Toward Autonomous Adaptive Intelligence: Building Upon Neural Models of How Brains Make Minds

Stephen Grossberg, *Fellow, IEEE*

***Abstract*—This article surveys the development of mathematical laws, circuits, and architectures that model how our brains make our minds, and shows how these contributions provide a blueprint for developing general-purpose autonomous adaptive algorithms and robots for intelligent applications in engineering and technology. The mathematical laws for short-term memory, medium-term memory, and long-term memory that provide the contemporary foundation for all subsequent biological neural network models were published in 1968, followed by a steady stream of model discovery and development to the present time. The Cohen–Grossberg model and theorem that was published in this journal in 1983 was one step in this series of developments. It proved global limit theorems for a general class of neural networks using a Lyapunov function as one tool. These theorems provide a guarantee that learning in these networks generates stable memories. This article surveys additional mathematical foundations for neural network design and applications and describes a modeling method for the incremental discovery of models with increasingly powerful and general functional capabilities. This method is illustrated by the adaptive resonance theory, or ART, architectures which explain how our brains autonomously learn to attend, recognize, and predict objects and events in a changing world, along the way explaining how our brains become conscious, as well as the computational constraints on learning that forced evolution to discover conscious states of mind. Multiple types of resonances support different kinds of conscious awareness and enable the explanation and prediction of large psychological and neurobiological databases about perception, cognition, emotion, and action. Because ART can be derived from a thought experiment about a universal problem of error correction in a changing world, its expanding applications to the development of autonomously intelligent systems should transform future technologies.**

***Index Terms*—Adaptive resonance theory, attention, autonomous adaptive intelligence, categorization, Cohen–Grossberg model, complementary computing, consciousness, content addressable memory, laminar computing, learning, long-term memory, Lyapunov function, medium-term memory, neural network, short-term memory, vigilance control.**

Manuscript received October 26, 2020; accepted November 23, 2020. Date of publication December 30, 2020; date of current version January 12, 2021. This article was recommended by Associate Editor D. C. Wunsch.

The author is with the Center for Adaptive Systems, Graduate Program in Cognitive and Neural Systems, Boston University, Boston, MA 02215 USA, also with the Department of Mathematics and Statistics, Boston University, Boston, MA 02215 USA, also with the Department of Psychological and Brain Sciences, Boston University, Boston, MA 02215 USA, and also with the Department of Biomedical Engineering, Boston University, Boston, MA 02215 USA (e-mail: steve@bu.edu).

Digital Object Identifier 10.1109/TSMC.2020.3041476

I. Introduction: 63 Years of Modeling Biological and Technological Autonomous Intelligence

THE INVITATION for me to contribute to this 50th-anniversary issue of the IEEE TRANSACTIONS OF SYSTEMS, MAN, AND CYBERNETICS encouraged me to look back in time at my personal scientific odyssey during this period. My odyssey began 63 years ago when, as a college Freshman in Dartmouth College taking a course in introductory psychology in 1957, I introduced foundational neural network equations and early models of how our brains make our minds, as reviewed in [1].

This scientific odyssey has continued unabated to the present time with the help of many gifted collaborators, as summarized at sites.bu.edu/steveg. The work has steadily progressed to include ever more comprehensive neural architectures to model essentially all the foundational capabilities of our biological intelligence and has spun off large-scale applications to engineering and technology along the way.

Our brains are unparalleled in generating autonomous adaptive intelligence. Many of the most exciting opportunities for transformative technological change during the coming century will, I believe, derive from developing increasingly autonomous adaptive algorithms and robots, with an understanding of how our own brains work as a reliable source of new design concepts and architectures.

These biological contributions are all amenable to application in engineering and technology because all the models try to explain how brains make minds; that is, how brain mechanisms give rise to psychological functions. To develop any application, one needs to know how it works and what it is for. Brain mechanisms and psychological functions provide that information. In addition, our brains are the best known examples of autonomous adaptively intelligent systems, and engineers and technologists are seeking increasingly autonomous solutions.

This article starts with one of my contributions to this journal in 1983 and ends with an example of the kinds of autonomous adaptive biological architectures that have already been developed and that are ready for technological implementation.

II. How Was the 1983 Cohen–Grossberg Liapunov Function Discovered?

A. Cohen–Grossberg Model and Lyapunov Function

One reason that I was asked to write an article for this special issue is that Michael Cohen and I published an

2168-2216 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

oft-cited article in this journal in 1983 [2]. Our article introduced a general model and Lyapunov function that includes many neural models that I discovered in the 1960s and 1970s, notably the additive and shunting models (e.g., [3]–[6], as well as many other models in the population biology and neural modeling literature. I earlier announced our results in my 1982 review paper [7] after our article was submitted for publication. This contribution has come to be called the Cohen–Grossberg model and Lyapunov function.

Many scientists who currently follow the neural networks literature did not know this literature in the early 1980s. They thus did not realize that the additive model had been introduced into the literature and used extensively in scores of articles starting in the 1960s. Nor did they know that Michael Cohen and I had published in 1983 our mathematical analyses of how the Lyapunov function of the additive model governs global converge of all its trajectories.

John Hopfield published our Lyapunov function for the additive model the next year [8]. Hopfield did not cite our earlier work, although we had reason to believe that he knew about it, as did many of his colleagues at CalTech and Bell Labs, who promoted it without acknowledging our priority. As a result, our discovery and analysis of the additive model and its Lyapunov function has been called the Hopfield model by many people.

To correct this historical error, while acknowledging Hopfield's effective popularization of it, various authors have called the model the Cohen–Grossberg–Hopfield model [9]–[14]. In this article, I will call the model the Cohen–Grossberg model because Hopfield did not contribute at all to the model in this generality, and did not prove any theorems about even the additive model.

The Cohen–Grossberg model is defined by the following system of nonlinear differential equations that describe the interactions among and between activities x_i, or short-term memory (STM) traces, of any finite number of cells or cell populations

$$dx_i/dt = a_i(x_i)\left[b_i(x_i) - \sum_j c_{ij}d_j(x_j)\right] \tag{1}$$

with symmetric interaction coefficients $c_{ij} = c_{ji}$ and weak assumptions on the state-dependent non-negative amplification functions $a_i(x_i)$, self-signal functions $b_i(x_i)$, and competitive interaction functions $d_j(x_j)$. Subscripts $i, j = 1, 2, \ldots, n$ and n can be chosen arbitrarily large. Each population in (1) can have its own, possibly very different, functions $a_i(x_i)$, $b_i(x_i)$, and $d_j(x_j)$. One possible physical interpretation of symmetric interaction coefficients $c_{ij} = c_{ji}$ is that the competitive interactions depend upon the distances between the populations. By (1), the ith population activity x can only grow, and thus momentarily be "winning the competition," at times when the competitive balance $[b_i(x_i) - \sum_j c_{ij}d_j(x_j)] > 0$ in (1). When $[b_i(x_i) - \sum_j c_{ij}d_j(x_j)] < 0$, the population is "losing the competition."

Michael Cohen and I defined the following Cohen–Grossberg Lyapunov function to prove convergence of all trajectories of the systems defined by (1)

$$V = -\sum_i \int b_i(u_i)d_i'(u_i)du_i + 1/2\sum_{j,k} c_{jk}d_j(x_j)d_k(x_k). \tag{2}$$

The function V in (2) is a Lyapunov function because, when it is differentiated along trajectories of (1), the result is

$$dV/dt = -\sum_i a_i d_i'\left[b_i - \sum_j c_{ij}d_j\right]^2. \tag{3}$$

If all the terms $a_i d_i'$ in (3) are non-negative—as they certainly are because each competitive signal function d_i is an increasing function with a non-negative derivative d_i'—then $dV/dt \leq 0$ along trajectories, so that V(t) can only decrease through time.

With this basic Lyapunov property ensured, it was possible to mathematically prove that every trajectory in all models of Cohen–Grossberg type converges through time to one of possibly infinitely many equilibrium points. This result applies to a very large class of models in mathematical biology and neurobiology, as reviewed in [15].

Our theorem guarantees, in particular, that no trajectories experience sustained oscillations in Cohen–Grossberg models. This fact raises important questions: What is the functional utility of converging trajectories and of sustained oscillations in our brains, or in technologies that are inspired by our brains? This is a question about the relationship between mechanism and function, or more specifically between brain mechanisms and psychological functions. I will discuss this set of issues in Section II.

A point of related historical interest is that Hopfield and Tank [16] described the Lyapunov function for the additive model as "A new concept for understanding the dynamics of neural circuitry" and went on to claim that "a nonsymmetric circuit…has trajectories corresponding to complicated oscillatory behaviors…but as yet we lack the mathematical tools to manipulate and understand them at a computational level." Both statements that the additive model was new, and that it has to oscillate if it has asymmetric coefficients, are incorrect.

Some additive models with asymmetric coefficients do oscillate. For example, in 1978, I proved theorems about how competitive networks with just three competing populations could persistently oscillate if their asymmetric coefficients were chosen correctly. I will say more about these "voting paradox" networks in Section II-F. By the time that the Hopfield and Tank [16] article appeared, a number of neural modelers were concerned about their incorrect priority claims, so *Science* published commentaries by several of us, including ours [17]. In [17], we cited our theorems from as early as 1969 about rather general classes of additive models that can learn with asymmetric coefficients while converging to global equilibria; e.g., [18]–[23].

B. Modeling Method and Cycle

My 1983 theorem with Michael Cohen was not my first effort to prove that global limit and oscillation theorems occur in specific classes of neural networks. Indeed, these theorems

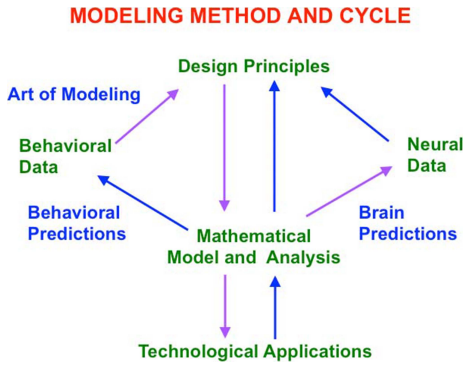

Fig. 1. My modeling method and cycle. See the text for details. (Reprinted with permission from [106].)

were inspired as part of a modeling method that has gone through multiple iterations, or cycles, during the past four decades (Fig. 1). This method led to a flood of discoveries about how brains make minds, as well as multiple large-scale applications in engineering and technology. I will mention various such applications in subsequent sections. The method also creates a conceptual pressure that facilitates future discoveries. What this method is and why it has not hit a brick wall is explained in Section III.

C. Adaptation Level Systems and the ABC Theorem

As part of this theoretical method and cycle, in 1978, I published two mathematical articles that laid the foundation for my discoveries with Michael Cohen. The first article proved a global limit theorem for a class of models that I called adaptation level systems [24]. These systems are in some ways more general, and in other ways less general, than the Cohen–Grossberg model. Together they pointed to open mathematical problems whose solutions are still not available. I mention one of them below.

Adaptation level systems are defined by the following equations:

$$dx_i/dt = a_i(x)[b_i(x_i) - c(x)] \tag{4}$$

where $x = (x_1, x_2, \ldots, x_n)$ is the vector of all the system's activities, and the function $c(x)$ can only increase in response to increases in any of its variables x_i. This is the key property that converts (4) into a competitive system, since increasing one population's activity can only reduce the rate of growth of the other populations' activities via (4).

Comparing (1) and (4) shows why (4) is in some ways more general than (1) and in other ways less general. For example, instead of amplification functions $a_i(x_i)$ in (1) that are influenced only by their own population, the amplification functions $a_i(x)$ in (4) can depend upon all the activities across the system in complex ways. Instead of the competitive interactions $-c_{ij}d_j(x_j)$ in (1) whose strength depends upon the symmetric interaction coefficients $c_{ij} = c_{ji}$, the competition term c(x) in (4) is the same for every population. This term plays the role of an *adaptation level* against which all the self-signal functions $b_i(x_i)$ are compared in (4) to determine which populations are winning or losing at any given time.

As (12) illustrates, the name adaptation level arose in on-center off-surround networks that occur in one form or another throughout our brains. In these networks, the size of the adaptation level, relative to a measure of a given cell's activity, determines if this activity increases or decreases at that time. In other words, an adaptation level is a kind of competition threshold. Accordingly, in multiple examples, the adaptation level is computed by a long-range influence that is shared equally by all the populations, even as it may change in complex ways through time, as can $c(x)$.

The famous mathematician Morris Hirsch, who was then at the University of California at Berkeley, began to call adaptation level systems the ABC model several years after I introduced them, when he worked to generalize my theorem [25]. Moe liked this name because the model equations are defined using three general terms that are denoted by A, B, and C in (4).

Proving the adaptation level theorem required multiple steps that took ten pages to carry out in print. I used a *Lyapunov functional* to help prove that all trajectories converge to limits. This Lyapunov functional was actually defined for a general class of competitive dynamical systems that includes adaptation level systems, namely

$$dx_i/dt = a_i(x)M_i(x). \tag{5}$$

In (5), state-dependent, non-negative amplification functions $a_i(x)$ are multiplied by state-dependent competition functions $M_i(x)$ which can only decrease as any of their variables increase. A Lyapunov functional for system (5) can be defined using the maximum function

$$M^+(x) = \max_j \{M_j(x) : j = 1, 2, \ldots, n\} \tag{6}$$

as follows:

$$L(x_t) = \int^t M^+(x(u))du. \tag{7}$$

Why (7) is a Lyapunov functional for all competitive systems (5) was proved as part of the mathematical analysis in [24] and [26]. The utility of this Lyapunov functional created the impetus to find a Lyapunov functional method to explain why all trajectories converge for the Cohen–Grossberg model. Although Michael Cohen and I failed to do this with Cohen–Grossberg systems, in 1982, we discovered the Cohen–Grossberg Lyapunov function instead.

D. After Competition Ignites, It Can Be Analyzed by Tracking Which Population Is Winning

My 1978 article about adaptation level systems is called *competition, decision, and consensus* because it analyzed the competitive adaptation level systems (4) and introduced a method of jumps to analyze the much larger class of competitive dynamical systems (5). Intuitively, a jump occurs when the system chooses a new winning population. This method quantifies the intuition that one can understand a competitive system by keeping track of who is winning it through time. Each jump is a kind of local "decision" where the system chooses a new winning population. More precisely, a jump

occurs from one population to another as soon as the latter population's activity starts to be the maximally enhanced one. The notion of "consensus" means that, even though the populations compete, they can gradually reach a consensus by all converging to equilibrium.

The method of jumps allowed me to analyze these systems by tracking both the maximum function in (6) and the minimum function

$$M^{-}(x) = \min_j \{M_j(x) : j = 1, 2, \ldots, n\}. \tag{8}$$

I proved that, if $M^{+}(x(T)) \geq 0$ at any time T, then $M^{+}(x(t)) \geq 0$ at all future times $t \geq T$. In other words, if there is a time $t = T$ when any population activity is growing—and thus potentially winning the competition—then one or more population activities will grow at all future times. I like to say that, at time T, the competition *ignites*. It is because ignition persists for all future times that (7) implies

$$dL(x_t)/dt = M^{+}(x(t)) \geq 0 \tag{9}$$

for all times $t \geq T$. Inequality (9) shows that (7) defines a Lyapunov functional after the competition ignites. If the system never ignites, then all population activities can only decrease through time. Since activities are bounded below, they all decrease to limits, thereby proving that all trajectories converge to equilibria in all cases.

E. Adaptation Levels in On-Center Off-Surround Networks and Competitive Economic Markets

Many processes in the real world can be written as an adaptation level system. Here are two that illustrate the wide range of possibilities.

First, consider recurrent on-center off-surround networks of neurons that obey the membrane equations, or shunting laws, of neurophysiology [27]. Such networks, which I first globally analyzed in [6], occur ubiquitously throughout our brains. Perhaps the simplest equation for a shunting network with a narrow on-center and a broad off-surround (Fig. 2) is

$$\begin{aligned} dx_i/dt = &-A_i x_i + (B - x_i)\big[f_i(x_i) + I_i\big] \\ &- (C + x_i)\sum_{j \neq i}\big[f_j(x_j) + I_j\big]. \end{aligned} \tag{10}$$

The excitatory and inhibitory shunting terms $(B - x_i)$ and $(C + x_i)$, respectively, keep the activities bounded and automatically gain-control the excitatory on-center signals $[f_i(x_i) + I_i]$ and the inhibitory off-surround signals $\sum_{ji}[f_j(x_j) + I_j]$, where the $f_i(x_i)$ are feedback signals within the network itself, and the I_i are external inputs.

The simplest adaptation level system is defined by eliminating the feedback signals $f_i(x_i)$ in (10). The resulting feedforward on-center off-surround shunting network can be rewritten in terms of the sum $I = \sum_j I_j$ of all the external inputs

$$dx_i/dt = -A_i x_i + (B - x_i)I_i - (C + x_i)[I - I_i]. \tag{11}$$

Rearranging terms in (11) and defining the ratio $\Theta_i = I_i/I$ of all the inputs leads to

$$dx_i/dt = -(A_i + I)x_i + I(B + C)\big[\Theta_i - C/(B + C)\big]. \tag{12}$$

Term $C/(B + C)$ in (12) is an adaptation level because x_i can only increase if the term $[\Theta_i - C/(B + C)] > 0$, and can only decrease if $[\Theta_i - C/(B + C)] < 0$.

The constraint of symmetric coefficients that was imposed to prove the Cohen–Grossberg theorem is not sufficient to prevent all systems with symmetric coefficients from oscillating. This was shown by Cohen in 1988 [28] for a class of networks with symmetric coefficients $C_{ij} = C_{ji}$ and $D_{ij} = D_{ji}$ that obey

$$dx_i/dt = -A_i x_i + (B_i - x_i)\sum_j C_{ij} f_j(x_j) - x_i \sum_j D_{ij} g_j(x_j). \tag{13}$$

A main difference between (13) and (10) is that the on-center of (13) is broader than a single population; that is, (13) defines a symmetric *cooperative*–competitive system. Cohen [28] proved that oscillations can occur in these systems, even though they have no external inputs. This result showed that symmetric coefficients are not *sufficient* to ensure that all trajectories converge to equilibria. The example of a "directed jump tree," or "generalized pecking order," in Section II-F also shows that symmetric coefficients are not *necessary* to ensure that all trajectories converge to equilibria.

Nor are cooperative interactions necessary to cause oscillations in a system with symmetric coefficients. Inhibitory signals from slowly varying interneuronal activities in the off-surround of a competitive system with a narrow on-center can also cause oscillations if the system also receives an external input, as often occurs *in vivo*. [29] used a Hopf bifurcation theorem [30] to show how such a system bifurcates into a persistent oscillation when the input is chosen sufficiently large. In this situation, the inhibitory signals $f_j(x_j)$ in the off-surround of (10) are replaced by $f_j(y_j)$, where y_j is the activity of an inhibitory interneuron that time averages signals from x_i. Oscillations can thus occur either in cooperative–competitive systems with symmetric coefficients, fast inhibitory interneurons, and no external inputs; or in competitive systems with symmetric coefficients, slow inhibitory interneurons, and external inputs; or a combination of the two.

A qualitatively different example of an adaptation level system defines a stable economic market in which arbitrarily many competing firms all produce the same commodity [15]. The theorem shows how all the firms can achieve price stability and balanced books, even if they have only their own selfish interests in mind, and even if the only information that they have about each other's activities is very indirect, via their effect on the publicly reported competitive market price. Such a competitive market may be defined by

$$dx_i/dt = A_i(x_i)[P(x) - C_i(x_i)]. \tag{14}$$

In (14), $P(x)$ is the competitive market price, where $P(x)$ can only decrease if any of the firms produces more of the commodity. The function $C_i(x_i)$ is the cost per item that each firm determines internally from multiple factors, and can thus be quite a general function of each firm's activities. Each firm produces more of the commodity only if $[P(x) - C_i(x_i)]$ is positive at that time. Since $A_i(x_i)$ is a non-negative state-dependent amplifier that each firm also defines internally, (14) says that

each firm produces more only at times when if it can make a net profit.

The adaptation level theorem proves that *if* all the firms agree to this general market-wide production strategy, then even though their individual production strategies are unknown to their competitors, and only the effects of their activities on the market price are known, not the price function itself, then the market as a whole achieves price stability and every firm balances its books. The theorem hereby provides a rigorous, albeit highly idealized, an example of the invisible hand that Adam Smith described in his classical 1776 book on The Wealth of Nations [31].

The theorem does not, however, determine which of the firms make money and which go out of business. It only provides a collectively agreed upon framework whose good effects on the market provide a context within additional strategies may be practiced in each firm to maximize its profits.

There are also many examples of adaptation level systems in population biology. The following section discusses some of these situations, but under different mathematical constraints.

F. Generalized Volterra–Lotka Systems: Directed Jump Trees Converge, Jump Cycles Oscillate

My second article in 1978 proved theorems about a related class of competitive dynamical systems that I call generalized Volterra–Lotka systems because they generalize the classical Volterra–Lotka systems of population biology, also known as predator–prey equations when they were introduced in 1910 by Alfred Lotka as part of his theory of chemical reactions [32]

$$dx_i/dt = x_i\left[A_i - \sum_j B_{ij}x_j\right]. \tag{15}$$

In each of these systems, global limits occur given certain parameters, but sustained oscillations occur given other parameters [26].

For example, May and Leonard [33] described a Volterra–Lotka system with three interacting variables v_1, v_2, and v_3 whose activities x_1, x_2, and x_3 persistently oscillate through time

$$\begin{aligned} dx_1/dt &= x_1[1 - x_1 - \alpha x_2 - \beta x_3] \\ dx_2/dt &= x_2[1 - \beta x_1 - x_2 - \alpha x_3] \\ dx_3/dt &= x_3[1 - \alpha x_1 - \beta x_2 - x_3]. \end{aligned} \tag{16}$$

They chose the parameters $\alpha < 1 < \beta$ and $\alpha + \alpha \geq 2$ to emulate the "voting paradox" that describes a global "contradiction" due to the fact that, in pairwise competition, population v_1 beats v_2, v_2 beats v_3, and v_3 beats v_1. When all three populations compete at the same time, there is thus no clear winner. As a result, when the activities x_1, x_2, and x_3 of all three populations in system (16) compete, they persistently oscillate, corresponding to the intuition that the system contains an internal contradiction and "cannot make up its mind."

I introduced systems of generalized Volterra–Lotka systems to include these, and many more, possibilities

$$dx_i/dt = A_i(x)\left[1 - \sum_j B_{ij}C_j(x_j)\right]. \tag{17}$$

Using the method of jumps, I was able to prove that if the jumps of an arbitrary competitive system, as in (5), form a "consistent decision scheme," then all trajectories converge, but if the jumps form an "inconsistent decision scheme," then there exist persistent oscillations in the system, as occurs during the voting paradox. By a consistent decision scheme, I mean that the jumps form a *directed decision tree*, or "generalized pecking order," that is constructed by drawing a directed arrow from one jump to the next one. After the last jump is made, the trajectory approaches equilibrium. By an inconsistent decision scheme, I mean that the jumps include one or more *directed jump cycles*. The corresponding trajectories then oscillate through time. I also imposed conditions to ensure that there was enough "energy" to keep the oscillation amplitudes from converging to zero. There is enough energy, for example, if $L(x_\infty) = \infty$, where $L(x_t)$ is defined by (7).

Not all oscillations are periodic in time. For example, May and Leonard [33] noted that their voting paradox oscillations had an ever longer cycle as time went on. They offered the qualitative explanation that the oscillating solutions approach a set of straight lines through time. I showed that, instead, the oscillations approach curved heteroclinic solutions, where a heteroclinic solution approaches one limit point as $t \to -\infty$ and another limit point as $t \to \infty$.

G. Unsolved Conjecture: Does Every Cohen–Grossberg System Induce Directed Jump Tree?

As Cohen [28] illustrates, many facts about how to design cooperative–competitive systems to achieve trajectory convergence or oscillations have not yet been mathematically proved. Here is one of them.

Given that all competitive systems with directed jump trees exhibit global convergence, one might conjecture that all Cohen–Grossberg systems in (1) with symmetric coefficients induce directed jump trees. Such a theorem would prove global convergence without needing a Lyapunov function. It could also be used to prove the more general theorem that Cohen–Grossberg systems whose coefficients may not be symmetric, but are close to being symmetric, also have only convergent trajectories. By studying the jump sets of such systems, one could prove much more by analyzing what parameter changes convert directed jump trees into directed jump cycles, and conversely. Such a unified analysis is much to be desired.

H. Can Competitive Systems Do Everything?

Cohen [28] considered cooperative–competitive systems in which each population can cooperate with other populations via the broad on-center coefficients C_{ij} in (11). This example raises the question: Are cooperative interactions *necessary* to achieve some kinds of system dynamics?

This question has been answered by another famous mathematician, Stephen Smale at the University of California in

Berkeley, and a collaborator of Morris Hirsch. Smale [34] showed by a simple construction "that the ordinary differential equation commonly used to describe competing species are compatible with any dynamical behavior provided the number of species is very large" (p. 5). Smale did this by showing how to embed an arbitrary n-dimensional autonomous system into an $(n + 1)$-dimensional competitive system. Using this construction, he was able to show that the class of competitive systems can "do everything" if there are sufficiently many interacting states. His theorem provided a mathematical way to think about Darwin's insight about the importance of competition in driving natural selection among many possibilities.

Smale's result does not imply that cooperative dynamics are not crucially important in biology. For starters, the mathematical definition of a competitive system only required that competition occur *between* populations. There can still be complex kinds of cooperation *within* each population of a competitive system, without undermining the convergence of all its trajectories to equilibria, as I proved in [24]. In addition, I showed that long-range cooperative interactions between populations often play crucial roles, and can occur in neural networks even if they do not have huge numbers of interacting populations.

A good example is the cortical processes during visual perception like perceptual grouping and surface filling-in which both use cooperative interactions; e.g., [35]–[45]. My webpage sites.bu.edu/steveg lists many downloadable articles that include examples of cooperative–competitive dynamics in all parts of our brains, as illustrated in the next section.

I. Population Biology and Neurobiology Articles in 1978

Not accidentally, 1978 was also the year when I published an article with the title *Communication, Memory, and Development* [46] about developmental processes within the natural world, and an article with the title *A Theory of Human Memory* [47] that focused on human activities. The latter article laid the foundation for much of the modeling work that I did in the subsequent decade.

I was able to understand how the population biology and neurobiology models in these articles work because of mathematical analyses such as those of adaptation level systems and generalized Volterra–Lotka systems. All of the above analyses consider how cell activities, or STM traces, behave through time. Also essential were early mathematical theorems about how humans and animals learn using adaptive weights, or long-term memory (LTM) traces, by interacting with STM traces that are activated by external or internal inputs. I will mention some of these developments below. The development of mathematical and physical understanding have always gone hand-in-hand, and have been mutually energizing, in my work.

The unifying power of these models in population biology is illustrated by the variety of examples that were analyzed in [46]. As noted in my book [48] "These examples include how *Hydra* regenerates a missing head [49]; how slime models aggregate in their search for food [50], [51]; how the folds in the cuticle of *Rhodnius*, a genus of assassin bugs, are determined [52]; how a sea urchin blastula becomes a gastrula [53]; how the French Flag problem is solved, namely, how an organ's shape can be maintained even as its size increases during the development of animals and their body parts; and how early neural development can be refined by adult learning because the laws for prenatal growth of cell connections during morphogenesis are formally similar to the STM and LTM laws that control postnatal synaptic learning in response to external inputs."

This 50th-anniversary issue of the IEEE SMC journal is a good place to review aspects of how my neural network models were discovered, and how their discovery and application to multiple interdisciplinary problems helped to trigger an explosion of additional discovery and application that has continued unabated to the present time, leading to biological neural networks that explain and predict large psychological and neurobiological databases about essentially all aspects of biological intelligence and constitute a blueprint for designing autonomous adaptive algorithms and mobile robots for engineering applications. Only a small subset of these discoveries, and their mathematical foundations and technological applications, can be summarized below, notably properties of adaptive resonance theory, or ART, that I will begin to discuss in Section V.

Many other discoveries have continued to be made to the present, and have been intermittently reviewed in recent articles. Articles during the past 20 years discuss many aspects of how our brains make our minds and describe neural models that unify explanations of psychological, anatomical, neurophysiological, biophysical, and biochemical data about perception, cognition, emotion, and action. These are the models that can be assembled into autonomous adaptive applications.

Model explanations include data about 3-D vision and figure-ground perception of both stationary and moving objects in scenes [35], [36], [42], [43], [54]–[56]; navigation based on visual or path integration signals [57], [58]; reactive and planned eye movements, active vision, numerical estimation, invariant category learning, and goal-oriented action [59]–[64]; speech and language perception, attention, working memory, learned planning, and cognitive information processing [48], [65]–[70]; cognitive–emotional dynamics, including reinforcement learning, motivated attention, adaptively timed learning, and goal-directed action [65], [71], [72]; social cognition [73]; breakdowns of brain mechanisms that cause behavioral symptoms of multiple mental disorders, including autism, Fragile X syndrome, medial temporal amnesia, Alzheimer's disease, visual and auditory neglect, and disordered sleep [73], [74]–[76]; designs for autonomous adaptive mobile robots whose controllers are reliable and explainable [62], [77], [78]; and the crown jewel of many studies of mind and brain: explaining *where* and *how* in our brains we become conscious, and *why* evolution was driven to discover conscious states of mind in the first place [41], [79].

As of this writing, data from hundreds of psychological and neurobiological experiments have been given unified mechanistic explanations, and scores of the predictions made within these models have been confirmed by subsequent psychological and neurobiological experiments, often years or decades after they were made. My book [48] provides

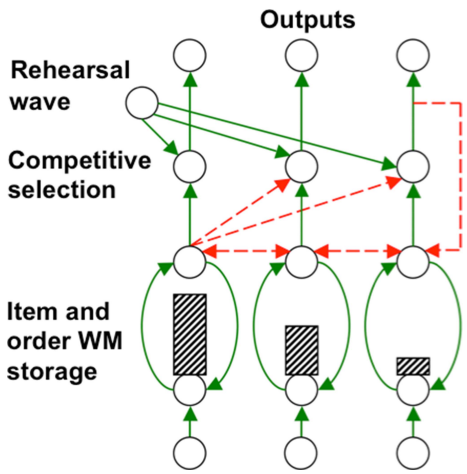

Fig. 2. Primacy gradient of activity, depicted by hatched bars proportional to cell activities, is stored in working memory within a recurrent shunting on-center off-surround network that is called an item and order working memory. The recurrent excitatory on-center signals are depicted by curved green arrows. The recurrent inhibitory off-surround signals are depicted by dashed red lines with double arrows. The items that are stored in working memory generate output signals via a feedforward competitive network whose inhibitory off-surround signals are depicted by diagonal dashed red arrows. This network chooses the currently most active cell activity at the competitive selection stage. Rehearsal of the chosen activity is controlled by a nonspecific rehearsal wave that enables the chosen activity to generate an output signal. Self-inhibitory feedback of the rehearsed item deletes it from working memory and enables the next most activity cell to be rehearsed, thereby preventing perseveration. Green = excitatory signals, and red = inhibitory signals. (Reprinted with permission from [93].)

a comprehensive, self-contained, and nontechnical introduction to all of these discoveries. It is written to be as accessible as possible to the broadest possible audience.

III. Why Can Models of How Brains Make Minds Be Discovered?

The title of this article raises immediate questions. The first question is: Why is it possible to discover mechanistic neural models of how brains make minds, let alone be able to mathematically analyze them? The second question is: Why can such models, once discovered, provide blueprints for designing autonomous adaptive systems that may be useful in engineering and technology, whether these systems are realized in software, hardware, or both.

As a principal founder and current research leader in the fields of natural and artificial neural networks [48], [78], [80] I can provide a historical and personal perspective about how it has been possible to discover mechanistic neural models of how brains make minds, as well as of the modeling method that I introduced in 1957 to do so, and which has led to steady model discovery and development to the present time.

A. Modeling Cycle, I: Brain Mechanisms Need to Achieve Behavioral Success

I derived my first neural network model in 1957 while I was a Freshman at Dartmouth College taking the introductory course in psychology. This discovery naturally led to the modeling method in Fig. 1. I discovered this method shortly after I became fascinated by classical psychological data about how humans and animals learn. The human learning data concern how we learn lists of events—whether a list of nonsense syllables, words, song lyrics, or dance movements—using methods of serial verbal learning or paired-associate learning. The animal learning data concern how many terrestrial animals learn to associate essentially arbitrary sensory events with subsequent rewards and punishments in order to later acquire valued goal objects, or avoid noxious ones, when they again experience those sensory events, or sufficiently similar ones. These cognitive–emotional links are typically learned using one of several methods, notably classical conditioning or instrumental conditioning.

Underlying both of these learning capabilities is the premise that "brain mechanisms need to achieve behavioral success," whether that success is measured by the ability to learn how to make a tool, or the ability to acquire a desired reward. This statement about behavioral success is just another way of saying that Darwinian selection operates on the behavioral level. Indeed, individuals who cannot carry out vital adaptive behaviors will not survive.

The modeling method in Fig. 1 begins with a systematic analysis of dozens or hundreds of behavioral experiments, because it is behavioral consequences that drive evolutionary success. Starting with one, or just a few experiments, would not create sufficiently many constraints to drive theoretical discovery. The art of modeling consists in being able to analyze dozens or hundreds of parametric behavioral experiments—whose data are given to us as static curves that plot one variable against another—as the result of an individual behaving autonomously in real time in a changing world. My analyses that translated static data into behavioral dynamics always led to the discovery of a small number of design principles that are capable of generating those data curves.

B. Method of Minimal Anatomies

Showing that these design principles can lead to behavioral success required a combination of experimental intuition and mathematical insight in order to convert the design principles into the *minimal* mathematical model that is capable of realizing them in real time. The *real time* requirement always leads to systems of differential equations. The *minimality* constraint means that the removal of any part of the model will reduce its explanatory range. Thus, the method embodies only mechanisms that are necessary to realize the design principles. That is why I call it the Method of Minimal Anatomies [81].

The first task of the mathematical model is to explain and predict a lot more behavioral data than were used to derive it (Fig. 1). These behavioral properties are emergent, or interactive, properties of model mechanisms acting together. Deriving such emergent properties often required mathematical analysis in the years before computers were powerful enough to simulate the dynamics of large nonlinear neural networks. Only decades later did computer simulations become a possible, indeed essential, theoretical tool.

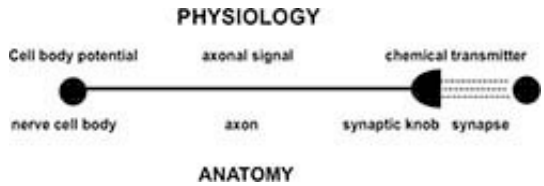

Fig. 3. Basic processes of STM and LTM at the cellular level. See the text for details. (Reprinted with permission from [48].)

C. Behavioral Success Is Computed by Neural Networks: STM and LTM

Because evolution acts on neural representations that compute indices of behavioral success, I realized from the start that the functional units of my derivation were not individual cells, but rather whole networks of cells, because these are the brain structures that are capable of computing emergent representations that govern behavioral success. That is why it is appropriate to talk about the field of neural *networks*.

My earliest derivation discovered two of the types of representations that operate on the network level: *patterns* of activities, or STM traces—that carry out fast information processing across multiple cell bodies—and *patterns* of adaptive weights, or LTM traces—that carry out learning and memory at multiple synaptic knobs. This fact led to a series of early theorems about how activity patterns are stored in STM [5], [6], [82] and about how LTM patterns are learned by interacting with STM patterns [3], [4], [18]–[20], [23], [83]–[88]. From a mathematical and engineering perspective, these patterns are just vectors. Thinking of them as patterns helps me to imagine them as embodied within dynamically evolving neural networks. Highlights of these interacting STM and LTM pattern properties are discussed in the next few sections.

D. From Mind to Brain: Behavioral Functions Impose Brain Mechanisms

The second task of the mathematical model came as a total surprise to me: each stage of the model could always be interpreted in terms of brain mechanisms (Fig. 1). That first happened during my initial model derivation as a college Freshman in 1957, at a time that I knew nothing about brain science. A chat with my pre-med friends clarified that this first model already had within it analogs of cell bodies, axons to carry signals between cell bodies, synaptic knobs and synapses at the ends of axons, cell potentials (or STM activities) in the cell bodies, spiking frequencies sent as signals along the axons, chemical transmitters in the synaptic knobs, and adaptive weights (or LTM traces) that helped to regulate the amounts of the available transmitter (Fig. 3).

The LTM traces were computed in the synaptic knobs, or possibly the abutting postsynaptic membranes, or both, because only here could associative learning be driven by correlating simultaneously available axonal signals from a presynaptic cell with the activity of the abutting postsynaptic cell to which these signals were sent. I also struggled at that time [89] to understand how these learning laws imply an intimate connection between associative learning and lateral inhibition, or competition, that I needed to model data about serial verbal learning [19] using only local computations.

Some of these neural processes were known, notably anatomical and neurophysiological facts, although the behavioral derivation of the model often led to a new functional and mechanistic understanding of them. Because previous models were mathematically primitive, at best, some key mechanisms were not known. For example, the associative learning law that I proposed in my 1969 articles was not yet known.

E. Pattern Processing Implies Gated Steepest Descent Learning Laws and Their Control Circuits

The discovery of this associative learning law happened because I knew that the functional unit of fast information processing in a network is a *spatial pattern* of activity, or STM traces, across multiple cells. This fact forced the conclusion that the functional unit of learning in a network is a *spatial pattern* of learned associative weights, or LTM traces in or abutting synaptic knobs. In other words, our brains are *self-organizing pattern processing machines*.

I also knew that signals from a presynaptic cell (population) to all of its axonal synaptic knobs act like *stimulus sampling signals* that could activate learning in their LTM traces. The stimulus sampling property clarifies how neurons can be packed tightly in brains and emit diverging learning signals to many other neurons (one-to-many), while receiving converging learning signals back from them (many-to-one). Stimulus sampling enables the learning of functionally independent subsets of neurons within the densely packed neuropile.

In order for an LTM *pattern* to learn to match a currently active STM *pattern*, or time average of STM patterns during multiple stimulus sampling episodes, each LTM trace has to be able to either increase or decrease, depending upon whether the abutting STM activity is small or large at that time. In other words, a single LTM trace could exhibit both Hebbian and anti-Hebbian properties. I call this kind of learning *gated steepest descent learning* because a stimulus sampling signal opens a learning gate, after which the LTM traces track current STM activities or signals, as illustrated by

$$dw_{ij}/dt = f_i(x_i)\big[g(x_j) - w_{ij}\big]. \tag{18}$$

In (18), w_{ij} is the LTM trace from cell v_i to cell v_j, $f_i(x_i)$ is the presynaptic signal along the axon from cell v_i to cell v_j, and $g_j(x_j)$ is a cell membrane potential, or local signal to the abutting synaptic knob, from the postsynaptic cell v_j. Stimulus sampling occurs whenever $f_i(x_i) > 0$. At these times, w_{ij} tracks $g_j(x_j)$ via steepest descent. This prediction started to receive experimental support from neurophysiological experiments in 1983 [90], [91]. The article [92] provides additional discussion about the birth of these learning laws.

A remarkable property of (18) is that LTM does not change unless the stimulus sampling signal is positive. This allows memories to last for a long time, even many years. One important factor that can enable a persisting memory to be changed is a *predictive disconfirmation* that can drive fast relearning. My book [48] explains how this can happen along with relevant psychological data.

Gated steepest descent learning was mathematically analyzed in a series of theorems covering ever more general classes of learning systems, just as competition was. The following associative learning system is illustrative [18], [20]

$$\frac{dx_i}{dt} = A_i x_i + \sum_{j \in J} B_j w_{ji} + C_i \tag{19}$$

and

$$\frac{dw_{ji}}{dt} = D_j w_{ji} + E_j x_i \tag{20}$$

where $i \in I, j \in J$, and the sets I and J of indices can be arbitrarily large, and can intersect in an arbitrary way. As a result, system (19), (20) includes feedforward and recurrent networks, as well as networks that are an arbitrary mixture of both.

The terms A_i, B_j, C_i, D_j, and E_j in (19) and (20) are all *functionals* of possibly highly nonlinear and time-dependent variables. Equation (19) defines the dynamics of STM in terms of generalized decay A_i, performance B_j, and input C_i functionals. Equation (20) defines the dynamics of LTM in terms of generalized memory D_j and learning (stimulus sampling) E_j functionals. These equations exhibit "local symmetry axes" because the decay, performance, memory, and learning functionals depend only upon the stimulus sampling cells v_j and not the postsynaptic cells that they influence. Under weak conditions, the relative LTM traces that are activated by the stimulus sampling cells all globally converge to the spatial patterns of the input functionals. In other words, learning is stable and complete as time goes on.

F. From Pattern Learning Laws to Neural Networks and Neural Architectures

The fact that neural networks learn LTM patterns, or vectors, also led to the discovery of the basic minimal network anatomies to learn an arbitrary STM spatial pattern (outstar), to categorize an arbitrary STM spatial pattern (instar), to learn an arbitrary space–time pattern (avalanche), to learn to transform an arbitrary input STM spatial pattern into an arbitrary output STM spatial pattern (associative map), to dynamically stabilize bottom-up categorization with top-down attentional matching (ART attentional system), and so on in an evolutionary progression of model discovery using the method of minimal anatomies (see [93] for a review of these foundational circuits).

G. Modeling Cycle, II: From Minimal Models to Autonomous Intelligence

No single cycle of this modeling method explains an entire brain. However, by clarifying what data the model can explain, what I like to call the "shape of the boundary between the known and the unknown" became clear. This "boundary" often helped me to identify a design principle that had not yet been incorporated into the model by focusing my attention upon phenomena that the current model could not explain. My task was then to incorporate an additional design principle in a self-consistent way that did not undermine the previous model, but rather expanded and deepened its interdisciplinary explanatory and predictive range.

Multiple cycles of this method have resulted in models that have explained such subtle issues as how and why the cerebral cortex organizes its cellular networks in layers—leading to the computational paradigm that I have called laminar computing that includes laminar neural models of all aspects of biological intelligence [36], [37], [40], [45], [54], [59], [66], [68], [69], [71], [75], [94]–[102]—and how and why evolution was driven to discover conscious states of mind—including the hypothesis that "all conscious states are resonant states" that I will discuss in Section V.

Also underlying both capabilities is the fact that *learning takes place in real time*, whether under unsupervised or supervised learning conditions. Indeed, to learn a list, the successive items need to occur within a limited time delay. Presenting one word each day, whether of the list of words or of a song's lyrics, will not lead to learning of the song. Typically, successive words are presented within hundreds of milliseconds or a few seconds to enable the song to be learned. The same is true for learning how to acquire a reward, since presenting a stimulus today and a rewarding outcome the next day will lead to no reinforcement learning. The stimulus typically precedes the reward by hundreds of milliseconds or a few seconds for reinforcement learning to occur.

Although substantial learning often occurs within a single learning trial, learning lists of items or events typically requires multiple learning and rehearsal trials, as illustrated by classical data about serial verbal learning. During this learning paradigm, items in the list are presented one at a time in a fixed order, followed by a rest period, after which the list is repeated in the same order and at the same rate. On each presentation of the list, a subject's task in response to each list item is to predict the next item before it is presented. This learning process guarantees that quite a few errors will be made on early learning trials. Learning proceeds until all the learned items are corrected predicted on a fixed number of consecutive trials.

During serial verbal learning, items at the beginning and end of the list are typically learned before items in the list middle. This property is called the classical "bowed serial position curve." Examples of bowing abound in our lives. For example, we may find it easier to remember how a relationship began and ended, with details in the middle often a muddle. These classical data were explained and simulated in [19] and [23] using the additive model whose learned coefficients become asymmetric because of the fixed order in which list items are learned. The many known facts about serial verbal learning when I began my work as a student was one of the main sources that guided by derivation of the additive model in the first place.

Despite their apparent simplicity, both list learning and reinforcement learning forced the analysis of deep conceptual issues. For starters, both require that the learner act as a "minimal adaptive predictor" that can distinguish between events that are causal versus events that are accidental [103], [104]. Additional complexities of learning include how learned memories become consolidated while we are asleep as a result of cortico–hippocampal interactions, among others. This consolidation process, and how its failure can lead to medial temporal amnesia, is modeled in [72].

IV. Equations, Modules, Modal Architectures: Autonomous Adaptive Intelligence and XAI

Some further remarks help to clarify why it is possible to create a mind–brain theory that enables links between brain mechanisms and psychological functions to be established, and how similar organizational principles and mechanisms, suitably specialized, can support conscious qualia across modalities.

A. Small Number of Fundamental Equations

Perhaps the most basic reason for this intermodality unity is that a small number of fundamental equations suffice to model all modalities, just as a small number of equations form the foundation of theoretical physics. These include equations for short-term memory, or STM; medium-term memory, or MTM; and long-term memory, or LTM, that were introduced in [3] and [81]; see [105] for a review. STM and LTM laws have been discussed above. MTM is realized by activity-dependent habituation, often mediated by a chemical transmitter, that occurs on a time scale of dozens or hundreds of milliseconds, slower than the time scale than STM, but much faster than the time scale of LTM, which can persist for years.

A typical habituative MTM trace at a synaptic knob between the ith and jth cells in a network obeys variations of the equation

$$dy_{ji}/dt = F\big(G - y_{ji}\big) - Hf_j(x_i)y_{ji}. \tag{21}$$

In (19), term $F(G-y_{ji})$ describes the rate of accumulation of the MTM transmitter y_{ji} from its current value to its maximum value G at rate F, while term $-Hf_j(x_i)y_{ji}$ describes the rate of gated, or mass action, release or inactivation of MTM due to the presynaptic signal $f_j(x_i)$.

B. Larger Number of Modules, or Microcircuits

These equations are used to define a somewhat larger number of *modules*, or microcircuits, that carry out different functions within multiple modalities. These modules include shunting on-center off-surround networks, gated dipole opponent processing networks, associative learning networks, spectral adaptively timed learning networks, and so on. Each type of module exhibits a diverse 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; STM and working memory storage; attentive matching of bottom-up input patterns with top-down learned expectations; and synchronous oscillations and traveling waves.

C. Combining Specialized Equations and Modules in Modal Architectures

These equations and modules are assembled and specialized within *modal architectures*, where "modal" stands for different modalities of biological intelligence, including architectures for vision, audition, cognition, cognitive–emotional interactions, and sensory–motor control. Modal architectures are less general than the von Neumann architecture that provides the mathematical foundation of modern digital computers, but much more general than a traditional AI algorithm. Their specializations of equations and modules have been selected by evolution over the millennia, guided by their behavioral success.

An integrated self is possible because it builds on a shared set of equations and modules within modal architectures that can interact seamlessly together. Grossberg [48] summarizes how such an integrated self can emerge through life-long processes of development and learning.

The adaptive resonance theory, or ART, networks that are reviewed in Section V form part of several different modal architectures, including modal architectures that enable conscious seeing, hearing, feeling, and knowing.

D. Explainable Artificial Intelligence

For such models to be useful in human societies, they must be able to interact harmoniously with humans. To accomplish that, their emergent properties need to be compatible with those of humans, and humans need to be able to explain how they work, notably how their representations lead to decisions and actions. Otherwise, no important decision, such as a financial or medical decision, could confidently be made using such a model.

This is a topic about explainable artificial intelligence, or XAI, which will also be touched on in the ART review. More generally, all known biological neural models of perception, cognition, emotion, and action are explainable [78], and thus all autonomous adaptive algorithms or mobile robots using their designs are also explainable.

V. Adaptive Resonance Theory: Rapid Category Learning and Prediction Throughout Life

A. Stability–Plasticity Dilemma and ART Matching Rule for Object Attention

Adaptive resonance theory, or ART, is a cognitive and neural theory of how our brains autonomously learn to attend, recognize, and predict objects and events in a changing world. ART currently enjoys the broadest explanatory and predictive range of available cognitive and neural theories. (See reviews in [41], [48], [62], [65], [74], [78], [79], and [106].)

ART's predictive power builds upon its ability to learn rapidly from a changing world, without forgetting previously learned memories. Humans can rapidly learn new facts without being forced to just as rapidly forget what they already know. We can thus go out into the world with confidence, without having to fear that by, say, learning to recognize a new person's face, we will be forced to forget the faces of our family members and friends. We are thus free from the nightmare of experiencing *catastrophic forgetting* [78], [88], [107]–[109], or the sudden and unexpected collapse of a subset of memories.

Many contemporary learning algorithms do experience catastrophic forgetting, particularly when they try to learn quickly in response to a changing world. These include the competitive learning, self-organizing map, backpropagation,

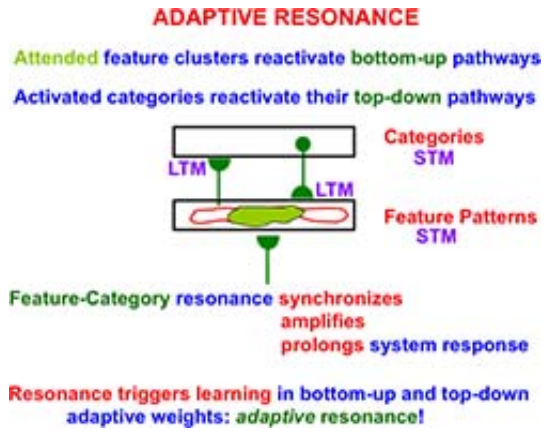

Fig. 4. Bottom-up and top-down interactions that lead to a feature–category resonance and learning in bottom-up adaptive filter LTM traces and top-down learned expectation LTM traces. Individual green pathways represent distributed patterns of signals across the entire network. See Fig. 5 and the text for details. (Adapted with permission from [41].)

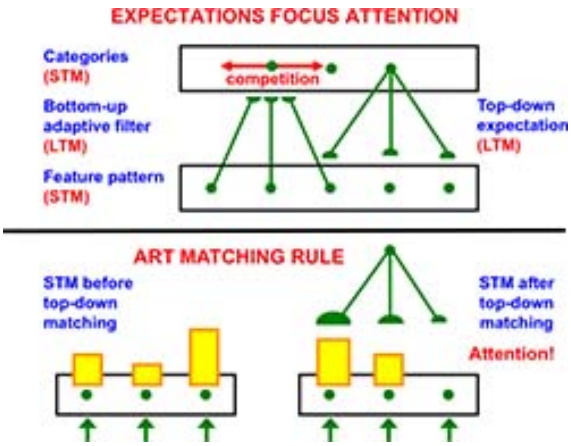

Fig. 5. Top of this figure depicts the many-to-one bottom-up connections from feature-selective cells to category cells, and the one-to-many top-down connections from category cells to feature-selective cells. The bottom of the figure illustrates how the ART matching rule (see Fig. 6) embodies object attention when an active category reads-out a top-down expectation. Active feature-selective cells that receive sufficiently strong top-down signals have their activities selected, amplified, and synchronized, thereby beginning to form an attentional focus, while cell activities that receive weak or no top-down signals are suppressed. See the text for details. (Reprinted with permission from [48].)

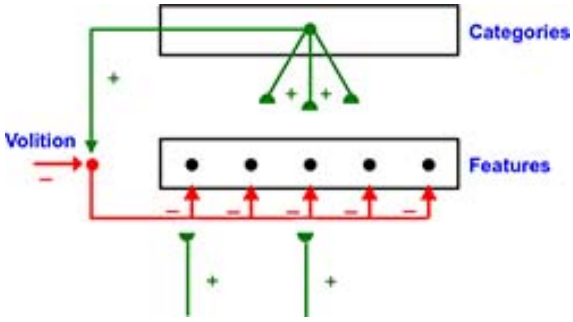

Fig. 6. ART matching rule for focusing attention upon critical feature patterns is embodied by a top-down, modulatory on-center, off-surround network. See the text for details. (Reprinted with permission from [41].)

deep learning, simulated annealing, neocognitron, support vector machine, regularization, and Bayesian models. The brain solves a challenging problem that many current learning models have not. Implementing the brain's way to avoid catastrophic forgetting is as important for understanding how our brains learn as it is for developing new autonomous adaptive algorithms and mobile robots for engineering and technology applications.

My 1980 article [88], in which I developed the foundations of ART, called the problem whereby the brain learns quickly and stably without catastrophically forgetting its past knowledge the *stability–plasticity dilemma*. ART solves the stability–plasticity dilemma by complementing its competitive learning category learning circuit with learned top-down expectations (Fig. 4) that obey the ART matching rule. The ART matching rule computationally instantiates how our brains carry out *object attention* whereby matching top-down expectations against bottom-up input patterns enables us to selectively attend to salient objects (Fig. 5). It is embodied by a top-down, modulatory on-center, off-surround circuit (Fig. 6) that mechanizes a kind of self-normalizing "biased competition".

ART explains how such top-down attentive matching helps to solve the stability–plasticity dilemma. In particular, when a good enough match occurs, a synchronous resonant state emerges that embodies an attentional focus and is capable of driving fast learning of bottom-up recognition categories and top-down expectations; hence the name *adaptive* resonance. The stability–plasticity dilemma is solved because learning occurs only between the attended critical feature pattern and its category. Within supervised ARTMAP models, the features in the critical feature pattern are attentively selected because of their ability to control successful predictions, while predictively irrelevant features are suppressed [110]–[112].

B. ART Universality: Thought Experiment About Error Correction in Changing World

ART circuit designs were derived from a thought, or Gedanken, experiment in Grossberg [88] about error correction in a changing world. This thought experiment sought to answer the question: How can a coding error be corrected if no individual cell knows that one has occurred? As [88, p. 7]) notes: "The importance of this issue becomes clear when we realize that erroneous cues can accidentally be incorporated into a code when our interactions with the environment are simple and will only become evident when our environmental expectations become more demanding. Even if our code perfectly matched a given environment, we would certainly make errors as the environment itself fluctuates."

The hypotheses from which ART circuits were derived are familiar facts from daily life. These facts are familiar to us because they are ubiquitous environmental constraints on the evolution of our brains. The thought experiment translates these hypotheses, step-by-step, into processes operating autonomously in real time with only locally computed quantities. The universality of ART as a cognitive and neural theory is demonstrated by the fact that, when a few familiar environmental constraints on incremental knowledge discovery are overcome in a self-organizing manner, then

ART circuits naturally emerge. This fact suggests that ART designs may, in some form, be embodied in all future *autonomous* adaptive intelligent devices, whether biological or artificial.

C. Large-Scale Applications in Engineering and Technology

ART's universality helps to explain why supervised ART, or ARTMAP systems, have done well in benchmark studies where they have been compared with other algorithms, and have been used in many large-scale engineering and technological applications. Initial ARTMAP benchmark studies demonstrated either better predictive accuracy, or many fewer learning trials, or both, when compared with leading machine learning, backpropagation, statistical learning, and genetic algorithm models [110].

Large-scale applications include the development of the Boeing Company's engineering design retrieval system that was used to rapidly learn millions of airplane parts in the Boeing airplane parts inventory [113], [114]. At the time of this application in the early 1990s, there were 16 000 000 parts, each defined by a 1 000 000-dimensional vector, in the Boeing parts inventory. After rapidly learning the part inventory, when a part description was input to the algorithm, it could choose the best matching part by direct access. This enabled the potential use of already fabricated parts to design new airplanes by slightly modifying their designs, and thereby potentially saving millions of dollars a year in fabrication costs. This algorithm was used to help design the Boeing 777. Because ARTMAP systems do not experience catastrophic forgetting, the Boeing system could continually be expanded as new parts entered its inventory.

Other large-scale applications include classification and prediction of sonar and radar signals, of medical, satellite, face imagery, social media data, and of musical scores; control of mobile robots and nuclear power plants, cancer diagnosis, air quality monitoring, strength prediction for concrete mixes, solar hot water system monitoring, chemical process monitoring, signature verification, electric load forecasting, tool failure monitoring, fault diagnosis of pneumatic systems, chemical analysis from ultraviolet and infrared spectra, decision support for situation awareness, vision-based driver assistance, user profiles for personalized information dissemination, frequency-selective surface design for electromagnetic system devices, Chinese text categorization, semiconductor manufacturing, gene expression analysis, sleep apnea and narcolepsy detection, stock association discovery, viability of recommender systems, power transmission line fault diagnosis, million city traveling salesman problem, identification of long-range aerosol transport patterns, product redesign based on customer requirements, photometric clustering of regenerated plants of gladiolus, manufacturing cell formation with production data, and discovery of hierarchical thematic structure in text collections, among others. Various of these applications are described at http://techlab.bu.edu/resources/articles/C5.htmlReferences. Discussions of these and other applications are summarized in [48].

Large-scale applications of ART in engineering and technology have continued unabated to the present. The December, 2019, issue of the journal *Neural Networks* contains recent summaries, notably the following two articles by Wunsch *et al.* (see https://arxiv.org/pdf/1910.13351.pdf and https://arxiv.org/pdf/1905.11437.pdf).

D. Explaining and Predicting Large Psychological and Neurobiological Databases

ART universality as a solution of a fundamental learning problem that confronts all humans and higher mammals may also explain why all of the foundational ART predictions have received increasing support from psychological and neurobiological data since ART was introduced in 1976 [115]–[117]. Since then, ART has undergone continual development to explain and predict increasingly large behavioral and neurobiological data bases, leading to a currently unrivalled explanatory and predictive range.

ART explanations range from normal and abnormal aspects of human and animal perception and cognition, to the spiking and oscillatory dynamics of hierarchically organized laminar thalamocortical networks in multiple modalities. Indeed, some ART models explain and predict behavioral, anatomical, neurophysiological, biophysical, and even biochemical data. ART currently provides functional and mechanistic explanations of such diverse topics as laminar cortical circuitry; invariant object and scenic gist learning and recognition; prototype, surface, and boundary attention; gamma and beta oscillations; learning of entorhinal grid cells and hippocampal place cells; computation of homologous spatial and temporal mechanisms in the entorhinal–hippocampal system; vigilance breakdowns during autism and medial temporal amnesia; cognitive–emotional interactions that focus attention on valued objects in an adaptively timed way; item–order–rank working memories and learned list chunks for the planning 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. Brain regions whose functions are clarified by ART 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. These accomplishments are surveyed in a self-contained nontechnical way in my book [48]. The original archival articles, as well as many conference lectures, can be downloaded from sites.bu.edu/steveg.

E. CLEARS: Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony

ART specifies mechanistic links between the processes of consciousness, learning, expectation, attention, resonance, and synchrony (the CLEARS processes). The CLEARS processes are all engaged when there is a good enough match between a bottom-up input pattern of features with a top-down learned

Fig. 7. Summary of resonances that support conscious states of mind, and the functions of those states in behavior. These resonances illustrate the general prediction that "all conscious states are resonant states." (Reprinted with permission from [48].)

expectation that is read out from an active recognition category. Then a *feature–category resonance* can occur that synchronizes, amplifies, and prolongs the activities of the critical feature pattern that is currently attended and the recognition category with which it is bound (Fig. 4). Such a feature–category resonance can drive fast learning within the bottom-up adaptive filter LTM weights and the top-down learned expectation LTM weights that join the currently active critical feature pattern and its category. A feature–category resonance can also support conscious recognition of attended visual objects and scenes (Fig. 7). This type of resonance is just one of several examples of the general hypothesis that "all conscious states are resonant states" [41].

These CLEARS links within ART clarify why many animals are intentional beings who pay attention to salient objects, why "all conscious states are resonant states, but not conversely," and how brains can learn both *many-to-one maps*—for example, representations whereby many object views, positions, and sizes all activate the same invariant object category—and *one-to-many maps*—for example, representations that enable us to expertly know many things about individual objects and events (Fig. 8).

F. Learning Many-To-One and One-To-Many Maps: Intermodal Prediction and Expertise

As Fig. 9 illustrates, a many-to-one map can be learned by two stages of learning within an ARTMAP, or fuzzy ARTMAP, learned mapping architecture (Fig. 10). In the top half of the figure, multiple fonts of the letter A are first categorized by an unsupervised ART category learning system into visual recognition categories that learn to cluster visual objects with similar visual features. As this is happening, the name "ay" is being learned in a separate unsupervised ART category learning system into auditory categories of learned object names. These two unsupervised ART systems are linked by associative learning into a supervised ARTMAP architecture as the system is trained to predict the names of the visual objects.

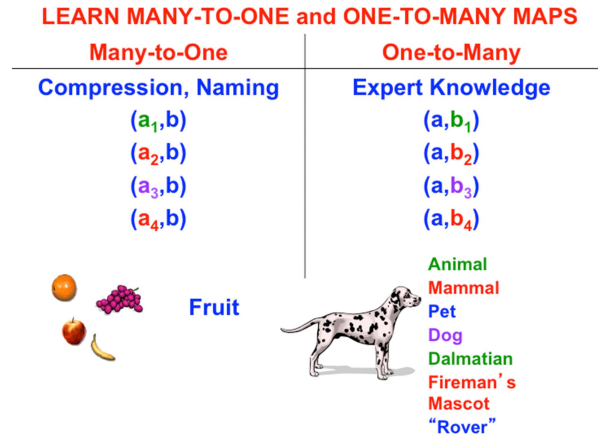

Fig. 8. Supervised ARTMAP systems can learn both many-to-one maps and one-to-many maps, as explained in greater detail in Figs. 9 and 10 and the text. (Reprinted with permission from [48].)

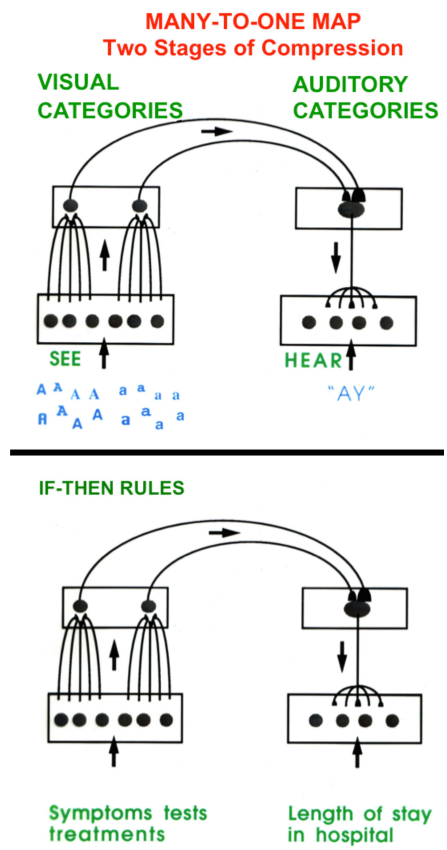

Fig. 9. Many-to-one maps can be learned by linking two unsupervised ART networks together with associative learning from the first network to the second, thereby converting the entire system into a supervised ARTMAP system, as depicted in Fig. 10. (Adapted with permission from [48].)

This kind of two-stage many-to-one ARTMAP architecture can learn an unlimited variety of different maps between arbitrary sets of n-dimensional input vectors and arbitrary sets of m-dimensional output vectors, for arbitrarily large finite numbers n and m. For example, in the bottom half of the figure, the visual features are replaced by symptoms, tests, and treatments from a medical database. These are associated with predictions about the length of stay in a hospital.

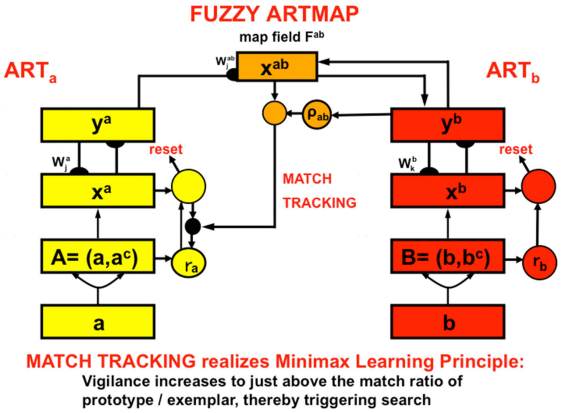

Fig. 10. Supervised ARTMAP system is capable of learning arbitrary maps from n-dimensional input vectors to the ART_a unsupervised learning model to m-dimensional output vectors from the ART_b unsupervised learning model, where the two ART models are linked by a learned associative map at a map field. Predictive errors, or mismatches at ART_b trigger a *match tracking* operation that causes vigilance in ART_a to increase until it just exceeds the analog match value between bottom-up and top-down signals in ART_a, thereby triggering hypothesis testing for a new category to be learned. See the text for details. (Adapted with permission from [111].)

This kind of supervised learning is controlled by a process of *match tracking* whereby the system as a whole minimizes predictive error while it maximizes learned category generality, thereby realizing a kind of *minimax learning*. I will explain how match tracking works in Section V-J.

G. Seventeen Differences Between Backpropagation (Also Deep Learning) and ART

Many neural network architectures, such as backpropagation and deep learning, are feedforward adaptive filters. Although ART does include a feedforward adaptive filter, it is much more than that. Grossberg [15] listed 17 major differences between ART and backpropagation (and thus deep learning). They are:

1) real-time (online) learning versus lab-time (offline) learning;
2) learning in nonstationary unexpected world versus in stationary controlled world;
3) self-organized unsupervised or supervised learning versus supervised learning;
4) dynamically self-stabilize learning to arbitrarily many inputs versus catastrophic forgetting;
5) maintain plasticity forever versus externally shut off learning when database gets too large;
6) effective learning of arbitrary databases versus statistical restrictions on learnable data;
7) learn internal expectations versus impose external cost functions;
8) actively focus attention to selectively learn critical features versus passive weight change;
9) closing versus opening the feedback loop between fast signaling and slower learning;
10) top-down priming and selective processing versus activation of all memory resources;
11) match learning versus mismatch learning: avoiding the noise catastrophe;
12) fast and slow learning versus only slow learning: avoiding the oscillation catastrophe;
13) learning guided by hypothesis testing and memory search versus passive weight change;
14) direct access to globally best match versus local minima;
15) asynchronous learning versus fixed duration learning: a cost of unstable slow learning;
16) autonomous vigilance control versus unchanging sensitivity during learning;
17) general-purpose self-organizing production system versus passive adaptive filter.

H. ART Is Explainable Self-Organizing Production System: It Is Trustworthy and Reliable

The last item in this list is that ART is a *general-purpose self-organizing production system* that is both *explainable* and *reliable* while it incrementally learns, using arbitrary combinations of unsupervised and supervised learning trials, to rapidly classify complex nonstationary databases without experiencing catastrophic forgetting. In particular, ART can learn an entire database using fast learning on a single learning trial (e.g., [118] and [119]).

ART is *general-purpose* because it can process any kind of inputs to its input and output modalities, whether from the external world or internally from other parts of the brain. ART is *self-organizing* because it can autonomously develop and learn in response to all these different kinds of inputs. ART's predictions are *explainable* using both its activity patterns, or STM traces, and its adaptive weights, or LTM traces.

Fig. 4 summarizes why both the patterns of STM activities and of LTM adaptive weights in ART systems are explainable in terms of the STM *critical feature patterns* that ART learns to attend based upon their predictive success. In brief, the STM critical feature patterns determine the information that controls ART information processing and action. These critical feature patterns can, in principle, be read by an observer. The critical feature patterns also determine the patterns of LTM adaptive weights that will be learned by ART bottom-up adaptive filters and top-down learned expectation. So it is also knowable what kind of learned information regulates fast information processing. In addition, at any stage of learning, the adaptive weights of the fuzzy ARTMAP algorithm can be translated into explicit fuzzy IF-THEN rules that explain what combinations of critical features, and within what numerical ranges, together predict successful outcomes [110].

The explainability of STM and LTM processes within ART are among the reasons that ART algorithms can be used with confidence to help solve large-scale real-world problems: practitioners can understand the basis of an ART explanation or prediction. ART learning and memory also do not experience catastrophic forgetting. Because of this combination of properties, ART is both *trustworthy* (because it is explainable) and *reliable* (because its learned memories do not experience catastrophic forgetting).

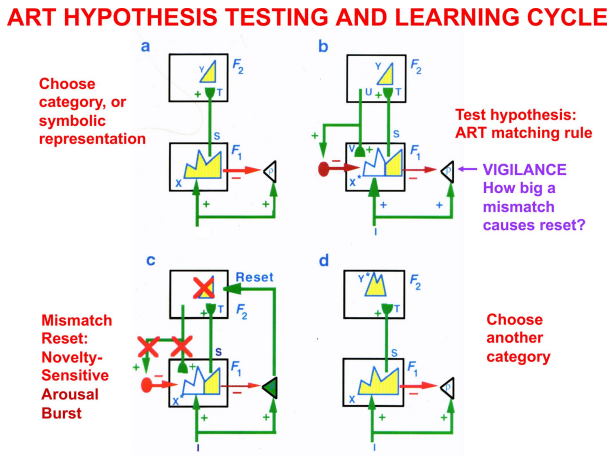

Fig. 11. ART hypothesis testing and learning cycle. See the text for details. (Reprinted with permission from [48].)

I. ART Is a Production System Because of Its Hypothesis Testing and Learning Cycle

Why is ART a production system? That is because of the way that it controls hypothesis testing to discover and learn recognition categories with which to classify and recognize objects, and use them to predict successful outcomes. The ART hypothesis testing and learning cycle is summarized in Fig. 11.

This search and learning cycle begins when an excitatory bottom-up input pattern I [green vertical arrow in Fig. 11(a)] is stored across feature detectors at level F_1 as an activity pattern X, which is colored yellow in Fig. 11(a). As in Fig. 4, multiple connections between processing stages are represented by a single connection, for simplicity. The bottom-up input pathway also sends excitatory inputs to the orienting system, which is represented by a triangle with a gain parameter ρ that is called the *vigilance* parameter. The orienting system adds the inputs that it receives at each time, multiplied by ρ. Activity pattern X also sends signals to the orienting system, but these signals are inhibitory [horizontal red arrow in Fig. 11(a)]. The orienting system hereby receives both excitatory and inhibitory inputs through time. At this moment, as many excitatory bottom-up inputs reach the orienting system as there are inhibitory inputs from X. The orienting system thus remains quiet.

Activity pattern X also generates a bottom-up excitatory input pattern S to the category level F_2. Before they reach F_2, the signals S are multiplied by adaptive weights, or LTM traces, that are represented by hemispherical synapses in Fig. 11(a) (see Fig. 3). These gated signals form the input pattern T to category level F_2. The inputs T are contrast enhanced and normalized within F_2 by a recurrent shunting on-center off-surround network [see (10)]. This competitive interaction selects a small number of the maximally activated cells and stores them in STM within F_2 (yellow activity pattern). The chosen cells are the category Y that codes the feature pattern at F_1. When only the maximally activated cells are stored, the network is said to have made a winner-take-all, or WTA, choice.

Fig. 11(b) shows that category Y within F_2 generates top-down excitatory signals U that are multiplied by adaptive weights. These gated signals V will learn the top-down expectation, or prototype, in LTM that embodies, and focuses attention upon, the STM *critical feature pattern* X^* (yellow activity pattern; inactive features in white) that survives the matching process. This top-down expectation input V embodies the ART matching rule (Fig. 6) by delivering a slightly excitatory, modulatory signal to F_1 cells in its on-center, while it strongly inhibits F_1 cells in its off-surround. Said in production system terminology, the top-down expectation *tests the hypothesis* Y using the *query* V.

If there is a mismatch between I and V at F_1, as in Fig. 11(b), then the ART matching rule creates a new STM activity pattern X^* (in yellow) at the cells where bottom-up and top-down inputs match. Mismatched features (white area) are inhibited. Pattern X^* is thus active at features I that are *confirmed by the query* V.

If the critical feature pattern X^* represents a good enough match between I and V, then it reactivates pattern Y at F_2 which, in turn, reactivates X^* at F_1. This excitatory feedback loop creates a feature–category resonance that dynamically links, or binds, X^* with Y. Such a resonance synchronizes, amplifies, and prolongs the system response of the attended critical features and the category with which it is bound (Fig. 4). The resonance either converges to a stable equilibrium or to a synchronous oscillation, depending upon the choice of network parameters, notably the relative sizes of excitatory and inhibitory cell response rates, as noted in Section II-E. A feature–category resonance triggers fast learning in the bottom-up adaptive filter and the top-down expectation LTM traces. It also supports conscious recognition of the object to which attention is being paid (Fig. 7).

Most mathematical theorems about ART properties consider this winner-take-all case. ART models with distributed categories have also been analyzed [120], [121]. Much more mathematical work needs to be done on how the most compact ART categories can be learned, without a loss of predictive success or memory stability.

J. Vigilance Control via Match Tracking: Minimax Learning Principle

As noted in Fig. 11(b), a parameter called vigilance determines whether a mismatch is big enough to activate the orienting system and thereby trigger hypothesis testing for a novel category to be learned, or a better matching category in case a category was already active when a new input arrives. Low vigilance permits large mismatches to occur before the orienting system fires. It leads to the learning of general categories with abstract prototypes (e.g., a face). High vigilance triggers hypothesis testing for a new category when even small mismatches occur, thereby leading to learning of specific categories with concrete prototypes (e.g., a frontal view of your mother's face). As a result, in the limit of high vigilance, the category prototype may encode an individual exemplar.

Vigilance is easy to compute in an ART system because it has an attentional system wherein learning and attention occur,

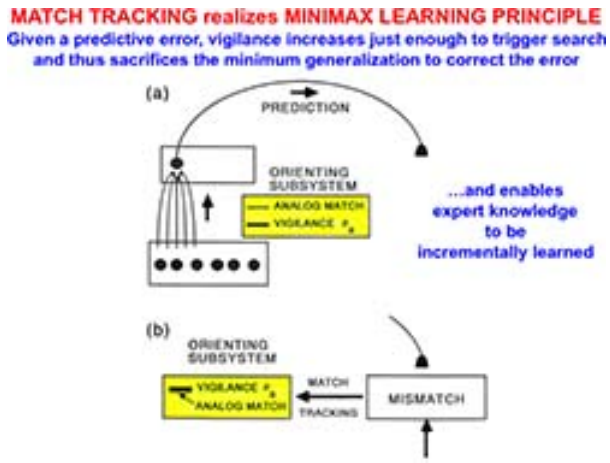

Fig. 12. How match tracking leads to a form of minimax learning that conjointly maximizes category generality while minimizing predictive error. (Reprinted with permission from [48].)

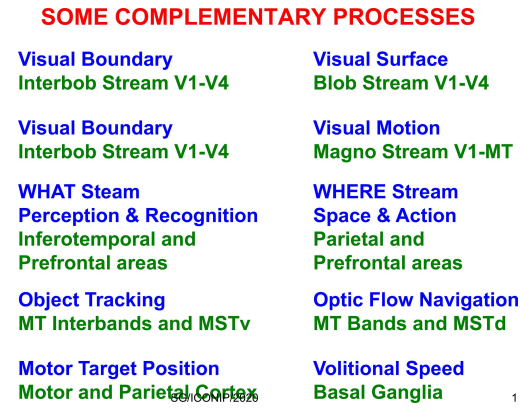

Fig. 13. Some pairs of computationally complementary processes. See the text for details. (Reprinted with permission from [48].)

that interacts with an orienting system which triggers reset and hypothesis testing in response to big enough mismatches between bottom-up and top-down signals within the attentional system. Said mathematically, hypothesis testing is triggered when the total excitation ρI of the orienting system is larger than the total inhibition from the critical features X^* across F_1 that survive top-down matching; that is, when $\rho|I| - |X^*| > 0$, where $|\cdot|$ denotes the number of positive inputs or matched features. Rewriting this inequality as $\rho > |X*||I|^{-1}$ shows that the orienting system is activated whenever ρ is chosen higher than the ratio of active matched features X^* in F_1 to the total features in I. Vigilance hereby controls how bad a mismatch is tolerated before hypothesis testing begins.

An efficient way to control vigilance is by a process of *match tracking:* When a prediction in an ARTMAP system [Figs. 10 and 12(a)] leads to a mismatch (e.g., "Rover" is the teaching signal when "dog" is the predicted outcome, as in Fig. 8), then a *match tracking* signal bumps vigilance up just enough to exceed the analog match between bottom-up and top-down signals that caused the predictive error [Fig. 12(b)].

The vigilance parameter ρ is then at the smallest value that can trigger category reset and hypothesis testing for a different category that will not lead to a mismatch. This process is called match tracking because vigilance "tracks" the degree of match between input exemplar and matched prototype [Fig. 12(b)]. Because low vigilance leads to the learning of general categories and high vigilance leads to the learning of concrete categories, match tracking realizes a minimax learning rule that conjointly *maximizes* category generality while it *minimizes* predictive error, thus using the minimal amount of memory that can correct predictive errors.

K. Memory Consolidation and Direct Access to the Globally Best Learned Match

As sequences of inputs are practiced over learning trials, the search process eventually converges upon stable categories, and hypothesis testing ends. Carpenter and Grossberg [118] proved mathematically that a familiar input directly activates the category whose prototype best matches it, without undergoing any search, while unfamiliar inputs continue to activate the orienting system to discover and learn categories until they become familiar. By always activating the globally best matching category, ART avoids getting stuck in local minima.

The process whereby search is automatically disengaged is a form of *memory consolidation* that emerges from network interactions. This ART process has been used to explain psychological and neurobiological data about how cortico–hippocampal interactions regulate memory consolidation within thalamocortical and corticocortical networks, both in normal learners and clinical patients who are afflicted by medial temporal amnesia due to hippocampal lesions [72], [122]. I call this process *emergent consolidation* because it is an emergent, or interactive, property of the entire ART system. Emergent consolidation can support structural consolidation at individual synapses, with resonance serving as a trigger for learning-dependent cellular processes such as protein synthesis, synapse formation, and transmitter production.

L. Complementary Computing and the Natural of Brain Specialization

I introduced the paradigm of complementary computing as I gradually realized over the years that many of the models that my colleagues and I were discovering exhibited computationally complementary properties. Complementary computing clarifies the nature of brain specialization by identifying computational units of brain design that can compute functionally significant behavioral properties [123].

To accomplish this, many brain processes occur in pairs that exhibit computationally complementary properties (Fig. 13). By themselves, these processes cannot control effective behaviors, but through their interactions they can. Along the way, complementary computing clarifies why there are so many parallel processing streams in our brains, with multiple processing stages. The processing stages often carry out a "hierarchical resolution of uncertainty" whereby complementary computational deficiencies are overcome. Fig. 14 shows a macrocircuit of the visual system with its many parallel processing streams and processing stages [124].

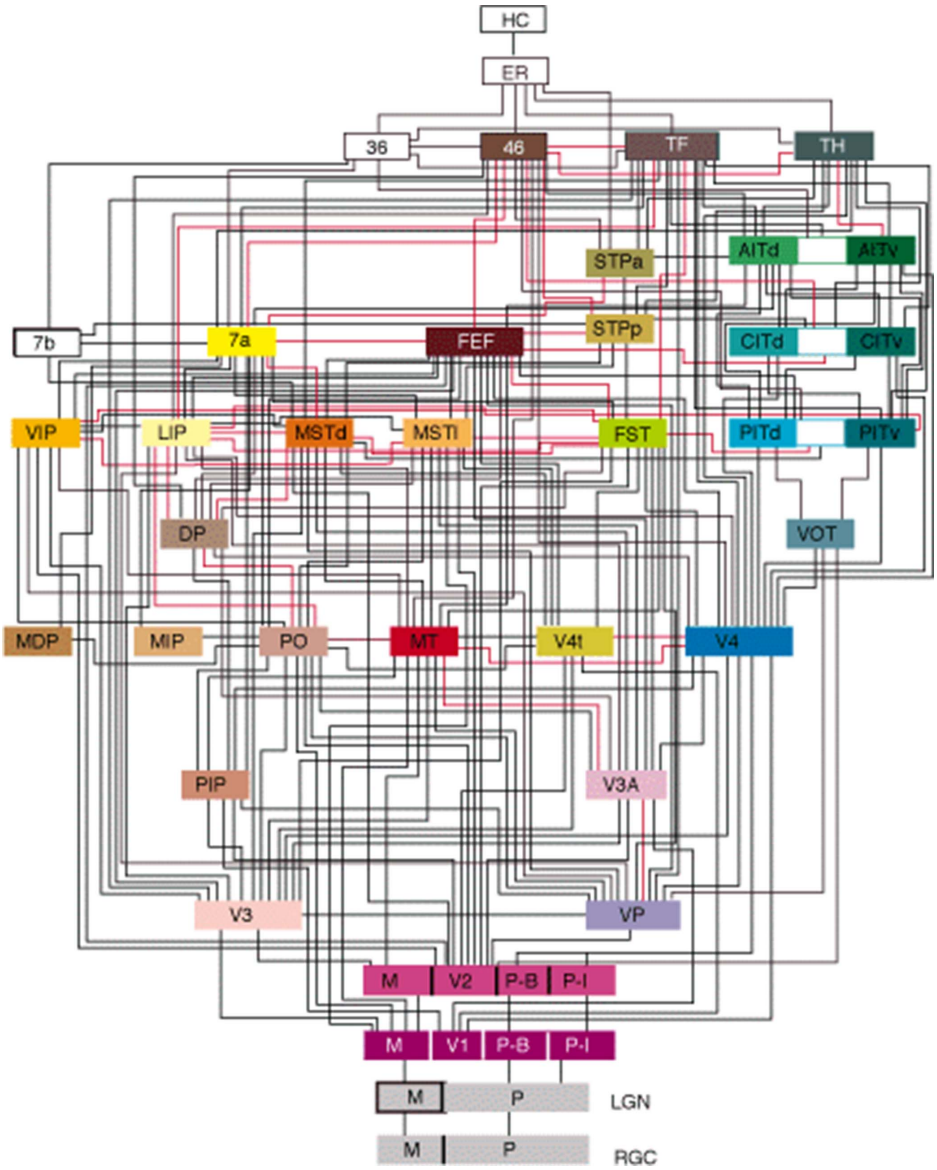

Fig. 14. Macrocircuit of the monkey visual system that shows its many parallel processing streams and multiple processing stages. (Reprinted with permission from [124].)

Large parts of the brain may exhibit complementary computing, such as the ventral cortical stream and the dorsal cortical stream (Fig. 15), with the ventral, or What, cortical stream carrying out processes of perception and categorization through cortical areas like inferotemporal cortex (IT), and the dorsal, or Where, cortical stream carrying out processes of spatial representation and action through cortical areas like the posterior parietal cortex (PPC).

These processes illustrate the fact that complementary computing describes mathematically rigorous differences in the computations within each of a pair of complementary processes (Fig. 16), with What stream processes carrying out excitatory matching and match-based learning processes to realize fast category learning without catastrophic forgetting, whereas Where stream processes carry out inhibitory matching and mismatch-based learning to continually update spatial representations and sensory–motor maps and gains in response to growth and other changes in our bodies. Fig. 16 shows that due to the computationally complementary organization of the brain, ART does not describe the many spatial and motor behaviors whose matching and learning laws differ from those of ART.

Fig. 17 refines the description in Fig. 16 of complementary What and Where stream processes by describing a macrocircuit of visual processes in the What and Where streams whose interactions realize visual intelligence. All of these processes and their interactions have been modeled, and many of these models have already been used in technological applications (see sites.bu.edu/steveg).

M. ART Embodies Complementary Computing

Fig. 13 does not exhaust the number of parallel processing streams that are known to be computationally complementary.

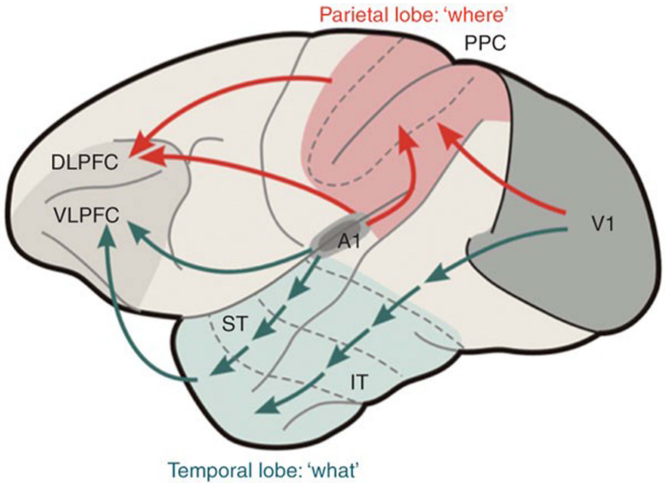

Fig. 15. Schematic of the ventral, or What, cortical stream for perception and categorization, and the dorsal, or Where, cortical stream for spatial representation and action. (Reprinted with permission from [157].)

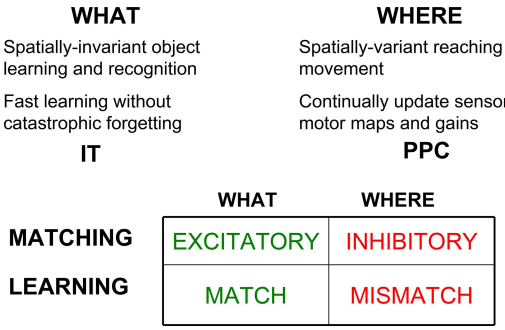

Fig. 16. Computationally complementary matching and learning laws of the What and Where cortical streams. See the text for details. (Reprinted with permission from [41].)

The attentional and orienting systems of ART also exhibit computationally complementary properties, as Fig. 18 illustrates in terms of two event-related potentials (ERPs) that are measured by scalp electrodes; namely, the processing negativity (PN) and N200 ERPs. The PN ERP [47], [103], [125], [126] can be measured when there is a good match between bottom-up and top-down signal patterns within the attentional system. The N200 ERP [127], [128] can be measured when there is a big enough mismatch to activate the orienting system. Inspection of PN and N200 properties at the bottom of Fig. 18 clearly shows their computationally complementary properties.

N. Laminar Computing: Why All Neocortical Circuits Are Laminar

I introduced the paradigm of laminar computing after many years of modeling visual processes such as those summarized in Fig. 17. Laminar computing describes how the cerebral cortex is organized into layered circuits whose specializations can support all forms of higher-order biological intelligence. In particular, laminar computing clarifies how bottom-up, horizontal, and top-down interactions are combined in a parsimonious and beautiful way at *all* levels of the perceptual and cognitive neocortex. These laminar circuits 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 [40], [129].

The LAMINART model (Fig. 19) of how ART mechanisms are realized in laminar cortical circuits explains how bottom-up, horizontal, and top-down interactions work together within the cortical layers to enable the visual cortex to realize: 1) the developmental and learning process whereby cortex shapes its circuits to match environmental constraints, and dynamically maintains them thereafter in a stable way; 2) the binding process whereby cortex groups distributed data into coherent object representations; and 3) the attentional process whereby cortex selectively processes important events.

To the present, ART mechanisms have been embodied in laminar cortical models of 3-D vision and figure-ground separation (3-D LAMINART model: [98], [100], [130], [131], audition, speech, and language (cARTWORD model: [66]); and cognitive information processing (LIST PARSE model: [69]).

O. SMART: Mismatch-Mediated Acetylcholine Release and Vigilance Control

The next step in the incremental development of laminar cortical models was the introduction of the synchronous matching ART, or SMART, model (Fig. 20). SMART is a laminar neural architecture for visual processing that includes a hierarchy of laminar cortical areas interacting with specific and nonspecific thalamic nuclei [68]. SMART introduced even more realism and functional capability into laminar cortical models by including, as noted in its Abstract: "single cell properties, such as spiking dynamics, spike-timing-dependent plasticity (STDP), and acetylcholine modulation; detailed laminar thalamic and cortical circuit designs and their interactions; aggregate cell recordings, such as current-source densities and local field potentials; and single cell and large-scale interareal oscillations in the gamma and beta frequency domains. In particular, the model predicts how laminar circuits of multiple cortical areas interact with primary and higher-order specific thalamic nuclei and nonspecific thalamic nuclei to carry out attentive visual learning and information processing. The model simulates how synchronization of neuronal spiking occurs within and across brain regions, and triggers STDP. Matches between bottom-up adaptively filtered input patterns and learned top-down expectations cause gamma oscillations that support attention, resonance, learning, and consciousness. Mismatches inhibit learning while causing beta oscillations during reset and hypothesis testing operations that are initiated in the deeper cortical layers. The generality of learned recognition codes is controlled by a vigilance process mediated by acetylcholine."

In particular, the model proposed how predictive mismatches activate nonspecific thalamic nuclei, which then activate the nucleus basalis of Meynert (Fig. 21). Nucleus basalis activation, in turn, causes acetylcholine release at layer 5 cells across the neocortex, thereby causing a reduction in

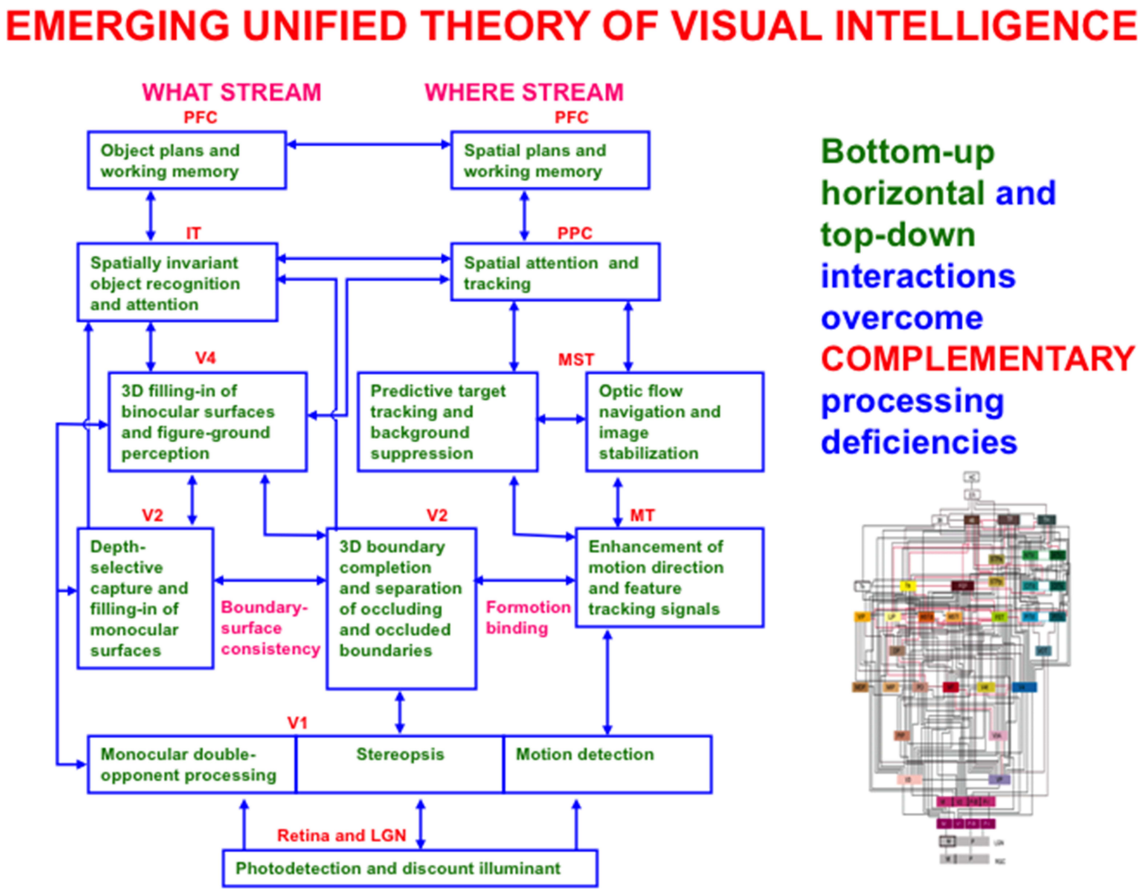

Fig. 17. Macrocircuit of the main visual processing stages and interactions within and across the What and Where cortical processing streams whose models together describe an emerging unified theory of visual intelligence. See the text for details. (Reprinted with permission from [106].)

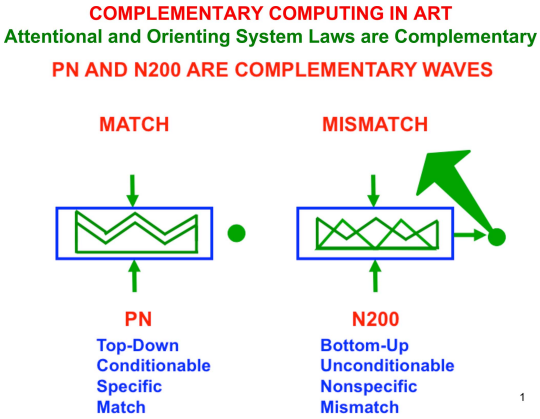

Fig. 18. Attentional and orienting systems of ART obey computationally complementary laws, as illustrated by the PN and N200 ERPs. See the text for details. (Reprinted with permission from [48].)

afterhyperpolarization currents, which causes an increase in vigilance, thereby initiating hypothesis testing to correct the error that caused the mismatch.

This cholinergic circuit plays a major role in learning, cognition, and consciousness. Reference [74] shows how breakdowns of this circuit can cause behavioral symptoms of mental disorders, such as Alzheimer's disease, autism, medial temporal amnesia, and disordered sleep.

P. Hierarchical Resolution of Uncertainty, Consciousness, and Goal-Oriented Action

Fig. 7 summarizes resonant states that support different kinds of conscious awareness. This section briefly discusses one of the organizational principles whereby our brains autonomously adapt to a changing world that may have driven evolution to discover conscious states. More complete descriptions are given in [41], [48], [62], and [79]. This principle is called *hierarchical resolution of uncertainty*. Hierarchical resolution of uncertainty means that multiple processing stages may be needed for our brains to generate a sufficiently complete, context-sensitive, and stable perceptual representation upon which to base a successful action.

How do our brains select the processing stage where such a sufficiently complete, context-sensitive, and stable perceptual representation is completed, so that it can be used to successfully control goal-oriented actions? Moreover, how are earlier processing stages, which compute more ambiguous and incomplete representations, prevented from triggering incorrect actions that could lead to disastrous consequences, before the higher processing stage is even activated?

I have predicted that the correct processing stage "lights up" into a conscious state due to a resonance with a subsequent processing stage that marks this representation as being a good enough one upon which to base a successful action. During visual perception, this perceptual representation

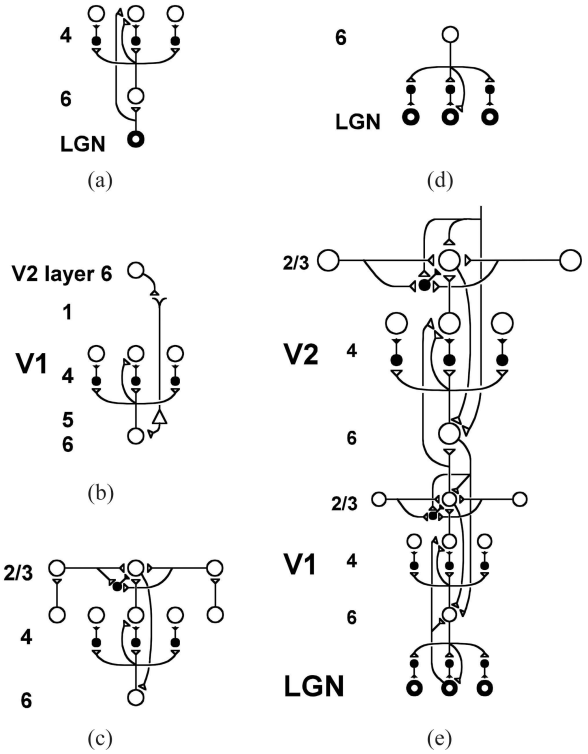

Fig. 19. LAMINART model clarifies how bottom-up, horizontal, and top-down interactions within and across cortical layers in V1 and V2 interblob and pale stripe regions, respectively, carry out bottom-up adaptive filtering, horizontal grouping, and top-down attention. Similar interactions seem to occur in all six-layered cortices. See text for details. (Reprinted with permission from [131].)

is a 3-D surface representation that can control actions like looking and reaching. The resonance that supports conscious seeing is called a *surface-shroud resonance* (Fig. 7) because the completed representation is a surface representation and it resonates with a form-fitting spatial attentional representation, called an *attentional shroud*, at the next processing stage [132], [133].

Surface-shroud resonances are predicted to be triggered by interactions between surface representations in the prestriate visual cortical area V4 and attentional shrouds in the PPC, before propagating top-down to lower cortical areas like V2 and V1—where the ART matching rule selects and focuses attention upon consistent critical feature patterns while inhibiting irrelevant features—and bottom-up to higher cortical areas like prefrontal cortex (PFC). Conscious states hereby provides an "extra degree of freedom" that enables our brains to avoid prematurely generating responses using inadequate perceptual representations. Fig. 7 summarizes some of the other known resonances that support different conscious states, each controlling their own types of actions.

ART predicts that "all conscious states are resonant states," but not that "all resonant states are conscious states." Resonant states that are not accessible to consciousness do not include feature detectors that are activated by external senses—such as those that support vision or audition—or internal senses—such as those that support the emotion. They include parietal–prefrontal resonances that trigger the selective opening of basal ganglia gates to enable the read out of contextually appropriate thoughts and actions [59], [61], [134] and entorhinal–hippocampal resonances that dynamically stabilize the learning of entorhinal grid cells and hippocampal place cells during spatial navigation [58], [135]–[137].

VI. Toward Intelligent Autonomous Adaptive Architectures: Predictive ART

Neural architectures that include increasingly large numbers of brain regions to carry out multiple functions in a coordinated way have been developed to the present time. One such architecture is the predictive ART, or pART, architecture [65] whose macrocircuit is summarized in Fig. 22. Fig. 22 is organized into distinct functional systems, all of which have been modeled along with their main interactions, and used to explain and predict large psychological and neurobiological databases.

The category learning and object attention processes within inferotemporal cortex, both anterior (ITa) and posterior (ITp), are highlighted with a red rectangle. ITa, in particular, learns view-, size-, and position-invariant recognition categories by interacting with ITp and PPC, among other brain regions. Surface-shroud resonances between IT and PPC play an important role in this learning process [60], [132], [138]. Said in another way, this system offers a self-organizing solution of the invariant pattern recognition problem.

The prefrontal cortical regions marked in green carry out higher-order intelligent processes that require the contextual information in stored sequences of previous events, whether objects or actions, in order to properly operate [69], [70], [79]. These processes include short-term storage of event sequences in working memories, learning of plans based on these sequences, and context-appropriate predictions and goal-oriented actions controlled by these plans.

The cortical and subcortical regions marked in red carry out value-based computations, such as reinforcement learning, emotion, motivated attention, and adaptively timed learning, attention, and goal-oriented action [71], [77], [81], [139]–[142].

The regions marked in black carry out multiple functions of visual perception (e.g., [35]–[39], [54], [94], and [143]–[146]). The distinction between IT in the What cortical stream and PPC in the Where cortical stream (Figs. 13 and 16) can be understood in pART within a much larger brain architecture for autonomous adaptive intelligence. Self-contained expositions of brain architectures for other aspects of autonomous intelligence, that have also been modeled and are ready for technological implication, are provided in [48].

VII. Some Final Thoughts About Future Work

Principled explanations and extensive computer simulations have been published over the years for all the pART processes. Other modeled brain processes that are not included in pART can be found in articles on my webpage sites.bu.edu/steveg. Despite this steady stream of modeling discoveries during the past few decades, it remains to develop a fully specified computational neural architecture that unifies all of these discoveries in a humanoid animat that would enable large-scale multimodal simulations to be carried out.

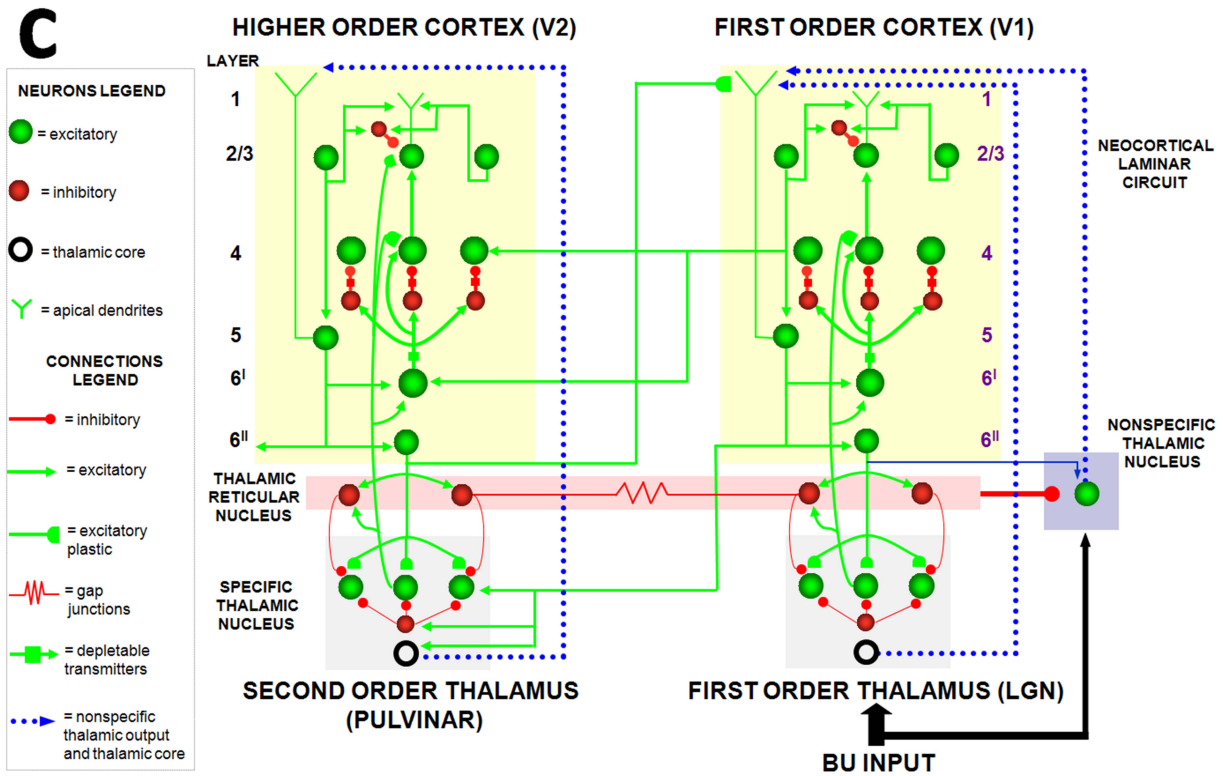

Fig. 20. Macrocircuit of the synchronized matching ART, or SMART, neural architecture. See the text for details. (Reprinted with permission from [68].)

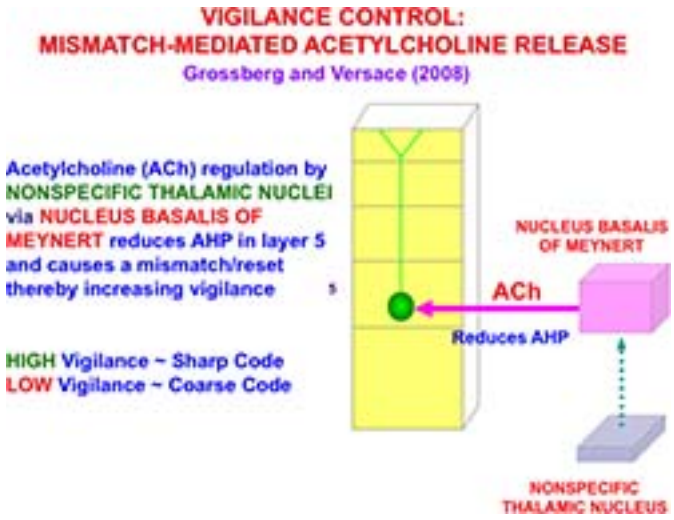

Fig. 21. Vigilance control is accomplished by release of acetycholine in cortical layer 5 cells in response to a big enough predictive mismatch. See the text for details. (Reprinted with permission from [68].)

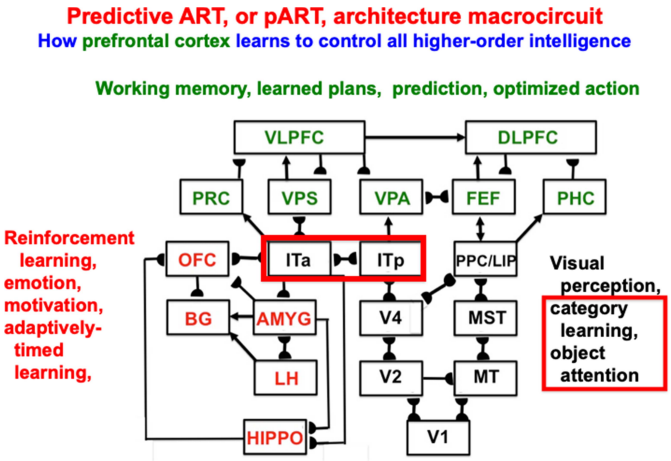

Fig. 22. *Predictive ART*, or pART, neural architecture combines many functionally different brain processes that together provide a possible design for aspects of intelligent autonomous adaptive applications in engineering and technology. See the text for details. (Reprinted with permission from [65].)

From Small-Scale Computer Simulations to Humanoid Animat Simulations and Robots: Over the years, my colleagues and I have carried out the largest computer simulations that could be accomplished in a reasonable amount of time with computers that were available to us; e.g., [37], [57], and [77]. As computational speed rapidly increases, and individual VLSI chips can carry out increasingly complex and multiplexed functions, working with a huminoid animat becomes increasingly plausible. It would be especially interesting to test the animat's behavior as it navigates a virtual reality environment while learning online about a stream of unexpected events. Eventually, such tests should be carried out on an embodied mobile robot navigating in the real world.

The Need for Small, Light, and Energy-Efficient General-Purpose Chips: Many technical challenges need to be overcome to fully realize this goal. One is to develop VLSI chips that are small, light, energy efficient, and general purpose enough to embody our brains' main equations, modules, and modal architectures—including laminar neocortical circuits such as those discussed in Section V-N—on board a mobile

robot, and that can be incrementally specialized to carry out all aspects of higher-order biological intelligence. This goal represents a huge intellectual and technical challenge.

While these developments are taking place, there remain many more scientific discoveries about how brains make minds that still have not been made, or at least not fully developed. Let me mention just two in ending.

Maintaining Balanced Systems: Homeostatic Plasticity. Each model neural network contains parameters that need to be approximately balanced in order to function properly in response to unexpected environmental challenges. An animat or mobile robot cannot strive to become autonomous if it cannot internally maintain balanced systems. Indeed, as I have noted in Section V-O, symptoms of several mental disorders can be traced to particular imbalances in complementary processes.

In vivo, achieving this balance occurs automatically via a process called *homeostatic plasticity* that has been intensely studied by Turrigiano *et al.* [147]–[150]. Our article [151] models aspects of homeostatic plasticity using recurrent shunting on-center off-surround networks with homeostatic modulations that automatically retune their optimal operating range through time. I consider this contribution to be just a small beginning of a major undertaking that will take quite a few ambitious young investigators to complete.

From Speech Perception to Language Understanding: Truly human-level intelligence requires the ability to learn and understand language by interacting with other humans. The work that my colleagues and I have done over the years includes modeling of foundational capabilities that are necessary for language, as well as other higher cognitive processes. These contributions include characterizing how laminar neocortical working memories are designed for the temporary storage of sequences of speech items, and the hierarchical chunking of those sequences into learned syllables, words, sequences of words, and sequences of sequences of words, as well as their variable-rate coarticulated performance (e.g., [66], [69], and [152]–[154]).

Much future work can profitably be done by building upon this conceptual and computational foundation to further model how rate-independent and speaker-independent language meanings are derived from rate-dependent and speaker-dependent speaker utterances. Ames and Grossberg [155] provided a pathway toward understanding how this happens by proposing the ARTSPEECH model to explain how speaker normalization is achieved. ARTSPEECH uses a specialization of the same cortical *strip map* neural circuit design that the ARTSTREAM model uses to separate the voices of different speakers in a noisy environment—that is, to solve the *cocktail party problem* [156]—and that is also used to represent numerical quantities, object orientations, and many other environmental events that may be coded by cortical maps. This design unity creates another exciting opportunity for future chip designers of humanoid animats.

Building upon these results will also require the efforts of many gifted young investigators for at least the next generation. More is written about these problems and all the other themes in this article in my forthcoming book [48].

REFERENCES

[1] *Stephen Grossberg*. [Online]. Available: https://en.wikipedia.org/wiki/Stephen_Grossberg

[2] M. A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, no. 5, pp. 815–826, Sep./Oct. 1983.

[3] S. Grossberg, "Some physiological and biochemical consequences of psychological postulates," *Proc. Nat. Acad. Sci.*, vol. 60, pp. 758–765, Jul. 1968.

[4] S. Grossberg, "Learning and energy-entropy dependence in some nonlinear functional-differential systems," *Bull. Amer. Math. Soc.*, vol. 75, no. 6, pp. 1238–1242, 1969.

[5] S. Grossberg, "Neural pattern discrimination," *J. Theor. Biol.*, vol. 27, pp. 291–337, May 1970.

[6] S. Grossberg, "Contour enhancement, short-term memory, and constancies in reverberating neural networks," *Stud. Appl. Math.*, vol. 52, pp. 213–257, Sep. 1973.

[7] S. Grossberg, "Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns," in *Competition and Cooperation in Neural Nets*, S. I. Amari and M. Arbib, Eds. New York, NY, USA: Springer, 1982.

[8] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Nat. Acad. Sci. USA*, vol. 81, pp. 3088–3092, May 1984.

[9] T. A. Burton, "Averaged neural networks," *Neural Netw.*, vol. 6, no. 5, pp. 677–680, 1993.

[10] T. Burwick, "Oscillatory networks: Pattern recognition without a superposition catastrophe," *Neural Comput.*, vol. 18, pp. 356–380, Feb. 2006.

[11] S. Guo, L. Huang, and L. Wang, "Linear stability and Hopf bifurcation in a two-neuron network with three delays," *Int. J. Bifurcat. Chaos*, vol. 14, no. 8, pp. 2799–2810, 2004.

[12] F. Hoppensteadt and E. Izhikevich, "Canonical neural models," in *Brain Theory and Neural Networks*, 2nd ed., M. A. Arbib, Ed. Cambridge, MA, USA: MIT Press, 2001, pp. 1–7.

[13] A. Menon, K. Mehrotra, C. K. Mohan, and S. Ranka, "Characterization of a class of sigmoid functions with applications to neural networks," *Neural Netw.*, vol. 9, pp. 819–835, Jul. 1996.

[14] J. Wu and X. Zou, "Patterns of sustained oscillations in neural networks with delayed interactions," *Appl. Math. Comput.*, vol. 73, pp. 55–75, Nov. 1995.

[15] S. Grossberg, "Nonlinear neural networks: Principles, mechanisms, and architectures," *Neural Netw.*, vol. 1, no. 1, pp. 17–61, 1988.

[16] J. J. Hopfield and D. W. Tank, "Computing with neural circuits: A model," *Science*, vol. 233, pp. 625–633, Aug. 1986.

[17] G. A. Carpenter, M. A. Cohen, and S. Grossberg, "Computing with neural networks," *Science*, vol. 235, pp. 1226–1227, Mar. 1987.

[18] S. Grossberg, "On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks," *J. Stat. Phys.*, vol. 1, pp. 319–350, Jun. 1969.

[19] S. Grossberg, "On the serial learning of lists," *Math. Biosci.*, vol. 4, pp. 201–253, Feb. 1969.

[20] S. Grossberg, "Pavlovian pattern learning by nonlinear neural networks," *Proc. Nat. Acad. Sci.*, vol. 68, pp. 828–831, Apr. 1971.

[21] S. Grossberg, "Pattern learning by functional-differential neural networks with arbitrary path weights," in *Delay Functional-Differential Equation and Their Application*, K. Schmitt, Ed. New York, NY, USA: Academic, 1972, pp. 121–160.

[22] S. Grossberg, *Studies of Mind and Brain*. Boston, MA, USA: D. Reidel Publishing Co., 1982.

[23] S. Grossberg and J. Pepe, "Spiking threshold and overarousal effects in serial learning," *J. Stat. Phys.*, vol. 3, pp. 95–125, Jun. 1971.

[24] S. Grossberg, "Competition, decision, and consensus," *J. Math. Anal. Appl.*, vol. 66, pp. 470–493, Nov. 1978.

[25] M. W. Hirsch, "Systems of differential equations that are competitive or cooperative I: Limit sets," *SIAM J. Math. Anal.*, vol. 13, pp. 167–179, Mar. 1982.

[26] S. Grossberg, "Decisions, patterns, and oscillations in nonlinear competitive systems with applications to Volterra-Lotka systems," *J. Theor. Biol.*, vol. 73, pp. 101–130, Jul. 1978.

[27] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, pp. 500–544, Aug. 1952.

[28] M. A. Cohen, "Sustained oscillations in a symmetric cooperative-competitive neural network: Disproof of a conjecture about content addressable memory," *Neural Netw.*, vol. 1, no. 3, pp. 217–221, 1988.

[29] S. A. Ellias and S. Grossberg, "Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks," *Biol. Cybern.*, vol. 20, pp. 69–98, Oct. 1975.

[30] E. Hopf, "Abzweigung einer periodischen Lösung von einer stationären Lösung eines differentialsystems," *Berlin Math. Phys. Sachs Acad. Wissen Leipzig*, vol. 94, pp. 1–22, Jan. 1942.

[31] A. Smith, *The Wealth of Nations*. London, U.K.: W. Strahan T. Cadell, 1776.

[32] A. J. Lotka, "Contribution to the theory of periodic reactions," *J. Phys. Chem.*, vol. 14, no. 3, pp. 271–274, 1910.

[33] R. M. May and W. J. Leonard, "Nonlinear aspects of competition between three species," *SIAM J. Appl. Math.*, vol. 29, no. 2, pp. 243–253, 1975.

[34] S. Smale, "On the differential equations of species in competition," *J. Math. Biol.*, vol. 3, pp. 5–7, Mar. 1976.

[35] R. Bhatt, G. A. Carpenter, and S. Grossberg, "Texture segregation by visual cortex: Perceptual grouping, attention, and learning," *Vis. Res.*, vol. 47, pp. 3173–3211, Nov. 2007.

[36] Y. Cao and S. Grossberg, "Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: A method for converting neural rate models into spiking models," *Neural Netw.*, vol. 26, pp. 75–98, Feb. 2012.

[37] Y. Cao and S. Grossberg, "A laminar cortical model for 3D boundary and surface representations of complex natural scenes," in *From Parallel to Emergent Computing*, A. Adamatzky, S. Akl, and G. Sirakoulis, Eds. Boca Raton, FL, USA: CRC, ch. 24, 2019.

[38] S. Grossberg, "3-D vision and figure-ground separation by visual cortex," *Percept. Psychophys.*, vol. 55, pp. 48–121, Jan. 1994.

[39] S. Grossberg, "Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures," *Psychol. Rev.*, vol. 104, pp. 618–658, Jul. 1997.

[40] S. Grossberg, "How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex," *Spat. Vis.*, vol. 12, no. 2, pp. 163–185, 1999.

[41] S. Grossberg, "Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support," *Neural Netw.*, vol. 87, pp. 38–95, Mar. 2017.

[42] S. Grossberg and S. Hong, "A neural model of surface perception: Lightness, anchoring, and filling-in," *Spat. Vis.*, vol. 19, nos. 2–4, pp. 263–321, 2006.

[43] S. Grossberg, A. Yazdanbakhsh, Y. Cao, and G. Swaminathan, "How does binocular rivalry emerge from cortical mechanisms of 3-D vision?" *Vis. Res.*, vol. 48, pp. 2232–2250, Sep. 2008.

[44] F. J. Kelly and S. Grossberg, "Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception," *Percept. Psychophys.*, vol. 62, pp. 1596–1618, Dec. 2000.

[45] R. Raizada and S. Grossberg, "Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system," *Cereb. Cort.*, vol. 13, pp. 100–113, Jan. 2003.

[46] S. Grossberg, "Communication, memory, and development," in *Progress in Theoretical Biology*, vol. 5, R. Rosen and F. Snell, Eds. New York, NY, USA: Academic, 1978, pp. 183–232.

[47] S. Grossberg, "A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans," in *Progress in Theoretical Biology*, vol. 5, R. Rosen and F. Snell, Eds. New York, NY, USA: Academic, 1978, pp. 233–374.

[48] S. Grossberg, *Conscious Mind/Resonant Brain: How Each Brain Makes a Mind*. New York, NY, USA: Oxford Univ. Press, 2021.

[49] A. Gierer and H. Meinhardt, "A theory of biological pattern formation," *Kybernetik*, vol. 12, pp. 30–39, Dec. 1972.

[50] J. T. Bonner, *On Development: The Biology of Form*. Cambridge, MA, USA: Harvard Univ. Press, 1974.

[51] H. Meinhardt and A. Gierer, "Applications of a theory of biological pattern formation based on lateral inhibition," *J. Cell Sci.*, vol. 15, pp. 321–346, Jul. 1974.

[52] P. A. Lawrence, F. H. C. Crick, and M. Munro, "A gradient of positional information in an insect, Rhodnius," *J. Cell Sci.*, vol. 112, pp. 815–853, Nov. 1972.

[53] T. Gustafson and L. Wolpert, "Cellular movement and contact in sea urchin morphogenesis," *Biol. Rev. Cambridge Philos. Soc.*, vol. 42, pp. 442–498, Aug. 1967.

[54] J. Berzhanskaya, S. Grossberg, and E. Mingolla, "Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception," *Spat. Vis.*, vol. 20, no. 4, pp. 337–395, 2007.

[55] S. Grossberg and A. Grunewald, "Temporal dynamics of binocular disparity processing with corticogeniculate interactions," *Neural Netw.*, vol. 15, pp. 181–200, Mar. 2002.

[56] S. Grossberg and L. Zajac, "How humans consciously see paintings and paintings illuminate how humans see," *Art Percept.*, vol. 5, pp. 1–95, Mar. 2017.

[57] A. Browning, S. Grossberg, and M. Mingolla, "Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance," *Neural Netw.*, vol. 22, pp. 1383–1398, Dec. 2009.

[58] S. Grossberg and P. K. Pilly, "Coordinated learning of grid cell and place cell spatial and temporal properties: Multiple scales, attention and oscillations," *Philos. Trans. Royal Soc. B.*, vol. 369, no. 1635, 2014, Art. no. 20120524. [Online]. Available: https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2012.0524

[59] J. W. Brown, D. Bullock, and S. Grossberg, "How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades," *Neural Netw.*, vol. 17, pp. 471–510, May 2004.

[60] H.-C. Chang, S. Grossberg, and Y. Cao, "Where's Waldo? How perceptual, cognitive, and emotional brain processes cooperate during learning to categorize and find desired objects in a cluttered scene," *Front. Integr. Neurosci.*, vol. 8, p. 43, Jun. 2014. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnint.2014.00043/full

[61] S. Grossberg, "Neural dynamics of the basal ganglia during perceptual, cognitive, and motor learning and gating," in *The Basal Ganglia: Novel Perspect. Motor Cognitive Functions*, J.-J. Soghomonian, Ed. Berlin, Germany: Springer, 2016, pp. 457–512.

[62] S. Grossberg, "The embodied brain of SOVEREIGN2: From space-variant conscious percepts during visual search and navigation to learning invariant object categories and cognitive-emotional plans for acquiring valued goals," *Front. Comput. Neurosci.,* vol. 13, p. 36, Jun. 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2019.00036/full

[63] S. Grossberg and D. Repin, "A neural model of how the brain represents and compares multi-digit numbers: Spatial and categorical processes," *Neural Netw.*, vol. 16, pp. 1107–1140, Oct. 2003.

[64] S. Grossberg, K. Srinivasan, and A. Yazdanbakhsh, "Binocular fusion and invariant category learning due to predictive remapping during scanning of a depthful scene with eye movements," *Front. Psychol. Percept. Sci.*, vol. 5, p. 1457, Jan. 2015. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01457/full

[65] S. Grossberg, "Desirability, availability, credit assignment, category learning, and attention: Cognitive-emotional and working memory dynamics of orbitofrontal, ventrolateral, and dorsolateral prefrontal cortices," *Brain Neurosci. Adv.*, vol. 2, pp. 1–50, May 2018. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/2398212818772179

[66] S. Grossberg and S. Kazerounian, "Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise," *J. Acoust. Soc. Amer.*, vol. 130, pp. 440–460, Jul. 2011.

[67] S. Grossberg, J. Palma, and M. Versace, "Resonant cholinergic dynamics in cognitive and motor decision-making: Attention, category learning, and choice in neocortex, superior colliculus, and optic tectum," *Front. Neurosci. Decis. Neurosci.*, vol. 9, p. 501, Jan. 2015. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2015.00501/full

[68] S. Grossberg and M. Versace, "Spikes, synchrony, and attentive learning by laminar thalamocortical circuits," *Brain Res.*, vol. 1218, pp. 278–312, Jul. 2008.

[69] S. Grossberg and L. Pearson, "Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works," *Psychol. Rev.*, vol. 115, pp. 677–732, Jul. 2008.

[70] M. R. Silver, S. Grossberg, D. Bullock, M. H. Histed, and E. K. Miller, "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 Netw.*, vol. 26, pp. 29–58, Feb. 2012.

[71] M. Dranias, S. Grossberg, and D. Bullock, "Dopaminergic and non-dopaminergic value systems in conditioning and outcome-specific revaluation," *Brain Res.*, vol. 1238, pp. 239–287, Oct. 2008.

[72] D. J. Franklin and S. Grossberg, "A neural model of normal and abnormal learning and memory consolidation: Adaptively timed conditioning, hippocampus, amnesia, neurotrophins, and consciousness," *Cogn. Affect. Behav. Neurosci.*, vol. 17, pp. 24–76, Feb. 2017. [Online]. Available: https://link.springer.com/article/10.3758/s13415-016-0463-y

[73] S. Grossberg and T. Vladusich, "How do children learn to follow gaze, share joint attention, imitate their teachers, and use tools during social interactions?" *Neural Netw.*, vol. 23, pp. 940–965, Oct./Nov. 2010.

[74] S. Grossberg, "Acetylcholine neuromodulation in normal and abnormal learning and memory: Vigilance control in waking, sleep, autism, amnesia and Alzheimer's disease," *Front. Neural Circuits*, vol. 11, p. 82, Nov. 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncir.2017.00082/full

[75] S. Grossberg and D. Kishnan, "Neural dynamics of autistic repetitive behaviors and fragile X syndrome: Basal ganglia movement gating and mGluR-modulated adaptively timed learning," *Front. Psychol. Psychopath.*, vol. 9, p. 269, Mar. 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00269/full

[76] S. Grossberg and D. Seidman, "Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates," *Psychol. Rev.*, vol. 113, pp. 483–525, Jul. 2006.

[77] W. Gnadt and S. Grossberg, "SOVEREIGN: An autonomous neural system for incrementally learning planned action sequences to navigate towards a rewarded goal," *Neural Netw.*, vol. 21, pp. 699–758, Jun. 2008.

[78] S. Grossberg, "A path toward explainable AI and autonomous adaptive intelligence: Deep learning, adaptive resonance, and models of perception, emotion, and action," *Front. Neurobot.*, vol. 14, p. 36, Jun. 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnbot.2020.00036/full

[79] S. Grossberg, "The resonant brain: How attentive conscious seeing regulates action sequences that interact with attentive cognitive learning, recognition, and prediction," *Attent. Percept. Psychophys.*, vol. 81, pp. 2237–2264, Jun. 2019.

[80] S. Grossberg, "Towards building a neural networks community," *Neural Netw.*, vol. 23, pp. 1135–1138, Dec. 2010.

[81] S. Grossberg, "Classical and instrumental learning by neural networks," in *Progress in Theoretical Biology*, R. Rosen and F. Snell, Eds. New York, NY, USA: Academic, 1974, pp. 51–141.

[82] S. Grossberg, "Biological competition: Decision rules, pattern formation, and oscillations," *Proc. Nat. Acad. Sci. USA*, vol. 77, no. 4, pp. 2338–2342, 1980.

[83] S. Grossberg, "A prediction theory for some nonlinear functional-differential equations, I: Learning of lists," *J. Math. Anal. Appl.*, vol. 21, no. 3, pp. 643–694. 1968.

[84] S. Grossberg, "A prediction theory for some nonlinear functional-differential equations, II: Learning of patterns," *J. Math. Anal. Appl.*, vol. 22, no. 3, pp. 490–522, 1968.

[85] S. Grossberg, "On the global limits and oscillations of a system of nonlinear differential equations describing a flow of a probabilistic network," *J. Differential Equ.*, vol. 5, no. 3, pp. 531–563, 1969.

[86] S. Grossberg, "Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, I," *J. Math. Mech.*, vol. 19, no. 1, pp. 53–91, 1969.

[87] S. Grossberg, "Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II," *Stud. Appl. Math.*, vol. 49, no. 1, pp. 135–166, 1970.

[88] S. Grossberg, "How does a brain build a cognitive code?" *Psychol. Rev.*, vol. 87, no. 1, pp. 1–51, 1980.

[89] S. Grossberg, "On learning, information, lateral inhibition, and transmitters," *Math. Biosci.*, vol. 4, nos. 3–4, pp. 255–310, 1969.

[90] W. B. Levy, S. E. Brassel, and S. D. Moore, "Partial quantification of the associative synaptic learning rule of the dentate gyrus," *Neuroscience*, vol. 8, no. 4, pp. 799–808, 1983.

[91] W. Singer, "Neuronal activity as a shaping factor in the self-organization of neuron assemblies," in *Synergetics of Brain*. E. Basar, H. Flohr, H. Haken, and A. J. Mandell, Eds. New York, NY, USA: Springer, 1983, pp. 89–101.

[92] S. Grossberg, "Birth of a learning law," *INNS/ENNS/JNNS Newslett.*, vol. 21, pp. 1–4, 1998.

[93] S. Grossberg. (2013). *Recurrent Neural Networks*. [Online]. Available: http://www.scholarpedia.org/article/Recurrent_neural_networks

[94] Y. Cao and S. Grossberg, "A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis," *Spatial Vis.*, vol. 18, no. 5, pp. 515–578, 2005.

[95] S. Grossberg, "How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex," *Behav. Cogn. Neurosci. Rev.*, vol. 2, no. 1, pp. 47–76, 2003.

[96] S. Grossberg J. Léveillé, and M. Versace, "How do object reference frames and motion vector decomposition emerge in laminar cortical circuits?" *Attention Perception Psychophys.*, vol. 73, no. 4, pp. 1147–1170, 2011.

[97] S. Grossberg and A. Seitz, "Laminar development of receptive fields, maps, and columns in visual cortex: The coordinating role of the subplate," *Cerebr. Cortex*, vol. 13, no. 8, pp. 852–863, 2003.

[98] S. Grossberg and G. Swaminathan, "A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: Development, attention and bistability," *Vis. Res.*, vol. 44, no. 11, pp. 1147–1187, 2004.

[99] S. Grossberg and J. R. Williamson, "A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings and learning," *Cerebr. Cortex*, vol. 11, no. 1, pp. 37–58, 2001.

[100] S. Grossberg and A. Yazdanbakhsh, "Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading," *Vis. Res.*, vol. 45, no. 13, pp. 1725–1743, 2005.

[101] J. Léveillé, M. Versace, and S. Grossberg, "Running as fast as it can: How spiking dynamics form object groupings in the laminar circuits of visual cortex," *J. Comput. Neurosci.*, vol. 28, no. 2, pp. 323–346, 2010.

[102] A. Yazdanbakhsh and S. Grossberg, "Fast synchronization of perceptual grouping in laminar visual cortical circuits," *Neural Netw.*, vol. 17, nos. 5–6, pp. 707–718, 2004.

[103] S. Grossberg, "Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory," in *Brain and Information Event-Related Potentials*, R. Karrer, J. Cohen, and P. Tueting, Eds. New York, NY, USA: Acad. Sci., 1984, pp. 58–142.

[104] S. Grossberg and D. S. Levine, "Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement," *Appl. Opt.*, vol. 26, no. 23, pp. 5015–5030, 1987.

[105] S. Grossberg, "Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world," *Neural Netw.*, vol. 37, pp. 1–47, Jan. 2013.

[106] S. Grossberg, "A half century of progress towards a unified neural theory of mind and brain with applications to autonomous adaptive agents and mental disorders," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, R. Kozma, C. Alippi, Y. Choe, and F. C. Morabito, Eds. Cambridge, MA, USA: Academic, 2018, pp. 31–52.

[107] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends Cogn. Sci.*, vol. 3, no. 4, pp. 128–135, 1999.

[108] M. McCloskey and N. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," *Psychol. Learn. Motivation*, vol. 24, pp. 109–164, 1989.

[109] R. Ratcliff, "Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions," *Psychol. Rev.*, vol. 97, no. 2, pp. 285–308, 1990.

[110] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 698–713, Sep. 1992.

[111] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Netw.*, vol. 4, no. 5, pp. 565–588, 1991.

[112] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems," *IEEE Trans. Neural Netw.*, vol. 6, no. 6, pp. 1330–1336, Nov. 1995.

[113] T. P. Caudell, S. D. G. Smith, R. Escobedo, and M. Anderson, "NIRS: Large scale ART-1 neural architectures for engineering design retrieval," *Neural Netw.*, vol. 7, no. 9, pp. 1339–1350, 1994.

[114] R. Escobedo, S. D. G. Smith, and T. P. Caudell, "A neural information retrieval system," *Int. J. Adv. Manuf. Technol.*, vol. 8, pp. 269–273, Jul. 1993.

[115] S. Grossberg, "On the development of feature detectors in the visual cortex with applications to learning and reaction–diffusion systems," *Biol. Cybern.*, vol. 21, pp. 145–159, Sep. 1976.

[116] S. Grossberg, "Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors," *Biol. Cybern.*, vol. 23, pp. 121–134, Sep. 1976.

[117] S. Grossberg, "Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions," *Biol. Cybern.*, vol. 23, pp. 187–202, Dec. 1976.

[118] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self organizing neural pattern recognition machine," *Comput. Vis. Graph. Image Process.*, vol. 37, no. 1, pp. 54–115, 1987.

[119] G. A. Carpenter and S. Grossberg, "The ART of adaptive pattern recognition by a self-organizing neural network," *Computer*, vol. 21, no. 3, pp. 77–88, 1988.
[120] G. A. Carpenter, "Distributed learning, recognition, and prediction by ART and ARTMAP neural networks," *Neural Netw.*, vol. 10, no. 8, pp. 1473–1494, 1997.
[121] G. A. Carpenter, B. L. Milenova, and B. W. Noeske, "dARTMAP: A neural network for fast distributed supervised learning," *Neural Netw.*, vol. 11, no. 5, pp. 793–813, 1998.
[122] G. A. Carpenter and S. Grossberg, "Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions," *Trends Neurosci.*, vol. 16, no. 4, pp. 131–137, 1993.
[123] S. Grossberg, "The complementary brain: Unifying brain dynamics and modularity," *Trends Cogn. Sci.*, vol. 4, no. 6, pp. 233–246, 2000.
[124] E. A. DeYoe and D. C. Van Essen, "Concurrent processing streams in monkey visual cortex," *Trends Neurosci.*, vol. 11, no. 5, pp. 219–226, 1988.
[125] J. P. Banquet and S. Grossberg, "Probing cognitive processes through the structure of eventrelated potentials during learning: An experimental and theoretical analysis," *Appl. Opt.*, vol. 26, no. 23, pp. 4931–4946, 1987.
[126] R. Näätänen, "Processing negativity: An evoked-potential reflection of selective attention," *Psychol. Bull.*, vol. 92, no. 3, pp. 605–640, 1982.
[127] R. Näätänen, M. Simpson, and N. E. Loveless, "Stimulus deviance and evoked potentials," *Biol. Psychol.*, vol. 14, nos. 1–2, pp. 53–98, 1982.
[128] M. Sams, P. Paavilainen, K. Alho, and R. Näätänen, "Auditory frequency discrimination and event-related potentials," *Electroencephalography Clin. Neurophysiol. Evoked Potentials Section*, vol. 62, no. 6, pp. 437–448, 1985.
[129] S. Grossberg, "Towards a unified theory of neocortex: Laminar cortical circuits for vision and cognition," in *Computational Neuroscience: From Neurons to Theory and Back Again*, P. Cisek, T. Drew, and J. Kalaska, Eds. Amsterdam, The Netherlands: Elsevier, 2007, pp. 79–104.
[130] S. Grossberg and R. D. S. Raizada, "Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex," *Vis. Res.*, vol. 40, nos. 10–12, pp. 1413–1432, 2000.
[131] R. Raizada and S. Grossberg, "Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast," *Vis. Cogn.*, vol. 8, nos. 3–5, pp. 431–466, 2001.
[132] A. Fazl, S. Grossberg, and E. Mingolla, "View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds," *Cogn. Psychol.*, vol. 58, no. 1, pp. 1–48. 2009.
[133] C. W. Tyler and L. L. Kontsevich, "Mechanisms of stereoscopic processing: Stereoattention and surface perception in depth reconstruction," *Perception*, vol. 24, no. 2, pp. 127–153, 1995.
[134] T. J. Buschman and E. K. Miller, "Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices," *Science*, vol. 315, no. 5820, pp. 1860–1862, 2007.
[135] C. G. Kentros, N. T. Agniotri, S. Streater, R. D. Hawkins, and E. R. Kandel, "Increased attention to spatial context increases both place field stability and spatial memory," *Neuron*, vol. 42, no. 2, pp. 283–295, 2004.
[136] R. G. M. Morris and U. Frey, "Hippocampal synaptic plasticity: Role in spatial learning or the automatic recording of attended experience?" *Philosoph. Trans. Roy. Soc. B Biol. Sci.*, vol. 352, no. 1360, pp. 1469–1503, 1997.
[137] P. K. Pilly and S. Grossberg, "How do spatial learning and memory occur in the brain? Coordinated learning of entorhinal grid cells and hippocampal place cells," *J. Cogn. Neurosci.*, vol. 24, no. 5, pp. 1031–1054, 2012.
[138] Y. Cao, S. Grossberg, and J. Markowitz, "How does the brain rapidly learn and reorganize view- and positionally-invariant object representations in inferior temporal cortex?" *Neural Netw.*, vol. 24, no. 10, pp. 1050–1061, 2011.
[139] J. Brown, D. Bullock, and S. Grossberg, "How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues," *J. Neurosci.*, vol. 19, no. 23, pp. 10502–10511, 1999.
[140] J. C. Fiala, S. Grossberg, and D. Bullock, "Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye blink response," *J. Neurosci.*, vol. 16, no. 11, pp. 3760–3774, 1996.
[141] S. Grossberg, "A neural theory of punishment and avoidance, I: Qualitative theory," *Math. Biosci.*, vol. 15, no. 1–2, pp. 39–67, 1972.
[142] S. Grossberg, "A neural theory of punishment and avoidance, I: Quantitative theory," *Math. Biosci.*, vol. 15, nos. 1–2, pp. 253–285, 1972.
[143] S. Grossberg, "Cortical dynamics of three-dimensional form, color, and brightness perception, I: Monocular theory," *Perception Psychophys.*, vol. 41, no. 2, pp. 87–116, 1987.
[144] S. Grossberg, "Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory," *Perception Psychophys.*, vol. 41, no. 2, pp. 117–158, 1987.
[145] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading," *Psychol. Rev.*, vol. 92, no. 2, pp. 173–211, 1985.
[146] S. Grossberg and E. Mingolla, "Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations," *Perception Psychophys.*, vol. 38, no. 2, pp. 141–171, 1985.
[147] G. G. Turrigiano, "Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same," *Trends Neurosci.*, vol. 22, no. 5, pp. 221–227, 1999.
[148] G. G. Turrigiano and S. B. Nelson, "Homeostatic plasticity in the developing nervous system," *Nat. Rev. Neurosci.*, vol. 5, pp. 97–107, Feb. 2004.
[149] G. G. Turrigiano and S. B. Nelson, "Hebb and homeostasis in neuronal plasticity," *Current Opin. Neurobiol.*, vol. 10, no. 3, pp. 358–364, 2000.
[150] C. J. Wierenga, K. Ibata, and G. G. Turrigiano, "Postsynaptic expression of homeostatic plasticity at neocortical synapses," *J. Neurosci.*, vol. 25, no. 11, pp. 2895–2905, 2005.
[151] B. Chandler and S. Grossberg, "Joining distributed pattern processing and homeostatic plasticity in recurrent on-center off-surround shunting networks: Noise, saturation, short-term memory, synaptic scaling, and BDNF," *Neural Netw.*, vol. 25, no. 1, pp. 21–29, 2012.
[152] I. Boardman, S. Grossberg, C. Myers, and M. A. Cohen, "Neural dynamics of perceptual order and context effects for variable-rate speech syllables," *Perception Psychophys.*, vol. 6, pp. 1477–1500, Dec. 1999.
[153] M. A. Cohen and S. Grossberg, "Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants," *IEEE Trans. Audio, Speech, Language Process.*, vol. 5, no. 4, pp. 301–318, Jul. 1997.
[154] S. Grossberg, I. Boardman, and M. A. Cohen, "Neural dynamics of variable-rate speech categorization," *J. Exp. Psychol. Human Perception Perform.*, vol. 23, no. 2, pp. 418–503, 1997.
[155] H. Ames and S. Grossberg, "Speaker normalization using cortical strip maps: A neural model for steady state vowel categorization," *J. Acoust. Soc. Amer.*, vol. 124, no. 6, pp. 3918–3936, 2008.
[156] S. Grossberg K. K. Govindarajan, L. L. Wyse, and M. A. Cohen, "ARTSTREAM: A neural network model of auditory scene analysis and source segregation," *Neural Netw.*, vol. 17, no. 4, pp. 511–536, 2004.
[157] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nat. Neurosci.*, vol. 12, no. 6, pp. 718–724, 2009.