ADAPTIVE RESONANCE THEORY: STABLE SELF-ORGANIZATION OF NEURAL RECOGNITION CODES IN RESPONSE TO ARBITRARY LISTS OF INPUT PATTERNS

Gail A. Carpenter† and Stephen Grossberg‡
Center for Adaptive Systems
Department of Mathematics
Boston University
Boston, Massachusetts 02215

1. SELF-ORGANIZATION OF NEURAL RECOGNITION CODES

A neural network, called an *adaptive resonance theory* (ART) architecture, for the learning of recognition categories is described herein. Real-time network dynamics for this model have been completely characterized through mathematical analysis and computer simulations. The architecture self-organizes and self-stabilizes its recognition codes in response to arbitrary orderings of arbitrarily many and arbitrarily complex binary input patterns. Top-down attentional and matching mechanisms are critical in self-stabilizing the code learning process. The architecture embodies a parallel search scheme which updates itself adaptively as the learning process unfolds. After learning self-stabilizes, the search process is automatically disengaged. Thereafter input patterns directly access their recognition codes without any search. Thus recognition time does not grow as a function of code complexity. A novel input pattern can directly access a category if it shares invariant properties with the set of familiar exemplars of that category. These invariant properties emerge in the form of learned critical feature patterns, or prototypes. The architecture possesses a context-sensitive self-scaling property which enables its emergent critical feature patterns to form. They detect and remember statistically predictive configurations of featural elements which are derived from the set of all input patterns that are ever experi-

† Supported in part by the Air Force Office of Scientific Research (AFOSR 85-0149 and AFOSR F49620-86-C-0037), the Army Research Office (ARO DAAG-29-85-K-0095), and the National Science Foundation (NSF DMS-84-13119).

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR 85-0149 and AFOSR F49620-86-C-0037) and the Army Research Office (ARO DAAG-29-85-K-0095).

Acknowledgements: We wish to thank Cynthia Suchta and Carol Yanakakis for their valuable assistance in the preparation of the manuscript.

enced. Four types of attentional process—priming, gain control, vigilance, and intermodal competition—are mechanistically characterized. Top-down priming and gain control are needed for code matching and self-stabilization. Attentional vigilance determines how fine the learned categories will be. If vigilance increases due to an environmental disconfirmation, then the system automatically searches for and learns finer recognition categories. A new nonlinear matching law (the 2/3 Rule) and new nonlinear associative laws (the Weber Law Rule, the Associative Decay Rule, and the Template Learning Rule) are needed to achieve these properties. All the rules describe emergent properties of parallel network interactions. The architecture circumvents the noise, saturation, capacity, orthogonality, and linear predictability constraints that limit the codes which can be stably learned by alternative recognition models. In addition, ART circuits have elsewhere been used to analyse data about speech perception, word recognition and recall, visual perception, olfactory coding, evoked potentials, thalamocortical interactions, attentional modulation of critical period termination, and amnesias (Banquet and Grossberg, 1986; Carpenter and Grossberg, 1985a, 1985b, 1986a, 1986b, 1986c; Grossberg, 1976a, 1976b, 1978a, 1980, 1986a; Grossberg and Stone, 1986a, 1986b). In the following pages, we describe intuitively some key properties of the model.

2. STABILITY-PLASTICITY DILEMMA: MULTIPLE INTERACTING MEMORY SYSTEMS

An adequate self-organizing recognition system must be capable of plasticity in order to learn about significant new events, yet it must also remain stable in response to irrelevant or often repeated events. In order to prevent the relentless degradation of its learned codes by the "blooming, buzzing confusion" of irrelevant experience, an ART system is sensitive to *novelty*. It is capable of distinguishing between familiar and unfamiliar events, as well as between expected and unexpected events.

Multiple interacting memory systems are needed to monitor and adaptively react to the novelty of events. Within ART, interactions between two functionally complementary subsystems are needed to process familiar and unfamiliar events. Familiar events are processed within an attentional subsystem. This subsystem establishes ever more precise internal representations of and responses to familiar events. It also builds up the learned top-down expectations that help to stabilize the learned bottom-up codes of familiar events. By itself, however, the attentional subsystem is unable simultaneously to maintain stable representations of familiar categories and to create new categories for unfamiliar patterns. An isolated attentional subsystem is either rigid and incapable of creating new categories for unfamiliar patterns, or unstable and capable of ceaselessly recoding the categories of familiar patterns.

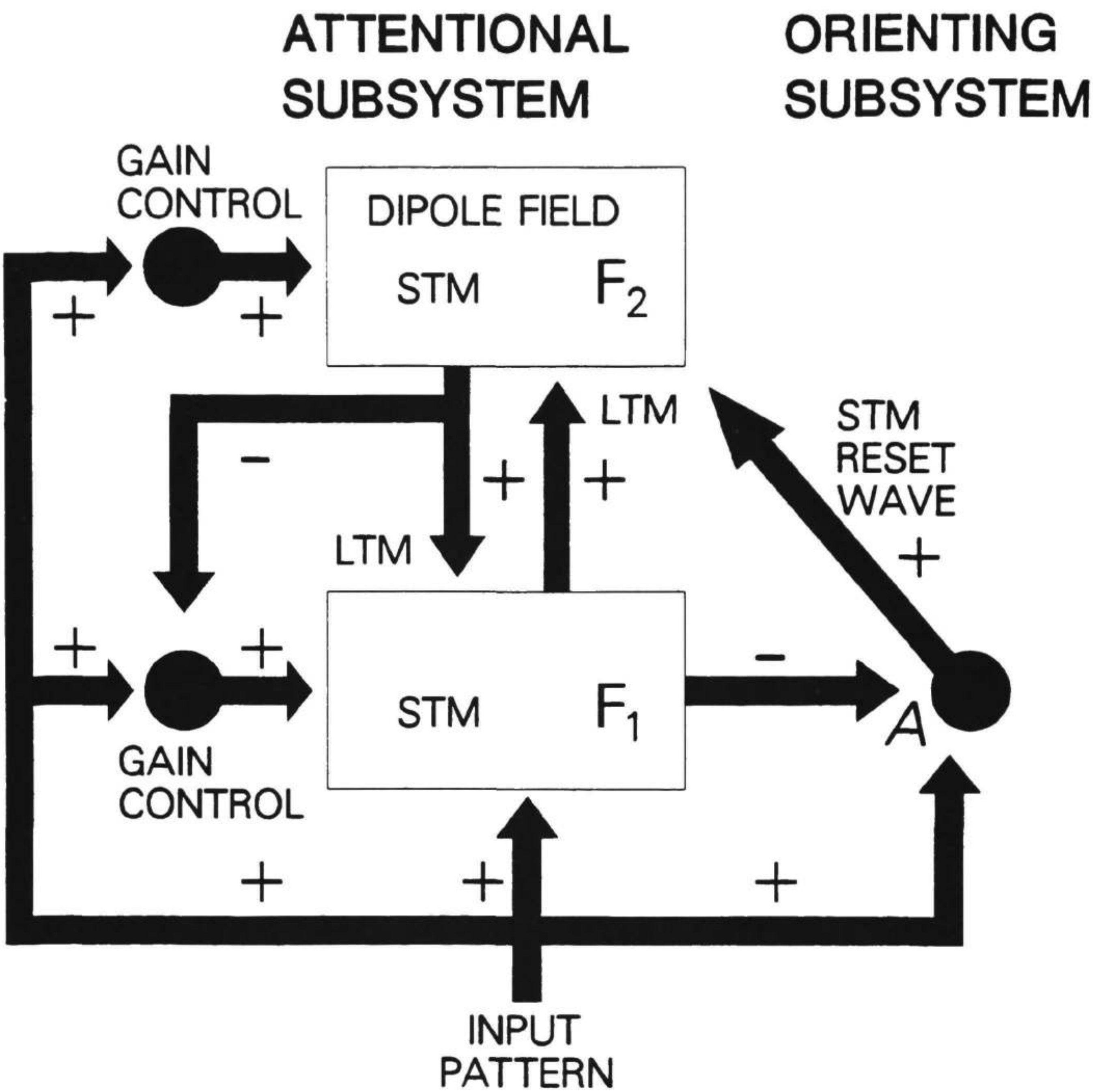
The second subsystem is an orienting subsystem that resets the attentional subsystem when an unfamiliar event occurs. The orienting subsystem is essential for expressing whether a novel pattern is familiar and well represented by an existing recognition code, or unfamiliar and in need of a new recognition code. Figure 1 schematizes the architecture that is analysed herein.

An ART system dynamically reorganizes its recognition codes to preserve its stability-plasticity balance as its internal representations become increasingly complex and differentiated through learning. By contrast, many classical adaptive pattern recognition systems become unstable when they are confronted by complex input environments. Unlike many alternative models the present model can deal with arbitrary combinations of binary input patterns. In particular, it places no orthogonality or linear predictability constaints upon its input patterns. The model computations remain sensitive no matter how many input patterns are processed. The model does not require that very small, and thus noise-degradable, increments in memory be made in order to avoid saturation of its cumulative memory. The model can store arbitrarily many recognition categories in response to input patterns that are defined on arbitrarily many input channels. Its memory matrices need not be square, so that no restrictions on memory capacity are imposed by the number of input channels. Finally, all the memory of the system can be devoted to stable recognition learning. It is not he case that the number of stable classifications is bounded by some fraction of the number of input channels or patterns. Thus a primary goal of the present article is to intuitively describe neural networks capable of self-stabilizing the self-organization of their recognition codes in response to an arbitrarily complex environment of input patterns.

Four properties are basic to the workings of the networks that we characterize herein.

A. Self-Scaling Computational Units: Critical Feature Patterns

Properly defining signal and noise in a self-organizing system raises a number of subtle issues. Pattern context must enter the definition so that input features which are treated as irrelevant noise when they are embedded in a given input pattern may be treated as informative signals when they are embedded in a different input pattern. The system's unique learning history must also enter the definition so that portions of an input pattern which are treated as noise when they perturb a system at one stage of its self-organization may be treated as signals when they perturb the same system at a different stage of its self-organization. The present systems automatically self-scale their computational units to embody context- and learning-dependent definitions of signal and noise.



1. Anatomy of the attentional-orienting system: Two successive stages, F_1 and F_2, of the attentional subsystem encode patterns of activation in short term memory (STM). Bottom-up and top-down pathways between F_1 and F_2 contain adaptive long term memory (LTM) traces which multiply the signals in these pathways. The remainder of the circuit modulates these STM and LTM processes. Modulation by gain control enables F_1 to distinguish between bottom-up input patterns and top-down priming, or template, patterns, as well as to match these bottom-up and top-down patterns. Gain control signals also enable F_2 to react supraliminally to signals from F_1 while an input pattern is on. The orienting subsystem generates a reset wave to F_2 when mismatches between bottom-up and top-down patterns occur at F_1. This reset wave selectively and enduringly inhibits active F_2 cells until the input is shut off.

One property of these self-scaling computational units is schematized in Figure 2. In Figure 2a, each of the two input patterns is composed of three features. The patterns agree at two of the three features, but disagree at the third feature. A mismatch of one out of three features may be designated as informative by the system. When this occurs, these mismatched features are treated as signals which can elicit learning of distinct recognition codes for the two patterns. Moreover, the mismatched features, being informative, are incorporated into these distinct recognition codes.

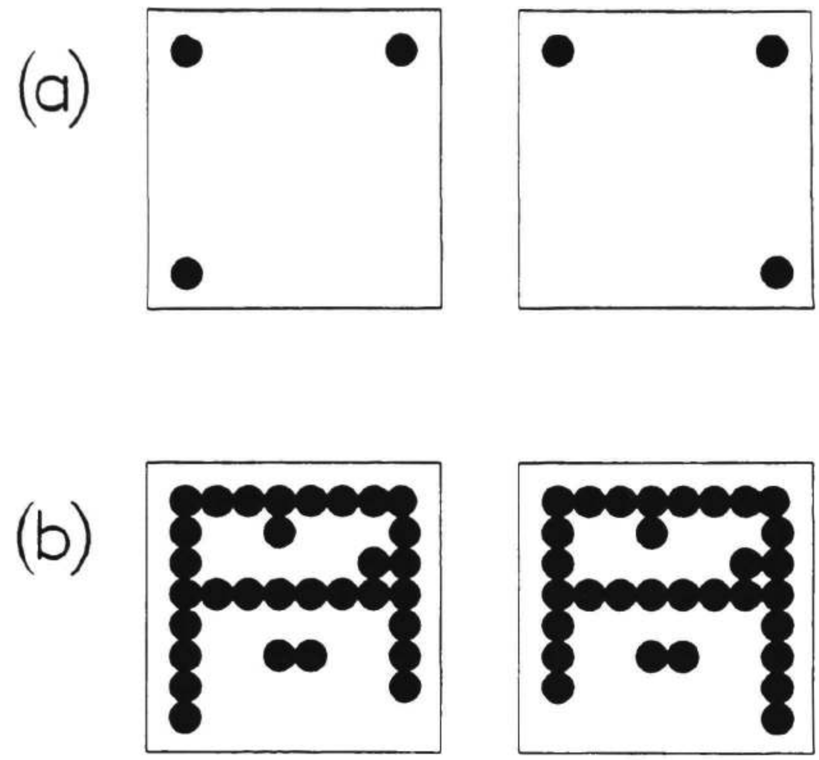
In Figure 2b, each of the two input patterns is composed of thirty-one features. The patterns are constructed by adding identical subpatterns to the two patterns in Figure 2a. Thus the input patterns in Figure 2b disagree at the same features as the input patterns in Figure 2a. In the patterns of Figure 2b, however, this mismatch is less important, other things being equal, than in the patterns of Figure 2a. Consequently, the system may treat the mismatched features as noise. A single recognition code may be learned to represent both of the input patterns in Figure 2b. The mismatched features would not be learned as part of this recognition code because they are treated as noise.

The assertion that *critical feature patterns* are the computational units of the code learning process summarizes this self-scaling property. The term *critical feature* indicates that not all features are treated as signals by the system. The learned units are *patterns* of critical features because the perceptual context in which the features are embedded influences which features will be processed as signals and which features will be processed as noise. Thus a feature may be a critical feature in one pattern (Figure 2a) and an irrelevant noise element in a different pattern (Figure 2b).

The need to overcome the limitations of featural processing with some type of contextually sensitive pattern processing has long been a central concern in the human pattern recognition literature. Experimental studies have led to the general conclusions that "The trace system which underlies the recognition of patterns can be characterized by a central tendency and a boundary" (Posner, 1973, p.54), and that "just listing features does not go far enough in specifying the knowledge represented in a concept. People also know something about the relations between the features of a concept, and about the variability that is permissible on any feature" (Smith and Medin, 1981, p.83). We illustrate herein how these properties may be achieved using self-scaling computational units such as critical feature patterns.

B. Self-Adjusting Memory Search

No pre-wired search algorithm, such as a search tree, can maintain its efficiency as a knowledge structure evolves due to learning in a unique input environment. A search order



2. Self-scaling property discovers critical features in a context-sensitive way: (a) Two input patterns of 3 features mismatch at 1 feature. When this mismatch is sufficient to generate distinct recognition codes for the two patterns, the mismatched features are encoded in LTM as part of the critical feature patterns of these recognition codes. (b) Identical subpatterns are added to the two input patterns in (a). Although the new input patterns mismatch at the same one feature, this mismatch may be treated as noise due to the additional complexity of the two new patterns. Both patterns may thus learn to activate the same recognition code. When this occurs, the mismatched feature is deleted from LTM in the critical feature pattern of the code.

that may be optimal in one knowledge domain may become extremely inefficient as that knowledge domain becomes more complex due to learning.

The ART system considered herein is capable of a parallel memory search that adaptively updates its search order to maintain efficiency as its recognition code becomes arbitrarily complex due to learning. This self-adjusting search mechanism is part of the network design whereby the learning process self-stabilizes by engaging the orienting subsystem (Section 5).

None of these mechanisms is akin to the rules of a serial computer program. Instead, the circuit architecture as a whole generates a self-adjusting search order and self-stabilization as emergent properties that arise through system interactions. Once the ART architecture is in place, a little randomness in the initial values of its memory traces, rather than a carefully wired search tree, enables the search to carry on until the recognition code self-stabilizes.

C. Direct Access to Learned Codes

A hallmark of human recognition performance is the remarkable rapidity with which familiar objects can be recognized. The existence of many learned recognition codes for alternative experiences does not necessarily interfere with rapid recognition of an unambiguous familiar event. This type of rapid recognition is very difficult to understand using models wherein trees or other serial algorithms need to be searched for longer and longer periods as a learned recognition code becomes larger and larger.

In an ART model, as the learned code becomes globally self-consistent and predictively accurate, the search mechanism is automatically disengaged. Subsequently, no matter how large and complex the learned code may become, familiar input patterns *directly access*, or activate, their learned code, or category. Unfamiliar patterns can also directly access a learned category if they share invariant properties with the critical feature pattern of the category. In this sense, the critical feature pattern acts as a prototype for the entire category. As in human pattern recognition experiments, an input pattern that matches a learned critical feature pattern may be better recognized than any of the input patterns that gave rise to the critical feature pattern (Posner, 1973; Posner and Keele, 1968, 1970).

Unfamiliar input patterns which cannot stably access a learned category engage the self-adjusting search process in order to discover a network substrate for a new recognition category. After this new code is learned, the search process is automatically disengaged and direct access ensues.

D. Environment as a Teacher: Modulation of Attentional Vigilance

Although an ART system self-organizes its recognition code, the environment can also modulate the learning process and thereby carry out a teaching role. This teaching role allows a system with a fixed set of feature detectors to function successfully in an environment which imposes variable performance demands. Different environments may demand either coarse discriminations or fine discriminations to be made among the same set of objects. As Posner (1973, pp.53–54) has noted:

> "If subjects are taught a tight concept, they tend to be very careful about classifying any particular pattern as an instance of that concept. They tend to reject a relatively small distortion of the prototype as an instance, and they rarely classify a pattern as a member of the concept when it is not. On the other hand, subjects learning high-variability concepts often falsely classify patterns as members of the concept, but rarely reject a member of the concept incorrectly...The situation

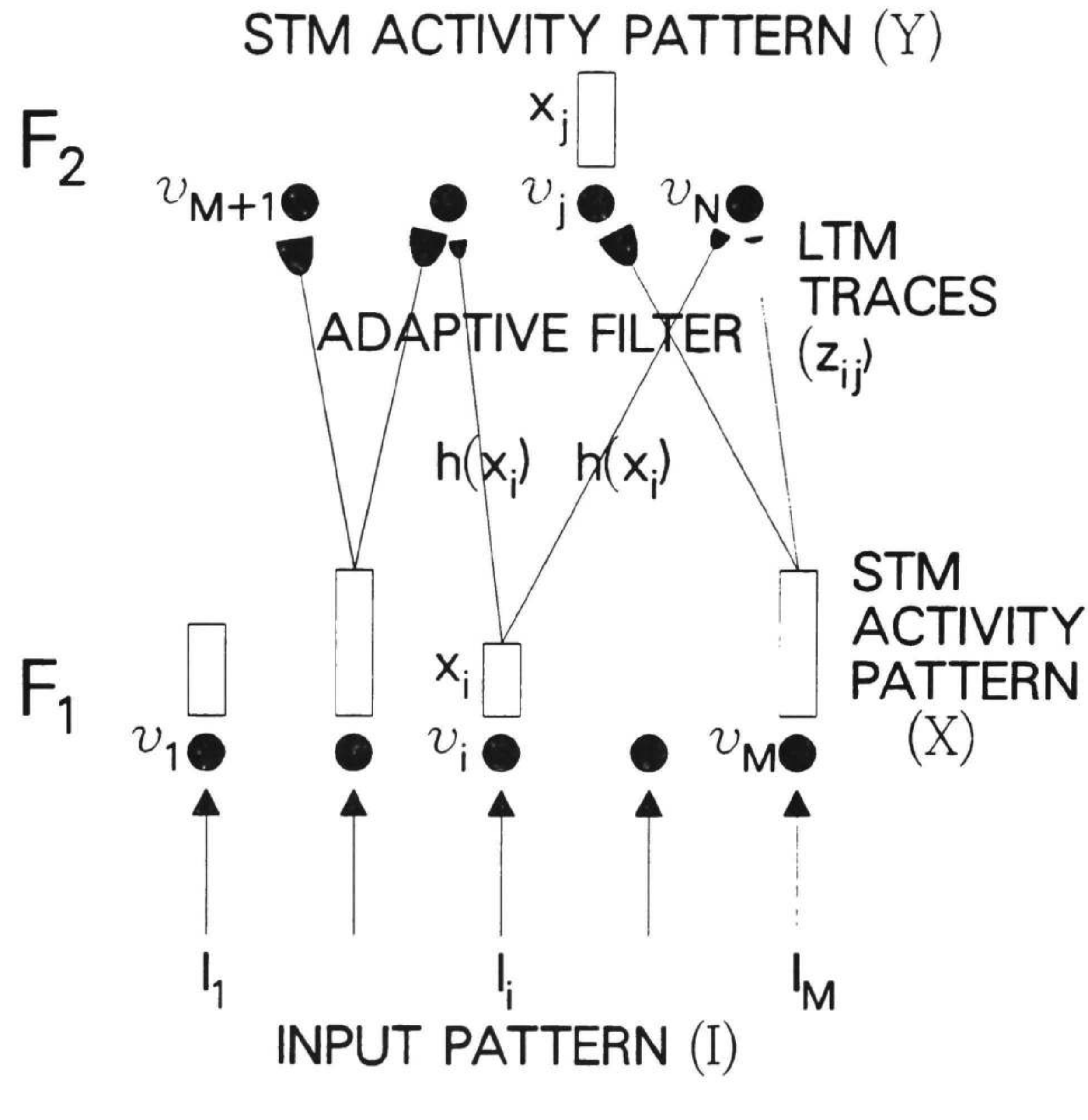
largely determines which type of learning will be superior."

In an ART system, if an erroneous recognition is followed by negative reinforcement, then the system becomes more *vigilant*. This change in vigilance may be interpreted as a change in the system's attentional state which increases its sensitivity to mismatches between bottom-up input patterns and active top-down critical feature patterns. A vigilance change alters the size of a single parameter in the network. The *interactions* within the network respond to this parameter change by learning recognition codes that make finer distinctions. In other words, if the network erroneously groups together some input patterns, then negative reinforcement can help the network to learn the desired distinction by making the system more vigilant. The system then behaves *as if* it has a better set of feature detectors.

The ability of a vigilance change to alter the course of pattern recognition illustrates a theme that is common to a variety of neural processes: a one-dimensional parameter change that modulates a simple nonspecific neural process can have complex specific effects upon high-dimensional neural information processing.

3. BOTTOM-UP ADAPTIVE FILTERING AND CONTRAST-ENHANCEMENT IN SHORT TERM MEMORY

The remainder of the article intuitively summarizes key model properties. We begin by considering the typical network reactions to a single input pattern I within a temporal stream of input patterns. Each input pattern may be the output pattern of a preprocessing stage. Different preprocessing is given, for example, to speech signals and to visual signals before the outcome of such modality-specific preprocessing ever reaches the attentional subsystem. The preprocessed input pattern I is received at the stage F_1 of an attentional subsystem. Pattern I is transformed into a pattern X of activation across the nodes, or abstract "feature detectors", of F_1 (Figure 3). The transformed pattern X represents a pattern in short term memory (STM). In F_1 each node whose activity is sufficiently large generates excitatory signals along pathways to target nodes at the next processing stage F_2. A pattern X of STM activities across F_1 hereby elicits a pattern S of output signals from F_1. When a signal from a node in F_1 is carried along a pathway to F_2, the signal is multiplied, or *gated*, by the pathway's long term memory (LTM) trace. The LTM gated signal (i.e., signal times LTM trace), not the signal alone, reaches the target node. Each target node sums up all of its LTM gated signals. In this way, pattern S generates a pattern T of LTM-gated and summed input signals to F_2 (Figure 4a). The transformation from S to T is called an *adaptive filter*.



3. Stages of bottom-up activation: The input pattern I generates a pattern of STM activation X across F_1. Sufficiently active F_1 nodes emit bottom-up signals to F_2. This signal pattern S is gated by long term memory (LTM) traces within the $F_1 \rightarrow F_2$ pathways. The LTM gated signals are summed before activating their target nodes in F_2. This LTM-gated and summed signal pattern T generates a pattern of activation Y across F_2. The nodes in F_1 are denoted by $v_1, v_2, \ldots, v_M$. The nodes in F_2 are denoted by $v_{M+1}, v_{M+2}, \ldots v_N$. The input to node v_i is denoted by I_i. The STM activity of node v_i is denoted by x_i. The LTM trace of the pathway from v_i to v_j is denoted by z_{ij}.

The input pattern T to F_2 is quickly transformed by interactions among the nodes of F_2. These interactions contrast-enhance the input pattern T. The resulting pattern of activation across F_2 is a new pattern Y. The contrast-enhanced pattern Y, rather than the input pattern T, is stored in STM by F_2.

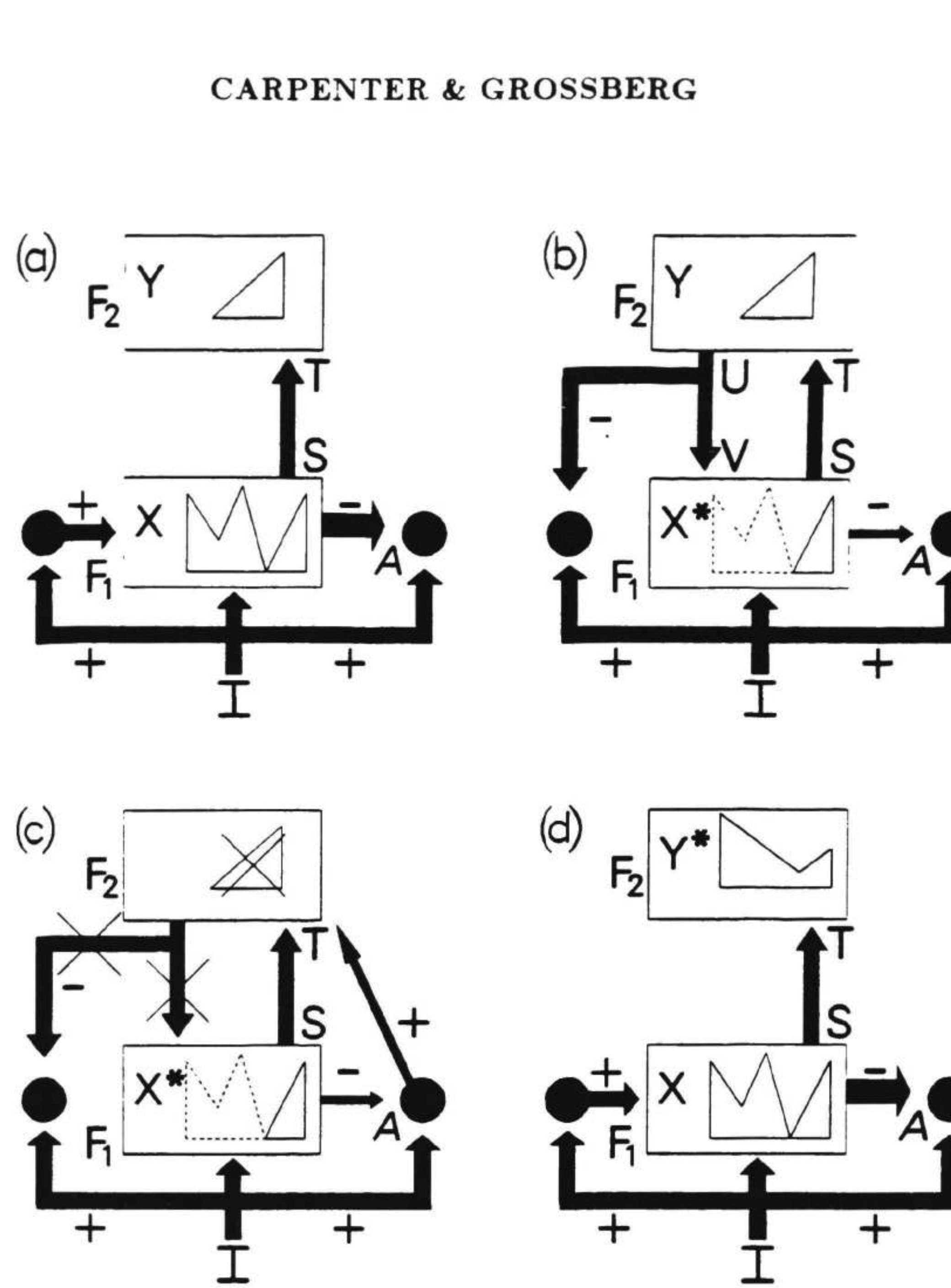
A special case of this contrast-enhancement process is one in which F_2 chooses the node which receives the largest input. The chosen node is the only one that can store activity in STM. In general, the contrast enhancing transformation from T to Y enables more than one node at a time to be active in STM. Such transformations are designed to simultaneously represent in STM several groupings, or chunks, of an input pattern (Cohen and Grossberg, 1986a, 1986b, 1986c; Grossberg, 1978a, 1986a). When F_2 is designed to make a choice in STM, it selects that global grouping of the input pattern which is preferred by the adaptive filter. This process automatically enables the network to partition all the input patterns which are received by F_1 into disjoint sets of recognition categories, each corresponding to a particular node (or "pointer," or "index") in F_2.

All the LTM traces in the adaptive filter, and thus all learned past experiences of the network, are used to determine the recognition code Y via the transformation $I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y$. However, only those nodes of F_2 which maintain stored activity in the STM pattern Y can elicit new learning at contiguous LTM traces. Because the recognition code Y is a more contrast-enhanced pattern than T, many F_2 nodes which receive positive inputs ($I \rightarrow X \rightarrow S \rightarrow T$) may not store any STM activity ($T \rightarrow Y$). The LTM traces in pathways leading to these nodes thus influence the recognition event but are not altered by the recognition event. Some memories which influence the focus of attention are not themselves attended.

4. TOP-DOWN TEMPLATE MATCHING AND STABILIZATION OF CODE LEARNING

As soon as the bottom-up STM transformation $X \rightarrow Y$ takes place, the STM activities Y in F_2 elicit a top-down excitatory signal pattern U back to F_1 (Figure 4b). Only sufficiently large STM activities in Y elicit signals in U along the feedback pathways $F_2 \rightarrow F_1$. As in the bottom-up adaptive filter, the top-down signals U are also gated by LTM traces and the LTM-gated signals are summed at F_1 nodes. The pattern U of output signals from F_2 hereby generates a pattern V of LTM-gated and summed input signals to F_1. The transformation from U to V is thus also an adaptive filter. The pattern V is called a *top-down template*, or *learned expectation*.

Two sources of input now perturb F_1: the bottom-up input pattern I which gave rise to the original activity pattern X, and the top-down template pattern V that resulted from



4. Search for a correct F_2 code: (a) The input pattern I generates the specific STM activity pattern X at F_1 as it nonspecifically activates A. Pattern X both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM pattern Y across F_2. (b) Pattern Y generates the top-down signal pattern U which is transformed into the template pattern V. If V mismatches I at F_1, then a new STM activity pattern X^* is generated at F_1. 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 A. (c) Then the input-driven activation of A can release a nonspecific arousal wave to F_2, which resets the STM pattern Y at F_2. (d) After Y is inhibited, its top-down template is eliminated, and X can be reinstated at F_1. Now X once again generates input pattern T to F_2, but since Y remains inhibited T can activate a different STM pattern Y^* at F_2. If the top-down template due to Y^* also mismatches I at F_1, then the rapid search for an appropriate F_2 code continues.

activating X. The activity pattern X^* across F_1 that is induced by I and V taken together is typically different from the activity pattern X that was previously induced by I alone. In particular, F_1 acts to match V against I. The result of this matching process determines the future course of learning and recognition by the network.

The entire activation sequence

$$I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y \rightarrow U \rightarrow V \rightarrow X^* \tag{1}$$

takes place very quickly relative to the rate with which the LTM traces in either the bottom-up adaptive filter $S \rightarrow T$ or the top-down adaptive filter $U \rightarrow V$ can change. Even though none of the LTM traces changes during such a short time, their prior learning strongly influences the STM patterns Y and X^* that evolve within the network by determining the transformations $S \rightarrow T$ and $U \rightarrow V$. We now discuss how a match or mismatch of I and V at F_1 regulates the course of learning in response to the pattern I, and in particular solves the stability-plasticity dilemma (Section 2).

5. INTERACTIONS BETWEEN ATTENTIONAL AND ORIENTING SUBSYSTEMS: STM RESET AND SEARCH

In Figure 4a, an input pattern I generates an STM activity pattern X across F_1. The input pattern I also excites the orienting subsystem A, but pattern X at F_1 inhibits A before it can generate an output signal. Activity pattern X also elicits an output pattern S which, via the bottom-up adaptive filter, instates an STM activity pattern Y across F_2. In Figure 4b, pattern Y reads a top-down template pattern V into F_1. Template V mismatches input I, thereby significantly inhibiting STM activity across F_1. The amount by which activity in X is attenuated to generate X^* depends upon how much of the input pattern I is encoded within the template pattern V.

When a mismatch attenuates STM activity across F_1, the total size of the inhibitory signal from F_1 to A is also attenuated. If the attenuation is sufficiently great, inhibition from F_1 to A can no longer prevent the arousal source A from firing. Figure 4c depicts how disinhibition of A releases an arousal burst to F_2 which equally, or nonspecifically, excites all the F_2 cells. The cell populations of F_2 react to such an arousal signal in a state-dependent fashion. In the special case that F_2 chooses a single population for STM storage, the arousal burst selectively inhibits, or resets, the active population in F_2. This inhibition is long-lasting. One physiological design for F_2 processing which has these properties is a *gated dipole field* (Grossberg, 1980, 1984a). A gated dipole field consists of opponent processing channels which are gated by habituating chemical transmitters. A nonspecific arousal burst induces selective and enduring inhibition of active populations within a gated dipole field.

In Figure 4c, inhibition of Y leads to removal of the top-down template V, and thereby terminates the mismatch between I and V. Input pattern I can thus reinstate the original activity pattern X across F_1, which again generates the output pattern S from F_1 and the input pattern T to F_2. Due to the enduring inhibition at F_2, the input pattern T can no longer activate the original pattern Y at F_2. A new pattern Y^* is thus generated at F_2 by I (Figure 4d). Despite the fact that some F_2 nodes may remain inhibited by the STM reset property, the new pattern Y^* may encode large STM activities. This is because level F_2 is designed so that its total suprathreshold activity remains approximately constant, or normalized, despite the fact that some of its nodes may remain inhibited by the STM reset mechanism. This property is related to the limited capacity of STM. A physiological process capable of achieving the STM normalization property is based upon on-center off-surround feedback interactions among cells obeying membrane equations (Grossberg, 1980, 1983).

The new activity pattern Y^* reads-out a new top-down template pattern V^*. If a mismatch again occurs at F_1, the orienting subsystem is again engaged, thereby leading to another arousal-mediated reset of STM at F_2. In this way, a rapid series of STM matching and reset events may occur. Such an STM matching and reset series controls the system's search of LTM by sequentially engaging the novelty-sensitive orienting subsystem. Although STM is reset sequentially in time via this mismatch-mediated, self-terminating LTM search process, the mechanisms which control the LTM search are all parallel network interactions, rather than serial algorithms. Such a parallel search scheme continuously adjusts itself to the system's evolving LTM codes. In general, the spatial configuration of LTM codes depends upon both the system's initial configuration and its unique learning history, and hence cannot be predicted *a priori* by a pre-wired search algorithm. Instead, the mismatch-mediated engagement of the orienting subsystem realizes the type of self-adjusting search that was described in Section 2B.

The mismatch-mediated search of LTM ends when an STM pattern across F_2 reads-out a top-down template which matches I, to the degree of accuracy required by the level of attentional vigilance, or which has not yet undergone any prior learning. In the latter case, a new recognition category is then established as a bottom-up code and top-down template are learned.

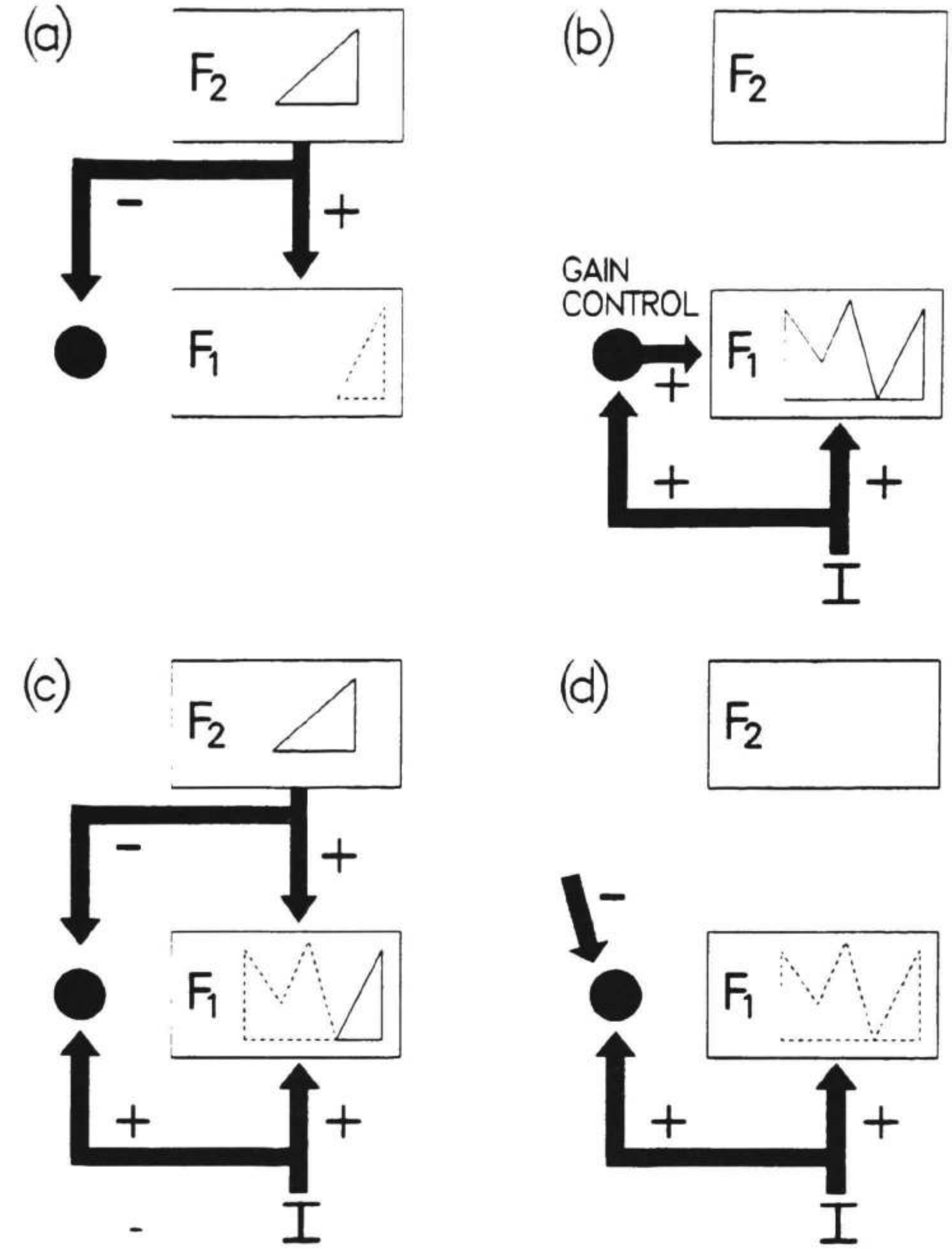
6. ATTENTIONAL GAIN CONTROL AND ATTENTIONAL PRIMING

Further properties of the top-down template matching process can be derived by considering its role in the regulation of attentional priming. Consider, for example, a situation in which F_2 is activated by a level other than F_1 before F_1 can be activated by a bottom-up input (Figure 5a). In such a situation, F_2 can generate a top-down template V to F_1. The level F_1 is then primed, or sensitized, to receive a bottom-up input that may or may not match the active expectancy. As depicted in Figure 5a, level F_1 can be primed to receive a bottom-up input without necessarily eliciting suprathreshold output signals in response to the priming expectancy.

On the other hand, an input pattern I must be able to generate a suprathreshold activity pattern X even if no top-down expectancy is active across F_1 (Figures 4a and 5b). How does F_1 know that it should generate a suprathreshold reaction to a bottom-up input pattern but not to a top-down input pattern? In both cases, excitatory input signals stimulate F_1 cells. Some auxiliary mechanism must exist to distinguish between bottom-up and top-down inputs. This auxiliary mechanism is called *attentional gain control* to distinguish it from *attentional priming* by the top-down template itself (Figure 5a). While F_2 is active, the attentional priming mechanism delivers *excitatory specific learned* template patterns to F_1. The attentional gain control mechanism has an *inhibitory nonspecific unlearned* effect on the sensitivity with which F_1 responds to the template pattern, as well as to other patterns received by F_1. The attentional gain control process enables F_1 to tell the difference between bottom-up and top-down signals.

7. MATCHING: THE 2/3 RULE

A rule for pattern matching at F_1, called the 2/3 Rule, follows naturally from the distinction between attentional gain control and attentional priming. It says that two out of three signal sources must activate an F_1 node in order for that node to generate suprathreshold output signals. In Figure 5a, during top-down processing, or priming, the nodes of F_1 receive inputs from at most one of their three possible input sources. Hence no cells in F_1 are supraliminally activated by the top-down template. In Figure 5b, during bottom-up processing, a suprathreshold node in F_1 is one which receives a specific input from both the input pattern I and a nonspecific excitatory signal from the gain control channel. In Figure 5c, during the matching of simultaneous bottom-up and top-down patterns, the nonspecific gain control signal to F_1 is inhibited by the top-down channel. Nodes of F_1 which receive sufficiently large inputs from both the bottom-up and the top-down signal patterns generate suprathreshold activities. Nodes which receive a bottom-up input or a top-down input, but not both, cannot become suprathreshold: mismatched inputs cannot generate suprathreshold activities. Attentional gain control thus leads to a matching process whereby the addition of top-down excitatory inputs to F_1 can lead



5. Matching by the 2/3 Rule: (a) A top-down template from F_2 inhibits the attentional gain control source as it subliminally primes target F_1 cells. (b) Only F_1 cells that receive bottom-up inputs and gain control signals can become supraliminally active. (c) When a bottom-up input pattern and a top-down template are simultaneously active, only those F_1 cells that receive inputs from both sources can become supraliminally active. (d) Intermodality inhibition can shut off the F_1 gain control source and thereby prevent a bottom-up input from supraliminally activating F_1. Similarly, disinhibition of the F_1 gain control source by an "act of will" may enable a top-down prime to become supraliminal.

to an overall decrease in F_1's STM activity (Figures 4a and 4b). Figure 5d shows how competitive interactions across modalities can prevent F_1 from generating a supraliminal reaction to bottom-up signals when attention shifts from one modality to another.

8. CONCLUDING REMARKS: SELF-STABILIZATION AND UNITIZATION WITHIN ASSOCIATIVE NETWORKS

The qualitative properties discussed herein are elsewhere supplemented by a complete set of mathematical theorems and many computer simulations (Carpenter and Grossberg, 1986a, 1986b, 1986c). Two main conclusions of our work are especially salient. First, the code learning process is one of progressive refinement of distinctions. The distinctions that emerge are the resultant of all the input patterns which the network ever experiences, rather than of some preassigned features. Second, the matching process compares whole patterns, not just separate features. It may happen that two different input patterns to F_1 overlap a template at the same set of feature detectors, yet the network will reset the F_2 node in response to one input but not the other. The degree of mismatch of template pattern and input pattern *as a whole* determines whether coding or reset will occur.

Thus the learning of categorical invariants resolves two opposing tendencies. As categories grow larger, and hence code increasingly global invariants, the templates which define them become smaller, as they discover and base the code on sets of critical feature patterns, or prototypes, rather than upon familiar pattern exemplars. This work shows how these two opposing tendencies can be resolved within a self-organizing system, leading to dynamic equilibration, or self-stabilization, of recognition categories in response to an arbitrary list of arbitrarily many binary input patterns. This self-stabilization property is of major importance for the further development of associative networks and the analysis of cognitive recognition processes.

REFERENCES

Banquet, J.P., & Grossberg, S. (1986) Structure of event-related potentials during learning: An experimental and theoretical analysis. Submitted for publication.

Carpenter, G.A., & Grossberg, S. (1985a) Category learning and adaptive pattern recognition: A neural network model. *Proceedings of the Third Army Conference on Applied Mathematics and Computing*, ARO Report **86-1**, 37–56.

Carpenter, G.A., & Grossberg, S. (1985b) Neural dynamics of adaptive pattern recognition: Priming, search, attention, and category formation. **Society for Neuroscience Abstracts, 11**, 1110.

Carpenter, G.A., & Grossberg, S. (1986a) Neural dynamics of category learning and recognition: Attention, memory consolidation, and amnesia. In J. Davis, R. Newburgh, and E. Wegman (Eds.), **Brain structure, learning, and memory**. AAAS Symposium Series.

Carpenter, G.A., & Grossberg, S. (1986b) Neural dynamics of category learning and recognition: Structural invariants, reinforcement, and evoked potentials. In M.L. Commons, S.M. Kosslyn, and R.J. Herrnstein (Eds.), **Pattern recognition and concepts in animals, people, and machines**. Hillsdale, NJ: Erlbaum.

Carpenter, G.A., & Grossberg, S. (1986c) A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing.*

Cohen, M.A., & Grossberg, S. (1986a) Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology*, **5**, 1–22.

Cohen, M.A., & Grossberg, S. (1986b) Unitized recognition codes for parts and wholes: The unique cue in configural discriminations. In M.L. Commons, S.M. Kosslyn, and R.J. Herrnstein (Eds.), **Pattern recognition and concepts in animals, people, and machines**. Hillsdale, NJ: Erlbaum

Cohen, M.A., & Grossberg, S. (1986c) Adaptive tuning of unitized perceptual groupings: Neural association, competition, and modulation.

Grossberg, S. (1976a) Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, **23**, 121–134.

Grossberg, S. (1976b) Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, **23**, 187–202.

Grossberg, S. (1978a) A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), **Progress in theoretical biology**, Vol.5, pp.233–374. New York: Academic Press.

Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review*, **87**, 1–51.

Grossberg, S. (1983) The quantized geometry of visual space: The coherent computation of depth, form, and lightness. *Behavioral Brain Sciences*, **6**, 625–692.

Grossberg, S. (1984a) Some psychophysiological and pharmacological correlates of a developmental, cognitive and motivational theory. In R. Karrer, J. Cohen, and P. Tueting (Eds.), **Brain and information: Event related potentials**, pp.58–151. New York: New York Academy of Sciences.

Grossberg, S. (1986a) The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E.C. Schwab and H.C. Nusbaum (Eds.), **Pattern recognition by humans and machines**, Vol.1. New York: Academic Press.

Grossberg, S., & Stone, G.O. (1986a) Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, **93**, 46–74.

Grossberg, S., & Stone, G.O. (1986b) Neural dynamics of attention switching and temporal order information in short term memory, *Memory and Cognition*.

Posner, M.I. (1973) **Cognition: an introduction**. Glenview, IL: Scott, Foresman, and Co.

Posner, M.I., & Keele, S.W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353–363.

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

Smith, E.E., & Medin, D.L. (1981) **Categories and concepts**. Cambridge, MA: Harvard University Press.