Adaptive Resonance Theory

Stephen Grossberg

September, 2000

Technical Report CAS/CNS-2000-024

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston, MA 02215

A 67
ENCYCLOPEDIA OF COGNITIVE SCIENCE
November, 2000
Publisher: Macmillan Reference Ltd., 25 Eccleston Place, London SW1W 9NF

ADAPTIVE RESONANCE THEORY

Contributor: Professor Stephen Grossberg[1]
Department of Cognitive and Neural Systems and Center for Adaptive Systems[2]
Boston University
677 Beacon Street
Boston, MA 02215
Send all correspondence to Professor Stephen Grossberg (above)
Phone: (617) 353-7857; Fax: (617) 353-7755
Email: steve@bu.edu

Technical Report CAS/CNS-2000-024
Boston, MA: Boston University

Keywords: adaptive resonance, recognition, learning, categorization, amnesia

List of first level subheadings:

Article definition:

Adaptive resonance theory, or ART, is a cognitive and neural theory about how the brain develops and learns to recognize and recall objects and events throughout life. ART shows how processes of learning, categorization, expectation, attention, resonance, synchronization, and memory search interact to enable the brain to learn quickly and to retain its memories stably, while explaining many data about perception, cognition, learning memory, and consciousness along the way.

[1] Supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409), the National Science Foundation (NSF IRI-97-20333), and the Office of Naval Research (ONR N00014-95-1-0657).
[2] The author thanks Diana Meyers for her valuable assistance in the preparation of the manuscript and figures.

Introduction

The processes whereby our brains continue to learn about, recognize, and recall a changing world in a stable fashion throughout life are among the most important for understanding cognition. These processes include the learning of top-down expectations, the matching of these expectations against bottom-up data, the focusing of attention upon the expected clusters of information, and the development of resonant states between bottom-up and top-down processes as they reach an attentive consensus between what is expected and what is there in the outside world. It is suggested that all conscious states in the brain are resonant states, and that these resonant states trigger learning of sensory and cognitive representations. The models which summarize these concepts are therefore called Adaptive Resonance Theory, or ART, models. ART was introduced in Grossberg (1976a, 1976b), along with rules for competitive learning and self-organizing maps. Since then, psychophysical and neurobiological data in support of ART have been reported in experiments on vision, visual object recognition, auditory streaming, variable-rate speech perception, somatosensory perception, and cognitive-emotional interactions, among others; e.g., Carpenter and Grossberg (1991, 1993), Grossberg (1999a, 1999b) and Grossberg and Merrill (1996). In particular, ART mechanisms seem to be operative at all levels of the visual system, and it is proposed how these mechanisms are realized by known laminar circuits of visual cortex. It is predicted that the same circuit realization of ART mechanisms will be found, suitably specialized, in the laminar circuits of all sensory and cognitive neocortex.

Although ART-style learning and matching processes seem to be found in many sensory and cognitive processes, another type of learning and matching are often found in spatial and motor processes. In particular, it is suggested that sensory and cognitive processing in the What processing stream of the brain obey learning and matching laws that are often *complementary* to those used for spatial and motor processing in the brain's Where processing stream. This enables our sensory and cognitive representations to maintain their stability as we learn more about the world, while allowing spatial and motor representations to forget learned maps and gains that are no longer appropriate as our bodies develop and grow from infanthood to adulthood. Procedural memories are proposed to be unconscious because the inhibitory matching process that supports these spatial and motor processes cannot lead to resonance.

The stability-plasticity dilemma: Rapid learning throughout life

We experience the world as a whole. Although myriad signals relentlessly bombard our senses, we somehow integrate them into unified moments of conscious experience that cohere together despite their diversity. Because of the apparent unity and coherence of our awareness, we can develop a sense of self that can gradually mature with our experiences of the world. This capacity lies at the heart of our ability to function as intelligent beings.

The apparent unity and coherence of our experiences is all the more remarkable when we consider several properties of how the brain copes with the environmental events that it processes. First and foremost, these events are highly context-sensitive. When we look at a complex picture or scene as a whole, we can often recognize its objects and its meaning at a glance, as in the picture of a familiar face. However, if we process the face piece-by-piece, as through a small aperture, then its significance may be greatly degraded. To cope with this context-sensitivity, the brain typically processes pictures and other sense data in parallel, as *patterns* of activation across a large number of feature-sensitive nerve cells, or neurons. The same is true for senses other than vision, such as audition. If the sound of the word GO is altered by clipping off the vowel O, then the consonant G may sound like a chirp, quite unlike its sound as part of GO.

During vision, all the signals from a scene typically reach the photosensitive retinas of the eyes at essentially the same time, so parallel processing of all the scene's parts begins at the retina itself. During audition, each successive sound reaches the ear at a later time. Before an entire pattern of sounds, such as the word GO, can be processed as a whole, it needs to be recoded, at a later processing stage, into a simultaneously available spatial pattern of activation. Such a processing stage is often called a working memory, and the activations that it stores are often called short term memory (STM) traces. For example, when you hear an unfamiliar telephone number, you can temporarily store it in working memory while you walk over to the telephone and dial the number.

In order to determine which of these patterns represents familiar events and which do not, the brain matches these patterns against stored representations of previous experiences that have been acquired through learning. Unlike the STM traces that are stored in a working memory, the learned experiences are stored in long term memory (LTM) traces. One difference between STM and LTM traces concerns how they react to distractions. For example, if you are distracted by a loud noise before you dial a new telephone number, its STM representation can be rapidly reset so that you forget it. On the other hand, if you are distracted by a loud noise, you (hopefully) will not forget the LTM representation of your own name.

How does learning of new information get stably stored in LTM? For example, after seeing an exciting movie just once, we can 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 surprising degree, we can rapidly learn new facts without being forced to just as rapidly forget what we already know. As a result, we do not need to avoid going out into the world for fear that, in learning to recognize a new friend's face, we will suddenly forget our parents' faces. This is sometimes called the problem of *catastrophic forgetting*.

Many contemporary learning algorithms *can* forget catastrophically. In contrast, the brain is capable of rapid yet stable autonomous learning of huge amounts of data in an ever-changing world. 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.

I have 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 we should expect to find similar design principles operating in all the brain systems that can stably learn an accumulating knowledge base in response to changing conditions throughout life. The discovery of such principles should also clarify how the brain unifies diverse sources of information into coherent moments of conscious experience.

The link between learning, expectation, attention, and resonance

Humans are *intentional* beings who learn expectation about the world and make predictions about what is about to happen. Humans are also *attentional* beings who focus processing resources upon a restricted amount of incoming information at any time. Why are we both intentional and attentional beings, and are these two types of processes related? The stability-plasticity dilemma and its solution using resonant states provides a unifying framework for understanding these issues.

To fix ideas about how we use a sensory or cognitive expectation, and how a resonant state is activated, suppose you were asked to "find the yellow ball within one-half second, and you will win a $10,000 prize". Activating an expectation of "yellow balls" enables more rapid detection of a yellow ball, and with a more energetic neural response, than if you were not looking for it. Sensory and cognitive top-down expectations hereby lead to *excitatory matching* with confirmatory bottom-up data. On the other hand, mismatch between top-down expectations and bottom-up data can suppress the mismatched part of the bottom-up data, and thereby start to focus attention upon the matched, or expected, part of the bottom-up data.

This sort of excitatory matching and attentional focusing on bottom-up data using top-down expectations is proposed to generate resonant brain states: When there is a good enough match between bottom-up and top-down signal patterns between two or more levels of processing, their positive feedback signals amplify and prolong their mutual activation, leading to a resonant state. The amplification and prolongation of the system's fast activations is sufficient to trigger learning in the more slowly varying adaptive weights that control the signal flow along pathways from cell to cell. Resonance hereby provides a global context-sensitive indicator that the system is processing data worthy of learning. That is why the theory which describes these processes is called *Adaptive* Resonance Theory, or ART.

ART thus predicts that there is an intimate connection between the mechanisms which enable us to learn quickly and stably about a changing world, and the mechanisms that enable us to learn expectations about such a world, test hypotheses about it, and focus attention upon information that we find interesting. ART also proposes that, in order to solve the stability-plasticity dilemma, only resonant states can drive rapid new learning, which gives the theory its name.

Learning within the sensory and cognitive domain is often *match learning*. Match learning occurs only if a good enough match occurs between bottom-up information and a learned top-down expectation that is read out by an active recognition category, or code. When such an approximate match occurs, previously learned knowledge can be refined. If novel information cannot form a good enough match with the expectations that are read-out by previously learned recognition categories, then a memory search, or hypothesis testing, is triggered that leads to selection and learning of a new recognition category, rather than catastrophic forgetting of an old one. Figure 1 illustrates how this happens in an ART model; it will be discussed in greater detail below. In contrast, learning within spatial and motor processes is proposed to be *mismatch learning* that continuously updates sensory-motor maps or the gains of sensory-motor commands. As a result, we can stably learn what is happening in a changing world, thereby solving the stability-plasticity dilemma, while adaptively updating our representations of where objects are and how to act upon them using bodies whose parameters change continuously through time.

It has been mathematically proved that match learning within an ART model leads to stable memories in response to arbitrary list of events to be learned (Carpenter and Grossberg, 1991). Match learning also has a serious potential weakness, however: If you can only learn when there is a good enough match between bottom-up data and learned top-down expectations, then how do you ever learn anything that you do not already know? ART proposes that this problem is solved by the brain by using another complementary interaction, this one between processes of *resonance* and *reset,* that are predicted to control properties of attention and memory search, respectively. These complementary processes help our brains to balance between the complementary demands of processing the familiar and the unfamiliar, the expected and the unexpected. One of these complementary processes is predicted to take place in the What cortical stream, notably in the visual, inferotemporal, and prefrontal cortex. It is here that top-down expectations are matched against bottom-up inputs (Chelazzi, *et al.,* 1998; Miller *et al,* 1996). When a top-down expectation achieves a good enough match with bottom-up data, this match process focuses attention upon those feature clusters in the bottom-up input that are expected. If the expectation is close enough to the input pattern, then a state of resonance develops as the attentional focus takes hold.

Figure 1 illustrates these ART ideas in a simple two-level example. Here, a bottom-up input pattern, or vector, I activates a pattern X of activity across the feature detectors of the first level F_1. For example, a visual scene may be represented by the features comprising its boundary and surface representations. This feature pattern represents the relative importance of different features in the inputs pattern I. In Figure 1A, the pattern peaks represent more activated feature detector cells, the troughs less activated feature detectors. This feature pattern sends signals S through an adaptive filter to the second level F_2 at which a compressed representation Y (also called a recognition category, or a symbol) is activated in response to the distributed input T. Input T is computed by multiplying the signal vector S by a matrix of adaptive weights that can be altered through learning. The representation Y is compressed by competitive interactions across F_2 that allow only a small subset of its most strongly activated cells to remain active in response to T. The pattern Y in the figure indicates that a small number of category cells may be activated to different degrees. These category cells, in turn, send top-down signals U to F_1. The vector U is converted into the top-down expectation V by being multiplied by another matrix of adaptive weights. When V is received by F_1, a matching process takes place between the input vector I and V which selects that subset X^* of F_1 features that were "expected" by the active F_2 category Y. The set of these selected features is the emerging "attentional focus".

Reconciling distributed and symbolic representations using resonance

If the top-down expectation is close enough to the bottom-up input pattern, then the pattern X^* of attended features reactivates the category Y which, in turn, reactivates X^*. The network hereby locks into a resonant state through a positive feedback loop that dynamically links, or binds, the attended features across X^* with their category, or symbol, Y.

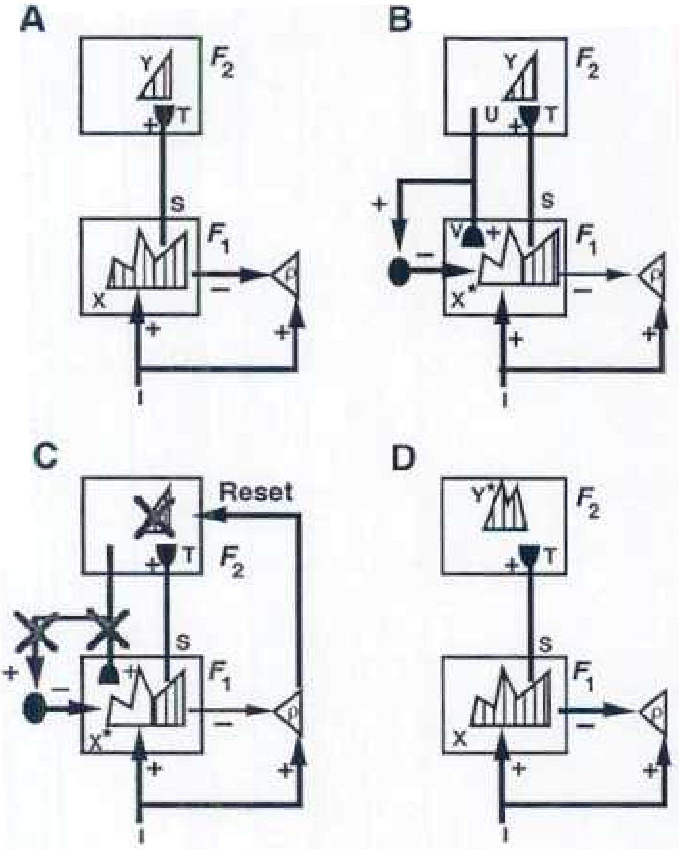

Figure 1. Search for a recognition code within an ART learning circuit: (A) The input pattern I is instated across the feature detectors at level F_1 as a short term memory (STM) activity pattern X. Input I also nonspecifically activates the orienting system ρ; that is, all the input pathways converge on ρ and can activate it. STM pattern X is represented by the hatched pattern across F_1. Pattern X both inhibits ρ and generates the output pattern S. Pattern S is multiplied by learned adaptive weights, also called long term memory (LTM) traces. These LTM-gated signals are added at F_2 cells, or nodes, to form the input pattern T, which activates the STM pattern Y across the recognition categories coded at level F_2. (B) Pattern Y generates the top-down output pattern U which is multiplied by top-down LTM traces and added at F_1 nodes to form a *prototype* pattern V that encodes the learned expectation of the active F_2 nodes. Such a prototype represents the set of commonly shared features in all the input patterns capable of activating Y. If V mismatches I at F$_1$, then a new STM activity pattern X^* is selected *at* F_1. X^* is represented by the hatched pattern. It consists of the features of I that are confirmed by V. Mismatched features are inhibited. The inactivated nodes corresponding to unconfirmed features of X are unhatched. The reduction in total STM activity which occurs when X is transformed into X^* causes a decrease in the total inhibition from F_1 to ρ. (C) If inhibition decreases sufficiently, ρ releases a nonspecific arousal wave to F_2; that is, a wave of activation that equally activates all F_2 nodes. This wave instantiates the intuition that "novel events are arousing". This arousal wave resets the STM pattern Y at F_2 by inhibiting Y. (D) After Y is inhibited, its top-down prototype signal is eliminated, and X can be reinstated at F_1. The prior reset event maintains inhibition of Y during the search cycle. As a result, X can activate a different STM pattern Y at F_2. If the top-down prototype due to this new Y pattern also mismatches I at F_1, then the search for an appropriate F_2 code continues until a more appropriate F_2 representation is selected. Such a search cycle represents a type of nonstationary hypothesis testing. When search ends, an attentive resonance develops and learning of the attended data is initiated. [Adapted with permission from Carpenter and Grossberg (1993).]

The individual features at F_1 have no meaning on their own, just like the pixels in a picture are meaningless one-by-one. The category, or symbol, in F_2 is sensitive to the global patterning of these features, but it cannot represent the "contents" of the experience, including their conscious qualia, due to the very fact that a category is a compressed, or "symbolic" representation. It has often been erroneously claimed that a single system is doomed to either process distributed features or symbolic representations, but not both. This is not true in ART. The resonance between these two types of information converts the *pattern* of attended features into a coherent context-sensitive state that is linked to its category through feedback. It is this coherent state, that joins together distributed features and symbolic categories, that can enter consciousness. ART predicts that *all conscious states are resonant states.* In particular, such a resonance binds spatially distributed features into either a stable equilibrium or a synchronous oscillation. Such synchronous oscillations have recently attracted much interest after being reported in neurophysiological experiments. This type of oscillation was predicted in the 1976 articles which introduced ART (see Grossberg, 1999b).

Resonance mediates between information processing and learning

In ART, the resonant state, rather than bottom-up activation, is predicted to drive the learning process. The resonant state persists long enough, and at a high enough activity level, to activate the slower learning processes in the adaptive weights that guide the flow of signals between bottom-up and top-down pathways between levels F_1 and F_2. This viewpoint helps to explain how adaptive weights that were changed through previous learning can regulate the brain's present information processing, without learning about the signals that they are currently processing unless they can initiate a resonant state. Through resonance as a mediating event, one can see from a deeper viewpoint why humans are intentional beings who are continually predicting what may next occur, and why we tend to learn about the events to which we pay attention.

How are learning and hypothesis testing related?

A sufficiently bad mismatch between an active top-down expectation and a bottom-up input, say because the input represents an unfamiliar type of experience, can drive a memory search. Such a mismatch within the attentional system is proposed to activate a complementary *orienting system*, which is sensitive to unexpected and unfamiliar events. ART suggests that this orienting system includes the hippocampal system, which has long been known to be involved in mismatch processing, including the processing of novel events (e.g., Otto and Eichenbaum, 1992). Output signals from the orienting system rapidly reset the recognition category that has been reading out the poorly matching top-down expectation (Figure 1B and 1C). The cause of the mismatch is hereby removed, thereby freeing the system to activate a different recognition category (Figure 1d). The reset event hereby triggers memory search, or hypothesis testing, which automatically leads to the selection of a recognition category that can better match the input.

If no such recognition category exists, say because the bottom-up input represents a truly novel experience, then the search process automatically activates an as yet uncommitted population of cells, with which to learn about the novel information. This learning process works well under both *unsupervised* and *supervised* conditions (e.g., Carpenter *et al.*, 1994). Unsupervised learning means that the system can learn how to categorize novel input patterns without any external feedback. Supervised learning uses predictive errors to let the system know whether it has categorized the information correctly. Supervision can force a search for new categories that may be culturally determined, and are not based on feature similarity alone. For example, separating the letters E and F into separate recognition categories is culturally determined; they are quite similar based on visual similarity alone. If the input pattern directly represented the pixels of E and F (which it, in general, would not), then both E and F might be classified in the same category with the category prototype of E, unless supervised feedback indicated that E is an incorrect response when F is the correct answer. Such error-based feedback enables variants of E and F to learn their own category and category prototype. Taken together, the interacting processes of attentive-learning and orienting-search hereby realize a type of error correction through hypothesis testing that can build an ever-growing, self-refining internal model of a changing world.

How is the generality of knowledge controlled? Exemplars and prototypes

A key problem about cognition concerns what combinations of features or other information are bound together into object or event representations. In particular, it is tempting to believe that exemplars, or individual experiences, are learned because humans can have very specific memories. For example, we can easily recognize the faces of each of our friends. On the other hand, storing every remembered experiences as exemplars can easily lead to a formidable combinatorial explosion of memory, as well as to difficult problems of memory retrieval. In addition, it is clear that we are also able to learn prototypes that can represent quite general properties of the environment (Posner and Keele, 1970). For example, we can recognize that everyone has a face. But then how do we learn specific episodic memories? ART provides a new answer to this question that overcomes problems faced by earlier models.

ART systems learn prototypes, but the generality of these prototypes can be controlled by a process of *vigilance* which can be influenced by environmental feedback or internal volition (Carpenter and Grossberg, 1991; Grossberg, 1999b). Low vigilance permits the learning of general categories with abstract prototypes. High vigilance forces a memory search to occur for a new category when even small mismatches exist between an exemplar and the category that it activates. As a result, in the limit of high vigilance, the category prototype may encode an individual exemplar. Vigilance is computed within the orienting system of an ART model (Figures 1B-D). It is here that bottom-up excitation from all the active features in an input pattern I are compared with inhibition from all the active features in a distributed feature representation across F_1. If the ratio of the total activity across the active features in F_1 (that is, the "matched" features) to the total activity due to all the features in I is less than a *vigilance parameter* ρ (Figure 1B), then a reset wave is activated (Figure 1C), which can drive the search for another category with which to classify the exemplar. In other words, the vigilance parameter controls how bad a match can be before search for a new category is initiated. If the vigilance parameter is low, then many exemplars can all influence the learning of a shared prototype, by chipping away at the features which are not in common to all the exemplars. If the vigilance parameter is high, then even a small difference between a new exemplar and a known prototype (e.g., F vs. E) can drive the search for a new category with which to represent F.

The simplest rule for controlling vigilance is called *match tracking*. Here a predictive error (e.g., E is predicted in response to F), the vigilance parameter increases until it is just higher than the ratio of active features in F_1 to total features in I. In other words, vigilance "tracks" the degree of match between input exemplar and matched prototype. This is the minimal level of vigilance that can trigger a reset wave and thus a memory search for a new category. It has been shown that match tracking realizes a Minimax Learning Rule that conjointly *maximizes* category generality while it *minimizes* predictive error. In other words, match tracking uses the least memory resources that can prevent errors from being made.

Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars, even a single exemplar. Thus a single ART system may be used, say, to learn abstract prototypes with which to recognize abstract categories of faces and dogs, as well as "exemplar prototypes" with which to recognize individual faces and dogs. ART models hereby try to learn the most general category that is consistent with the data. This tendency can, for example, lead to the type of overgeneralization that is seen in young children until further learning leads to category refinement. Many benchmark studies of how ART uses vigilance control to classify complex data bases have shown that the number of ART categories that is learned scales well with the complexity of the input data; see Carpenter and Grossberg (1994) for illustrative benchmark studies.

Memory consolidation and the emergence of rules

As sequences of inputs are practiced over learning trials, the search process eventually converges upon stable categories. It has been mathematically proved (Carpenter and Grossberg, 1987a) that familiar inputs directly access the category whose prototype provides the globally best match, while unfamiliar inputs engage the orienting subsystem to trigger memory searches for better categories until they become familiar. This process continues until the memory capacity, which can be chosen arbitrarily large, is fully utilized. The process whereby search is automatically disengaged is a form of *memory consolidation* that emerges from network interactions. Emergent consolidation does not preclude structural consolidation at individual cells, since the amplified and prolonged activities that subserve a resonance may be a trigger for learning-dependent cellular processes, such as protein synthesis and transmitter production. It has also been shown that the adaptive weights which are learned by some ART models can, at any stage of learning, be translated into IF-THEN rules (e.g., Carpenter and Grossberg, 1994). Thus the ART model is a self-organizing rule-discovering production system as well as a neural network. These examples show that the claims of some cognitive scientists and AI practioners that neural network models cannot learn rule-based behaviors are incorrect.

Corticohippocampal interactions and medial temporal amnesia

As noted above, the attentional subsystem of ART has been used to model aspects of inferotemporal (IT) cortex, and the orienting subsystem models part of the hippocampal system. The interpretation of ART dynamics in terms of IT cortex led Miller, Li, and Desimone (1991) to successfully test the prediction that cells in monkey IT cortex are reset after each trial in a working memory task. To illustrate the implications of an ART interpretation of IT-hippocampal interactions, I will review how a lesion of the ART model's orienting subsystem creates a formal memory disorder with symptoms much like the medial temporal amnesia that is caused in animals and human patients after hippocampal system lesions. In particular, such a lesion *in vivo* causes unlimited anterograde amnesia; limited retrograde amnesia; failure of consolidation; tendency to learn the first event in a series; abnormal reactions to novelty, including perseverative reactions; normal priming; and normal information processing of familiar events. Unlimited anterograde amnesia occurs because the network cannot carry out the memory search to learn a new recognition code. Limited retrograde amnesia occurs because familiar events can directly access correct recognition codes. Before events become familiar, memory consolidation occurs which utilizes the orienting subsystem (Figure 1C). This failure of consolidation does not necessarily prevent learning *per se*. Instead, it is predicted to learn coarser categories due to the failure of vigilance control and memory search. For the same reason, learning may differentially influence the first recognition category activated by bottom-up processing, much as amnesics are particularly strongly wedded to the first response they learn. Perseverative reactions can occur because the orienting subsystem cannot reset sensory representations or top-down expectations that may be persistently mismatched by bottom-up cues. The inability to search memory prevents ART from discovering more appropriate stimulus combinations to attend. Normal priming occurs because it is mediated by the attentional subsystem. Data which support these predictions are summarized in Grossberg and Merrill (1996), who also note that these are not the only problems that can be caused by such a lesion due to the predicted role of hippocampal structures in learned spatial navigation and adaptive timing functions.

Knowlton and Squire (1993) have reported that amnesics can classify items as members of a large category even if they are impaired on remembering the individual items themselves. To account for these results, the authors proposed that item and category memories are formed by distinct brain systems. Grossberg and Merrill (1996) suggested that their data could be explained by a single ART system in which the absence of vigilance control caused only coarse categories to form. Recently, Nosofsky and Zaki have quantitatively simulated the Knowlton and Squire data using a single-system model in which category sensitivity is low.

Cortical substrates of ART matching

How are ART top-down matching rules implemented in the cerebral cortex of the brain? An answer to this question has been recently proposed as part of a rapidly developing theory of why the cerebral cortex is typically organized into six distinct layers of cells (Grossberg, 1999a). Earlier mathematical work had predicted that such a matching rule would be realized by a *modulatory top-down on-center off-surround network* (e.g., Carpenter and Grossberg, 1991; Grossberg, 1999b). Figure 2 shows how such a matching circuit may be realized in the cortex. In Figure 2, the top-down circuit generates outputs from cortical layer 6 of V2 that activate layer 6 of V1 via the vertical pathway between these layers that ends in an open triangle (which designates an excitatory connection). Cells in layer 6 of V1, in turn, activate an "on-center off-surround" circuit to layer 4 of V1. In this circuit, an excitatory cell (open circle) in layer 6 excites the excitatory cell (open circle) immediately above it in layer 4 via the vertical pathway from layer 6 to 4 that ends in an open triangle. This excitatory interaction constitutes the "on-center". The same excitatory cell in layer 6 also excites nearby inhibitory cells (closed black circles) which, in turn, inhibit cells in layer 4. This spatially distributed inhibition constitutes the "off-surround" of the layer 6 cell. The on-center is predicted to have a modulatory, or sensitizing, effect on layer 4, due to the balancing of excitatory and inhibitory inputs to layer 4 within the on-center. The inhibitory signals in the off-surround can strongly suppress unattended visual features. This arrangement clarifies how top-down attention can sensitize the brain to get ready for expected information that may or may not actually occur, without actively firing the sensitized target cells and thereby inadvertently creating hallucinations that the information is already there. When this balance breaks down, hallucinations may, indeed, occur that have many of the properties reported by schizophrenic patients.

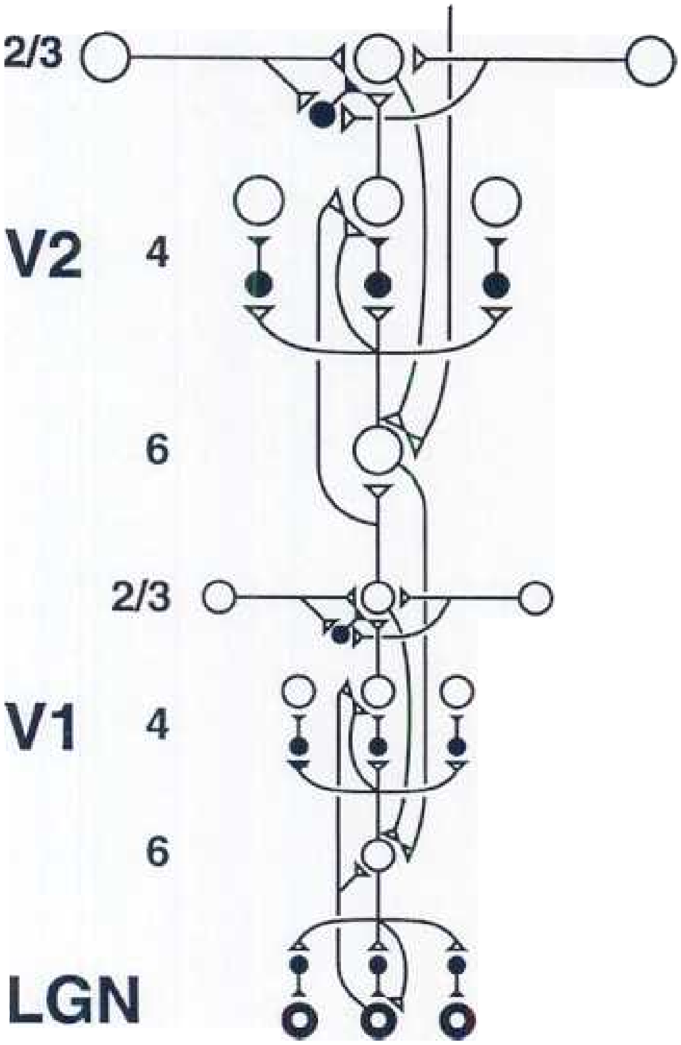

Figure 2. The LAMINART model: The model is a synthesis of feedforward (or bottom-up), feedback (or top-down), and horizontal interactions within and between the lateral geniculate nucleus (LGN) and visual cortical areas V1 and V2. Cells and connections with open symbols indicate excitatory interactions, and closed symbols indicate inhibitory interactions. The stippled top-down connections indicate attentional feedback. See Grossberg (1999a) and Grossberg and Raizada (2000) for further discussion of how these circuits work. [Adapted with permission from Grossberg and Raizada (2000).]

References

Carpenter, G.A. and Grossberg, S. (1991). **Pattern Recognition by Self-organizing Neural Networks. Cambridge**, MA: MIT Press.

Carpenter, G.A. and Grossberg, S. (1994). Integrating symbolic and neural processing in a self-organizing architecture for pattern recognition and prediction. In V. Honavar and L. Uhr (Eds.), **Artificial intelligence and neural networks: Steps towards principled prediction.** San Diego: Academic Press, 387-421.

Chelazzi, L., Duncan, J., Miller, E.K., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, **80**, 2918-2940.

Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision*, **12** 163-186.

Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, **8**, 1-44.

Grossberg, S. and Merrill, J.W.L. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience,* **8**, 257-277.

Knowlton, B.J. and Squire, L.R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, **262**, 1747-1749

Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, **16**, 5154-5167.

Posner, M.I. and Keele, S.W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology,* **83**, 304-308.

Otto, T. and Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: Evidence for hippocampal processing in recognition memory. *Hippocampus*, **2**, 323-334.

Bibliography

Clark, E.V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T.E. Morre (Ed.), **Cognitive development and the acquisition of language**. New York: Academic Press, 65-110.

Goodale, M.A. and Milner, D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, **15**, 20-25.

Grossberg, S. (2000). How hallucinations may arise from brain mechanisms of learning, attention, and volition. *Journal of the International Neuropsychological Society*, **6**, 583-592.

Lynch, G., McGaugh, J.L., and Weinberger, N.M. (Eds.) (1984). **Neurobiology of learning and memory**. New York: Guilford Press.

Miller, E.K., Li, L., and Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, **254**, 1377-1379.

Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, **6**, 414-417.

Nosofsky, R.M. and Zaki, S.R. (2000). Category learning and amnesia: An exemplar model perspective. **Proceedings of the 2000 Memory Disorders Research Society Annual Meeting**, Toronto, Canada.

Sokolov, E.N. (1968). **Mechanisms of memory**. Moscow University Press.

Squire, L.R. and Butters, N. (Eds.) (1984). **Neuropsychology of memory**. New York: Guilford Press.

Vinogradova, O.S. (1975). Functional organization of the limbic system in the process of registration of information: Facts and hypotheses. In R.L. Isaacson and K.H. Pribram (Eds.), **The hippocampus, Vol. 2.** Plenum Press, 3-69.