Neural Networks 37 (2013) 1–47

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world ☆

Stephen Grossberg *

Center for Adaptive Systems, Graduate Program in Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, United States
Center for Computational Neuroscience and Neural Technology, and Department of Mathematics, Boston University, 677 Beacon Street, Boston, MA 02215, United States

ARTICLE INFO

Dedicated to Gail Carpenter in appreciation of her many fundamental contributions to Adaptive Resonance Theory.

Keywords:
Adaptive Resonance Theory
Consciousness
Learning
Expectation
Attention
Recognition
Synchrony
Working memory
Reinforcement learning
Adaptive timing
Working memory
Speech perception
Gamma and beta oscillations
Inferotemporal cortex
Parietal cortex
Prefrontal cortex
Entorhinal cortex
Hippocampal cortex
Amygdala
Basal ganglia

ABSTRACT

Adaptive Resonance Theory, or ART, is a cognitive and neural theory of how the brain autonomously learns to categorize, recognize, and predict objects and events in a changing world. This article reviews classical and recent developments of ART, and provides a synthesis of concepts, principles, mechanisms, architectures, and the interdisciplinary data bases that they have helped to explain and predict. The review illustrates that ART is currently the most highly developed cognitive and neural theory available, with the broadest explanatory and predictive range. Central to ART's predictive power is its ability to carry out fast, incremental, and stable unsupervised and supervised learning in response to a changing world. ART specifies mechanistic links between processes of consciousness, learning, expectation, attention, resonance, and synchrony during both unsupervised and supervised learning. ART 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 that are functionally described include visual and auditory neocortex; specific and nonspecific thalamic nuclei; inferotemporal, parietal, prefrontal, entorhinal, hippocampal, parahippocampal, perirhinal, and motor cortices; frontal eye fields; supplementary eye fields; amygdala; basal ganglia: cerebellum; and superior colliculus. Due to the complementary organization of the brain, ART does not describe many spatial and motor behaviors whose matching and learning laws differ from those of ART. ART algorithms for engineering and technology are listed, as are comparisons with other types of models.

☆ Invited article submitted to the 25th anniversary special issue of *Neural Networks*.

* Correspondence to: Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, United States. Tel.: +1 617 353 7858; fax: +1 617 353 7755.
E-mail address: steve@bu.edu.
URL: http://cns.bu.edu/~steve.

0893-6080/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2012.09.017

1. Introduction: the stability–plasticity dilemma and rapid learning throughout life

1.1. Fast recognition learning without catastrophic forgetting

Adaptive Resonance Theory, or ART, is a cognitive and neural theory of how the brain autonomously learns to categorize, recognize, and predict objects and events in a changing world. The problem of learning makes the unity of conscious experience hard to understand, if only because humans are able to rapidly learn enormous amounts of new information, on their own, throughout life. How do humans integrate all this information into unified experiences that cohere into a sense of self? One has only to see an exciting movie just once to marvel at this capacity, since we can then tell our friends many details about it later on, even though the individual scenes flashed by very quickly. More generally, we can quickly learn about new environments, even if no one tells us how the rules of each environment differ. To a remarkable degree, humans can rapidly learn new facts without being forced to just as rapidly forget what they already know. As a result, we can confidently go out into the world without fearing that, in learning to recognize a new friend's face, we will suddenly forget the faces of our family and friends. This is sometimes called the problem of *catastrophic forgetting*.

1.2. Some models that experience catastrophic forgetting

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, back propagation, simulated annealing, neocognitron, support vector machine, regularization, and Bayesian models. The brain solves a challenging problem that many current biological and technological learning models have not yet solved: It is a self-organizing system that is capable of rapid, yet stable, autonomous learning in real time of huge amounts of data from a changing environment that can be filled with unexpected events. Discovering the brain's solution to this key problem is as important for understanding ourselves as it is for developing new pattern recognition and prediction applications in technology.

Grossberg (1980) has called the problem whereby the brain learns quickly and stably without catastrophically forgetting its past knowledge the *stability–plasticity dilemma*. The stability–plasticity dilemma must be solved by every brain system that needs to rapidly and adaptively respond to the flood of signals that subserves even the most ordinary experiences. If the brain's design is parsimonious, then similar design principles should operate in all brain systems that can stably learn an accumulating knowledge base in response to changing conditions throughout life. The discovery of such principles should clarify how the brain unifies diverse sources of information into coherent moments of conscious experience. ART has attempted to articulate these principles and the neural mechanisms that realize them. The next sections summarize aspects of how this is proposed to occur.

1.3. Linking consciousness, learning, expectation, attention, resonance, and synchrony

ART clarifies key brain processes from which conscious experiences emerge. It predicts a functional link between processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony (the CLEARS processes). ART predicted that all brain representations that solve the stability–plasticity dilemma use variations of CLEARS mechanisms (Grossberg, 1978a, 1980, 2007a). Synchronous resonances are, in particular, expected to occur between multiple cortical and subcortical areas. Various data support this prediction; e.g., see Buschman and Miller (2007), Engel, Fries, and Singer (2001), Grossberg (2009b), and Pollen (1999).

Through these CLEARS connections, ART clarifies why many animals are intentional beings who pay attention to salient objects, why "all conscious states are resonant states", and how brains can learn both *many-to-one maps* (representations whereby many object views, positions, and sizes all activate the same invariant object category) and *one-to-many maps* (representations that enable us to expertly know many things about individual objects and events).

ART accomplishes these properties by proposing how top-down expectations focus attention on salient combinations of cues, and characterizes how attention may operate via a form of self-normalizing "biased competition" (Desimone, 1998). ART explains how such top-down attentive matching may help 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.

All of the main ART predictions have received increasing support from psychological and neurobiological data since ART was introduced in Grossberg (1976a, 1976b). ART has undergone continual development to explain and predict increasingly large behavioral and neurobiological data bases, ranging 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. In this sense, they provide a growing set of examples capable of partially solving the classical mind/body problem. All the author's major articles, including those that develop ART, may be downloaded from http://cns.bu.edu/~steve.

1.4. Equations for short-term memory, medium-term memory, and long-term memory

How does ART sit within the corpus of all neural models? In particular, is the brain just a bag of tricks, as some authors have proposed (e.g., Ramachandran (1990))? This article illustrates a contrary view based on the author's view after developing models of mind and brain for 55 years (Grossberg, 1988, http://cns.bu.edu/Profiles/Grossberg/GrossbergNNeditorial2010.pdf). During this period, I led the discovery and development of a small number of equations (e.g., equations for short-term memory, or STM; medium-term memory, or MTM; and long-term memory, or LTM) and a somewhat larger number of modules or microcircuits (e.g., shunting on-center off-surround networks, gated dipole opponent processing networks, associative learning networks, reinforcement learning networks, spectral timing networks, and the like), which have been specialized and assembled into modal architectures, where the term "modal" stands for modality (e.g., architectures for vision, audition, cognition, cognitive–emotional interactions, sensory–motor control, and the like). Modal architectures are less general than a Turing or von Neumann architecture for general computing, but far more general than a traditional AI algorithm. They are designed to be capable of general-purpose self-organizing processing of a particular modality of biological intelligence and their particular specializations of the basic equations and modules have been selected over the millennia by evolutionary pressures.

ART networks form part of such modal architectures. Modal architectures, in turn, embody new paradigms for brain computing that I have called Complementary Computing (Grossberg, 2000b)