechanistic explanations of clinical data *automatically* emerge from these theories. In a similar way, the discovery of key brain mechanisms, circuits, and architectures to explain one kind of data has often thrust me into

explanations of other, seemingly quite different, kinds of data where variations and specialization of these mechanisms, circuits, and architectures are also operative. In this sense, by getting the theoretical foundations of biological intelligence right, one can then begin to reap the benefits of the gift that never stops giving.

## MODELING METHOD AND CYCLE



**Figure 1.** A modeling method and cycle that clarifies how increasingly refined neural models can explain and predict increasingly large interdisciplinary behavioral and neurobiological data bases.

### **A theoretical method for linking brain to mind: The method of minimal anatomies**

One cannot hope to derive a unified theory of an entire brain in one step, and one should not try to do so. Rather, this grand goal can be achieved incrementally, in stages, starting with a large behavioral database that excites a theorist's imagination (Figure 1). The derivation begins with behavioral data because *brain* evolution needs to achieve *behavioral* success. Starting with behavioral data enables models to be derived whose brain mechanisms have been shaped during evolution by behavioral success. Starting with a large database helps to rule out incorrect, but otherwise seemingly plausible, models of how a brain works.

Such a derivation has always led in the past to the discovery of novel design principles and mechanisms (Figure 1) with which to explain how an individual, behaving in real time, can generate the behavioral data as emergent properties. This conceptual leap from data to design is the art of modeling. Once derived, despite being based on psychological constraints, the minimal mathematical model that realizes the behavioral design principles has always

looked like part of a brain (Figure 1). I first experienced such a derivation of brain mechanisms from psychological hypotheses when I was a Freshman at Dartmouth College in 1957-1958. It was a transformative experience that shaped the rest of my life <https://youtu.be/9n5AnvFur7I>.

The past 60 years of modeling have abundantly supported the hypothesis that brains look the way that they do because they embody natural computational designs whereby individuals autonomously adapt to changing environments in real time. The revolution in understanding biological intelligence is thus, more specifically, a revolution in understanding *autonomous adaptive intelligence*. The link from behavior-to-principle-to-model-to-brain has, in addition, often disclosed unexpected functional roles of the derived brain mechanisms that are not clear from neural data alone.

At any stage of this modeling cycle, the goal is to first derive the *minimal model* that embodies the psychological hypotheses that drive the model derivation. Such a “minimal” model is one for which, if any model mechanism is removed, or “lesioned”, then the remaining model can no longer explain a key set of previously explained data. A wise theorist should, I believe, strongly resist “throwing in” known neural mechanisms that are not yet in the minimal model if there is no functional understanding of why they are needed. Once the link between mechanism and function is broken in this way, the ability of the current minimal model to drive further model refinements will be lost.

In particular, once a connection is made top-down from behavior to brain by such a minimal model, mathematical and computational analysis discloses what data the minimal model, and its individual and species variations, can and cannot explain. The data that cannot be explained are as important as those that can be explained, because they demarcate a “boundary between the known and the unknown”. Analysis of this boundary focuses a theorist’s attention upon design principles that the current model does not yet embody. These new design principles and their mechanistic realizations are then consistently incorporated into the model to generate a more realistic model, and one that has always been able to explain and predict a lot more psychological and neurobiological data. If the model cannot be refined, or unlumped, in this way, then that is strong evidence that the current model contains a serious error, and must be discarded.

This theoretical cycle has been successfully repeated multiple times, and has led to models with an increasingly broad explanatory and predictive range, including models that can individually explain psychological, neurophysiological, neuroanatomical, biophysical, and biochemical data. In this specific sense, the classical Mind/Body problem is being solved through principled, albeit incremental, refinements and expansions of theoretical understanding. One can think of these incremental refinements as a way that a theory can try to carry out a kind of “conceptual evolution” by analyzing how various environmental pressures may have driven the biological evolution of our brains.

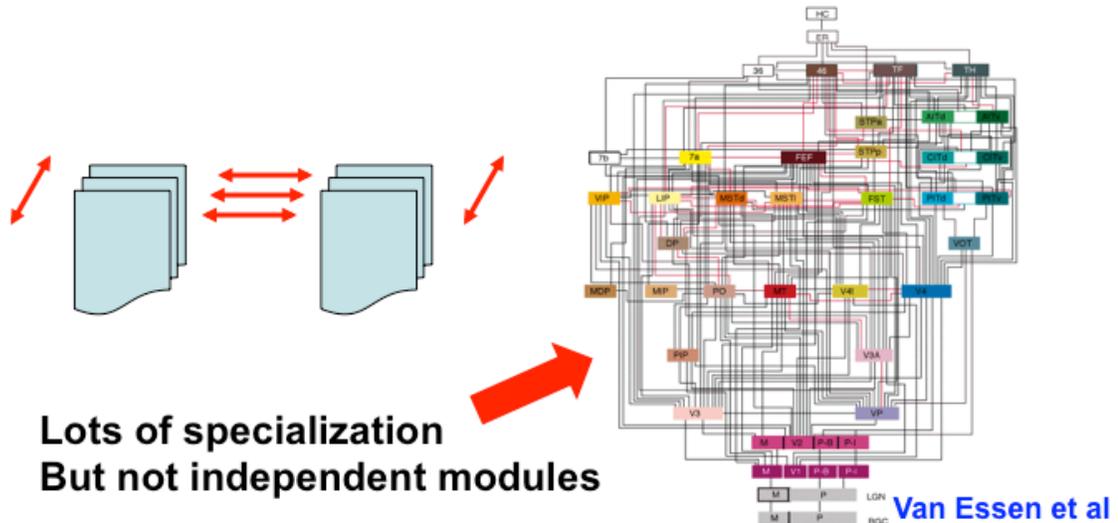
### **Revolutionary brain paradigms: Complementary Computing and Laminar Computing**

The possibility of deriving a unified theory of mind and brain has built upon the discovery that advanced brains embody novel computational paradigms in order to achieve autonomous adaptive intelligence. Two of these paradigms are Complementary Computing and Laminar Computing.

# COMPLEMENTARY COMPUTING

New principles of  
**UNCERTAINTY** and **COMPLEMENTARITY**  
clarify why

**Multiple Parallel Processing Streams Exist in the Brain**

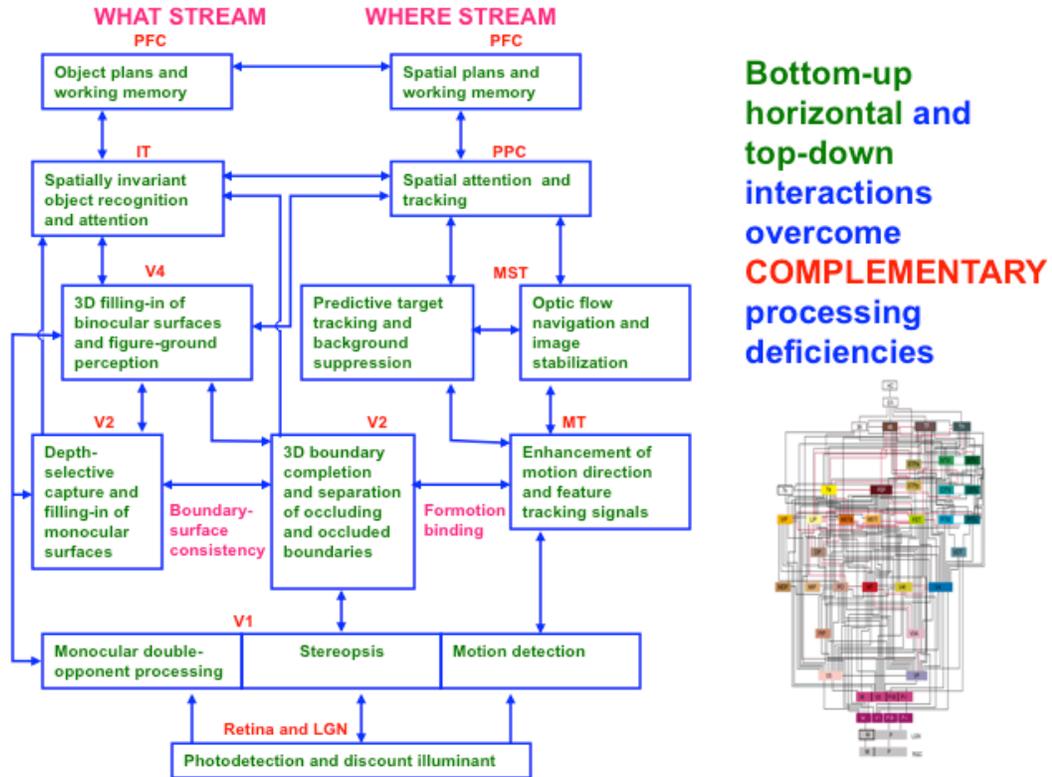


**Figure 2.** Complementary Computing clarifies why there are multiple parallel processing streams in the brain, each with multiple processing stages to resolve computational uncertainties that cannot be overcome by just one processing stream or stage. The anatomical macrocircuit of the visual system dramatically illustrates this state of affairs.

Complementary Computing (Grossberg, 2000a) describes how the brain is organized into complementary parallel processing streams whose interactions generate biologically intelligent behaviors (Figure 2). A single cortical processing stream can individually compute some properties well, but cannot, by itself, process other computationally complementary properties. Pairs of complementary cortical processing streams interact to generate emergent properties that overcome their complementary deficiencies to compute complete information with which to represent or control some faculty of intelligent behavior. Complementary Computing hereby clarifies how different brain regions can achieve a great deal of specialization without being independent modules.

Figure 2 includes an anatomical macrocircuit of the monkey visual system that illustrates its multiple brain regions, and the dense connections between them (Felleman and van Essen, 1991). Figure 3 summarizes a macrocircuit of some of the main brain regions that are modeled in an emerging unified theory of visual intelligence, and the perceptual processes that they carry out. This macrocircuit also includes bottom-up, horizontal, and top-down connections that are needed to overcome computational weaknesses due to Complementary Computing that each brain region would experience if it acted alone.

## EMERGING UNIFIED THEORY OF VISUAL INTELLIGENCE



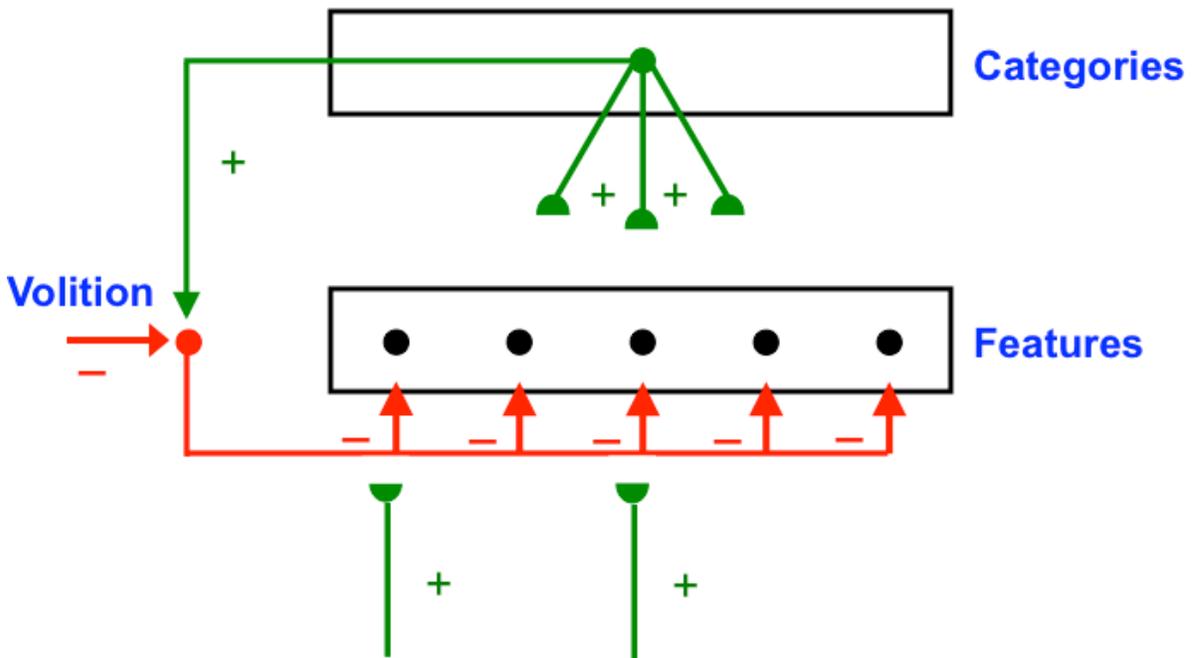
**Figure 3.** A model macrocircuit of an emerging unified theory of visual intelligence. Its processing stages begin at the retina and end in the prefrontal cortex, and include both the What and the Where cortical processing streams. The bottom-up, horizontal, and top-down interactions between model processing stages are needed to overcome the computational weaknesses that each processing stage would experience if it acted alone, due to Complementary Computing.

### The WHAT and WHERE cortical streams are complementary

The category learning, attention, recognition, and prediction circuits of the ventral, or What, cortical processing stream for perception and cognition (Mishkin, 1982; Mishkin, Ungerleider, and Macko, 1983) are computationally complementary to those of the dorsal, or Where and How, cortical processing stream for spatial representation and action (Goodale and Milner, 1992; Mishkin, 1982; Mishkin, Ungerleider, and Macko, 1983). One reason for this What-Where complementarity is that the What stream learns object recognition categories that are substantially invariant under changes in an object's view, size, and position. These invariant object categories enable our brains to recognize valued objects without experiencing a combinatorial explosion. They cannot, however, locate and act upon a desired object in space. Cortical Where stream spatial and motor representations can locate objects and trigger actions towards them, but cannot recognize them. By interacting together, the What and Where streams can recognize valued objects and direct appropriate goal-oriented actions towards them.

### Adaptive Resonance Theory

Abundant psychological and neurobiological data have confirmed all of the foundational predictions concerning how perceptual/cognitive processes in the What stream use excitatory matching and match-based learning to create self-stabilizing categorical representations of objects and events, notably recognition categories that can be learned quickly without experiencing catastrophic forgetting during subsequent learning. In other words, this learning process solves the *stability-plasticity dilemma*. They thereby enable increasing expertise, and an ever-expanding sense of self, to emerge throughout life. See Grossberg (2013, 2017b) for reviews.



**Figure 4.** ART Matching Rule. Bottom-up inputs can activate their target featural cells, other things being equal. A top-down expectation, by itself, can only modulate, prime, or sensitize cells in its excitatory on-center (green pathways with hemicircular adaptive synapses) because of the wider off-surround (red pathways) that tends to balance the top-down excitation (“one-against-one”) within the on-center, while causing driving inhibition in the off-surround. When bottom-up inputs and a top-down expectation are both active, only cells where bottom-up excitation and the top-down excitatory prime converge in the on-center can fire (“two-against-one”), while other featural cells are inhibited.

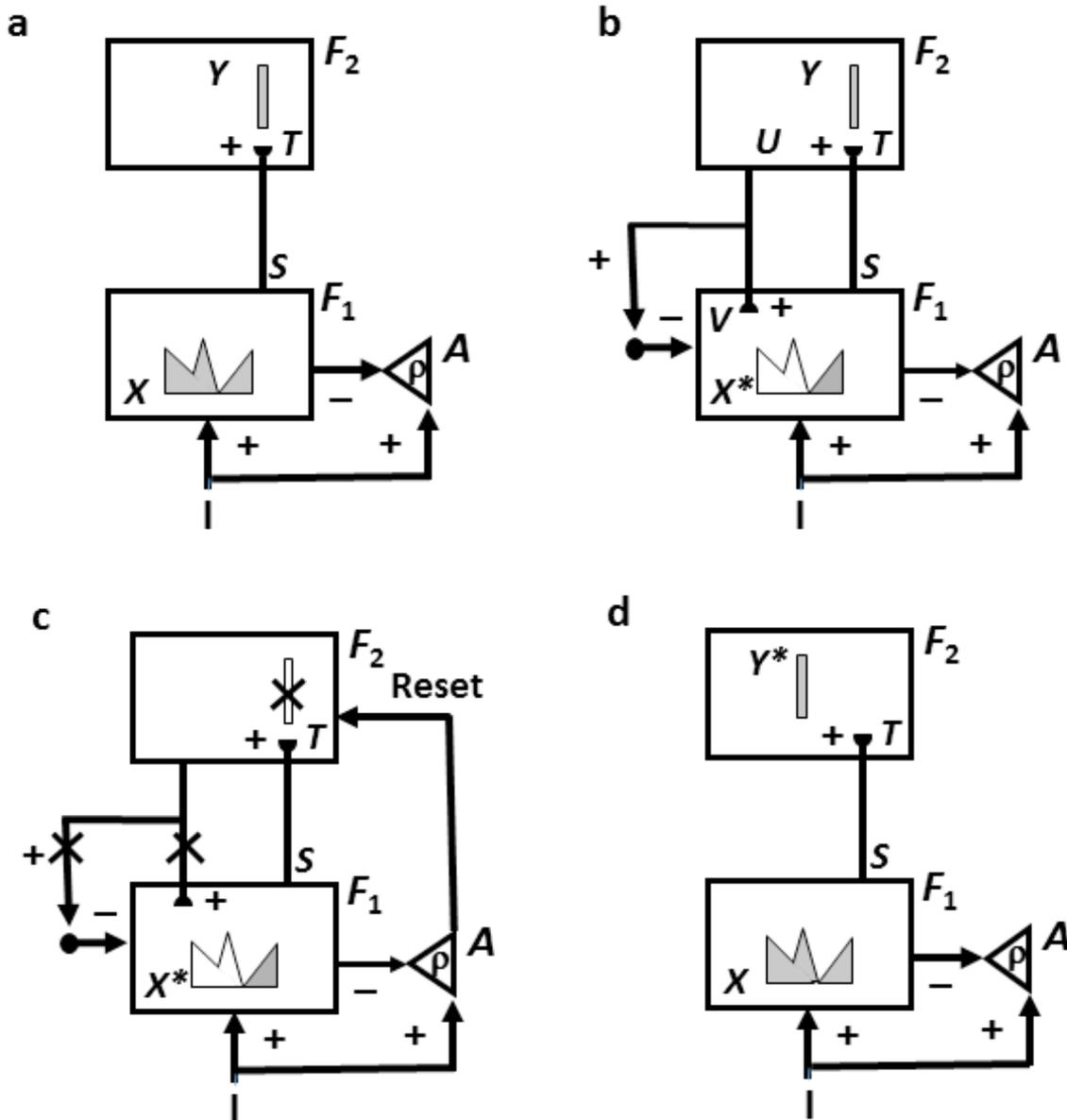
Excitatory matching by object attention is embodied by the ART Matching Rule (Figure 4). This type of attentional circuit enables us to prime our expectations to anticipate objects and events before they occur, and to focus attention upon expected objects and events when they do occur. Good enough matches between expected and actual events trigger resonant states that can support learning of new recognition categories and refinement of old ones, while also triggering conscious recognition of the *critical feature patterns* that are attended and enable recognition to occur. Excitatory matching also controls reset of the attentional focus

when bottom-up inputs significantly mismatch currently active top-down expectations. Cycles of resonance and reset underlie much of the brain's perceptual and cognitive dynamics (Figure 5).

These matching and learning laws have been articulated as part of Adaptive Resonance Theory, or ART, which has been systematically developed since it was first reported in 1976 (Grossberg, 1976a, 1976b). ART is a cognitive and neural theory of how the brain autonomously learns to attend, recognize, and predict objects and events in a changing world. ART is currently the most highly developed cognitive and neural theory available, with the broadest explanatory and predictive range. Central to ART's predictive power is its ability to carry out fast, incremental, and stable unsupervised and supervised learning in response to a changing world. ART specifies mechanistic links between processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony (the CLEARs processes) during both unsupervised and supervised learning. I have predicted that all brains that can solve the stability-plasticity dilemma do so using these predicted links between CLEARs processes. Indeed, my 41-year old prediction that "all conscious states are resonant states" is consistent with all the data that I know, and has helped to explain many data about consciousness, as will be briefly noted below.

ART hereby contributes to functional and mechanistic explanations of such diverse topics as 3D vision and figure-ground perception in natural scenes; optic-flow based navigation in natural scenes towards goals around obstacles and spatial navigation in the dark; invariant object and scenic gist learning, recognition, and search; prototype, surface, and boundary attention; gamma and beta oscillations during cognitive dynamics; learning of entorhinal grid cells and hippocampal place cells, including the use of homologous spatial and temporal mechanisms in the medial entorhinal-hippocampal system for spatial navigation and the lateral stream for adaptively timed cognitive-emotional learning; breakdowns in attentive vigilance during autism, medial temporal amnesia, and Alzheimer's disease; social cognitive abilities such as the learning of joint attention and the use of tools from a teacher, despite the different coordinate systems of the teacher and learner; a unified circuit design for all item-order-rank working memories that enable stable learning of recognition categories, plans, and expectations for the representation and control of sequences of linguistic, spatial, and motor information; conscious speech percepts that are influenced by future context; auditory streaming in noise during source segregation; and speaker normalization that enables language learning from adults after a critical period of babbled sounds by a child; cognitive-emotional dynamics that direct motivated attention towards valued goals; and adaptive sensory-motor control circuits, such as those that coordinate predictive smooth pursuit and saccadic eye movements, and coordinate looking and reaching movements. Brain regions that are functionally described include visual and auditory neocortex; specific and nonspecific thalamic nuclei; inferotemporal, parietal, prefrontal, entorhinal, hippocampal, parahippocampal, perirhinal, and motor cortices; frontal eye fields; supplementary eye fields; amygdala; basal ganglia; cerebellum; and superior colliculus.

ART does not, however, describe many spatial and motor behaviors. These processes typically use different matching and learning laws. ART is thus not "a theory of everything".



**Figure 5.** How ART searches for and learns a new recognition category using cycles of match-induced resonance and mismatch-induced reset. (a) Input pattern  $I$  is instated across feature detectors at level  $F_1$  as an activity pattern  $X$ , at the same time that it generates excitatory signals to the orienting system  $A$  with a gain  $\rho$  that is called the *vigilance* parameter. Activity pattern  $X$  generates inhibitory signals to the orienting system  $A$  as it generates a bottom-up input pattern  $S$  to the category level  $F_2$ . A dynamic balance within  $A$  between excitatory inputs from  $I$  and inhibitory inputs from  $S$  keeps  $S$  quiet. The bottom-up signals in  $S$  are multiplied by learned adaptive weights to form the input pattern  $T$  to  $F_2$ . The inputs  $T$  are contrast-enhanced and normalized within  $F_2$  by recurrent lateral inhibitory signals that obey the membrane equations of neurophysiology, otherwise called shunting interactions. This competition leads to selection and activation of a small number of cells within  $F_2$  that receive the largest inputs. The chosen cells represent the category  $Y$  that codes

for the feature pattern at  $F_1$ . In this figure, a winner-take-all category is shown. (b) The category activity  $Y$  generates top-down signals  $U$  that are multiplied by adaptive weights to form a prototype, or critical feature pattern,  $V$  that encodes the expectation that the active  $F_2$  category has learned for what feature pattern to expect at  $F_1$ . This top-down expectation input  $V$  is added at  $F_1$  cells. If  $V$  mismatches  $I$  at  $F_1$ , then a new STM activity pattern  $X^*$  (the hatched pattern), is selected at cells where the patterns match well enough. In other words,  $X^*$  is active at  $I$  features that are confirmed by  $V$ . Mismatched features (white area) are inhibited. When  $X$  changes to  $X^*$ , total inhibition decreases from  $F_1$  to  $A$ . (c) If inhibition decreases sufficiently,  $A$  releases a nonspecific arousal burst to  $F_2$ ; that is, “novel events are arousing”. The vigilance parameter  $\rho$  determines how bad a match will be tolerated before a burst of nonspecific arousal is triggered. This arousal burst triggers a memory search for a better-matching category, as follows: Arousal resets  $F_2$  by inhibiting  $Y$ . (d) After  $Y$  is inhibited,  $X$  is reinstated and  $Y$  stays inhibited as  $X$  activates a different category, that is represented by a different activity pattern  $Y^*$ , at  $F_2$ . Search continues until a better matching, or novel, category is selected. When search ends, an attentive resonance triggers learning of the attended data in adaptive weights within both the bottom-up and top-down pathways. As learning stabilizes, inputs  $I$  can activate their globally best-matching categories directly through the adaptive filter, without activating the orienting system. [Adapted with permission from Carpenter and Grossberg (1993).]

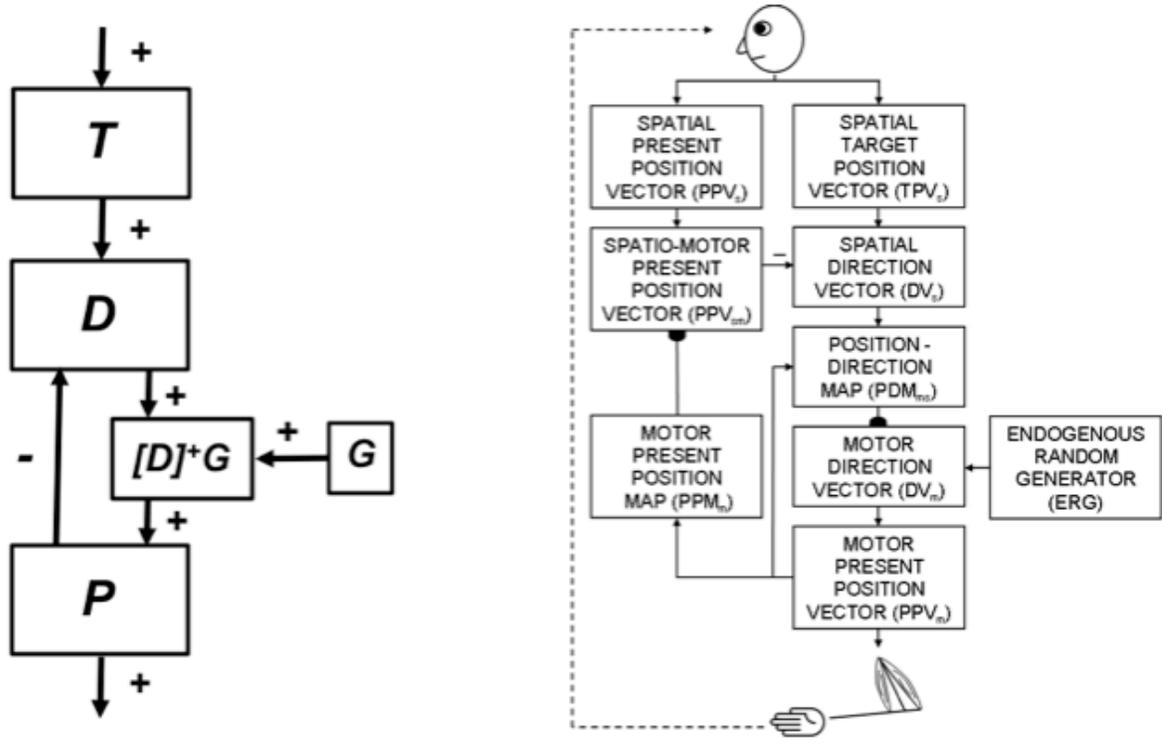
### **Vector Associative Maps for spatial representation and action**

Complementary spatial/motor processes in the Where stream often use inhibitory matching and mismatch-based learning to continually update spatial maps and sensory-motor gains that can effectively control our changing bodies throughout life. Inhibitory matching can take place between representations of where we want to move and where we are now (Figure 6), so that when we arrive at where we want to be, the match equals zero. Inhibitory matching by the Vector Associative Map, or VAM, Matching Rule thus cannot solve the stability-plasticity dilemma (Gaudiano and Grossberg, 1991, 1992). That is why spatial and motor representations cannot support conscious qualia. Instead, spatial maps and motor gains experience catastrophic forgetting as they learn how to accurately control our changing bodies throughout life.

Together these complementary processes create a self-stabilizing perceptual/cognitive front end in the What stream for learning about the world and becoming conscious of it, while it intelligently commands more labile spatial/motor processes in the Where stream that control our changing bodies.

### **Homologous laminar cortical circuits for all biological intelligence: Beyond Bayes**

The second computational paradigm is called Laminar Computing (Grossberg, 1999, 2013b, 2017b; Grossberg, Mingolla, and Ross, 1997). Laminar Computing describes how the cerebral cortex is organized into layered circuits whose specializations support all higher-order biological intelligence. Indeed, the laminar circuits of cerebral cortex seem to realize a revolutionary computational synthesis of the best properties of feedforward and feedback processing, digital and analog processing, and data-driven bottom-up processing and hypothesis-driven top-down processing (Grossberg, 2007, 2013b). For example, in response to an unambiguous scene, a fast feedforward sweep can occur through the entire visual hierarchy, leading to rapid recognition, as reported by Thorpe, Fize, and Marlot (1996). Such



**Figure 6.** (left panel) Vector Integration To Endpoint, or VITE, model circuit for reaching. A present position vector ( $P$ ) is subtracted from a target position vector ( $T$ ) to compute a difference vector ( $D$ ) that represents the distance and direction in which the arm must move. The rectified difference vector ( $[D]$ ) is multiplied by a volitional GO signal ( $G$ ) before the velocity vector  $[D]G$  is integrated by  $P$  until  $P$  equals  $T$ , hence the model name Vector Integration to Endpoint. [Adapted with permission from Bullock and Grossberg (1988).] (right panel) DIRECT model circuit. This refinement of VITE processing enables the brain to carry out what is called motor equivalent reaching, in particular to move a tool under visual guidance to its correct endpoint position on the first try, without measuring the dimensions of the tool or the angle that it makes with the hand. DIRECT hereby clarifies how a spatial affordance for tool use may have arisen from the ability of the brain to learn how to reach during infant development. In DIRECT, this developmental process uses an endogenous random generator, or ERG, to provide the “energy” during a critical period of motor babbling to activate a motor direction vector ( $DV_m$ ) that moves the hand/arm via the motor present position vector ( $PPV_m$ ). As the hand/arm moves, the eyes reactively track the position of the moving hand, giving rise to the visually-activated spatial target position vector ( $TPV_s$ ) and spatial present position vector ( $PPV_s$ ), which coincide during reactive tracking. These vectors are used to compute the spatial difference vector ( $DV_s$ ). It is this spatial transformation, along with the mapping from spatial directions into motor directions, that gives the model its motor equivalent reaching capabilities. To compute these transformations, the  $PPV_s$  first activates the spatio-motor present position vector ( $PPV_{sm}$ ), which is then subtracted from the  $TPV_s$ . As a result, the  $PPV_s$  signal that reaches the  $TPV_s$  is slightly delayed, thereby enabling the  $DV_s$  computation to occur. The  $PPV_{sm}$  stage is one of two stages in the model where spatial and motor representations are combined. The subscripts “s” and “m” denote spatial and motor, respectively. A transformation, called a *circular reaction*

(Piaget, 1945, 1951, 1952), is learned from spatial-to-motor and motor-to-spatial representations at the two adaptive pathways in the model, which are denoted by hemispherical synapses. In particular, the spatial direction vector ( $DV_s$ ) is adaptively mapped into the motor direction vector ( $DV_m$ ), thereby carrying out the transformation from visual *Direction Into* joint *Rotation* that gives the DIRECT model its name. [Reprinted with permission from Bullock, Grossberg, and Guenther (1993).]

a feedforward sweep can occur, for example, in the LAMINART architecture (Grossberg and Raizada, 2000; Raizada and Grossberg, 2001) that is shown in Figure 7 by leaping from retina to the lateral geniculate nucleus (LGN), then through layers 6, 4, and 2/3 in cortical area V1 to layers 6, 4, and 2/3 in cortical area V2, and beyond.

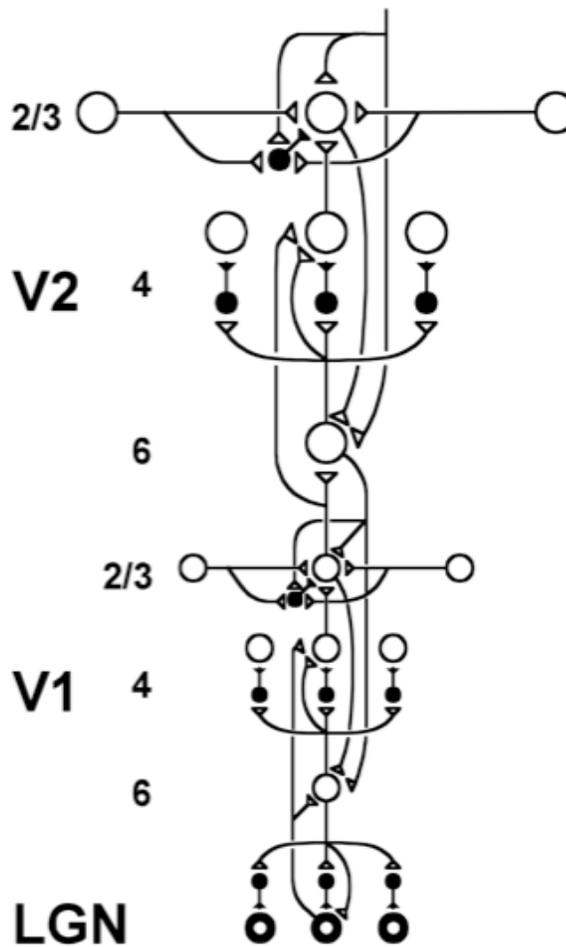
If, however, a scene contains ambiguous information, for example in the form of multiple possible groupings of the same sets of features in a complex textured scene, then the network can automatically use its feedback loops to make the best decision in the face of this uncertainty. In particular, competition among these groupings can occur due to inhibitory interneurons in layers 4 and 2/3 (black cells and synapses in Figure 7). This competition can cause all cell activities to become smaller because the competitive circuits in the model are *self-normalizing*; that is, they tend to conserve the total activity of the circuit. This self-normalizing property arises from the ability of the shunting on-center off-surround networks that realize the competitive circuits to process input contrasts over a large dynamic range without saturation, and thereby solve what I have called the *noise-saturation dilemma* (Grossberg, 1973, 1980).

Self-normalizing competition among alternative cortical interpretations of the data may hereby reduce the activation amplitude and coherence of each grouping alternative, thereby slowing down its processing. This slowing down of processing rate occurs as interlaminar, but intracortical, feedback between layers 2/3-to-6-to-4-to-2/3 (Figure 7), among other feedback pathways, contrast-enhances and amplifies the grouping that is supported by the most evidence. The amplification of the winning grouping's activity automatically speeds up its ability to send output signals to the next cortical region.

This example illustrates an important sense in which the cortex “runs as fast as it can” in response to the degree of uncertainty in the data, automatically switching from fast feedforward processing in response to unambiguous data to slower feedback processing to resolve uncertainties in the data to the degree that the data allow. Our brains hereby go beyond current Bayesian models to implement a kind of real-time probability theory and hypothesis testing that trades uncertainty against speed to make the best decisions in response to probabilistic environments whose rules can change rapidly through time.

Figure 7 also depicts how the ART Matching Rule circuit in Figure 4 is realized within the laminar circuits of neocortex. For example, the top-down pathway from layer 6 in V2 projects to layer 6 in V1, which sends bottom-up signals to layer 4. These bottom-up signals are sent via a modulatory on-center (note the balanced excitatory and inhibitory pathways to layer 4) surrounded by a driving off-surround network. The top-down signals from V2 are hereby “folded” at layer 6 in V1 in order to reach layer 4. I have accordingly called this property *folded feedback*.

Because the ART Matching Rule is realized within laminar neocortical circuits, they can solve the stability-plasticity dilemma and support rapid learning and stable memory.



**Figure 7.** The LAMINART model clarifies how bottom-up, horizontal, and top-down interactions within and across cortical layers in V1 and V2 interblob and pale stripe regions, respectively, carry out bottom-up adaptive filtering, horizontal grouping, and top-down attention to carry out perceptual grouping, including boundary completion. Similar interactions seem to occur in all six-layered cortices. See text for details. [Reprinted with permission from Raizada and Grossberg (2001).]

Figure 7 also shows that bottom-up signals from the LGN use the same modulatory on-center, off-surround network to activate layer 4 in V1 that is used by the top-down attentional feedback pathway. In addition, there is a direct bottom-up excitatory pathway from LGN to layer 4 so that the LGN can activate V1 in response to inputs from the retina. Taken together, the direct LGN-to-4 pathway and the LGN-to-6-to-4 modulatory on-center, off-surround network ensure that bottom-up inputs from the LGN to V1 are contrast-normalized at layer 4 cells.

The sharing of the layer 6-to-4 modulatory on-center, off-surround network by bottom-up and top-down pathways converts this network into a decision interface where pre-attentive automatic bottom-up processing and attentive task-selective top-down processing can cooperate and compete to choose the combination of signals that is most salient at any given moment.

Such a cooperative-competitive decision interface exists in every granular neocortical area. As a result, a top-down task-selective priming signal from a higher cortical area can propagate through multiple lower cortical areas via their layers 6, which can then activate their layer 6-to-4 modulatory on-center, off-surround networks. In this way, an entire cortical hierarchy may get ready to process incoming bottom-up signals to accommodate the bias imposed by the prime.

Figure 7 also shows that layer 2/3 in each cortical area also projects back to layer 6, and then up to layer 4 via the folded feedback network. The horizontal connections in layer 2/3 carry out a variety of functions in different cortical areas. In V2, they carry out perceptual grouping and boundary completion (von der Heydt, Peterhans, and Baumgartner, 1984), a process whose so-called *bipole grouping* properties were predicted before the neurophysiological data of von der Heydt et al. were reported (Cohen and Grossberg, 1984; Grossberg, 1984) and which were subsequently extensively modeled by LAMINART (e.g., Grossberg and Raizada, 2000; Grossberg and Yazdanbakhsh, 2005; Leveille, Versace, and Grossberg, 2010; Raizada and Grossberg, 2001). In cognitive processing regions, such as the ventrolateral prefrontal cortex, it has been suggested that such horizontal connections enable learning of categories, also called *list chunks*, that respond selectively to sequences of items that are stored in working memory (Grossberg and Pearson, 2008; Kazerounian and Grossberg, 2014).

The development of these horizontal connections begins before birth and continues in response to the statistics of visual environments after birth. The fact that the layer 2/3-to-6-to-4-to-2/3 pathway satisfies the ART Matching Rule enables this development, as well as that of other cortical circuits, to dynamically self-stabilize even before higher cortical areas are developed enough to send reliable top-down intercortical attentional signals with which to further stabilize it. Thus “cells that fire together can wire together” without risking catastrophic forgetting in these laminar cortical circuits. I like to describe this property by saying that “the preattentive perceptual grouping is its own attentional prime” (Grossberg, 1999).

The above combination of properties illustrates how parsimoniously and elegantly laminar cortical circuits carry out their multi-faceted functions.

**The same canonical laminar design models vision, speech, and cognition: VLSI!** Even elegant model designs must also support intelligent behavioral functions in order to provide compelling explanations of how brains work, and a guide for new technological developments. In fact, variations of the LAMINART cortical design have, to the present, been naturally embodied in laminar cortical models of vision, speech, and cognition that explain and predict psychological and neurobiological data that other models have not yet handled. These models include the 3D LAMINART model of 3D vision and figure-ground separation (e.g., Cao and Grossberg, 2005, 2012; Fang and Grossberg, 2009; Grossberg and Howe, 2003; Grossberg and Swaminathan, 2004; Grossberg and Yazdanbakhsh, 2005; Grossberg et al., 2008; Leveille, Versace, and Grossberg, 2010), the cARTWORD model of conscious speech perception (Grossberg and Kazerounian, 2011; Kazerounian and Grossberg, 2014), and the LIST PARSE model of cognitive working memory and chunking (Grossberg and Pearson, 2008; Silver et al., 2011).

These models illustrate how *all* neocortical areas combine bottom-up, horizontal, and top-down interactions that embody variations of the same canonical laminar cortical circuitry

that is illustrated by Figure 7. These specialized laminar architectures hereby provide a blueprint for a general-purpose VLSI chip set whose specializations may be used to embody different kinds of biological intelligence as part of an autonomous adaptive agent. From the perspective of ART as a biological theory, they also illustrate how different resonances may use similar circuits to support different conscious experiences, as I will note in greater detail below.

### **Why a unified theory is possible: Equations, modules, and architectures**

There are several fundamental mathematical reasons why it is possible for human scientists to discover a unified mind-brain theory that links brain mechanisms and psychological functions, and to demonstrate how similar organizational principles and mechanisms, suitably specialized, can support conscious qualia across modalities.

One reason for such inter-modality unity is that a small number of equations suffices to model all modalities. These include equations for short-term memory, or STM; medium-term memory, or MTM; and long-term memory, or LTM, that I published in *The Proceedings of the National Academy of Sciences* in 1968. See Grossberg (2013a, 2013b) for recent reviews of these equations.

These equations are used to define a somewhat larger number of modules, or microcircuits, that are also used in multiple modalities where they can carry out different functions within each modality. These modules include shunting on-center off-surround networks, gated dipole opponent processing networks, associative learning networks, spectral adaptively-timed learning networks, and the like. Each of these types of modules exhibits a rich, but not universal, set of useful computational properties. For example, shunting on-center off-surround networks can carry out properties like contrast-normalization, including discounting the illuminant; contrast-enhancement, noise suppression, and winner-take-all choice; short-term memory and working memory storage; attentive matching of bottom-up input patterns and top-down learned expectations; and synchronous oscillations and traveling waves.

Finally, these equations and modules are specialized and assembled into modal architectures, where “modal” stands for different modalities of biological intelligence, including architectures for vision, audition, cognition, cognitive-emotional interactions, and sensory-motor control.

An integrated self or agent, with autonomous adaptive capabilities, is possible because it builds on a shared set of equations and modules within modal architectures that can interact seamlessly together.

Modal architectures are *general-purpose*, in the sense that they can process any kind of inputs to that modality, whether from the external world or from other modal architectures. They are also *self-organizing*, in the sense that they can autonomously develop and learn in response to these inputs. Modal architectures are thus less general than the von Neumann architecture that provides the mathematical foundation of modern computers, but much more general than a traditional AI algorithm. ART networks form part of several different modal architectures, including modal architectures that enable seeing, hearing, feeling, and knowing.

### **All conscious states are resonant states**

ART resonances clarify questions such as the following, which have been raised by distinguished philosophers (Grossberg, 2017b): What kind of “event” occurs in the brain

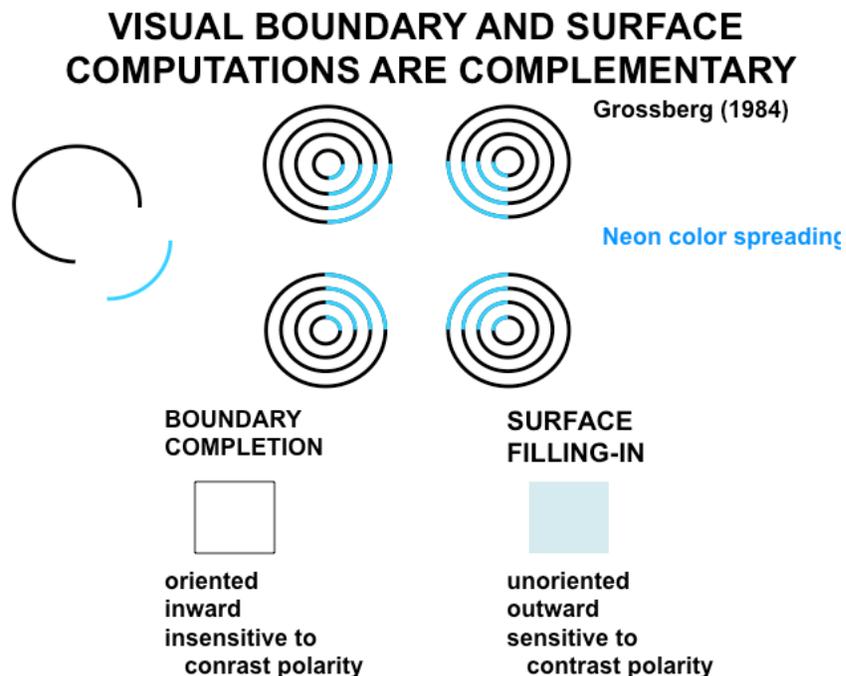
during a conscious experience that is anything more than just a "whirl of information-processing"? What happens when conscious mental states "light up" and directly appear to the subject? ART explains that, over and above "just" information processing, our brains sometimes go into a context-sensitive *resonant state* that can involve multiple brain regions. Abundant experimental evidence support the ART prediction that "all conscious states are resonant states". Not all brain dynamics are "resonant", and thus consciousness is not just a "whirl of information-processing".

Second, when does a resonant state embody a conscious experience? And how do different resonant states support different kinds of conscious qualia? The other side of the coin is equally important: When does a resonant state fail to embody a conscious experience? ART explains (Grossberg, 2017b) how various evolutionary challenges that advanced brains face in order to adapt to changing environments in real time have been met with particular conscious states, which form part of larger adaptive behavioral capabilities. ART sheds new mechanistic light on the fact that humans are not conscious just to Platonically contemplate the beauty of the world. Rather, humans are conscious in order to enable them to better adapt to the world's changing demands. To illustrate these claims, ART explains how resonances for conscious seeing help to ensure effective looking and reaching, resonances for conscious hearing help to ensure effective speaking, and resonances for conscious feeling help to ensure effective goal-directed action.

### **The varieties of brain resonances and the conscious experiences that they support**

Towards this end, ART has explained six different types of neural representations of conscious qualia, and has provided enough theoretical background and data explanations based on these representations to illustrate their explanatory and predictive power (Grossberg, 2017b). These explanations also suggest multiple kinds of experiments to deepen our mechanistic understanding of the brain mechanisms for generating conscious resonances.

For example, *surface-shroud resonances* are predicted to support conscious percepts of visual qualia. *Feature-category resonances* are predicted to support conscious recognition of visual objects and scenes. Both kinds of resonances may synchronize during conscious seeing and recognition, so that we know what a familiar object is when we consciously see it. *Stream-shroud resonances* are predicted to support conscious percepts of auditory qualia. *Spectral-pitch-and-timbre resonances* are predicted to support conscious recognition of sources in auditory streams. Stream-shroud and spectral-pitch-and-timbre resonances may synchronize during conscious hearing and recognition of auditory streams, so that we know what the familiar sounds are that are segregated in a stream. *Item-list resonances* are predicted to support recognition of speech and language. They may synchronize with stream-shroud and spectral-pitch-and-timbre resonances during conscious hearing of speech and language, and build upon the selection of auditory sources by spectral-pitch-and-timbre resonances in order to recognize the acoustical signals that are grouped together within these streams. *Cognitive-emotional resonances* are predicted to support conscious percepts of feelings, as well as recognition of the source of these feelings. Cognitive-emotional resonances can also synchronize with resonances that support conscious qualia and knowledge about them. All of these resonances have distinct anatomical substrates that are explained in Grossberg (2017b), which also explains various psychological and neurobiological data from typical and clinical individuals.



**Figure 8.** Complementary computational properties of visual boundaries and surfaces. Visual boundaries and surfaces are computed by the interblob and blob cortical processing streams, respectively, that occur within and between cortical areas V1, V2, and V4. An illusory square is completed in response to the configuration of black and blue arcs that form the image. When this happens, breaks, called end gaps, occur in the boundaries where the black arcs touch the blue arcs. Blue color can then flow out of the blue arcs to fill-in the interior of the illusory square. The resulting percept of neon color spreading illustrates complementary properties of boundary completion and surface filling-in (Grossberg, 1984; Grossberg and Mingolla, 1985), namely: Boundaries are completed in an *oriented* way, *inward* between pairs or greater numbers of inducers with similar orientational preferences, and are *insensitive* to contrast polarity, in the sense that they pool over opposite contrast polarities using V1 complex cells at each position. In contrast, surfaces fill-in brightness and surface color in an *unoriented* way, *outward* from individual contrastive inducers, and are *sensitive* to contrast polarity, indeed support the visible qualia that observers can consciously see. These three pairs of boundary and surface properties (oriented vs. unoriented, inward vs. outward, insensitive vs. sensitive to contrast polarity) are manifestly complementary. Properties that are needed to complete a boundary cannot be used to fill-in a surface, and conversely. On the other hand, boundaries and surfaces need to reciprocally interact across the interblob and blob stream for either of them to generate a useful percept.

### Why does resonance trigger consciousness?

Detailed analyses of psychological and neurobiological data by ART clarify why resonance is necessary for consciousness. As one example: In order to fully compute visual boundaries and surfaces whereby to see the world, the brain computes three pairs of complementary

computational properties of boundaries and surfaces (Figure 8), along with three hierarchical resolutions of uncertainty that require multiple processing stages to overcome. This example illustrates that there is a great deal of uncertainty in the early stages of visual processing by the brain. Only after all three hierarchical resolutions of uncertainty are complete, and after boundaries are completed and surfaces filled-in, has the brain constructed a contextually informative and temporally stable enough representation of scenic objects on which to base adaptive behaviors.

If this is indeed the case, then why do not the earlier stages undermine behavior? The proposed answer is that *brain resonance, and with it conscious awareness, is triggered at the processing stage that represents visual boundary and surface representations, after they are complete and stable enough to control visually-based behaviors like attentive looking and reaching*. ART also explains how, after such a resonance is triggered between prestriate visual cortex and parietal cortex, it can propagate bottom-up to higher cortical areas, such as prefrontal cortex, and top-down to earlier cortical and LGN processing areas, using the ART Matching Rule (Figure 4) to select data that consistent with the triggering resonance and to suppress inconsistent information.

### **Towards autonomous adaptive intelligent agents and clinical therapies in society**

The above summary suggests that a firm foundation has been build over the past 50 years whereby discoveries about mind and brain can greatly influence the development of technologies that can have a profound impact on society, and can facilitate a deeper mechanistic understanding of several major mental disorders. The technological developments will include increasingly autonomous adaptive agents, whereas new clinical therapies for mental disorders will benefit from understanding the neural mechanisms can cause their behavioral symptoms.

## References

- Bullock, D., and Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review*, 95, 49-90.
- Bullock, D., Grossberg, S., and Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, 5, 408-435.
- Cao, Y., and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515-578.
- Cao, Y., and Grossberg, S. (2012). Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: A method of converting neural rate models into spiking models. *Neural Networks*, 26, 75-98.
- Carpenter, G. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16, 131-137.
- Cohen, M.A., and Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, 36, 428-456.
- Fang, L. and Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45-82.
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing by the primate cerebral cortex. *Cerebral Cortex*, 1, 1-47.
- Gaudiano P., and Grossberg S. (1991). Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories. *Neural Networks*, 4, 147-183.
- Gaudiano, P., and Grossberg, S. (1992). Adaptive vector integration to endpoint: Self-organizing neural circuits for control of planned movement trajectories. *Human Movement Science*, 11, 141-155.
- Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20-15.
- Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121-134.
- Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187-202.
- Grossberg, S. (1984). Outline of a theory of brightness, color, and form perception. In E. Degreef and J. van Buggenhaut (Eds.), *Trends in mathematical psychology*. Amsterdam: North-Holland, pp. 59-86.
- Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12, 163-186.
- Grossberg, S. (2000a). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4, 233-246.
- Grossberg, S. (2000b). The imbalanced brain: From normal behavior to schizophrenia. *Biological Psychiatry*, 48, 81-98.
- Grossberg, S. (2007). Consciousness CLEARs the mind. *Neural Networks*, 20, 1040-1053.
- Grossberg, S. (2013a). Adaptive resonance theory. *Scholarpedia*, [http://www.scholarpedia.org/article/Adaptive\\_resonance\\_theory](http://www.scholarpedia.org/article/Adaptive_resonance_theory).

- Grossberg, S. (2013b). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 37, 1-47.
- Grossberg, S. (2017a). Acetylcholine neuromodulation in normal and abnormal learning and memory: Vigilance control in waking, sleep, autism, amnesia, and Alzheimer's disease. *Frontiers in Neural Circuits*, November 2, 2017, <https://doi.org/10.3389/fncir.2017.00082>
- Grossberg, S. (2017b). Towards solving the Hard Problem of Consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks*, 87, 38-95.
- Grossberg, S., and Howe, P.D.L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, 43, 801-829.
- Grossberg, S., and Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: A neural model of phonemic restoration using subsequent context in noise. *Journal of the Acoustical Society of America*, 130, 440-460.
- Grossberg, S., and Kishnan, D. (2017). Neural dynamics of autistic repetitive behaviors and Fragile X syndrome: Basal ganglia movement gating and mGluR-modulated adaptively-timed learning. Submitted for publication.
- Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92, 173-211.
- Grossberg, S., Mingolla, E., and Ross, W.D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences*, 20, 106-111.
- Grossberg, S., and Pearson, L. (2008). 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.
- Grossberg, S., and Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
- Grossberg, S., and Seidman, D. (2006). Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates. *Psychological Review*, 113, 483-525.
- Grossberg, S., and Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention and bistability. *Vision Research*, 44, 1147-1187.
- Grossberg, S., and Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725-1743.
- Grossberg, S., Yazdanbakhsh, A., Cao, Y., and Swaminathan, G. (2008). How does binocular rivalry emerge from cortical mechanisms of 3-D vision? *Vision Research*, 48, 2232-2250.
- Kazerounian, S., and Grossberg, S. (2014). Real-time learning of predictive recognition categories that chunk sequences of items stored in working memory. *Frontiers in Psychology: Language Sciences*, doi: 10.3389/fpsyg.2014.01053.
- Leveille, J., Versace, M., and Grossberg, S. (2010). Running as fast as it can: How spiking dynamics form object groupings in the laminar circuits of visual cortex. *Journal of Computational Neuroscience*, 28, 323-346.

- Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions Royal Society of London B*, 298, 85–95.
- Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414-417.
- Piaget, J. (1945). *La Formation du Symbole Chez L'enfant*. Paris: Delachaux Niestle, S.A.
- Piaget, J. (1951). Play, dreams and imitation in childhood (C. Gattegno and C. F. M. Hodgson, Trans.). London: Roudedge and Kegan Paul.
- Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International Universities Press.
- Raizada, R., and Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, 8, 431-466.
- Silver, M.R., Grossberg, S., Bullock, D., Histed, M.H., and Miller, E.K. (2011). A neural model of sequential movement planning and control of eye movements: Item-order-rank working memory and saccade selection by the supplementary eye fields. *Neural Networks*, 26, 29-58.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- von der Heydt., R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260–1262