The Hippocampus and Cerebellum in Adaptively Timed Learning, Recognition, and Movement

Stephen Grossberg
Boston University

John W.L. Merrill
Microsoft Corporation

Abstract

The concepts of declarative memory and procedural memory have been used to distinguish two basic types of learning. A neural network model suggests how such memory processes work together as recognition learning, reinforcement learning, and sensorimotor learning take place during adaptive behaviors. To coordinate these processes, the hippocampal formation and cerebellum each contains circuits that learn to adaptively time their outputs. Within the model, hippocampal timing helps to maintain attention on motivationally salient goal objects during variable task-related delays, and cerebellar timing controls the release of conditioned responses. This property is part of the model's description of how cognitive-emotional interactions focus attention on motivationally valued cues, and how this process breaks down due to hippocampal ablation. The model suggests that the hippocampal mechanisms that help to rapidly draw attention to salient cues could prematurely release motor commands were not the release of these commands adaptively timed by the cerebellum. The model hippocampal system modulates cortical recognition learning without actually encoding the representational information that the cortex encodes. These properties avoid the difficulties faced by several models that propose a direct hippocampal role in recognition learning. Learning within the model hippocampal system controls adaptive timing and spatial orientation. Model properties hereby clarify how hippocampal ablations cause amnesic symptoms and difficulties with tasks which combine task delays, novelty detection, and attention toward goal objects amid distractions. When these model recognition, reinforcement, sensorimotor, and timing processes work together, they suggest how the brain can accomplish conditioning of multiple sensory events to delayed rewards, as during serial compound conditioning.

SOME NEURAL SUBSTRATES OF DECLARATIVE AND PROCEDURAL MEMORY

A central problem in cognitive neuroscience concerns how humans and other animals learn to recognize objects, to predict and attend to their rewarding or punishing consequences, and to perform appropriately timed actions capable of realizing or avoiding these consequences. Multiple brain regions participate in these processes, including inferotemporal cortex, amygdala, hippocampal formation, and cerebellum. The complexity of these processes has led to the development of neural models that might shed light on their cellular and network properties. A neural model is described herein to suggest why both the hippocampus and the cerebellum contain circuits that are specialized for adaptive timing. Although the two timing circuits may share cellular and circuit properties, the model predicts that they carry out distinct functional roles during the learning and memory processes that subserve recognition and movement tasks.

These distinct roles are used to clarify several of the conceptual dichotomies that have been useful in research about normal and amnesic learning and memory. One such dichotomy concerns the distinctions between declarative memory and procedural memory, knowing that and knowing how, memory and habit, or memory with record and memory without record (Bruner, 1969; Mishkin, 1982, 1993; Ryle, 1949; Squire & Cohen, 1984). The amnesic patient HM exemplified this distinction by learning and remembering motor skills like assembly of the Tower of Hanoi without being able to recall having done so (Bruner, 1969; Cohen & Squire, 1980; Mishkin, 1982; Ryle, 1949; Scoville & Milner, 1957; Squire & Cohen, 1984). HM's surgical lesion included extensive parts of the hippocampal formation and amygdala. Subsequent animal studies have shown that damage to the hippocampal formation (Ammon's horn, dentate gyrus, subiculum, fornix) and the parahippocampal region (entorhinal, perirhinal, and parahippocampal cortices) can reproduce analogous amnesic symptoms (Mishkin, 1978; Squire & Zola-Morgan, 1991). These results impli-
cate this aggregate hippocampal system in the processes that regulate declarative memory, or "knowing that." Such processes support a competence for learning recognition categories and being able to flexibly access them in a task-specific way (Eichenbaum, Otto, & Cohen, 1994).

A parallel line of research has implicated the cerebellum in the processing of procedural memory, or "knowing how." The cerebellum is an essential circuit for conditioning discrete adaptive responses during eye movements, arm movements, nictitating membrane movements, and jaw movements (Ebner & Bloedel, 1981; Gilbert & Thach, 1977; Ito, 1984; Lisberger, 1988; Optican & Robinson, 1980; Thompson, 1988; Thompson et al., 1984, 1987). Models of cerebellar learning have been developed over the years to help explain these motor conditioning data (Albus, 1971; Bullock, Fiala, & Grossberg, 1994; Fujita, 1982a, 1982b; Grossberg, 1969b, 1972b; Grossberg & Kuperstein, 1986; Ito, 1984; Lisberger, 1988; Marr, 1969).

A third line of research on learning and memory concerns cognitive-emotional interactions, including how a conditioned stimulus (CS) such as a tone or light, when paired with an unconditioned stimulus (US) such as a shock, can learn to generate conditioned responses (CR), such as fear or limb withdrawal, that were originally elicited only by the US. Such learning is optimal at a range of positive interstimulus intervals (ISI) that are characteristic of the animal and the task, and is greatly attenuated at zero ISI and long ISIs (Smith, 1968). Although the amygdala has been identified as a primary site in the expression of emotion and stimulus–reward association (Aggleton, 1993), the hippocampal formation has also been implicated in the processing of cognitive-emotional interactions. In particular, Thompson et al. (1987) distinguished two types of learning that go on during conditioning of the rabbit NMR: "conditioned fear" learning linked to the hippocampus and "learning of the discrete adaptive response" within the cerebellum (p. 82). In addition, removal of the hippocampal formation greatly attenuates attentional blocking (Rickert, Bennett, Lane, & French, 1978; Schmajuk, Spear, & Isaacson, 1983; Solomon, 1977). Blocking is the process whereby conditioning of a cue CS1 to a US prevents a second cue CS2 from being conditioned to US when it is later presented before US as part of a simultaneous CS1 + CS2 stimulus compound. Much experimental and theoretical work has suggested that CS2 loses its ability to be conditioned to US because it is an irrelevant cue that predicts no more about the US than does CS1 when presented alone (Grossberg, 1975, 1982; Kamin, 1969). Blocking enables a learning subject to attend selectively to relevant cues.

The present article synthesizes, into a single neural architecture, models that have been developed to explain data from each of these three areas. This synthesis clarifies how the various models work together to control behavior. In particular, it suggests why both the cerebellum and the hippocampal system may need adaptive timing circuits for their normal functioning. We suggest that the hippocampal mechanisms that help to rapidly draw attention to salient cues could prematurely release motor commands were these commands not adaptively timed by the cerebellum. To reach such conclusions as efficiently as possible, the article provides just enough information about the component models to understand how they can work together to explain key data. Mathematical equations and computer simulations of these models are described in detail in articles cited below.

Why should a single, albeit complex, brain region like the hippocampal system be involved in so many processes: recognition learning, reinforcement learning, and motivated attention? A clue is provided by neural data and models about how each of these processes work. In particular, both recognition learning and reinforcement learning are regulated by a matching process whereby bottom-up stimuli from the outside world are matched against top-down learned expectations to determine whether attentive learning or memory search will occur. The unblocking paradigm illustrates this matching process for the case of reinforcement learning (Kamin, 1969). The unblocking paradigm is a variant of the blocking paradigm in which the US changes intensity in the two learning episodes. Thus if CS1 is followed by one US intensity (US1), and the compound stimulus CS1 + CS2 is followed by a different US intensity (US2), then CS2 can become conditioned to the US, unlike in the blocking paradigm, and does so with an emotional valence that depends upon the sign of the difference US1 - US2 between US1 and US2 (Kamin, 1969). The mismatch between the actual intensity US2 and the expected intensity US1 triggers a memory search that attentionally "unblocks" the representation of CS2 that is stored in short-term memory, and enables it to learn to predict the change in US intensity (Grossberg, 1975). This memory search helps to focus attention upon that subset of sensory cues that predicts motivationally salient outcomes in a given context, and to block those that do not.

Recognition learning is accomplished by interactions between inferotemporal cortex (IT) and hippocampal formation, among other brain areas (Desimone, 1991; Desimone & Ungerleider, 1989; Eichenbaum, Otto, & Cohen, 1994; Gochin, Miller, Gross, & Gerstein, 1991; Harries & Perrett, 1991; Mishkin, 1978, 1982; Mishkin & Appenzeller, 1987; Perrett, Mistlin, & Chitty, 1987; Schwartz, Desimone, Albright, & Gross, 1983; Squire & Zola-Morgan, 1991). These interactions include the matching process that modulates the course of recognition learning in IT cortex and the course of reinforcement learning in thalamocorticoamygdala circuits. Some models are analyzed below of how these recognition
and reinforcement learning circuits interact with motor learning circuits. It is shown that the behavioral success of this interaction requires both types of circuits to be adaptively timed.

STABLE LEARNING THROUGHOUT LIFE USING ADAPTIVE RESONANCE

The first type of model results from an analysis of how humans and animals rapidly learn to categorize and name events and their contexts in real time. These adaptive resonance theory (ART) models have been used to help explain and predict a large body of cognitive and neural data about recognition learning, recall, attention, priming, and memory search (Carpenter & Grossberg, 1991, 1993; Grossberg, 1982b, 1987, 1988a). ART systems realize this synthesis by incorporating mechanisms that solve a fundamental problem about learning and memory that is called the stability-plasticity dilemma. An adequate self-organizing recognition system must be capable of plasticity in order to rapidly learn about significant new events, yet its memory must also remain stable in response to irrelevant or often repeated events. Thus we can learn to recognize many new faces without risking the unselective forgetting of our parents’ faces. In ART, interactions between an attentional subsystem and an orienting subsystem, or novelty detector, self-stabilize the learning process as the network becomes familiar with an environment by categorizing the information within it in a way that leads to behavioral success (Grossberg, 1980).

Learning takes place in the attentional subsystem. Its processes include activation of short-term memory (STM) traces, incorporation through learning of STM information into a longer-lasting long-term memory (LTM) traces, and interactions between pathways that carry specific information with nonspecific pathways that modulate the specific pathways. These interactions between specific STM and LTM processes and nonspecific modulatory processes regulate the stability-plasticity balance during normal learning, as follows.

The attentional subsystem undergoes both bottom-up learning and top-down learning between processing levels such as those denoted by $\mathcal{F}_1$ and $\mathcal{F}_2$ in Figure 1. Level $\mathcal{F}_1$ contains a network of nodes, or cell populations, each of which is activated by a particular combination of sensory features. Level $\mathcal{F}_2$ contains a network of nodes that represent recognition codes, or categories, which are selectively activated by the activation patterns across $\mathcal{F}_1$. Each $\mathcal{F}_1$ node sends output signals to a subset of $\mathcal{F}_2$ nodes. Each $\mathcal{F}_2$ node thus receives inputs from many $\mathcal{F}_1$ nodes. The thick pathway from $\mathcal{F}_1$ to $\mathcal{F}_2$ in Figure 1 represents the array of diverging and converging pathways. Learning takes place at the synapses denoted by semicircular endings in the $\mathcal{F}_1 \rightarrow \mathcal{F}_2$ pathways. Ways that end in arrowheads do not undergo learning. This bottom-up learning enables $\mathcal{F}_2$ nodes to become selectively tuned to particular combinations of activation patterns across $\mathcal{F}_1$ by changing their LTM traces.

Why is bottom-up learning insufficient in a system that can autonomously solve the stability-plasticity dilemma? This analysis was carried out in that part of the ART model that combines bottom-up associative learning and lateral inhibition for purposes of learned categorization. This type of model is often called a self-organizing feature map, competitive learning, or learned vector quantization. In such a model, as shown in Figure 2A, an input pattern registers itself as a pattern of activity, or STM, across the feature detectors of level $\mathcal{F}_1$. Each $\mathcal{F}_1$ output signal is multiplied or gated, by the adaptive weight, or LTM trace, in its respective pathway. All these LTM-gated inputs are added up at their target $\mathcal{F}_2$ nodes. Competitive interactions, mediated by lateral inhibition within $\mathcal{F}_2$, contrast-enhance this input pattern. Even if many $\mathcal{F}_2$ nodes may receive inputs from $\mathcal{F}_1$, lateral inhibition acts to cause a much smaller set of $\mathcal{F}_2$ nodes to store their activation in STM.

It is useful to think of all the STM signals that converge on an $\mathcal{F}_2$ node as an STM pattern, or vector. Likewise, all the LTM traces that multiply these signals on their way to a prescribed $\mathcal{F}_2$ node form an LTM vector. The operation of adding up the LTM-gated signals at each $\mathcal{F}_2$ node is called the inner product, or dot product, of the two vectors. It measures how similar the two vectors are, and increases as a function of their similarity. The LTM traces thereby filter the STM signal pattern and generate larger inputs to those $\mathcal{F}_2$ nodes whose LTM patterns are most similar to the STM pattern.

As noted above, the lateral inhibition among $\mathcal{F}_2$ nodes selects just a few of the more active $\mathcal{F}_2$ nodes for STM storage. This contrast-enhancing operation enables many input patterns at $\mathcal{F}_1$ that share similar input features to be classified by a small set of $\mathcal{F}_2$ nodes. The $\mathcal{F}_2$ nodes hereby become category nodes that are capable of classifying the inputs to $\mathcal{F}_1$.

In a self-organizing feature map, only the $\mathcal{F}_2$ nodes that win the contrast-enhancing competition and store their activity in STM can influence the learning process. STM activity at the winning $\mathcal{F}_2$ nodes selectively opens a learning gate at the LTM traces that abut those nodes. These LTM traces can then approach, or track, the input signals in their pathways, a process called steepest descent, or instar learning. In its simplest form, this learning law can be expressed by the equation

$$\frac{d}{dt} w_{ij} = f(x_{ij}) (w_{ij} + S_i)$$

where $\frac{d}{dt}w_{ij}$ is the time rate of change of the LTM trace, or adaptive weight, $w_{ij}$ from the $i$th $\mathcal{F}_1$ node to $\mathcal{F}_2$ node.
the $j$th $\mathcal{F}_2$ node, $f(x_j)$ is the learning gating signal that becomes positive only if the postsynaptic activity, or potential, $x_j$ of the $j$th $\mathcal{F}_2$ node becomes sufficiently large, and $S_i$ is the $i$th bottom-up signal. This learning rule was introduced into neural network models in Grossberg (1969a) and is the learning rule that was used to introduce ART (Grossberg, 1976b). While tracking the signals in its pathway, such an LTM trace $w_j$ can either increase (if the signal $S_i$ is large) or decrease (if the signal $S_i$ is small). It thus combines Hebbian and anti-Hebbian learning properties in a way that has been used to model neurophysiological data about hippocampal LTP and LTD (Artola & Singer, 1993; Levy, 1985; Levy & Desmond, 1985) and adaptive tuning of cortical feature detectors during the visual critical period (Rauschecker & Singer, 1979; Singer, 1983).

In particular, as Table 1 shows, significant postsynaptic activity, mediated by the gating signal $f(x_j)$, is needed to cause any change in $w_j$. If this modulatory gate opens, then $w_j$ may increase or decrease, depending upon the relative size of $S_i$. Since $S_i$, in turn, may influence the amount of postsynaptic activity $x_j$ via the presynaptic signal $S_i w_j$, various secondary effects can occur that are beyond the scope of this discussion (but see Carpenter & Grossberg, 1990). It is perhaps worth noting, however, that an early prediction (Grossberg, 1968b, 1969c, 1974) suggested that synaptic learning would be mediated by a postsynaptic process of protein synthesis and receptor sensitization that controls a coordinated presynaptic process of transmitter production. The postsynaptic signal process was predicted to be triggered by an inward $\text{Ca}^{2+}$ current that is antagonistic to $\text{Mg}^{2+}$. Coordinated presynaptic and postsynaptic changes were predicted to depend upon the inward $\text{Ca}^{2+}$ current in synergy with an inward $\text{Na}^+$ current and an outward $\text{K}^+$ current. Similar concepts have been used, in greatly elaborated form,
Figure 2. ART search for a recognition code: (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 subsys-

system $A$ (see Fig. 1). STM pattern $X$ is represented by the hatched pattern across $F_1$. Pattern $X$ both inhibits $A$ and generates the output pattern $S$. Pattern $S$ is multiplied by long-term memory (LTM) traces and added at $F_2$ 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 the prototype pattern $V$ that encodes the learned expectation of the active $F_2$ nodes. If $V$ mismatches I at $F_1$ then a new STM activity pattern $X'$ is generated at $F_1$. $X'$ is represented by the hatched pattern. It includes the features of I that are confirmed by $V$. Inactivated nodes corresponding to unconfirmed features of X are unhatched. The reduction in total STM activity that occurs when $X$ is transformed into $X'$ causes a decrease in the total inhibition from $F_1$ to $A$. (C) If inhibition decreases sufficiently, $A$ releases a nonspecific arousal wave to $F_2$, which resets the STM pattern $Y$ at $F_2$. (D) After $Y$ is inhibited, its top-down prototype signal is eliminated, and $X$ can be reinstated at $F_1$. Enduring traces of the prior reset lead $X$ to activate a different STM pattern $Y'$ at $F_2$. If the top-down prototype due to $Y'$ 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. Then an attentive resonance develops and learning of the attended data is initiated.

to explain recent data about LTP and LTD; e.g., see Artola and Singer (1993) and Kuno (1995). Gated steepest descent learning may thus be viewed as a first approximation to a much more complex cascade of biochemical events.

The net effect of such learning is to train the LTM vectors of the winning $F_2$ category nodes to become more similar to the STM patterns that they filter. As a result, the winning $F_2$ categories sharpen their tuning curves to respond more selectively to the STM patterns that they have experienced.

Self-organizing feature map models were introduced and computationally characterized in Malsburg (1973) and Grossberg (1976a, 1978). These models were subsequently applied and further developed by many authors, notably Kohonen (1984). They exhibit many useful properties, especially if not too many input patterns, or clusters of input patterns, perturb level $F_1$ relative to the number of categorizing nodes in level $F_2$. Grossberg (1976a) proved under these sparse environmental conditions that category learning is stable, with LTM traces that track the statistics of the environment, are self-normalizing, and oscillate a minimum number of times. Also, the $F_2$ category selection rule, like a Bayesian classifier, tends to minimize error.

It was also proved, however, that under more general environmental conditions, learning becomes unstable and subject to catastrophic forgetting. Such a model
Table 1. The instar learning, or gated steepest descent learning rule, embodies both Hebbian (LTP) and anti-Hebbian (LTD) properties within a single process.

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>State of S</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>State of x</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>State of w</td>
<td>↓</td>
<td>↓</td>
<td>↔</td>
<td>↔</td>
</tr>
</tbody>
</table>

+$+$ \quad \text{active;} \quad \uparrow \quad \text{increase;} \quad 
- $-$ \quad \text{inactive;} \quad \downarrow \quad \text{decrease;} \quad 
\leftrightarrow \quad \text{no change;} 

Table 2. & \text{S}_i & \text{w}_{ij} & \text{x}_j & \bullet & \text{+} & \text{+} & \text{+} & \text{-} & \text{-} 

could forget the faces of your parents while learning a new face. This memory instability is due to basic properties of associative learning and lateral inhibition. Although a gradual switching off of plasticity can partially overcome the problem, such a mechanism cannot work in a learning system whose plasticity is maintained throughout adulthood. These results put into sharp focus the problem of how the brain dynamically self-stabilizes its memory while remaining open to new experiences throughout life, a topic that has attracted increasing interest (Kandel & O’Dell, 1992). An analysis of this instability, together with data about categorization, conditioning, and attention, led to the introduction of ART models that self-stabilize the memory of self-organizing feature maps in response to an arbitrary stream of input patterns (Grossberg, 1976b).

In an ART model, learning does not occur when some winning $\mathcal{F}_2$ activities are stored in STM. Instead activation of $\mathcal{F}_2$ nodes may be interpreted as “making a hypothesis” about an input at $\mathcal{F}_1$. When $\mathcal{F}_2$ is activated, it quickly generates an output pattern that is transmitted along the top-down adaptive pathways from $\mathcal{F}_2$ to $\mathcal{F}_1$. These top-down signals are multiplied in their respective pathways by LTM traces at the semicircular synaptic knobs of Figure 2B. The LTM-gated signals from all the active $\mathcal{F}_2$ nodes are added to generate the total top-down feedback pattern from $\mathcal{F}_2$ to $\mathcal{F}_1$. This pattern plays the role of a learned expectation. Activation of this expectation may be interpreted as “testing the hypothesis,” or “reading out the prototype,” of the active $\mathcal{F}_2$ category. As shown in Figure 2B, ART networks are designed to match the “expected prototype” of the category against the bottom-up input pattern, or exemplar, to $\mathcal{F}_1$. Nodes that are activated by this exemplar are suppressed if they do not correspond to large LTM traces in the top-down prototype pattern. The resultant $\mathcal{F}_1$ pattern encodes the cluster of input features that the network deems relevant to the hypothesis based upon its past experience. This resultant activity pattern, called $\mathbf{x}'$ in Figure 2B, encodes the pattern of features to which the network “pays attention.”

If the expectation is close enough to the input exemplar, then a state of resonance develops as the attentional focus takes hold. The pattern $\mathbf{x}'$ of attended features reactivates the $\mathcal{F}_2$ category $\mathbf{y}$, which, in turn, reactivates $\mathbf{x}'$. The network locks into a resonant state through a positive feedback loop that dynamically links, or binds, $\mathbf{x}'$ with $\mathbf{y}$. Damasio (1989) has used the term “convergence zones” to describe such a resonant process. Such resonances are capable of binding spatially distributed features into synchronous and coherent states, both in corticocortical and thalamocortical feedback networks (Grossberg, 1976b; Grossberg & Somers, 1991).

Neurophysiological data that are consistent with the prediction that ART-like resonances exist between LGN and V1 have recently been reported (Sillito, Jones, Gerstein, & West, 1994). In particular, it was suggested in Grossberg (1980) that top-down corticogeniculate feedback would selectively amplify monocular LGN activations that are consistent with the oriented binocular cortical cells that activate the feedback, while inhibiting LGN cells that are not. In addition, top-down feedback by itself, as in all ART systems, was suggested not to be fully able to activate LGN cells. In support of this prediction, Sillito et al. (1994) reported that “cortically induced correlation of relay cell activity produces coherent firing in those groups of relay cells with receptive field alignments appropriate to signal the particular orientation of the moving contour to the cortex . . . this increases the gain of the input for feature-linked events detected by the cortex . . . the cortico-thalamic input is only strong enough to exert an effect on those LGN cells that are additionally polarized by their retinal input . . . the feedback circuit searches for correlations that support the ‘hypothesis’ represented by a particular pattern of cortical activity” (pp. 479-482). Gove, Grossberg, and Mingolla (1995) have shown how this type of corticogeniculate feedback and resonance can be used as part of a larger model of cortical visual processing to simulate data about brightness perception and illusory contours.

Similar ART matching and resonance rules have been used to explain and predict behavioral and brain data from other task domains. For example, Carpenter and Grossberg (1993) have used ART matching and resonance rules to explain data about visual object recognition and medial temporal amnesia (see below). Govindarajan, Grossberg, Wyse, and Cohen (1994) have used ART matching and resonance rules to simulate auditory psychophysical data about acoustic source segregation when multiple sources harmonically overlap, as during a cocktail party. Grossberg, Boardman, and Cohen (1994) have used ART matching and resonance rules to
simulate psychophysical data about variable-rate speech categorization. Grossberg and Stone (1986a) have used such rules to explain data about lexical priming and decision making. Roberts, Aguilar, Bullock, and Grossberg (1994) have used ART matching and resonance rules to explain neural data about multimodal control of saccadic eye movements. Why should similar matching and resonance rules be used in so many brain systems?

ART shows how these matching and resonance rules can be used to help solve the noise-saturation dilemma in any brain system that dynamically adjusts and maintains its parameters to cope with changing environmental conditions throughout life. The matched resonant state, rather than bottom-up activation, is predicted to drive the learning process. The resonant state persists long enough, at a high enough activity level, to activate the slower learning process; hence the term adaptive resonance theory. ART systems learn prototypes, rather than exemplars, because the attended feature vector $X^*$, rather than the input exemplar itself, is learned. Both the bottom-up LTM traces that tune the category nodes and the top-down LTM traces that filter the learned expectation learn to correlate activation of $F_2$ nodes with the set of all attended $X^*$ vectors that they have ever experienced. These attended STM vectors assign less STM activity to features in the input vector $I$ that mismatch the learned top-down prototype $V$ than to features that match $V$.

**PROTOTYPE LEARNING OR EXEMPLAR LEARNING?**

A similar type of matching by similarity across arrays of features has been used to quantitatively fit categorization data from human subjects (Estes, 1994). Models of this type assume that every input exemplar that a subject has ever experienced is stored, leading to formidable problems of memory storage and retrieval. Such models have not yet been shown capable of real-time autonomous categorization of complex databases. ART models computationally elaborate the idea that humans learn prototypes (Posner & Keele, 1968, 1970), which save greatly on memory resources by allowing many exemplars to be represented by a single category prototype. ART models have also been used for real-time autonomous categorization of complex databases (e.g., Asfour, Carpenter, & Grossberg, 1995; Asfour et al., 1993; Bachelder, Waxman, & Seibert, 1993; Baloch & Waxman, 1991; Bradski & Grossberg, 1995; Carpenter et al., 1992; Carpenter, Grossberg, & Reynolds, 1991, 1995; Carpenter & Ross, 1994, 1995; Carpenter & Tan, 1995; Cauiliss, Smith, Escobedo, & Anderson, 1994; Dubrawski & Crowley, 1994; Gjerdingen, 1990; Goodman et al., 1992; Ham & Han, 1993; Harvey, 1993; Kasperkiewicz, Racz, & Dubrawski, 1994; Keyvan, Durg, & Rabelo, 1993; Metha, Vij, & Rabelo, 1993; Moya, Koch, & Hostetler, 1993; Seibert & Waxman, 1992; Suzuki, Abe, & Ono, 1994; Suzuki, 1995; Wienke, Xie, & Hopke, 1994).

Given that ART systems learn prototypes, how can they also learn to recognize unique experiences, such as a particular view of a friend's face? The prototypes learned by ART systems accomplish this by realizing a qualitatively different concept of prototype than that offered by previous models. In particular, ART prototypes form in a way that is designed to conjointly maximize category generalization while minimizing predictive error (Carpenter et al., 1991, 1992). As a result, ART prototypes can automatically learn individual exemplars when environmental conditions require highly selective discriminations to be made. How the matching process achieves this is discussed below.

Before describing how this is achieved, let us note what happens if the mismatch between bottom-up and top-down information is too great for a resonance to develop. Then the $F_2$ category is quickly reset and a memory search for a better category is initiated. This combination of top-down matching, attention focusing, and memory search is what stabilizes ART learning and memory in an arbitrary input environment. The attentional focusing by top-down matching prevents inputs that represent irrelevant features at $F_1$ from eroding the memory of previously learned LTM prototypes. In addition, the memory search resets $F_2$ categories so quickly when their prototype $V$ mismatches the input vector $I$ that the more slowly varying LTM traces do not have an opportunity to correlate the attended $F_1$ activity vector $X^*$ with them. Conversely, the resonant event, when it does occur, maintains and amplifies the matched STM activities for long enough and at high enough amplitudes for learning to occur in the LTM traces.

Whether or not a resonance occurs depends upon the level of mismatch, or novelty, that the network is prepared to tolerate. Novelty is measured by how well a given exemplar matches the prototype that its presentation evokes. The criterion of an acceptable match is defined by an internally controlled parameter $p$ called vigilance (Carpenter & Grossberg, 1987a, 1991). The vigilance parameter is computed in the orienting subsystem $A$ (see Fig. 1). Vigilance weighs how similar an input exemplar $I$ must be to a top-down prototype $V$ in order for resonance to occur. Resonance occurs if $p \|I\| - \|X^*\| \leq 0$. This inequality says that the $F_1$ attentional focus $X^*$ inhibits $A$ more than the input $I$ excites it. If $A$ remains quiet, then an $F_1 \leftrightarrow F_2$ resonance can develop.

Either a larger value of $p$ or a smaller match ratio $\|X^*\|/\|I\|$ makes it harder to satisfy the resonance inequality. When $p$ grows large or $|X^*/I|$ is so small that $p\|I\| - |X^*/I| > 0$, then $A$ generates an arousal burst, or novelty wave, that resets the STM pattern across $F_2$ and initiates a bout of hypothesis testing, or memory search. During search, the orienting subsystem interacts with the attentional subsystem (Fig. 2C and D) to rapidly reset mismatched categories and to select better $F_2$.
representations with which to categorize novel events at $F_1$, without risking unselective forgetting of previous knowledge. Search may select a familiar category if its prototype is similar enough to the input to satisfy the resonance criterion. The prototype may then be refined by attentional focusing. If the input is too different from any previously learned prototype, then an uncommitted population of $F_2$ cells is selected and learning of a new category is initiated.

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. A single system can learn both, as the need arises, by increasing vigilance just enough to activate $A$ if a previous categorization leads to a predictive error (Carpenter et al., 1991, 1992).

CORTICO-hippocampal interactions and medial temporal amnesia

As sequences of inputs are practiced over learning trials, the search process eventually converges upon stable categories. It has been mathematically proved 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 (Carpenter and Grossberg, 1987a, 1991). 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.

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, Carpenter and Grossberg (1993) described 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 patient HM after hippocampal system lesions. In particular, such a lesion in vitro 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 (Cohen, 1984; Graf, Squire, & Mandler, 1984; Lynch, McGaugh, & Weinberger, 1984; Squire & Butters, 1984; Squire & Cohen, 1984; Warrington & Weiskrantz, 1974; Zola-Morgan & Squire, 1990).

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 that utilizes the orienting subsystem (Fig. 2C). This failure of consolidation does not necessarily prevent learning per se. Instead, learning influences the first recognition category activated by bottom-up processing, much as "amnesics are particularly strongly wedded to the first response they learn" (Gray, 1982, p. 253). 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.

Similar behavioral problems have been identified in hippocampectomized monkeys. Gaffan (1985) noted that fornix transection "impairs ability to change an established habit . . . in a different set of circumstances that is similar to the first and therefore liable to be confused with it" (p. 94). In ART, a defective orienting subsystem prevents the memory search whereby different representations could be learned for similar events. Pribram (1986) called such a process a "competence for recombinant context-sensitive processing" (p. 362). These ART mechanisms illustrate how memory consolidation and novelty detection may be mediated by the same neural structures (Zola-Morgan & Squire, 1990), why hippocampectomized rats have difficulty orienting to novel cues (O'Keefe & Nadel, 1978), and why there is a progressive reduction in novelty-related hippocampal potentials as learning proceeds in normal rats (Deadwyler, West, & Lynch, 1979; Deadwyler, West, & Robinson, 1981). In ART, the orienting system is automatically disengaged as events become familiar during the memory consolidation process.

A COMPARISON OF HIPPOCAMPAL LEARNING MODELS

This review of ART properties enables us to comment on recent data and models about the hippocampal sys-
tem. In particular, the novelty-sensitive matching and memory search properties that are modulated by the ART orienting subsystem suggest how the hippocampal system may contribute to flexible expression of memories in novel contexts and why hippocampal neurons respond differently to match and nonmatch conditions (Otto & Eichenbaum, 1992). Indeed, mismatches within the attentional system trigger memory searches for better recognition categories by activating the orienting subsystem.

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 parallel brain systems. This hypothesis does not, however, explain what these systems are, how they interact, or how some large categories may form even though item memories, that may be viewed as "specific" or "concrete" categories, do not. These authors also noted that "the possibility must be considered that classification learning is dependent on declarative knowledge...amnesic patients did perform numerically worse than the control subjects" (Knowlton & Squire, 1993, p. 1748). Within an ART model, coarse categories tend to form when the orienting subsystem is inoperative because there is no vigilance control or memory search. Thus the coarse categories and a tendency to perseveration go together in this case. Carpenter and Grossberg (1987b; see also Carpenter & Grossberg, 1991) provided simulation examples of coarse category learning with zero vigilance in which each category can be activated by multiple exemplars. Finer item-specific categories that match their structure to environmental demands can form when the orienting subsystem is active. These model properties enable the amnesic data pattern to be rationalized without requiring that item and category memories be coded by parallel brain systems. ART does not, however, deny that categories for individual events and for sequences of events may form at distinct levels of a single hierarchically organized memory system (Grossberg, 1978, 1987), rather than in a pair of parallel memory systems.

ART properties also provide an alternative to the popular hypothesis that the hippocampal formation somehow temporarily stores recognition codes from all sensory modalities before the temporal cortex can more permanently do so (Eichenbaum et al., 1994; Marr, 1971; McClelland, McNaughton, & O'Reilly, 1994; Milner, 1989). This hypothesis faces formidable obstacles as soon as one seriously tries to model how such a process could work. For example, how could the hippocampal system rapidly store all the information that one can recall after seeing an exciting movie? McClelland, McNaughton, and O'Reilly (1994) admit that their model cannot do this. In fact, not only is fast learning impossible, but also "the sequential acquisition of new data...can lead to catastrophic interferences with what has previously been learned." Only if learning is slow and carefully interleaved on sufficiently small and regular databases can it occur at all in this type of model. Such a model fails to solve the stability-plasticity dilemma.

A more general concern is that these models do not consider the nature of the representations that are learned in any realistic behavioral experience. For example, how could the known anatomy of the hippocampal formation rapidly learn to represent all the types of sensory information—visual, auditory, touch, etc.—that specialized thalamocortical systems have evolved to represent? How could it then selectively transfer this information back into the respective cortical systems? This is a problem about the nature of the proposed hippocampal representation and about the capacity of the hippocampus to rapidly store vast amounts of information. These models seem to seriously underestimate the complexity and subtlety of these issues by discussing small-scale toy problems that do not attempt to represent any nontrivial sensory information in a real-time learning environment.

To clarify some of the representational difficulties in broad strokes, let us consider several illustrative options: (1) the mapping between cortex-to-hippocampal system is one-to-one; (2) the mapping from cortex to hippocampal system is many-to-one and the reverse mapping is one-to-many; (3) the temporary hippocampal storage is in some form of STM; (4) the temporary hippocampal storage is in some form of LTM. Options (1) and (2) may be combined with options (3) and (4) into four cases: (1,3), (1,4), (2,3), and (2,4).

Consider cases (1,3) and (1,4). These seem impossible because the hippocampal system does not have nearly enough cells to represent in a one-to-one fashion all of the cellular activation patterns over all of the sensory cortices. Consider case (2,3). This implies that the sensory cortices transform sensory inputs into activations of sensory feature detectors of various sorts, after which these cortical STM activation patterns are compressed by many-to-one pathways into STM activations of multimodal hippocampal categories. Apart from the general problem that STM does not have nearly a large enough memory capacity to store the amounts of information in question (Atkinson & Shiffrin, 1971; Miller, 1956), the reverse one-to-many mapping from hippocampal system to cortex cannot recover the full dimension of the original cortical STM patterns without the intervention of some sort of LTM that organizes the signal traffic between the hippocampal system and cortex.

This leaves case (2,4) in which, after the cortex filters incoming sensory information, multimodal hippocampal categories are learned by reciprocal interactions between cortex and the hippocampal system. Here the pathways between cortex and hippocampus would rapidly store LTM traces to organize the reciprocal signal traffic. The hippocampal system would later read out this...
information so that the slower cortical learning could somehow catch up. Such a system experiences the full burden of the stability-plasticity dilemma, which these models were not designed to do. Fast learning of such a system in response to a rich and varying input environment can cause catastrophic forgetting (Carpenter & Grossberg, 1987a; Grossberg, 1976a, 1988b). It forgets what it has learned even as it is trying to learn more. However, if the hippocampal system cannot learn the data quickly, then it cannot impart this knowledge to the more slowly learning cortical system. Could this problem be avoided by incorporating ART dynamics into such a model?

This does not seem possible, because another problem faces such a hypothesis; namely, that all the system's learning is trapped in LTM traces that lie between the cortex and the hippocampal system. This learned information could not be directly transferred to other cortical learning systems by any local operation. For direct transfer of this LTM to happen, nonlocal transport of LTM traces would be needed from one corticohippocampal pathway to a spatially disjoint thalamocortical or corticocortical pathway. Such an operation is physically inconceivable. An LTM trace is not a number in a register, to be simply copied from one place to another. It is a complex metabolic interaction between parts of cells that captures the degree of interaction between those neurons. That degree of interaction depends upon the internal states of the neurons involved, as well as upon a variety of other factors. To transfer that learned information to some other part of cortex, it would need to be "read-out" back into the original cortical memory store. Thus the hippocampally stored LTM information could only be read back to cortex indirectly by using the LTM traces to reactivate the original cortical STM patterns. If, however, these reactivated cortical STM patterns can then incite corticocortical learning, why could they not do so originally?

The only plausible alternative remaining is that the hippocampal system reads these STM patterns into the cortex many times so that the slow cortex can gradually learn them over many learning trials. This is the type of assumption that McClelland, McNaughton, and O'Reilly (1994) make. This hypothesis implies the existence of a highly sophisticated hippocampal controller that could reproduce the experiences of a whole day many times across the entire cortex without interfering with the processing of other experiences. This cannot, by the nature of cortical representation, happen during waking hours without interfering with the STM registration of ongoing experiences. Moreover, there is simply not enough time during sleep to reproduce multiple recollections of a previous day's experiences unless one is asleep much longer than one is awake. Nor do sleep EEGs reproduce waking cortical patterns most of the time.

These problems are avoided in ART systems. The ART orienting subsystem is eventually disengaged as novel inputs become incorporated through practice into the context of other learned knowledge and eventually become familiar to the attentional subsystem. Such a model consolidation process is consistent with the temporary nature of hippocampal engagement during learning and the temporally graded nature of retrograde amnesia. However, no recognition codes are ever stored within an ART orienting subsystem. Rather, interactions between the orienting and attentional subsystems enable the latter to stably learn new recognition categories whose structure is sensitive to environmental relationships and the global organization of previously learned knowledge. Thus ART models suggest how learning within thalamocortical and corticocortical systems may be modulated by hippocampal interactions without requiring that the hippocampal system actually store the learned representations.

A PREDICTION ABOUT PROTOTYPE LEARNING

The ART conception of temporal-hippocampal interactions suggests the following prediction. Level $F_2$ properties may be compared with properties of cell activations in inferotemporal cortex (IT) during recognition learning in monkeys. The ability of $F_2$ nodes to learn categories with different levels of generalization clarifies how some IT cells can exhibit high specificity, such as selectivity to views of particular faces, while other cells respond to broader features of the animal's environment (Desimone & Ungerleider, 1989; Gochin et al., 1991; Harries and Perrett, 1991; Mishkin, 1982; Mishkin & Appenzeller, 1987; Perrett et al., 1987; Schwartz et al., 1983; Seibert & Waxman, 1991). Moreover, when monkeys are exposed to easy and difficult discriminations (Spitzer, Desimone, & Moran, 1988), "in the difficult condition the animals adopted a stricter internal criterion for discriminating matching from nonmatching stimuli . . . the animals' internal representations of the stimuli were better separated, independent of the criterion used to discriminate them . . . increased effort appears to cause enhancement of the responses and sharpened selectivity for attended stimuli" (pp. 339-340). These are also properties of model cells in $F_2$ due to the role of vigilance control. ART prototypes represent smaller sets of exemplars at higher vigilance levels, so a stricter matching criterion is learned. These exemplars match their finer prototypes better than do exemplars that match a coarser prototype. This better match more strongly activates the corresponding $F_2$ nodes.

This property suggests that operations that make the novelty-related potentials of the hippocampus more sensitive to input changes may trigger the formation of more selective inferotemporal recognition categories. Can such a correlation between IT discrimination and hippocampal potentials be recorded, say, when monkeys
learn easy and difficult discriminations? Conversely, operations that progressively block the expression of hippocampal novelty potentials are suggested to cause learning of coarser recognition categories, with amnesic symptoms as a limiting case.

The conclusion that no learning occurs in the ART orienting system does not force the theory to deny that some types of learning do occur in the hippocampal system. The model suggests that these learning processes are involved in adaptively timed modulation of reinforcement learning and aspects of spatial orientation, as discussed below.

A FRAMEWORK FOR TEMPORAL LEARNING

Before turning to this discussion, it is appropriate to comment upon how an ART-based system could rapidly learn the information in a movie. There are many levels on which such a problem could be approached, and it seems fair to say that no available theory proposes a complete explanation of this competence. On the other hand, the critique of alternative models has been made on the level of their inability to rapidly and stably learn large amounts of information, notably temporally ordered information. This is not a problem in an ART-based system.

A framework for accomplishing this was described in Grossberg (1978) using a combination of ART category learning, working memories, temporal associative learning networks, and predictive feedback within the system. A great deal of work has since been done to further carry out this program. For example, ART-based architectures, called VIEWNET systems, are capable of rapidly and stably learning to recognize three-dimensional (3-D) objects by categorizing their two-dimensional (2-D) views and learning to associate their 2-D view categories with 3-D object categories that are invariant under changes of familiar 2-D view (Bradski & Grossberg, 1994, 1995). Properties of these 2-D view and 3-D object category nodes may be compared with neural responses from distinct cell populations in monkey inferotemporal cortex (Logothetis et al., 1994).

The 3-D object categories may, in turn, be stored in a working memory (Baddeley, 1986) that can encode both object representations and their temporal order in STM. This type of working memory is designed so that its contents may rapidly and stably be learned and categorized by another ART network, whose active nodes are said to code list categories. This list categorization process has been proved to retain its stability even as new information continues to be stored in the working memory through time (Bradski, Carpenter, & Grossberg, 1992, 1994; Cohen & Grossberg, 1986, 1987; Grossberg, 1978; Grossberg & Stone, 1986a). Interactions between such a working memory and its list categories have been used to explain data from experiments about the sequential performance of stored motor commands (Boardman & Bullock, 1991; Grossberg & Kuperstein, 1989), about errors in serial item and order recall due to rapid visual attention shifts (Grossberg & Stone, 1986a), about errors and reaction times during lexical priming and episodic memory experiments (Grossberg & Stone, 1986b), and about data concerning word superiority, phonemic restoration, and backward effects on speech perception (Cohen & Grossberg, 1986; Grossberg, 1986). Such a working memory design thus seems to be used in several modalities. This is plausible when one realizes that the design embodies a few simple principles that enable its temporally evolving STM patterns to be stably categorized in LTM.

Temporal cortex provides a likely neural substrate for such a working memory (Goldman-Rakic, 1994). Here, information from multiple sensory modalities converges and may interact with subcortical reward mechanisms to sustain an attentional focus upon salient goals (Gaffan, 1994; Knight, 1994). Can ART systems learn multimodal list categories and focus attention on predictively successful ones?

Multimodal information distributed across a working memory may indeed be integrated into ART categories (Asfour, 1994; Asfour et al., 1993). Such an ART system, called fusion ARTMAP, is designed to solve the credit assignment problem of selectively resetting those input channels that are causing predictive errors. In addition, ART models of cognitive-emotional interactions have been described to suggest how attention may be selectively allocated to event categories that have high salience due to prior reinforcement and how less salient events may be attentionally blocked (Grossberg, 1975, 1982a, 1984; Grossberg & Levine, 1987; Grossberg & Merrill, 1992); also see below. They have also been used to explain and predict cognitive data about human decision making under risk as a manifestation of cognitive-emotional neural mechanisms (Grossberg & Gutowski, 1987), and to shed some light upon how these cognitive-emotional interactions may break down during mental depression (Grossberg, 1972a, 1984).

The motivationally modulated list categories may, in turn, be recurrently linked together by an associative learning network that helps to predict the categories most likely to occur in a given temporal context. Such networks have been used to model the position-dependent error gradients and learning rates that are observed during human verbal learning and to predict how this process breaks down in schizophrenic subjects (Grossberg, 1969d, 1982b; Grossberg & Pepe, 1970, 1971). Finally, the attended list categories may be used to predict the next images that are expected by the system, a one-to-many process called outstar learning (Grossberg, 1968a, 1978, 1980). One possible anatomical substrate of this type of predictive learning is frontaltemporal projections (Gaffan, 1994).

Taken together, these architectural elements may be called a resonant avalanche. This name acknowledges
the role of resonance in stabilizing the learning process, and of the avalanche of temporal associations in predicting the events that the system next expects to experience. (For a summary of avalanches at different levels of complexity, see Grossberg, 1978.) Although the theory of resonant avalanches has not yet been completely developed, there are enough mathematical, computational, and data simulation results available to conclude that ART systems escape the critique of other models that was proposed above.

ADAPTIVELY TIMED COGNITIVE–EMOTIONAL AND SENSORIMOTOR INTERACTIONS

Let us now return to the question of what sorts of learning are predicted to occur in the hippocampal system by an ART-based model. As in our remarks about frontotemporal interactions, this discussion will include an analysis of issues concerning reinforcement and temporal processing. The model frontotemporal interactions that were reviewed above concern a type of macro-timing that integrates information across a series of events. The model frontotemporal–hippocampal interactions now to be discussed consider a type of micro-timing that calibrates how long motivated attention may be allocated to a single predicted event.

Some authors (e.g., Eichenbaum et al., 1994) have dichotomized the representational properties of hippocampal memory processing—namely, those relating to recognition learning and memory—as being "orthogonal functional properties" from hippocampal temporal processing properties. It is unclear why a single brain structure should combine properties if they are indeed "orthogonal." The adaptive timing model described below suggests how these representational and temporal processes may be linked. The timing model is part of a larger model system that controls how cognitive–emotional and sensorimotor interactions are coordinated, including how classical and instrumental conditioning are adaptively timed and modulated by cognitive recognition processes (Baloch & Waxman, 1991; Grossberg, 1971, 1972a, 1975, 1982a, 1987; Grossberg & Levine, 1987; Grossberg & Merrill, 1992; Grossberg & Schmajuk, 1987).

This cognitive–emotional model suggests that (at least) three types of internal representation interact during conditioning: sensory representations S, drive representations D, and motor representations M (Fig. 3). The S representations are categorical thalamocortical representations of external events, including the object recognition categories that are learned by IT cortex and linked to frontal cortex via frontotemporal interactions. The D representations are categorical thalamocortical representations of external events, including the object recognition categories that are learned by IT cortex and linked to frontal cortex via frontotemporal interactions. The internal motivational signals include hypothalamic and amygdala cir-
circuits, at which homeostatic and reinforcing cues converge to generate emotional reactions and motivational decisions. The M representations include cerebellar circuits that control discrete adaptive responses. Three types of learning take place among these representations: S → D conditioned reinforcer learning that converts a CS into a reinforcer by pairing activation of its sensory representation S with activation of the drive representation D that receives input from a salient US or other conditioned reinforcer CS; D → S incentive motivational learning whereby an activated drive representation D may learn to prime the sensory representations S of all cues, including CSs, that have consistently been activated when it has; and S → M habit, or motor, learning whereby the sensorimotor maps, vectors, and gains that are involved in motor control may be adaptively calibrated.

These processes contribute to the modulation of declarative memory by motivational feedback and to the learning and performance of procedural memory. Thus learned S → D → S positive feedback quickly draws attention to motivationally salient cues and blocks activation of less salient cues via lateral inhibition among the S categories. D → S motivational feedback also energizes the release of discrete adaptive S → M responses. Based on a theoretical analysis, the final common path of the drive representations D, at or after the stage at which motivational decisions are made, was predicted to intersect or be modulated by the hippocampal formation (Grossberg, 1975, 1982a). In support of this prediction, Thompson et al. (1984, 1987) have shown that emotional conditioning (as in the S → D circuit) influences hippocampal sites, whereas motor conditioning (as in the S → M circuit) occurs within the cerebellum. In addition, hippocampal ablation attenuates blocking (Rickert et al., 1978; Schmajuk et al., 1983; Solomon, 1977). Blocking fails in the model when D → S feedback is impaired, as follows. In the complete model, when the S population activities that categorize conditioned reinforcers are amplified by strong conditioned S → D → S attentional feedback, they can block activation of other S populations via S → S lateral inhibition. When D → S feedback is removed, amplification and its blocking effect are eliminated. See Grossberg and Levine (1987) for blocking simulations. These model properties clarify how damage to the hippocampal system that involves both its drive-modulatory and orienting functions can result in either impaired or abnormally strong utilization of contextual cues, and a failure of flexible reset and memory search for appropriate cues to attend.

Why should a single brain region, like the hippocampal system, modulate both recognition learning and reinforcement learning? We suggest that this is so in part because the same adaptive timing and orienting processes modulate both types of learning (Grossberg & Merrill, 1992; Grossberg & Schmajuk, 1989). This linkage clarifies how the hippocampal system may mediate tasks like delayed nonmatch to sample (DNMS) wherein both temporal delays and novelty-sensitive recognition processes are involved (Gaffan, 1974; Mishkin & Delacour, 1975). The proposed adaptive timing and orienting properties of the hippocampal system are envisaged to cooperate in the following way. As shown in Figures 3 and 4, S → D → S feedback can rapidly focus attention on motivationally salient cues, as inhibition from D to the orienting subsystem inhibits orienting reactions that would otherwise occur in response to irrelevant situational cues. The inhibition from D to the orienting subsystem helps to model competition between consummatory and orienting behaviors (Staddon, 1983).

Another process is, however, needed to prevent the premature reset of attention by potentially distracting irrelevant cues during variable task-specific delays. For example, suppose that an animal inspects a food box right after a signal occurs that has regularly predicted food delivery in 6 sec. Why is not the mismatch between the learned expectation of food and the percept of no-food treated like a predictive failure? Why, as often occurs when a previously rewarded cue is no longer rewarded, does the mismatch not trigger reset of attention, frustration, forgetting, and exploratory behavior? Were this to happen, humans and animals would restlessly explore their environments without being able to wait for delayed rewards.

**SPECTRAL TIMING IN THE HIPPOCAMPUS AND DEFICITS DUE TO ITS REMOVAL**

We suggest that a "spectral timing" circuit S → T operates in parallel with the fast S → D → S emotional conditioning circuit (Fig. 4) to maintain attention on salient cues during variable task-specific delays. Different populations of cells in T can be conditioned to respond selectively to different ISI intervals. The total population output sums the output from all cells in the spectrum. Remarkably, this population response accurately models the ISI, even though no single cell does (Fig. 5). Learned S → T timing maintains inhibition of the orienting subsystem and, in the example noted above, enables attention to be maintained on motivationally salient goal-related cues within the 6 sec delay. If food does not occur even after 6 or more seconds have elapsed, then the adaptive timing circuit becomes quiet, and subsequent ART mismatches can trigger attentional reset, frustration, forgetting, and exploration in a manner modeled in Grossberg (1987).

We predicted in Grossberg and Merrill (1992) that this spectral timing circuit T exists in the hippocampal dentate-CA3 region in order to explain neurophysiological data showing that hippocampal CA3 pyramidal cell firing often reflects the temporal delays observed in the conditioned nictitating membrane response (Berger, Berry, & Grossberg and Merrill, 269)
Figure 4. A spectrally timed conditioning model with feedback pathways $D \rightarrow S^{(2)} \rightarrow S^{(1)}$ that are capable of focusing attention in an adaptively timed fashion on reinforcing events. The sensory representations $S$ of Figure 3 are here broken into two successive levels $S^{(1)}$ and $S^{(2)}$. Levels $S^{(1)}$ and $S^{(2)}$ interact via reciprocal excitatory pathways. The excitatory pathways $S^{(1)} \rightarrow D$ and $D \rightarrow S^{(2)}$ are, as in Figure 3, adaptive. Representations in $S^{(2)}$ can, however, fire only if they receive convergent signals from $S^{(1)}$ and $D$. Then they deliver positive feedback to $S^{(1)}$ and bias the competition to focus attention on their respective features and to attentionally block inhibited features. Prior to conditioning, a CS can be stored in STM only at $S^{(1)}$ and can subliminally prime $S^{(2)}$ and D representations without supraliminally firing these representations. After conditioning, the CS can trigger strong conditioned $S^{(1)} \rightarrow D \rightarrow S^{(2)} \rightarrow S^{(1)}$ feedback and rapidly draw attention to itself as it activates the emotional representations and motivational pathways controlled by D. Representation D can also inhibit the orienting subsystem as it focuses attention upon motivationally valued sensory events. The sensory representations $S^{(1)}$ send parallel pathways to a spectral timing circuit $T$ whose adaptive weights $z$ sample the Now Print, or teaching signal, $N$ that is transiently activated by changes in the activity of the drive representation $D$. After conditioning of $T$ takes place, adaptively timed output from $T$ can maintain attention on task-relevant cues for a learned duration via the $T \rightarrow D \rightarrow S$ feedback pathway. Timed signals also inhibit the orienting subsystem via the $T \rightarrow D \rightarrow A$ pathway and thereby help to prevent distracting events from interfering with planned consummatory acts. [Reprinted with permission from Grossberg and Merrill (1992).]

Figure 5. A computer simulation of spectral timing: (a) In response to a CS input $I$ in Figure 4, a spectrum of population activities $x_{ij}$ react at different rates and generate signals $f_I = f_I(x_{ij})$; (b) each signal causes a transmitter $y_{ij}$ in its pathway to become inactivated, or habituate, at a different rate; (c) the transmitters $y_{ij}$ multiply, or gate, the signals $f_I$ to generate net signals $g_{ij}$ and the sum, expressed via the teaching signal $N$, jointly activate adaptive weights, or LTM traces, $z_{ij}$, which generate adaptively gated output signals $h_{ij} = g_{ij} y_{ij}$. (d) Although individual signals $h_{ij}$ do not time the ISI well, the population sum $R = \sum_i h_{ij}$ of the adaptive signals does accurately time the ISI (dotted vertical lines). [Reprinted with permission from Grossberg and Merrill (1992).] and humans always have problems with DNMS and related tasks that involve stimulus delays? In the model, when the timing circuit $T$ is removed, attention may more easily be distracted from goal objects during task-related delays. On the other hand, if the orienting subsystem is also removed, then flexible reset of attention in response to novel events is impaired, thereby eliminating a key mechanism whereby a distracting event could undermine performance. If the attentional system remains intact, then direct activations of individual recognition codes in response to a familiar event are still possible, and the matching process per se can partially update short-term memory. However, the network can
no longer flexibly search for the proper configuration of targets to attend, especially in the presence of complex spatial layouts that include distracting cues. The lack of timed control over variable delays can thus harm behavior more when it is necessary to shift attention among different sets of cues. Gaffan (1992) has described analogous data from hippocampectomized monkeys.

Both DNMS performance at brief delays and single-pair object discrimination learning with brief intertrial intervals are spared in hippocampal subjects (Eichenbaum et al., 1994). In the model, this is also true because the fast $S \rightarrow D \rightarrow S$ attentional circuit remains intact. Long interstimulus delays, say of a day, also spare the performance of animals in some conditions (Mishkin, Malamut, & Bachevalier, 1984). These results have led some investigators to claim that the hippocampal system subserves a memory store of *intermediate* duration (Eichenbaum et al., 1994). As noted above, how the hippocampal system could create such a representation before it is transferred to the appropriate neocortical representations across several modalities has never been explained, and faces serious conceptual difficulties.

The ART model does not need to posit any such hippocampal memory store. At short delays, the fast feedback $S \rightarrow D \rightarrow S$ system helps to focus attention on motivationally salient objects and to initiate attentional blocking. The failure of blocking at intermediate delays due to removal of the $S \rightarrow T$ circuit leads to abnormally strong utilization of contextual cues. This processing failure causes little problem at long delays because potentially disruptive cues, being so widely separated in time, decay before they can compete for attention. These properties can be inferred from the model simulations of blocking by Grossberg and Levine (1987). It has not, to our knowledge, yet been tested whether the spectral timing circuit that is proposed to exist in dentate-CA3 plays the role described above in the DNMS paradigm.

**SPECTRALLY TIMED GAIN CONTROL IN THE CEREBELLUM**

Why is adaptive timing also needed in the motor conditioning circuit? This need is clarified by the fact that the $S \rightarrow D \rightarrow S$ circuit focuses attention quickly on motivationally salient cues and can thereby just as quickly activate the motor circuit (Fig. 3). Without adaptive timing within the motor circuit itself, the conditioned response could be prematurely released. Thus the clear survival advantage of attending quickly to motivationally important sensory events could disrupt the properly timed execution of responses contingent upon these events. The model suggests that this problem does not occur during normal behaviors because the hippocampal dentate-CA3 circuit and the cerebellar motor circuit are both adaptively timed. These distinct timing functions have been dissociated through ablation (Ebner & Bloedel, 1981; Gilbert & Thach, 1977; Optican & Robinson, 1980; Thompson, 1988; Thompson et al., 1984, 1987) and ISI shift experiments during which the peak time of the hippocampal trace can change before the peak time of the discrete adaptive response (Hochler & Thompson, 1980). The model suggests that orienting responses may be inhibited by the hippocampal dentate-CA3 timing circuit during the same time intervals when conditioned responses are disinhibited by the cerebellar timing circuit. This coordinated action extends the classical idea that consummatory and orienting responses are mutually inhibitory.

Recent experiments on conditioning the rabbit NMR suggest that response learning occurs within a subcortical cerebellar pathway, whereas response timing occurs within the cerebellar cortex (Perrett, Ruiz, & Mauk, 1993). If the cortical timing circuit is ablated, then motor responses are, indeed, prematurely released. These experimental results are consistent with the classical hypothesis that a fast cerebellar motor pathway—here interpreted to be subcortical (Lisberger, 1988)—can learn a conditioned gain appropriate to the response using climbing fiber inputs as a teaching signal (Albus, 1971; Fujita, 1982a, 1982b; Grossberg, 1969b, 1972b; Grossberg & Kuperstein, 1986; Marr, 1969).

We hypothesize, in addition, that adaptive timing is learned by a spectral timing circuit in which parallel fiber-Purkinje cell cortical synapses use climbing fiber inputs as a teaching signal (Fig. 6). In this conception, cortical learning opens a timed gate by removing Purk-
gate opens, the subcortical motor pathway can read-out its learned gain with the correctly timed ISI between CS and US. Learned suppression of Purkinje cell output may be accomplished by conditioned long-term depression, or LTD (Hochler & Thompson, 1980; Ito, 1984). Eight key data properties have been simulated by this model (Bullock et al., 1994): Model Purkinje cell activity decreases in the interval following the onset of the CS, model neural cell responses match CR topography, CR peak amplitude occurs at the US onset, a discrete CR peak shift occurs with a change in ISI, peak CR acquisition and response rates depend unimodally on the ISI, CR onset latency decreases during training, and maladaptively timed small-amplitude CRs result from ablation of cerebellar cortex.

Some striking cellular and circuit homologs exist between these model cerebellum and hippocampal timing mechanisms. Both control an inhibitory gate that modulates another learning process, and both occur on dendrites whose summed output across a spectrum of rate-sensitive cell sites determines the collective timed response. These similarities suggest the prediction that both the hippocampal dentate cell and cerebellar Purkinje cell dendrites may undergo similar biophysical events during conditioning.

COOPERATIVE HIPPOCAMPAL AND CEREBELLAR TIMING DURING SERIAL COMPOUND CONDITIONING

How do the hippocampal and cerebellar timing circuits cooperate during timed behaviors? We illustrate such cooperation below by explaining paradoxical data about serial compound conditioning, during which a sequence CS1-CS2-US of two Cs precedes a US (Kehoe & Morrow, 1984; Kehoe et al., 1979, 1987). Robust serial compound conditioning to CS1 can occur even if primary CS1-US conditioning at the same ISI, in the absence of CS2, is ineffective. This happens, for example, if the CS1-CS2 ISI = 2400 msec and the CS2-US ISI = 400 msec (Kehoe & Morrow, 1984). How does the occurrence of CS2 enable CS1 to bridge the 2800 msec ISI before US occurs?

We suggest that CS2 can reactivate the sensory representation S1 of CS1 via the drive representation D along the feedback pathway CS1 -> S1 -> D -> S2, and thereby restart the S1 -> T and S1 -> M timing circuits. In particular, on the first learning trial, the activity of S1 does not persist until US occurs, but the activity of S2 does. As a result, S2 -> D and D -> S2 conditioning start to occur. On later learning trials, S1 is active when CS2 occurs. Thus S1 is active when S2 activates D. S1 can hereby also learn to activate D, and D can be reciprocally conditioned to both S1 and S2 via the D -> S1 and D -> S2 feedback pathways. In this way, activation of D by CS2 reactivates S1 and restarts its timing circuits, so that they are active when the US occurs. As a result, S1 -> M conditioning of the NMR is possible, but is released earlier than the 2800 msec ISI between CS1 and US.

This explanation clarifies why, if the ISI between CS1 and CS2 is short enough, then CS2 elicits less NMR conditioning than it does when it is conditioned to the US at the same ISI without the occurrence of CS1 (Kehoe et al., 1979). If the CS1-CS2 delay is short enough, S1 can partially block S2 because S1 -> D -> S1 feedback is still strong when CS2 occurs. Conversely, if the total CS1-US ISI is increased, then CS2 can elicit more NMR conditioning than it would in the absence of CS1. Here, S1's activity subsides by the time S2 occurs, but it primes D with residual activity that can amplify S2 -> D -> S2 and S2 -> T conditioning when CS2 and US occur. Kehoe et al. (1993) have shown that a spectral timing model can, indeed, be used to simulate key properties of serial compound conditioning data.

CONCLUDING REMARKS

The neural model described herein suggests how the hippocampal system and cerebellum may cooperate to control adaptively timed recognition learning, motivated attention, and conditioned responding. The model clarifies how the hippocampal system may combine novelty-based modulation of recognition learning and reinforcement learning with a competence for adaptively timed attention and inhibition of orienting responses. In particular, it suggests how orienting responses may be inhibited by the hippocampal dentate-CA3 timing circuit during the same time interval during which goal-oriented conditioned responses are released by adaptively timed opening of the cerebellar Purkinje cell gate.

The model distinguishes between the micro-timing that is needed to determine how long motivated attention needs to be focused on a single predicted goal event, and the macro-timing whereby attention is maintained during the planned performance of a sequence of actions leading to a goal. Both sorts of timing would appear to be at work during many behaviors. A partially developed theory of how they are coordinated clarifies some aspects of the complex pattern of connections that exists between the temporal cortex, frontal cortex, and hippocampal system.

Why the hippocampal system should play a role in spatial orientation is also consistent with this modeling framework. This link is established when one poses the question of how an animal can direct its goal-oriented attentive behaviors among sets of environmental landmarks that vary in their motivational salience. Such a perspective is consistent with the proposal that the hippocampal system can play a role as a cognitive map (Leonard & McNaughton, 1990; O'Keefe, 1990; O'Keefe
suitably defined, without denying its relevance for the control of approach-avoidance behaviors (Amsel, 1993). How to computationally integrate the steering role of reinforcement and motivation into a self-organizing network for spatial orientation remains an open problem. Despite these theoretical gaps, the ART models that have already been developed put mechanistic flesh on the metaphorical bones of declarative memory and procedural memory by articulating new behavioral principles, neural mechanisms, and experimental explanations and predictions that can be used to clarify how a freely moving individual flexibly learns about and acts upon valued goal objects in a timely fashion.

Acknowledgments
S. G. was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), ARPA (ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100 and ONR N00014-92-J-1309). J. W. L. M. was supported in part by the Air Force Office of Scientific Research (AFOSR F4962086-C0037 and AFOSR 90-0128), the National Institute of Mental Health (NIMH MH-42900), and the Office of Naval Research (ONR N00014-92-J-1904). Present address: Microsoft Corporation, One Microsoft Way, Richmond, Washington 98052. We wish to thank Cynthia E. Bradford and Diana Meyers for their valuable assistance in the preparation of the manuscript.

Reprint request should be sent to Professor Stephen Grossberg, Department of Cognitive and Neural Systems, Boston University, 111 Cummington Street, Boston, MA 02215.

REFERENCES
MAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. 
*Neural Networks, 4*, 565-588.


Grossberg, S. (1972a). A neural theory of punishment and


Grossberg and Merrill 277