

NEURAL DYNAMICS OF CATEGORY LEARNING  
AND RECOGNITION: ATTENTION,  
MEMORY CONSOLIDATION, AND AMNESIA

Gail A. Carpenter† and Stephen Grossberg‡

**Abstract**

A theory is developed of how recognition categories can be learned in response to a temporal stream of input patterns. Interactions between an attentional subsystem and an orienting subsystem enable the network to self-stabilize its learning, without an external teacher, as the code becomes globally self-consistent. Category learning is thus determined by global contextual information in this system. The attentional subsystem learns bottom-up codes and top-down templates, or expectancies. The internal representations formed in this way stabilize themselves against recoding by matching the learned top-down templates against input patterns. This matching process detects structural pattern properties in addition to local feature matches. The top-down templates can also suppress noise in the input patterns, and can subliminally prime the network to anticipate a set of input patterns. Mismatches activate an orienting subsystem, which resets incorrect codes and drives a rapid search for new or more appropriate codes. As the learned code becomes globally self-consistent, the orienting subsystem is automatically disengaged and the memory consolidates. After the recognition categories for a set of input patterns self-stabilize, those patterns directly access their categories without any search or recoding on future recognition trials. A novel pattern exemplar can directly access an established category if it shares invariant properties with the set of familiar exemplars of that category. Several attentional and nonspecific arousal mechanisms modulate the course of search and learning. Three types of attentional mechanism—priming, gain control, and vigilance—are distinguished. Three types of nonspecific arousal are also mechanistically characterized. The nonspecific vigilance process determines how fine the learned categories will be. If vigilance increases due, for example, to a negative reinforcement, then the system automatically searches for and learns finer recognition categories. The learned top-down expectancies become more abstract as the recognition categories become broader. The learned code is a property of network interactions and the entire history of input pattern presentations. The interactions generate emergent rules such as a Weber Law Rule, a 2/3 Rule, and

---

† Supported in part by the National Science Foundation (DMS-84-13119) and the Office of Naval Research (ONR N00014-83-K0337)

‡ Supported in part by the Air Force Office of Scientific Research (AFOSR 85-0149) and the Office of Naval Research (ONR N00014-83-K0337).

an Associative Decay Rule. No serial programs or algorithmic rule structures are used. The interactions explain and predict properties of evoked potentials (processing negativity, mismatch negativity, P300). Malfunction of the orienting system causes a formal amnesic syndrome analogous to that caused by malfunction of medial temporal brain structures: limited retrograde amnesia, long-range anterograde amnesia, failure of memory consolidation, effective priming, and defective reactions to novel cues. Comparisons with alternative theories of amnesia are made.

## 1. Introduction: Self-Organization of Recognition Categories

A fundamental problem of perception and learning concerns the characterization of how recognition categories emerge as a function of experience. When such categories spontaneously emerge through an individual's interaction with an environment, the processes are said to undergo *self-organization* (Basar, Flohr, Haken, and Mandell, 1983). This article develops a theory of how recognition categories can self-organize, and relates these results to recent data about evoked potentials and about amnesias due to malfunction of medial temporal brain structures. Results of evoked potential and clinical studies suggest which macroscopic brain structures could carry out the theoretical dynamics (Section 19). The theory also specifies microscopic neural dynamics, with local processes obeying membrane equations (Appendix).

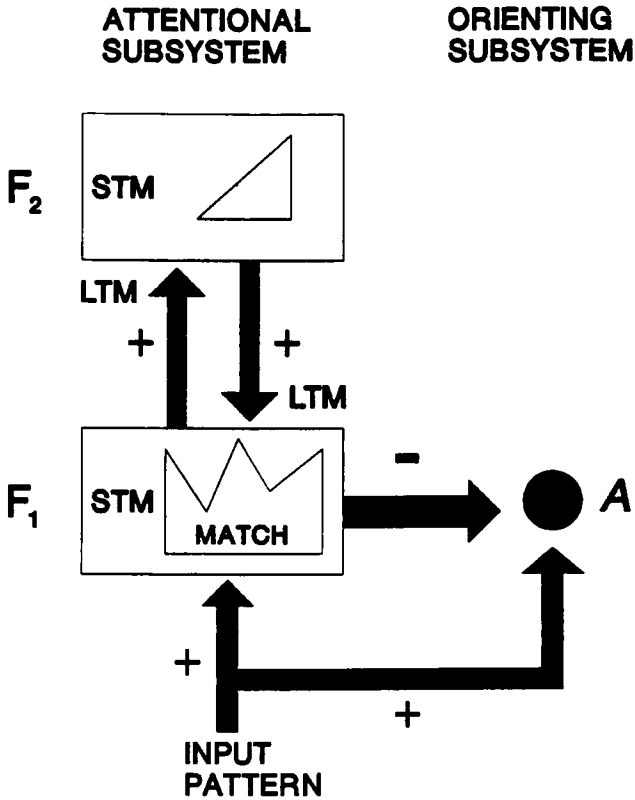
We focus herein upon principles and mechanisms that are capable of self-organizing stable recognition codes in response to arbitrary temporal sequences of input patterns. These principles and mechanisms lead to the design of a neural network whose parameters can be specialized for applications to particular problem domains, such as speech and vision. In these domains, preprocessing stages prepare environmental inputs for the self-organizing category formation and recognition system. Work on speech and language preprocessing has characterized those stages after which such a self-organizing recognition system can build up codes for phonemes, syllables, and words (Grossberg, 1978, 1985a; Grossberg and Stone, 1985). Work on form and color preprocessing has characterized those stages after which such a self-organizing recognition system can build up codes for visual object recognition (Grossberg and Mingolla, 1985a, 1985b).

### *Code Stabilization by Top-Down Expectancies*

Mathematical analysis and computer simulations of the neural network described in the present article show how the network can learn bottom-up codes and top-down expectancies in response to a temporal stream of input patterns. The internal representations formed in this way stabilize themselves against recoding in response to irrelevant input patterns by using the matching properties of the learned top-down expectancies. This code-stabilizing mechanism also suppresses noise in the input patterns, and can attentionally prime a network to anticipate an input pattern or category of input patterns. Moreover, the network automatically rescales its noise criterion to each pattern context: A particular mismatched feature which is processed as noise in a complex pattern with many features may, in the context of a simple pattern with few features, signal a pattern mismatch. Thus the theory shows that a definition of signal vs. noise which is sensitive to the global structure of input patterns is an intrinsic property of the mechanisms whereby recognition codes for these patterns are learned in a self-stabilizing fashion.

### *Attentional and Orienting Subsystems*

The class of networks that we consider develops the *adaptive resonance theory*. The theory's relationships to a wide variety of interdisciplinary data and other models is described in Grossberg (1976b, 1980, 1982, 1984a) and Grossberg and Stone (1985). In this theory, an interaction between two functionally complementary subsystems is needed to process familiar and unfamiliar events. Familiar events are processed within a consummatory, or attentional, subsystem. This subsystem establishes ever more precise internal representations of and responses to familiar events. It also builds up the

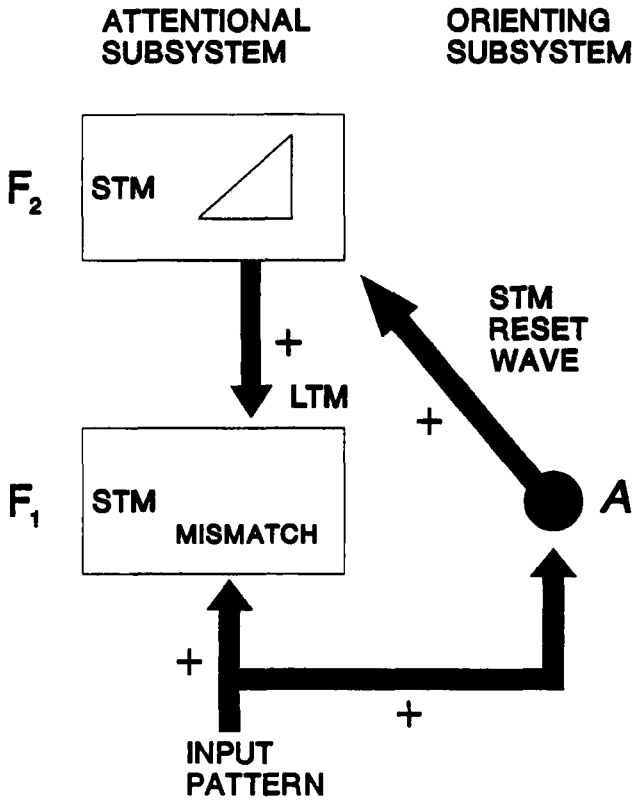


**Figure 1.** Interactions between the attentional subsystem and the orienting subsystem: Adaptive bottom-up signals and top-down signals between levels  $F_1$  and  $F_2$  determine whether the input pattern will be matched or mismatched at  $F_1$ . A match inhibits the orienting subsystem  $A$ .

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 for familiar patterns (Section 12).

The second subsystem is an orienting subsystem that overcomes the rigidity of the attentional subsystem when unfamiliar events occur and enables the attentional subsystem to learn from these novel experiences. The orienting subsystem is essential for expressing whether a novel pattern is “familiar” and well represented by an existing category, or “unfamiliar” and in need of a new category.

All input events start to be processed by the attentional subsystem. A familiar



**Figure 2.** A mismatch at  $F_1$  between the bottom-up input pattern and the top-down template, or expectancy, reduces inhibition from  $F_1$  to the orienting subsystem *A*. The orienting subsystem can then release a burst of nonspecific arousal capable of resetting short term memory (STM) at  $F_2$ .

event can activate a top-down template, or expectancy, which it tries to match within the attentional subsystem (Figure 1). A successful approximate match can deform, amplify, and sustain in short-term memory (STM) the activity pattern that was initially activated by the input within the attentional subsystem. Amplified, or resonant, STM activities constitute the fully elaborated recognition event. They inhibit the orienting subsystem and engage the learning, or long-term memory (LTM), process. A familiar event can maintain or modify its prior learning as its recognition takes place.

An unfamiliar event also starts to be processed by the attentional subsystem. Such an event may also activate a category which thereupon reads-out a top-down template. If the unfamiliar event can approximately match this template, then it can be recognized as an exemplar of the category on its first presentation. If the unfamiliar event is too different from familiar exemplars of the sampled category, then it cannot approximately

match this template (Figure 2). A mismatch within the attentional subsystem activates the orienting subsystem. Activation of the orienting subsystem functionally expresses the novelty, or unexpectedness, of the unfamiliar event. The orienting subsystem, in turn, rapidly resets the active representation within the attentional subsystem as it simultaneously energizes an orienting response.

The reset of the attentional subsystem by the orienting subsystem leads to the selection of a new representation within the attentional subsystem. This new representation may cause yet another mismatch, hence another STM reset event and the selection of yet another representation. In this way, the orienting subsystem mediates a rapid search which continues until a representation is found that does not cause a large mismatch. Then the search ends, an STM resonance develops, and the learning process can encode the active representation to which the search led. The system's recognition categories are hereby altered in either of two ways. If the search leads to an established category, then learning may change the criteria for accessing that category. If the search leads to uncommitted cells, then learning can add a new category to the recognition code.

This search process, although unfolding serially in time, is not controlled by a serial mechanism. Rather, it is driven by the successive release of nonspecific orienting bursts that are triggered by automatic processing of mismatch events. The entire history of learning determines the order of search in the network and, in turn, the new learning which can occur at the end of a search. Thus the search process adaptively modifies itself as the knowledge encoded by the network evolves. By contrast, a prewired search tree could not, in principle, maintain its efficiency after unpredictable changes in knowledge occurred. Instead, the novelty-sensitive orienting subsystem, through its interactions with the evolving knowledge of the attentional subsystem, defines an efficient, self-adjusting search routine.

#### *Tuning of Categories by Attention*

The criterion of mismatch is also determined by a parallel mechanism. In particular, a nonspecific vigilance, or attentional, parameter determines how fine the learned categories will be. If, for example, vigilance increases due to negative reinforcement or other attention-focusing agents, then the system will automatically search for and learn finer recognition categories.

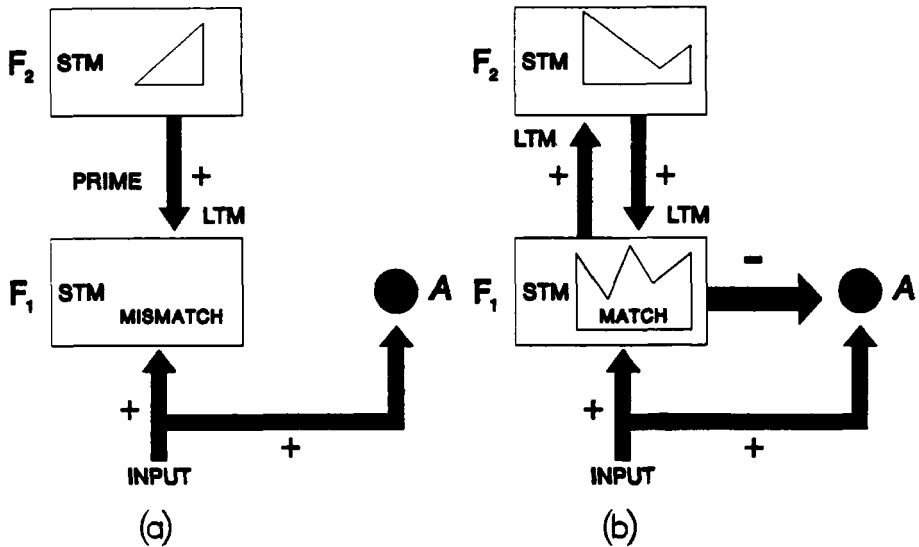
#### *Direct Access to Familiar Categories and Memory Consolidation*

Although an unfamiliar event may initially drive a search for an internal representation, after this representation is learned, future presentation of the input pattern need not engage the search process. Instead, the memory consolidates and a familiar input pattern can directly access its recognition category. That is, the familiar pattern can directly activate its code with neither search nor recoding.

#### *Top-Down Subliminal Priming*

A familiar event may, however, also engage the search process (Figure 3). This can occur when the system is primed to expect a different familiar event, so that a top-down expectancy is already active when the familiar event occurs. The familiar input event may mismatch this expectancy. A search will then be elicited leading to activation of the familiar event's bottom-up code and top-down expectancy. Such a search resets the erroneous code so that the correct code can be activated, but does not lead to learning of a new category. By contrast, if the system is primed to expect a familiar event that then occurs, a resonance can develop more rapidly than in an unprimed network. Consequently, anticipation of a familiar event can enhance recognition of that event by the network.

The model's flexible and dynamic relationship between matching, orienting, attention, and learning proves its worth by enabling efficient learning and self-stabilization of recognition categories with any prescribed refinement. The coarseness of the categories is not prewired. Nor is an identity match performed. In fact, the learned top-down expectancies become more abstract as the categories become broader. Moreover, the



**Figure 3.** Reset of a subliminal prime: (a) The top-down expectancy, or prime, subliminally activates  $F_1$  before the input pattern arrives. If the input pattern mismatches the prime, then an arousal burst from  $A$  can reset STM at  $F_2$  and thereby deactivate the prime. (b) Then  $F_1$  can access its correct  $F_2$  code. The subsequent match at  $F_1$  between the input pattern and a compatible top-down template prevents the input pattern from activating  $A$  and thereby erroneously resetting the correct  $F_2$  code.

network automatically rescales its matching criterion so that even with a fixed level of attentional vigilance, the network can both differentiate finer details of simple input patterns and tolerate larger mismatches of complex input patterns. This same rescaling property defines the difference between irrelevant noise and significant pattern mismatches. As with many other network properties, the rescaling property also emerges from interactions between the attentional subsystem and the orienting subsystem. If a mismatch within the attentional subsystem does not generate a search, then the mismatched features are treated as noise in the sense that they are eliminated from the critical feature pattern learned by the template. If the mismatch does generate a search, then the mismatched features may be included in the template of the category to which the search leads. Since the orienting subsystem is sensitive to the *relative* degree of match between an input pattern and a template, finer template mismatches with simple input patterns may drive a search, whereas larger mismatches with complex input patterns may not. Thus whole activity *patterns* across a field of feature-selective cells, rather than activations of single cells or feature detectors, are the computational units of the network.

#### *Short Term Memory and Long Term Memory*

Although the top-down expectancies, or templates, that are learned by the network are computed using deterministic laws, they support the recognition of categories whose

degree of fuzziness can be tuned by altering the level of vigilance. The coexistence of deterministic computations with fuzzy, or seemingly probabilistic, recognitions is made possible in the network through interactions between short term memory (STM) and long term memory (LTM) mechanisms. Using its fuzzy recognition criteria, the network can transform a continuum of possible input patterns into a discrete set of recognition categories.

Interaction of STM and LTM processes also enables the entire past learning experience of the network to influence each of its future recognition and learning events. Thus the apparently evanescent moment of recognition, or resonance, embodies all the knowledge that the network has accumulated to that time. Recognition in such a network is intrinsically context-sensitive.

#### *Reconciling Local Features and Context-Sensitive Interactions*

Using its context-sensitive interactions the network is able both to maintain stable internal representations against erosion by irrelevant environmental fluctuations and to learn rapidly in a new environment. Although local properties of feature detection are necessary for building up such internal representations, local properties alone are insufficient to distinguish between relevant and irrelevant environmental inputs. The network's ability to stabilize its learned codes against adventitious recoding is due to the same context-sensitive mechanisms that make every recognition event reflect the network's global history of learning.

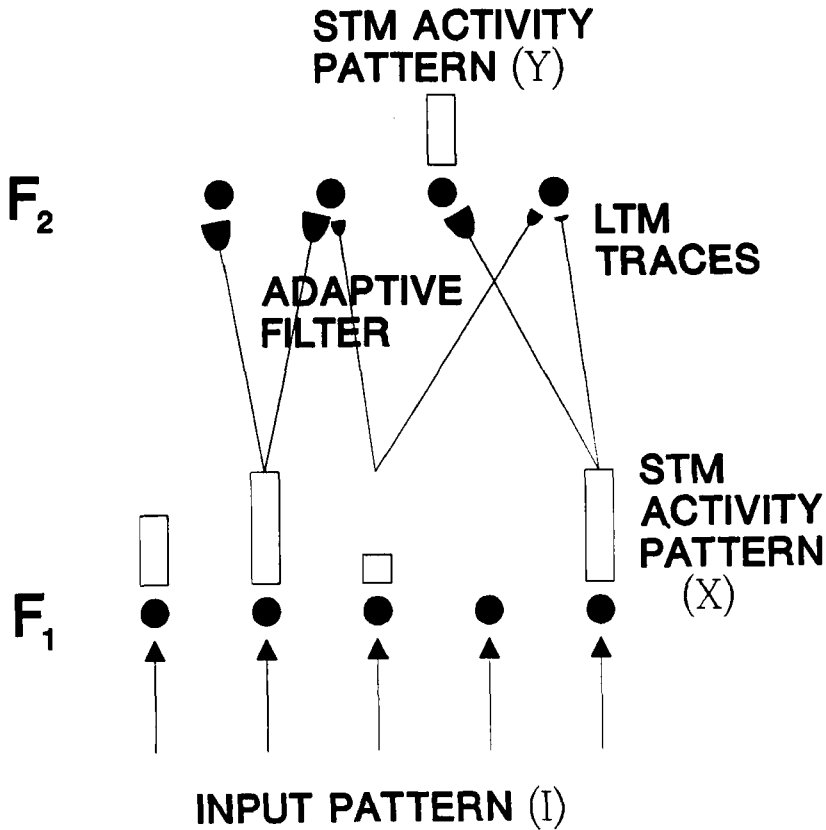
Thus we are led to consider how a single network can reconcile local features with global context-sensitivity, serial search with parallel processing, discrete categories with continuously varying events, deterministic computations with fuzzy sets, and stable memory with rapid learning.

## **2. Bottom-Up Adaptive Filtering and Contrast-Enhancement in Short Term Memory**

We now introduce in a qualitative way the main mechanisms of the theory. We do so 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. The input pattern  $I$  is received at the stage  $F_1$  of the attentional subsystem. Pattern  $I$  is transformed into a pattern  $X$  of activation across the nodes of  $F_1$  (Figure 4). 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 5a). The transformation from  $S$  to  $T$  is called an *adaptive filter*.

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, in which  $F_2$  chooses the node which receives the largest input, is here considered in detail. The chosen node is the only one that can store activity in STM. In more general versions of the theory, 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 many subsets, or groupings, of an input pattern (Cohen and Grossberg, 1985; Grossberg, 1985a). When  $F_2$  is designed to make a choice in STM, it selects that global

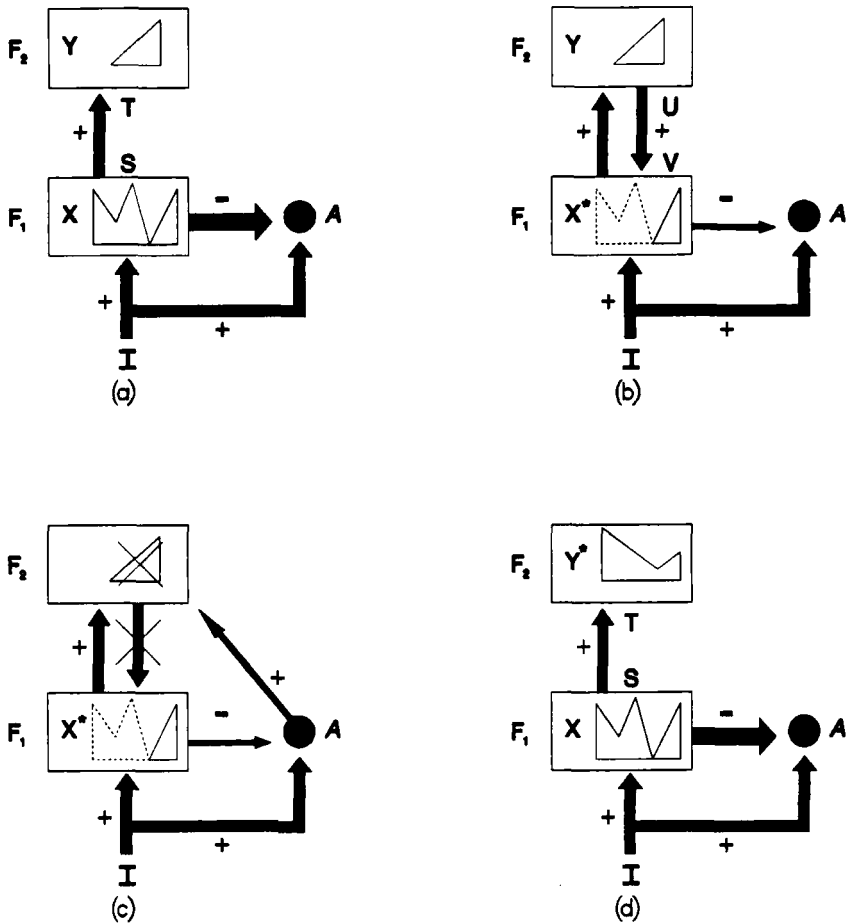


**Figure 4.** Stages of bottom-up activation: The input pattern  $I$  generates a pattern of STM activation  $X$  across  $F_1$ . Sufficiently active  $F_1$  nodes emit bottom-up signals to  $F_2$ . This signal pattern  $S$  is gated by long term memory (LTM) traces within the  $F_1 \rightarrow F_2$  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  $F_2$ .

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 in  $F_2$ . The present article analyses in detail the design of such a categorical mechanism. This special case is both interesting in itself and a necessary prelude to the analysis of recognition codes in which multiple groupings of  $X$  are simultaneously represented by  $Y$ .

Only those nodes of  $F_2$  which maintain stored activity in STM can elicit new learning





**Figure 5.** 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.

at contiguous LTM traces. Whereas all the LTM traces in the adaptive filter, and thus all learned past experiences of the network, are used to determine recognition via the transformation  $I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y$ , only those LTM traces whose STM activities in  $F_2$  survive the contrast-enhancement process can learn in response to the activity pattern  $X$ .

The bottom-up STM transformation  $I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y$  is not the only process that regulates network learning. In the absence of top-down processing, the LTM traces within the adaptive filter  $S \rightarrow T$  (Figure 5a) can respond to certain sequences of input patterns by being ceaselessly recoded in such a way that individual events are never eventually encoded by a single category no matter how many times they are presented. An infinite class of examples in which temporally unstable codes evolve is described in Section 12. It was the instability of bottom-up adaptive coding that led Grossberg (1976a, 1976b) to introduce the adaptive resonance theory.

In the adaptive resonance theory, a matching process at  $F_1$  exists whereby learned top-down expectancies, or templates, from  $F_2$  to  $F_1$  are compared with the bottom-up input pattern to  $F_1$ . This matching process stabilizes the learning that emerges in response to an arbitrary input environment. The constraints that follow from the need to stabilize learning enable us to choose among the many possible versions of top-down template matching and STM processes. These learning constraints upon the adaptive resonance top-down design have enabled the theory to explain data from visual and auditory information processing experiments in which learning has not been a manipulated variable (Grossberg, 1980, 1985a; Grossberg and Stone, 1985). The present article develops these mechanisms into a rigorously characterized learning system whose properties have been quantitatively analysed (Carpenter and Grossberg, 1985a, 1985b). This analysis has revealed new design constraints within the adaptive resonance theory. The system that we will describe for learned categorical recognition is one outcome of this analysis.

### 3. Top-Down Template Matching and Stabilization of Code Learning

We now begin to consider how top-down template matching can stabilize code learning. In order to do so, top-down template matching at  $F_1$  must be able to prevent learning at bottom-up LTM traces whose contiguous  $F_2$  nodes are only momentarily activated in STM. This ability depends upon the different rates at which STM activities and LTM traces can change. The STM transformation  $I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y$  takes place very quickly. By "very quickly" we mean much more quickly than the rate at which the LTM traces in the adaptive filter  $S \rightarrow T$  can change. 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  $F_1$ . 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 before 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 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

$$I \rightarrow X \rightarrow S \rightarrow T \rightarrow Y \rightarrow U \rightarrow V \rightarrow X^* \quad (1)$$

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. 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  $I$ .

#### 4. Interactions Between Attentional and Orienting Subsystems: STM Reset and Search

This section outlines how a mismatch at  $F_1$  regulates the learning process. With this general scheme in mind, we will be able to consider details of how bottom-up filters and top-down templates are learned and how matching takes place.

Level  $F_1$  can compute a match or mismatch between a bottom-up input pattern  $I$  and a top-down template pattern  $V$ , but it cannot compute which STM pattern  $Y$  across  $F_2$  generated the template pattern  $V$ . Thus the outcome of matching at  $F_1$  must have a nonspecific effect upon  $F_2$  that can potentially influence all of the  $F_2$  nodes, any one of which may have read-out  $V$ . The internal organization of  $F_2$  must be the agent whereby this nonspecific event, which we call a *reset wave*, selectively alters the stored STM activity pattern  $Y$ . The reset wave is one of the three types of nonspecific arousal that exist within the network. In particular, we suggest that a mismatch of  $I$  and  $V$  within  $F_1$  generates a nonspecific arousal burst that inhibits the active population in  $F_2$  which read-out  $V$ . In this way, an erroneous STM representation at  $F_2$  is quickly eliminated before any LTM traces can encode this error.

The attentional subsystem and the orienting subsystem work together to carry out these interactions. All learning takes place within the attentional subsystem. All matches and mismatches are computed within the attentional subsystem. The orienting subsystem is the source of the nonspecific arousal bursts that reset STM within level  $F_2$  of the attentional subsystem. The outcome of matching within  $F_1$  determines whether or not such an arousal burst will be generated by the orienting subsystem. Thus the orienting system mediates reset of  $F_2$  due to mismatches within  $F_1$ .

Figure 5 depicts a typical interaction between the attentional subsystem and the orienting subsystem. In Figure 5a, an input pattern  $I$  instates an STM activity pattern  $X$  across  $F_1$ . The input pattern  $I$  also excites the orienting population  $A$ , but pattern  $X$  at  $F_1$  inhibits  $A$  before it can generate an output signal.

Activity pattern  $X$  also generates an output pattern  $S$  which, via the bottom-up adaptive filter, instates an STM activity pattern  $Y$  across  $F_2$ . In Figure 5b, 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  $F_1$ , this activity no longer prevents the arousal source  $A$  from firing. Figure 5c depicts how disinhibition of  $A$  releases a nonspecific arousal burst to  $F_2$ . This arousal burst, in turn, selectively inhibits the active population in  $F_2$ . This inhibition is long-lasting. One physiological design for  $F_2$  processing which has these necessary properties is a *dipole field* (Grossberg, 1980, 1984a). A dipole field consists of opponent processing channels which are gated by habituating chemical transmitters. A nonspecific arousal burst induces selective and enduring inhibition within a dipole field. In Figure 5c, inhibition of  $Y$  leads to inhibition of the top-down template  $V$ , and thereby terminates the mismatch between  $I$  and  $V$ . Input pattern  $I$  can thus reinstate the 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 same pattern  $Y$  at  $F_2$ . A new pattern  $Y^*$  is thus generated at  $F_2$  by  $I$  (Figure 5d). 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, based upon on-center off-surround interactions among cells obeying membrane equations, is described in 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, the mechanisms which control the LTM search are all parallel network interactions, rather than serial algorithms. Such a parallel search scheme is necessary in a system whose LTM codes do not exist *a priori*. In general, the spatial configuration of codes in such a system depends upon both the system's initial configuration and its unique learning history. Consequently, no prewired serial algorithm could possibly anticipate an efficient order of search.

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

We now begin to consider details of the bottom-up/top-down matching process across  $F_1$ . The nature of this matching process is clarified by a consideration of how  $F_1$  distinguishes between activation by bottom-up inputs and top-down templates.

## 5. Attentional Gain Control and Attentional Priming

The importance of the distinction between bottom-up and top-down processing becomes evident when one observes that the same top-down template matching process which stabilizes learning is also a mechanism of attentional priming. Consider, for example, a situation in which  $F_2$  is activated by a level other than  $F_1$  before  $F_1$  is itself activated. In such a situation,  $F_2$  can generate a top-down template  $V$  to  $F_1$ . The level  $F_1$  is then primed, or ready, to receive a bottom-up input that may or may not match the active expectancy. Level  $F_1$  can be primed to receive a bottom-up input without necessarily eliciting suprathreshold output signals in response to the priming expectancy. If this were not possible, then every priming event would lead to suprathreshold consequences. Such a property would prevent subliminal anticipation of a future event.

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  $F_1$  (Figure 5). 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, an input pattern stimulates  $F_1$  cells. Some auxiliary mechanism must exist to distinguish between bottom-up and top-down inputs. We call this auxiliary mechanism *attentional gain control* to distinguish it from *attentional priming* by the top-down template itself. The attentional priming mechanism delivers *specific* template patterns to  $F_1$ . The attentional gain control mechanism has a *nonspecific* effect on the sensitivity with which  $F_1$  responds to the template pattern, as well as to other patterns received by  $F_1$ . Attentional gain control is one of the three types of nonspecific arousal that exist within the network. With the addition of attentional gain control, we can explain qualitatively how  $F_1$  can tell the difference between bottom-up and top-down signal patterns.

The need to dissociate attentional priming from attentional gain control can also be seen from the fact that top-down priming events do not lead necessarily to subliminal reactions at  $F_1$ . Under certain circumstances, top-down expectancies can lead to suprathreshold consequences. We can, for example, experience internal conversations or images at will. Thus there exists a difference between the read-out of a top-down template, which is a mechanism of attentional priming, and the translation of this operation into suprathreshold signals due to attentional gain control. An "act of will" can amplify attentional gain control signals to elicit a suprathreshold reaction at  $F_1$  in response to an attentional priming pattern from  $F_2$ .

Figures 6 and 7 depict two schemes whereby supraliminal reactions to bottom-up signals, subliminal reactions to top-down signals, and supraliminal reactions to matched bottom-up and top-down signals can be achieved. Figures 6d and 7d show, in addition, how competitive interactions across modalities can prevent  $F_1$  from generating a supraliminal reaction to bottom-up signals, as when attention shifts from one modality to another.

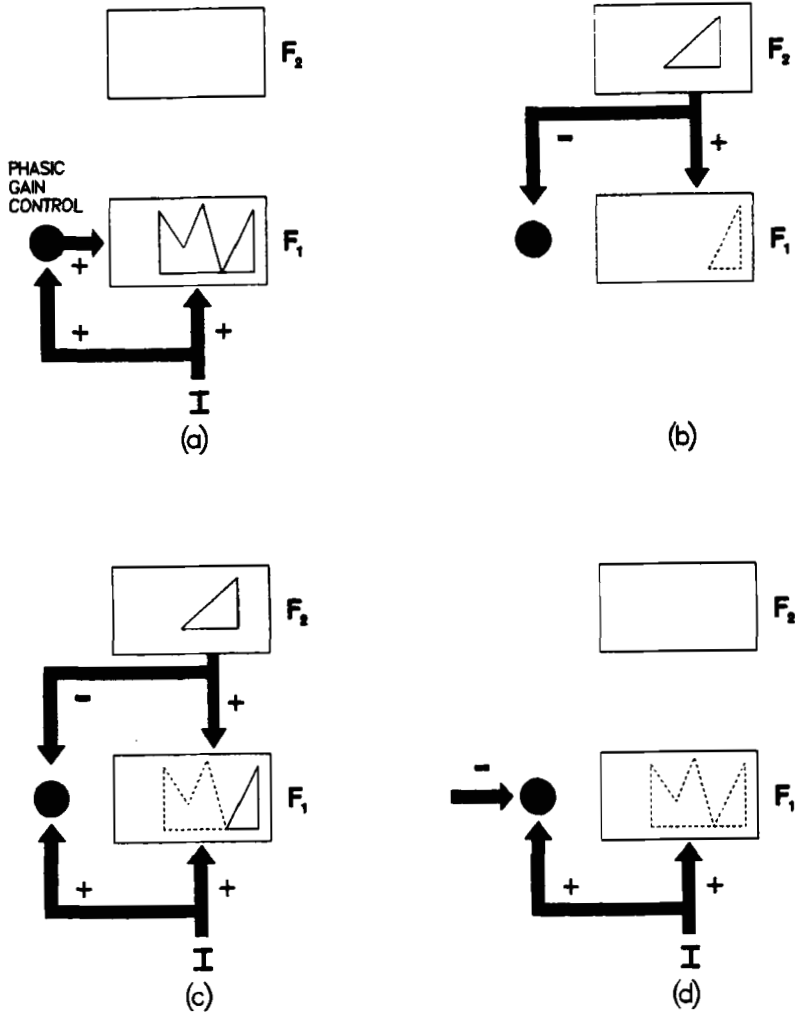
Both of the attentional gain control schemes in Figures 6 and 7 satisfy the same functional requirements. Both schemes are formally equivalent; that is, they obey the same system of differential equations. Both schemes can also explain the same body of psychological data. Each scheme can, for example, be used to clarify and modify the distinction between "automatic activation" and "conscious attention" that has arisen from psychological experiments on word recognition and related phenomena concerning human information processing (Grossberg and Stone, 1985). Physiological data are needed to choose one scheme over the other. In particular, within Figure 7, but not Figure 6, the bottom-up input pattern activates an attentional gain control channel. Thus in the scheme of Figure 6, bottom-up inputs activate two nonspecific processing channels, the attentional gain control channel within the attentional subsystem and the nonspecific arousal channel within the orienting subsystem. Herein, we will often motivate our formal constructions by considering the scheme in Figure 6, but it should not be forgotten that both schemes are formally, if not physiologically, equivalent.

## 6. Matching: The 2/3 Rule

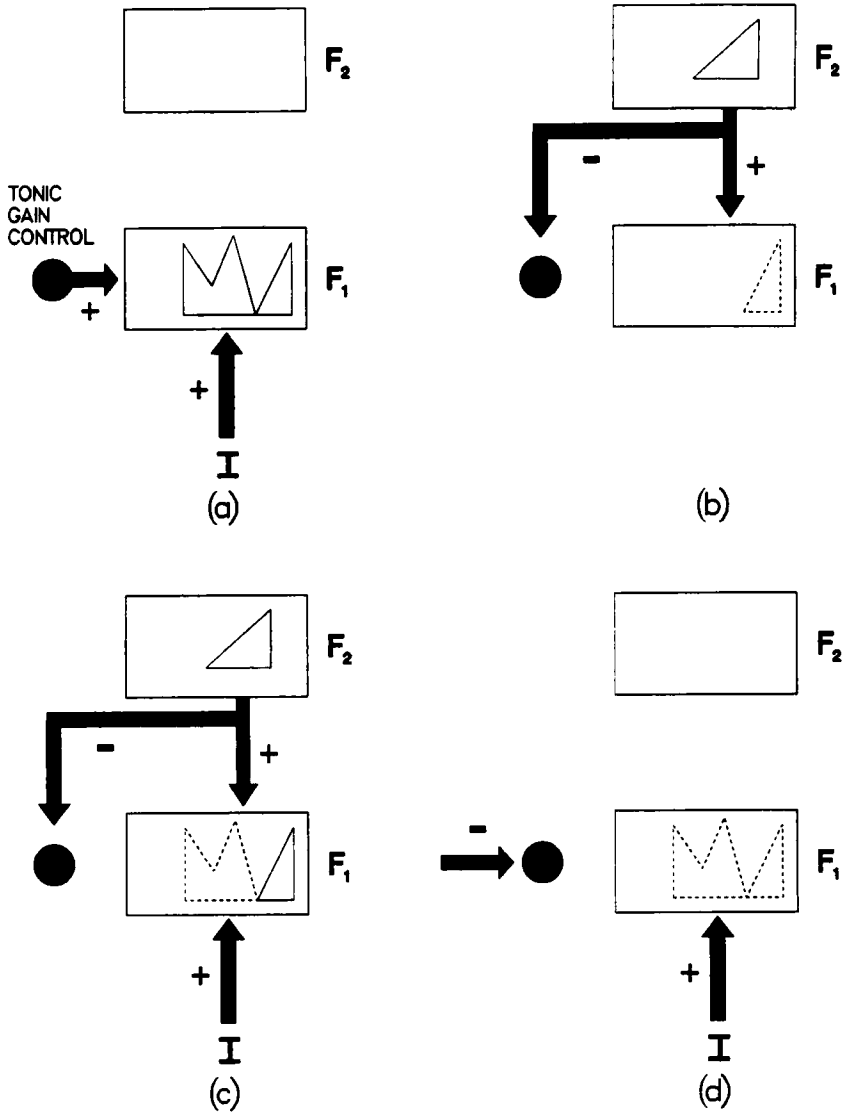
We can now outline the matching and coding properties that are used to generate learning of self-stabilizing recognition categories. Two different types of properties need to be articulated: the bottom-up coding properties which determine the order of search, and the top-down matching properties which determine whether an STM reset event will be elicited. Order of search is determined entirely by properties of the attentional subsystem. The choice between STM reset and STM resonance is dependent upon whether or not the orienting subsystem will generate a reset wave. This computation is based on inputs received by the orienting subsystem from both the bottom-up input pattern  $I$  and the STM pattern which  $F_1$  computes within the attentional subsystem (Figure 5). Both the order of search and the choice between reset and resonance are sensitive to the matched patterns *as a whole*. This global sensitivity is key to the design of a single system capable of matching patterns in which the number of coded features, or details, may vary greatly. Such global context-sensitivity is needed to determine whether a fixed amount of mismatch should be treated as functional noise, or as an event capable of eliciting search for a different category. For example, one or two details may be sufficient to differentiate two small but functionally distinct patterns, whereas the same details, embedded in a large, complex pattern may be quite irrelevant.

We first discuss the properties which determine the order of search. Network interactions which control search order can be described in terms of three rules: the 2/3 Rule, the Weber Law Rule, and the Associative Decay Rule.

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 6a,



**Figure 6.** Matching by 2/3 Rule: (a) In this example, nonspecific attentional gain control signals are phasically activated by the bottom-up input. In this network, the bottom-up input arouses two different nonspecific channels: the attentional gain control channel and the orienting subsystem. Only  $F_1$  cells that receive bottom-up inputs and gain control signals can become supraliminally active. (b) A top-down template from  $F_2$  inhibits the attentional gain control source as it subliminally primes target  $F_1$  cells. (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, since the gain control source is inhibited. (d) Intermodality inhibition can shut off the gain control source and thereby prevent a bottom-up input from supraliminally activating  $F_1$ .



**Figure 7.** This figure differs from Figure 6 only in that the attentional gain control source is tonically active.

for example, during bottom-up processing, a suprathreshold node in  $F_1$  is one which receives a specific input from the input pattern  $I$  and a nonspecific attentional gain control signal. All other nodes in  $F_1$  receive only the nonspecific gain control signal. Since these cells receive inputs from only one pathway they do not fire.

In Figure 6b, during top-down processing, or priming, some nodes in  $F_1$  receive a template signal from  $F_2$ , whereas other nodes receive no signal whatsoever. All 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 a top-down template.

During simultaneous bottom-up and top-down signaling, the attentional gain control signal is inhibited by the top-down channel (Figure 6c). Despite this fact, some nodes of  $F_1$  may receive sufficiently large inputs from both the bottom-up and the top-down signal patterns to generate suprathreshold outputs. Other nodes may receive inputs from the top-down template pattern or the bottom-up input pattern, but not both. These nodes receive signals from only one of their possible sources, hence do not fire. Cells which receive no inputs do not fire either. Thus only cells that are conjointly activated by the bottom-up input and the top-down template can fire when a top-down template is active. The 2/3 Rule clarifies the apparently paradoxical process whereby the addition of top-down excitatory inputs to  $F_1$  can lead to an overall decrease in  $F_1$ 's STM activity (Figures 5a and 5b).

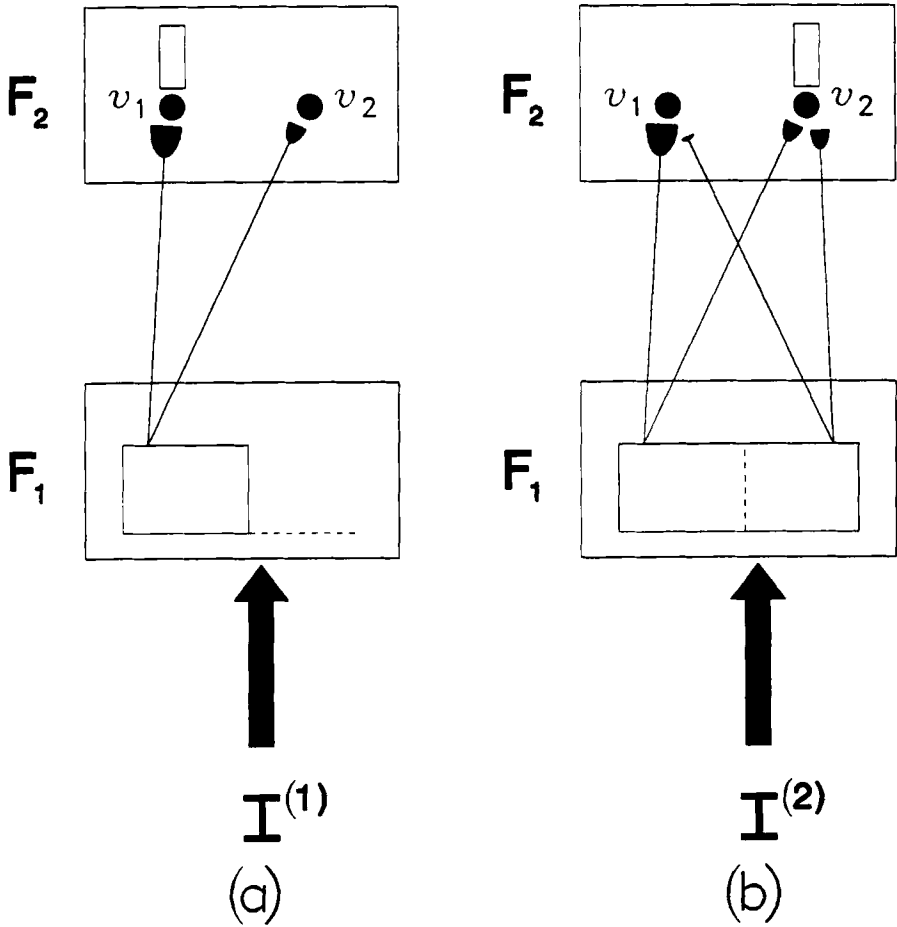
## 7. Direct Access To Subsets and Supersets

The Weber Law Rule can be motivated by considering the following situation. Suppose that a bottom-up input pattern  $I^{(1)}$  activates a network in which pattern  $I^{(1)}$  has already been perfectly coded by the adaptive filter from  $F_1$  to  $F_2$ . Suppose, moreover, that another pattern  $I^{(2)}$  has also been perfectly coded and that  $I^{(2)}$  contains  $I^{(1)}$  as a subset; that is,  $I^{(2)}$  equals  $I^{(1)}$  at all the nodes where  $I^{(1)}$  is positive. If  $I^{(1)}$  and  $I^{(2)}$  are sufficiently different, they should have access to distinct categories at  $F_2$ . However, since  $I^{(2)}$  equals  $I^{(1)}$  at their intersection, and since all the  $F_1$  nodes where  $I^{(2)}$  does not equal  $I^{(1)}$  are inactive when  $I^{(1)}$  is presented, how does the network decide between the two categories when  $I^{(1)}$  is presented? This question suggests that, in response to an input pattern  $I^{(1)}$  that is perfectly coded, the node  $v_1$  in  $F_2$  which codes  $I^{(1)}$  should receive a bigger signal from the adaptive filter than the node  $v_2$  in  $F_2$  which codes a superset  $I^{(2)}$  of  $I^{(1)}$  (Figure 8a). In order to realize this constraint, the LTM traces at  $v_2$  which filter  $I^{(1)}$  should be smaller than the LTM traces at  $v_1$  which filter  $I^{(1)}$ . Since the LTM traces at  $v_2$  were coded by the superset pattern  $I^{(2)}$ , this constraint suggests that larger input patterns are encoded by smaller LTM traces. Thus the absolute sizes of the LTM traces projecting to the different nodes  $v_1$  and  $v_2$  reflect the overall sizes of the input patterns  $I^{(1)}$  and  $I^{(2)}$  coded by these nodes.

The relative sizes of the LTM traces projecting to a single node reflect the internal structuring of the input patterns coded by that node. Consider, for example, the LTM traces in pathways between  $F_1$  cells where  $I^{(1)}$  equals zero and the  $F_2$  node  $v_1$  (Figure 8b). During learning of  $I^{(1)}$ , these LTM traces decay toward zero. By contrast, consider the LTM traces to  $v_2$  in pathways from  $F_1$  cells that are activated by  $I^{(2)}$  but not  $I^{(1)}$ . These LTM traces become large as learning of  $I^{(2)}$  proceeds.

The preceding discussion suggests a constraint that enables a subset  $I^{(1)}$  to selectively activate its node  $v_1$  rather than the node corresponding to a superset  $I^{(2)}$ . On the other hand, the superset  $I^{(2)}$  should be able to directly activate its node  $v_2$  rather than the node  $v_1$  of a subset  $I^{(1)}$ . However, the positive LTM traces of  $v_1$  are larger than the corresponding LTM traces of  $v_2$ , and presentation of  $I^{(2)}$  activates the entire





**Figure 8.** Weber law and associative decay rules for long term memory: When input  $I^{(1)}$  activates  $F_1$ , node  $v_1$  at  $F_2$  is chosen. When input  $I^{(2)}$  activates  $F_1$ , node  $v_2$  at  $F_2$  is chosen. (a) Because  $I^{(2)}$  is a superset of  $I^{(1)}$ , the LTM traces in pathways to  $v_1$  from  $F_1$  nodes that are activated by  $I^{(1)}$  are larger than the LTM traces to  $v_2$  in pathways from these same  $F_1$  nodes. (b) Consider  $F_1$  nodes that are activated by  $I^{(2)}$  but not  $I^{(1)}$ . The LTM traces in their pathways to  $v_1$  are small. In contrast, the LTM traces in their pathways to  $v_2$  are large, as are all the other LTM traces to  $v_2$  whose pathways are activated by  $I^{(2)}$ .

subset pattern  $I^{(1)}$ . The fact that  $I^{(2)}$  is filtered by more positive LTM traces at  $v_2$  than it is at  $v_1$  must be able to compensate for the larger size of the LTM traces at  $v_1$ . By establishing a proper balance between the size and the number of positive LTM traces, the Weber Law Rule allows both  $I^{(1)}$  and  $I^{(2)}$  to have direct access to their respective nodes  $v_1$  and  $v_2$ .

### 8. Weber Law Rule and Associative Decay Rule for Long Term Memory

We now describe more precisely the two learning rules whereby the LTM traces allow direct access to both subset and superset  $F_2$  codes. The conjoint action of a Weber Law Rule and an Associative Decay Rule for the learned sizes of LTM traces has the desired properties. To fix ideas, suppose that each input pattern  $I$  to  $F_1$  is a pattern of 0's and 1's. Let  $|I|$  denote the number of 1's in the input pattern  $I$ . The two rules can be summarized as follows.

#### Associative Decay Rule

After learning of  $I$  has taken place, LTM traces in the bottom-up coding pathways and the top-down template pathways between an inactive  $F_1$  node and an active  $F_2$  node equal 0, or at least are very small. Associative learning within the LTM traces can thus cause decreases as well as increases in the sizes of the traces. This is a non-Hebbian form of associative learning.

#### Weber Law Rule

After learning input pattern  $I$ , LTM traces in bottom-up coding pathways corresponding to active  $F_1$  and  $F_2$  nodes equal

$$\frac{\alpha}{\beta + |I|}. \quad (2)$$

By (2), the size of each positive LTM trace which codes  $I$  decreases as  $|I|$  increases.

Consider again the subset  $I^{(1)}$  and the superset  $I^{(2)}$ . By (2), the positive LTM traces which code  $I^{(1)}$  have size

$$\frac{\alpha}{\beta + |I^{(1)}|} \quad (3)$$

and the positive LTM traces which code  $I^{(2)}$  have size

$$\frac{\alpha}{\beta + |I^{(2)}|}, \quad (4)$$

where  $|I^{(1)}| < |I^{(2)}|$ . When  $I^{(1)}$  is presented at  $F_1$ ,  $|I^{(1)}|$  nodes in  $F_1$  are suprathreshold. Thus the *total* input to  $v_1$  has size

$$J_{11} = \frac{\alpha |I^{(1)}|}{\beta + |I^{(1)}|} \quad (5)$$

and the *total* input to  $v_2$  has size

$$J_{12} = \frac{\alpha |I^{(1)}|}{\beta + |I^{(2)}|}. \quad (6)$$

Because  $|I^{(1)}| < |I^{(2)}|$ , it follows that  $J_{11} > J_{12}$ . Thus  $I^{(1)}$  activates  $v_1$  instead of  $v_2$ .

When  $I^{(2)}$  is presented at  $F_1$ ,  $|I^{(2)}|$  nodes in  $F_1$  are suprathreshold. Thus the *total* input to  $v_2$  is

$$J_{22} = \frac{\alpha |I^{(2)}|}{\beta + |I^{(2)}|}. \quad (7)$$

We now invoke the Associative Decay Rule. Because  $I^{(2)}$  is a superset of  $I^{(1)}$ , only those  $F_1$  nodes in  $I^{(2)}$  that are also activated by  $I^{(1)}$  project to positive LTM traces at  $v_1$ . Thus the *total* input to  $v_1$  is

$$J_{21} = \frac{\alpha |I^{(1)}|}{\beta + |I^{(1)}|}. \quad (8)$$

Both  $J_{22}$  and  $J_{21}$  are expressed in terms of the function

$$W(\xi) = \frac{\alpha \xi}{\beta + \xi}, \quad (9)$$

which is an increasing function of  $\xi$ . Since  $|I^{(1)}| < |I^{(2)}|$ ,  $J_{22} > J_{21}$ . Thus the superset  $I^{(2)}$  activates its node  $v_2$  more than the subset node  $v_1$ .

Thus the conjoint action of a Weber Law Rule and an Associative Decay Rule for bottom-up learning permits direct access to the  $F_2$  nodes of both subset and superset input patterns. The Weber Law Rule is the outcome of mass action competitive interactions, as we will illustrate in the Appendix. These competitive interactions may occur among the nodes of  $F_1$  or among the LTM traces abutting each  $F_2$  node. We hereby suggest how the functional problem of direct access to subset and superset codes may be mechanistically solved by nonlinear neural interactions.

### 9. Fast Learning and Slow Learning: The Direct Access Rule

In order to characterize the course of learning, the rate of change of the LTM traces on each learning trial must be specified. In this article, we consider cases in which, on every learning trial, the LTM traces can reach the new equilibrium values imposed by the input pattern on that trial. We call these *fast learning* cases. We have also considered cases in which the LTM traces change too slowly to reach the new equilibrium values imposed by the input pattern on a single trial. We call these the *slow learning* cases.

During both *fast learning* and *slow learning*, the STM traces change more quickly than the LTM traces, and the learning process eventually self-stabilizes. However, the system is more sensitive to the ordering of the input patterns during *fast learning* than during *slow learning*. During *slow learning*, each LTM trace averages across time intervals that are much longer than a single trial, and thereby becomes less sensitive to the ordering of the inputs. In the next section, we will show how the input order can influence the choice of coding categories in the *fast learning* case. *Slow learning* is considered in Carpenter and Grossberg (1985b).

We note, finally, that the 2/3 Rule and the Weber Law Rule suggest how the initial values of STM traces and LTM traces should be chosen. The choice of initial STM traces is simple: the system starts out at equilibrium, or with zero STM traces, and the STM traces quickly return to equilibrium after each input pattern shuts off.

Initial LTM traces need to be chosen differently in the bottom-up adaptive filter than in the top-down adaptive filter. Due to the Weber Law Rule, the individual bottom-up LTM traces that are learned in response to large input patterns will be relatively small. In order for presentation of a perfectly coded large pattern to directly access its coded node, rather than an uncoded node, the initial values of the bottom-up LTM traces must be smaller than the learned LTM values corresponding to large input patterns. In addition, although some bottom-up LTM traces may initially equal zero, other LTM

traces abutting each  $F_2$  node must initially be positive in order for  $F_1$  to excite that node at all.

Due to the 2/3 Rule, the initial top-down LTM traces cannot be too small. When an input pattern first chooses an  $F_2$  node, the LTM traces that gate the top-down template of that node must satisfy the 2/3 Rule even before any template learning occurs. If the top-down LTM traces started out too small, no  $F_1$  node would receive enough top-down input to satisfy the 2/3 Rule. Consequently, the whole system would shut down. Top-down learning is thus a type of learning-by-selection.

In summary, bottom-up LTM traces start out small, whereas top-down LTM traces start out large. Bottom-up learning and top-down learning sculpt the spatial distribution of their LTM traces, as well as their overall sizes, through time. The constraint that the initial sizes of the top-down LTM traces be large is a consequence of the 2/3 Rule. The constraint that the initial sizes of the bottom-up LTM traces be small is needed to guarantee direct access to perfectly coded  $F_2$  nodes. We therefore call this latter constraint the Direct Access Rule.

### 10. Stable Choices in Short Term Memory

We can now begin to characterize the order of search in a network that obeys the following constraints: (1) Fast learning occurs (Section 9); (2) Input patterns are composed of 0's and 1's; (3) The 2/3 Rule holds (Section 6); (4) The Weber Law Rule holds (Section 8); (5) The Direct Access Rule holds (Section 9).

This discussion of search order does not analyse whether or not an STM reset event will stop the search at any given step. The criteria for STM reset are provided in Section 15. Other things being equal, a network with a higher level of vigilance will require better  $F_1$  matches, and hence will search more deeply, in response to each input pattern. Thus when an input pattern is presented, the set of learned filters and templates depends upon the prior levels of vigilance. The same ordering of input patterns may thus generate different LTM encodings due to the prior settings of the nonspecific vigilance parameter. The present discussion considers the order in which search will occur in response to a single input pattern which is presented after an arbitrary set of filters and templates has been learned.

A simple function determines the order in which encoded  $F_2$  nodes  $v_j$  are searched in response to an input pattern  $I$ . This function, which we call the Order Function, is defined as follows.

*Order Function*

$$T_j = \frac{\alpha |V^{(j)} \cap I|}{\beta + |V^{(j)}|}. \quad (10)$$

In equation (10),  $V^{(j)}$  denotes the top-down template pattern that is read-out by node  $v_j$  of  $F_2$ . Since only one node at a time is active in  $F_2$ , the total template read-out by  $F_2$  is the template corresponding to the node which is active at that time.

After  $I$  has been presented to  $F_1$ , but before  $F_2$  becomes active, function  $T_j$  in (10) is the total bottom-up input to node  $v_j$ . As in Section 8, term  $\alpha(\beta + |V^{(j)}|)^{-1}$  in (10) is a consequence of the Weber Law Rule. This term describes the size of the positive learned LTM traces which abut  $v_j$ . Term  $|V^{(j)} \cap I|$  describes the number of pathways abutting node  $v_j$  which have positive learned LTM traces and which carry positive signals when input  $I$  is presented. The total number of pathways abutting  $v_j$  which have positive learned LTM traces is  $|V^{(j)}|$ . This is true because a bottom-up LTM trace from node  $v_i$  in  $F_1$  to node  $v_j$  in  $F_2$  grows due to learning if and only if the corresponding top-down LTM trace from  $v_j$  to  $v_i$  grows due to learning. There are as

many positive learned LTM traces in pathways leading to  $v_j$  as there are in pathways leading from  $v_j$ . At times when input  $I$  is registered by  $F_1$ , only  $|V^{(j)} \cap I|$  of these  $|V^{(j)}|$  pathways are activated. The total input to node  $v_j$  in  $F_2$  is thus given by  $T_j$  in (10).

Level  $F_2$  chooses that node  $v_j$  which receives the largest input  $T_j$ . If we order the inputs in terms of decreasing size, as in

$$T_{j_1} > T_{j_2} > T_{j_3} > \dots, \quad (11)$$

then node  $v_{j_1}$  is initially chosen by  $F_2$ . After  $v_{j_1}$  is chosen, it reads-out template  $V^{(j_1)}$  to  $F_1$ . When  $V^{(j_1)}$  and  $I$  both perturb  $F_1$ , a new activity pattern  $X^*$  is registered at  $F_1$ , as in (1) and Figure 5b. A new bottom-up signal pattern from  $F_1$  to  $F_2$  may then be registered at  $F_2$ . How can we be sure that  $v_{j_1}$  will continue to receive the largest input from  $F_1$  after its template  $V^{(j_1)}$  is processed by  $F_1$ ? The 2/3 Rule provides this guarantee as follows.

The 2/3 Rule shuts off those active  $F_1$  nodes whose top-down LTM traces from  $v_{j_1}$  are zero due to prior learning of  $V^{(j_1)}$ . A top-down LTM trace becomes zero if and only if the corresponding bottom-up LTM trace becomes zero. Thus  $F_1$  nodes which are deactivated by the 2/3 Rule connect to bottom-up pathways whose LTM traces abutting  $v_{j_1}$  are zero. Hence, these pathways make no contribution to the total input  $T_{j_1}$  to node  $v_{j_1}$ . Thus the total input  $T_{j_1}$  is not altered due to read-out of the template  $V^{(j_1)}$ .

All other inputs  $T_j$  are either unchanged or decrease due to deactivation of some  $F_1$  nodes by the 2/3 Rule. In general, after the template  $V^{(j_1)}$  acts at  $F_1$ , the total input to node  $v_j$  at  $F_2$  is

$$\frac{\alpha |V^{(j)} \cap V^{(j_1)} \cap I|}{\beta + |V^{(j)}|}. \quad (12)$$

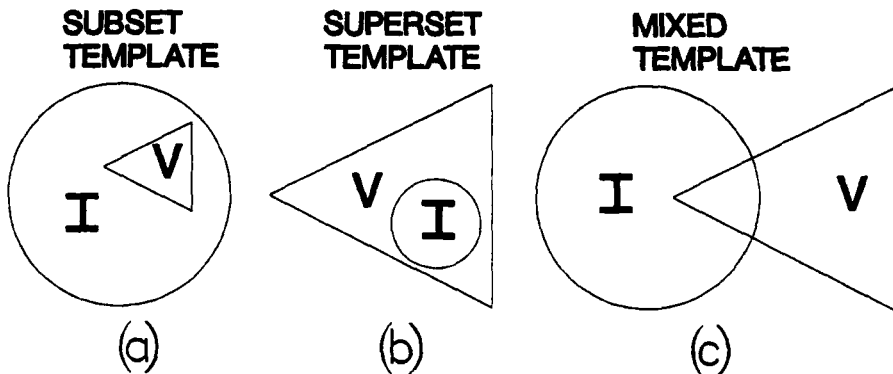
By (11),  $T_{j_1}$  was the maximal input to  $F_2$  before template  $V^{(j_1)}$  was read-out. By (10) and (12)  $T_{j_1}$  remains the maximal input to  $F_2$  after  $V^{(j_1)}$  is read-out. In summary, the 2/3 Rule stabilizes the STM choice at  $F_2$  before and after read-out of a top-down template.

Were the 2/3 Rule not operative, read-out of the template  $V^{(j_1)}$  might supraliminally activate many  $F_1$  nodes that had not previously been activated by the input  $I$  alone. These new  $F_1$  activations could cause a different  $F_2$  node to be chosen, and its template could cause yet another  $F_2$  node to be chosen. A rapid and non-terminating series of  $F_2$  choices could hereby be generated by an input  $I$ . Later  $F_2$  choices in this series could be activated by  $F_1$  nodes which receive no inputs whatsoever from  $I$ . The 2/3 Rule prevents this type of chaotic result from occurring. In other words, it instates a type of pattern matching within  $F_1$  which ensures that the choice of  $F_2$  nodes remains linked to the input pattern  $I$ .

## 11. Order of Search and the Subset Recoding Property

Because  $F_2$  can make choices which are not changed by read-out of the chosen node's template, the ordering of the bottom-up signals

$$T_j = \frac{\alpha |V^{(j)} \cap I|}{\beta + |V^{(j)}|} \quad (10)$$



**Figure 9.** Three types of relationships between input pattern  $I$  and template pattern  $V$ : (a) Subset template. (b) Superset template. (c) Mixed template.

by size, namely

$$T_{j_1} > T_{j_2} > T_{j_3} > \dots, \quad (11)$$

determines the order

$$v_{j_1}, v_{j_2}, v_{j_3}, \dots \quad (13)$$

of search. Thus simple algebraic computations enable one to predict the order of search in this network.

To discuss the order of search in response to the input pattern  $I$ , we define three types of learned templates: subset templates, superset templates, and mixed templates. The LTM traces of a subset template  $V$  are large only at a subset of the  $F_1$  nodes which are activated by the input pattern  $I$  (Figure 9a). The LTM traces of a superset template  $V$  are large at all the  $F_1$  nodes which are activated by the input pattern  $I$ , as well as at some  $F_1$  nodes which are not activated by  $I$  (Figure 9b). The LTM traces of a mixed template  $V$  are large at some, but not all, the  $F_1$  nodes which are activated by the input pattern  $I$ , as well as at some  $F_1$  nodes which are not activated by  $I$  (Figure 9c).

If a search ends when a prescribed template  $V^{(j)} = V$  is being read-out by the  $F_2$  node  $v_j$ , then this template's LTM traces recode to the new template  $V^{(j)} = V \cap I$ . This conclusion follows from the conjoint action of the 2/3 Rule and the Associative Decay Rule. Only  $F_1$  nodes in the set  $V \cap I$  can remain supraliminal due to the 2/3 Rule, and the LTM traces of pathways between  $v_j$  and inactive  $F_1$  nodes converge to zero due to the Associative Decay Rule. Thus, after learning occurs, the active template  $V^{(j)} = V$ , whether it began as a subset template, a superset template, or a mixed template, is recoded into the subset template  $V^{(j)} = V \cap I$  by the input pattern  $I$ . This subset recoding property is a key requirement for code stability.

## 12. Example of Code Instability

We now illustrate the importance of the subset recoding property by describing how its absence can lead to a temporally unstable code. In the simplest type of code instability example, the code becomes unstable because neither top-down template nor reset mechanisms exist (Grossberg, 1976a). Then, in response to certain input sequences that are repeated through time, a given input pattern can be ceaselessly recoded into more

than one category. In the example that we will now describe, the top-down template signals are active and the reset mechanism is functional. However, the inhibitory top-down attentional gain control signals (Figures 6c and 7c) are chosen too small for the 2/3 Rule to hold at  $F_1$ . We show also that a larger choice of attentional gain control signals restores code stability by reinstating the 2/3 Rule. These simulations also illustrate three other points: how a novel exemplar can directly access a previously established category; how the category in which a given exemplar is coded can be influenced by the categories which form to encode very different exemplars; and how the network responds to exemplars as coherent groupings of features, rather than to isolated feature matches or mismatches.

Figure 10a summarizes a computer simulation of unstable code learning. Figure 10b summarizes a computer simulation that illustrates how reinstatement of the 2/3 Rule can stabilize code learning. The format used in this figure will also be used in displaying our other computer simulations. We therefore describe this figure in detail.

The first column of Figure 10a describes the four input patterns that were used in the simulation. These input patterns are labeled A, B, C, and D. Patterns B, C, and D are all subsets of A. The relationships among the inputs that make the simulation work are as follows:

*Code Instability Example*

$$D \subset C \subset A, \quad (14)$$

$$B \subset A, \quad (15)$$

$$B \cap C = \phi, \quad (16)$$

$$|D| < |B| < |C|. \quad (17)$$

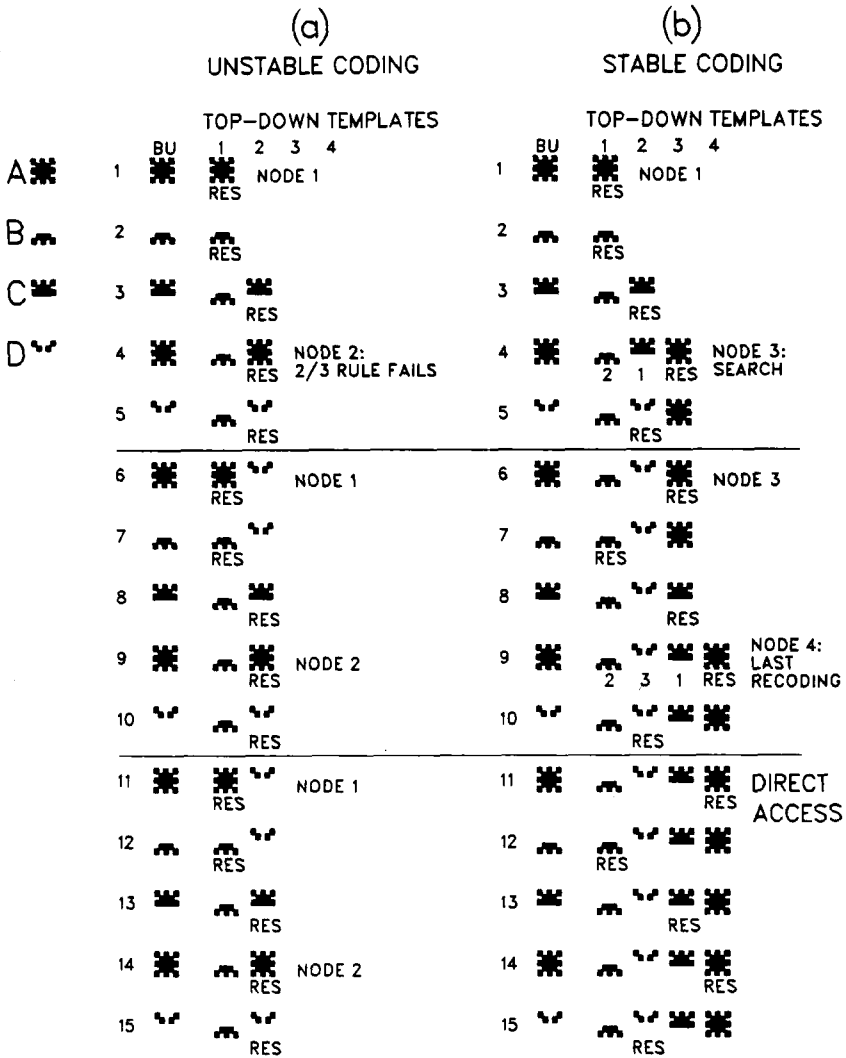
These results thus provide infinitely many examples in which an alphabet of just four input patterns cannot be stably coded without the 2/3 Rule. The numbers 1, 2, 3, ... listed in the second column itemize the presentation order. The third column, labeled BU for Bottom-Up, describes the input pattern that was presented on each trial. In both Figures 10a and 10b, the input patterns were periodically presented in the order ABCAD.

Each of the Top-Down Template columns in Figure 10 corresponds to a different node in  $F_2$ , with column 1 corresponding to node  $v_1$ , column 2 corresponding to node  $v_2$ , and so on. Each row summarizes the network response to its input pattern. The symbol RES, which stands for *resonance*, designates the node in  $F_2$  which codes the input pattern on that trial. For example,  $v_2$  codes pattern C on trial 3, and  $v_1$  codes pattern B on trial 7. The patterns in a given row describe the templates after learning has occurred on that trial.

In Figure 10a, input pattern A is periodically recoded: On trial 1, it is coded by  $v_1$ ; on trial 4, it is coded by  $v_2$ ; on trial 6, it is coded by  $v_1$ ; on trial 9, it is coded by  $v_2$ . This alternation in the nodes  $v_1$  and  $v_2$  which code pattern A repeats indefinitely.

Violation of the 2/3 Rule occurs on trials 4, 6, 8, 9, and so on. This violation is illustrated by comparing the template of  $v_2$  on trials 3 and 4. On trial 3, the template of  $v_2$  is coded by pattern C, which is a subset of pattern A. On trial 4, pattern A is presented and directly activates node  $v_2$ . Because the 2/3 Rule does not hold, pattern A remains supraliminal in  $F_1$  even after the subset template C is read-out from  $v_2$ . Thus no search is elicited by the mismatch of pattern A and its subset template C. Consequently the template of  $v_2$  is recoded from pattern C to its superset pattern A.

In Figure 10b, by contrast, the 2/3 Rule does hold due to a larger choice of the attentional gain control parameter. Thus the network experiences a sequence of recodings that ultimately stabilizes. In particular, on trial 4, node  $v_2$  reads-out the subset template C, which mismatches the input pattern A. The numbers beneath the template



**Figure 10.** Stabilization of categorical learning by the 2/3 Rule: In both (a) and (b), four input patterns A, B, C, and D are presented repeatedly in the list order ABCAD. In (a), the 2/3 Rule is violated because the top-down inhibitory gain control mechanism is weak (Figures 6c and 7c). Pattern A is periodically coded by  $v_1$  and  $v_2$ . It is never coded by a single stable category. In (b), the 2/3 Rule is restored by strengthening the top-down inhibitory gain control mechanism. After some initial recoding during the first two presentations of ABCAD, all patterns directly access distinct stable categories.



symbols in row 4 describe the order of search. First,  $v_2$ 's template C mismatches A. Then  $v_1$ 's template B mismatches A. Finally A activates the uncommitted node  $v_3$ , which resonates with  $F_1$  as it learns the template A.

Scanning the rows of Figure 10b, we see that pattern A is coded by  $v_1$  on trial 1; by  $v_3$  on trials 4 and 6; and by  $v_4$  on trial 9. On all future trials, input pattern A is coded by  $v_4$ . Moreover, all the input patterns A, B, C, and D have learned a stable code by trial 9. Thus the code self-stabilizes by the second run through the input list ABCAD. On trials 11 through 15, and on all future trials, each input pattern chooses a different node ( $A \rightarrow v_4$ ;  $B \rightarrow v_1$ ;  $C \rightarrow v_3$ ;  $D \rightarrow v_2$ ). Each pattern belongs to a separate category because the vigilance parameter was chosen to be large in this example. Moreover, as explained in Section 7, after code learning stabilizes, each input pattern directly activates its node in  $F_2$  without undergoing any additional search. Thus after trial 9, only the "RES" symbol appears under the top-down templates. The patterns shown in any row between 9 and 15 provide a complete description of the learned code. Examples of how a novel exemplar can activate a previously learned category are found on trials 2 and 5 in Figures 10a and 10b. On trial 2, for example, pattern B is presented for the first time and directly accesses the category coded by  $v_1$ , which was previously learned by pattern A on trial 1. In terminology from artificial intelligence, B activates the same categorical "pointer," or "marker," or "index" as in A. In so doing, B does not change the categorical "index," but it may change the categorical template, which determines which input patterns will also be coded by this index on future trials. The category does not change, but its invariants may change.

An example of how presentation of very different input patterns can influence the category of a fixed input pattern is found through consideration of trials 1, 4, and 9 in Figure 10b. These are the trials on which pattern A is recoded due to the intervening occurrence of other input patterns. On trial 1, pattern A is coded by  $v_1$ . On trial 4, A is recoded by  $v_3$  because pattern B has also been coded by  $v_1$  and pattern C has been coded by  $v_2$  in the interim. On trial 9, pattern A is recoded by  $v_4$  both because pattern C has been recoded by  $v_3$  and pattern D has been coded by  $v_2$  in the interim.

In all of these transitions, the global structure of the input pattern determines which  $F_2$  nodes will be activated, and global measures of pattern match at  $F_1$  determine whether these nodes will be reset or allowed to resonate in STM.

### 13. Search of Subsets, Supersets, and Mixed Sets

Before the code in Figure 10b finally stabilizes, it searches the network in the order characterized by (13). We now describe implications of this search order in a case of special interest, which includes the example described in Figure 10b. This is the case wherein parameter  $\beta$  in (10) is "small." By small, we mean that parameter  $\beta$  satisfies the inequality

$$\beta < \frac{1}{|I|_{\max} - 1}, \quad (18)$$

where  $|I|_{\max}$  is the largest number of  $F_1$  nodes that are activated by any input pattern I. The following assertions are proved in Carpenter and Grossberg (1985b).

#### A. Subset Templates

Suppose that there exist learned templates which are subsets of the input pattern I (Figure 9a). Then, if inequality (18) holds, the first node in  $F_2$  to be chosen corresponds to the largest subset template V. Whether or not template V can match the input I well enough to prevent STM reset of  $F_2$  depends upon the choice of the vigilance parameter, as well as upon how much smaller V is than I. If  $V = I$ , then reset never occurs. In this case, the Direct Access Rule (Section 9) implies that the node corresponding to V is chosen first. This node's template V covers  $\bar{I}$  at  $F_1$ . Consequently, no reduction in  $F_1$  activity is caused by the 2/3 Rule, and STM reset does not occur.

If the first chosen node does not cover  $I$ , then reset may occur. If reset does occur, then the network continues to search  $F_2$  nodes which possess subset templates. Search order proceeds from larger to smaller subset templates. This search order follows from (10), (11), and (13), because, whenever  $V^{(j)} \subset I$ , then  $V^{(j)} \cap I = V^{(j)}$ , so that the order function  $T_j$  satisfies

$$T_j = \frac{\alpha |V^{(j)}|}{\beta + |V^{(j)}|}. \quad (19)$$

Thus the order in which subset templates are searched is determined by the relative sizes of  $|V^{(j)}|$  across all subset templates. Figure 10b illustrates these subset search properties. On trial 9, for example, in response to the input pattern A, the nodes corresponding to the subset templates C, B, and D are searched in order of decreasing template size, as in (17).

#### B. Superset Templates and No Mixed Templates

Suppose that the network has searched all learned subset templates corresponding to the input pattern  $I$ . We now consider the subsequent search order by breaking up the possibilities into several cases. In this section, we suppose that no mixed templates have been learned, but that at least one superset template has been learned.

Our main conclusion is that, if all subset templates have already been reset, then the system will code input  $I$  using the  $F_2$  node  $v_j$  with the smallest superset template  $V^{(j)} = V$ . Due to this coding event,  $V^{(j)}$  will be recoded to

$$V^{(j)} = V \cap I = I. \quad (20)$$

The network chooses the smallest superset template first because

$$T_j = \frac{\alpha |I|}{\beta + |V|} \quad (21)$$

whenever  $V \supset I$ . Thus the smallest of the superset templates generates the largest bottom-up input  $T_j$ . The network does not reset this choice because the superset template  $V$  completely covers the input pattern  $I$  at  $F_1$ . By the 2/3 Rule, the  $F_1$  activity pattern caused by  $I$  alone persists after the superset template takes effect. No reduction of  $F_1$  activity is caused by the superset template. Hence its  $F_2$  code is not reset by the orienting subsystem. Thus the same property which guarantees stable choices in STM (Section 10) also implies that search ends if it can reach the smallest superset template.

It remains to explain why subsets are searched before supersets, and why supersets are searched before uncommitted nodes.

Given a subset template  $V^{(i)}$  and a superset template  $V^{(j)}$  of the input pattern  $I$ ,

$$|V^{(i)}| \leq |I| < |V^{(j)}|, \quad (22)$$

$$T_i = \frac{\alpha |V^{(i)}|}{\beta + |V^{(i)}|}, \quad (23)$$

and

$$T_j = \frac{\alpha |I|}{\beta + |V^{(j)}|}. \quad (24)$$

It follows from (18), (22), (23), and (24) that

$$T_i > T_j, \quad (25)$$

and hence that subset templates are searched before superset templates. This property depends critically on the small choice of  $\beta$  in (18).

Nodes with superset templates are searched before uncommitted nodes due to the same property that guarantees direct access to perfectly coded nodes. In Section 9 we noted that initial bottom-up LTM values must be chosen small enough to permit direct access to nodes which perfectly code any input pattern. In particular,

$$z_0 < \frac{\alpha}{\beta + |V^{(j)}|}, \tag{26}$$

where  $z_0$  is the maximal size of any initial bottom-up LTM trace, and  $\alpha(\beta + |V^{(j)}|)^{-1}$  is the learned LTM value corresponding to the superset template  $V^{(j)}$ . The total bottom-up input to an uncommitted node in response to input  $I$  is thus at most  $z_0 |I|$ , which is less than the total bottom-up input  $\alpha |I| (\beta + |V^{(j)}|)^{-1}$  to a superset node  $v_j$ .

C. Superset Templates and Mixed Templates

Suppose that the network has already searched its subset templates. Suppose also that both superset templates and mixed templates have previously been learned. Section 13B showed that, if a node with a superset template is activated, then the input pattern will be coded by that node. In particular, the node's template will be recoded to match the input pattern perfectly. We now characterize the circumstances under which the network will search mixed templates before it searches superset templates.

Consider nodes  $v_i$  which code mixed templates  $V^{(i)}$  with respect to the input pattern  $I$ . Also let  $V^{(J)}$  be the smallest superset template corresponding to  $I$ . Then

$$T_i = \frac{\alpha |V^{(i)} \cap I|}{\beta + |V^{(i)}|} \tag{27}$$

and

$$T_J = \frac{\alpha |I|}{\beta + |V^{(J)}|}. \tag{28}$$

A mixed template  $V^{(i)}$  will be searched before the superset template  $V^{(J)}$  if and only if

$$T_i > T_J. \tag{29}$$

When parameter  $B$  satisfies (18), inequality (29) holds if and only if

$$\frac{|V^{(i)} \cap I|}{|V^{(i)}|} > \frac{|I|}{|V^{(J)}|}. \tag{30}$$

This fact is proved in Carpenter and Grossberg (1985b).

Since a search always ends when a superset node is chosen, only nodes  $v_i$  whose mixed templates satisfy (30) can possibly be searched. These nodes are searched in order of decreasing  $|V^{(i)} \cap I| |V^{(i)}|^{-1}$ . If two nodes have the same ratio, then the one with the larger mixed template is searched first. If the search reaches the node  $v_j$  with the smallest superset template, then it terminates at  $v_j$ .

D. Mixed Templates But No Superset Templates

Suppose that the network has already searched its subset templates. Suppose that mixed templates, but no superset templates, have previously been learned. In this situation, the search can end by choosing either a node  $v_i$  with a mixed template  $V^{(i)}$

or a node which has not previously been chosen. For example, a node  $v_i$  with mixed template will be chosen before a new node if

$$\frac{\alpha |V^{(i)} \cap I|}{\beta + |V^{(i)}|} > z_0 |I|, \quad (31)$$

where  $z_0$  is the maximal initial size of the bottom-up LTM traces. Recall that

$$\frac{\alpha}{\beta + |V^{(i)}|} > z_0 \quad (26)$$

for all templates  $V^{(i)}$  in order to enable perfectly coded nodes to be directly accessed. Inequality (31) can thus hold when  $|V^{(i)} \cap I|$  is not too much smaller than  $|I|$ .

#### E. Neither Mixed Templates Nor Superset Templates

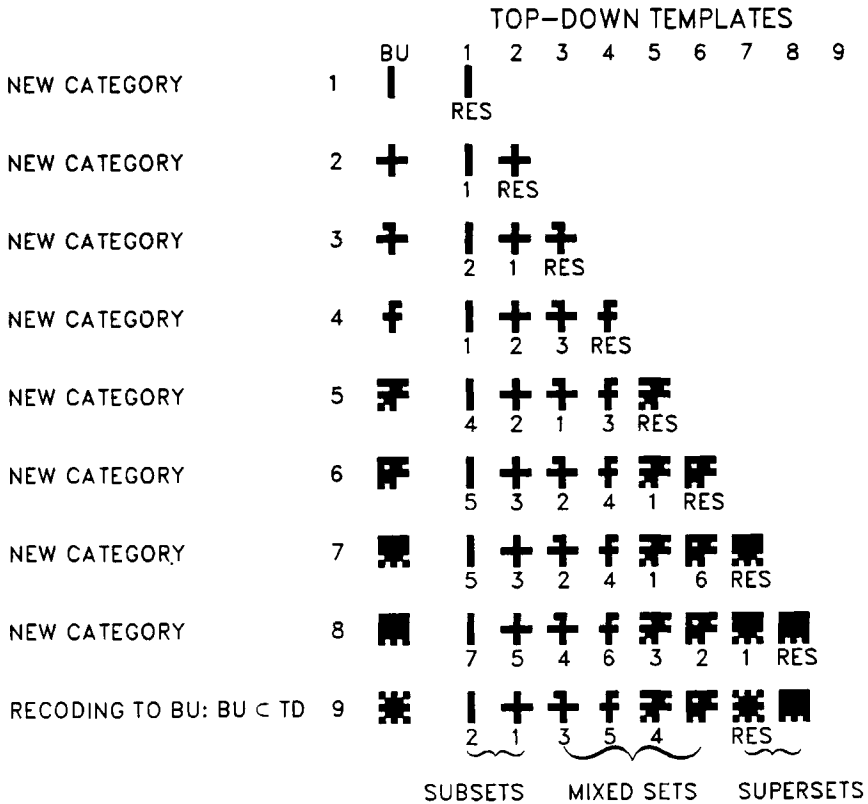
In this case, after all subset nodes are searched, the previously uncommitted nodes are searched. Their initial bottom-up input sizes to  $F_2$  depend upon the choice of initial LTM traces. Thus the order of search among the uncommitted nodes is determined by a random factor. The first uncommitted node that is activated ends the search and codes the input pattern  $I$ . This is true because all initial top-down LTM traces are chosen large enough to satisfy the 2/3 Rule (Section 9).

In case there are no uncommitted nodes to be searched after all committed nodes are rejected, then the input pattern cannot be coded by the network. This property is a consequence of the network's ability to buffer, or protect, its codes against persistent recoding by unappropriate events.

Figures 11 and 12 depict two coding sequences that illustrate the main points in the preceding discussion. In Figure 11, each of nine input patterns was presented once. We consider the order of search that occurred in response to the final input pattern  $I$  that was presented on trial 9. By trial 8, nodes  $v_1$  and  $v_2$  had already encoded subset templates of this input pattern. On trial 9, these nodes were therefore searched in order of decreasing template size. Nodes  $v_3$ ,  $v_4$ ,  $v_5$ , and  $v_6$  had encoded mixed templates of the input pattern. These nodes were searched in the order  $v_3 \rightarrow v_5 \rightarrow v_4$ . This search order was not determined by template size *per se*, but was rather governed by the ratio  $|V^{(i)} \cap I| |V^{(i)}|^{-1}$  in (30). These ratios for nodes  $v_3$ ,  $v_5$ , and  $v_4$  were  $9/10$ ,  $14/16$ , and  $7/8$ , respectively. Since  $14/16 = 7/8$ , node  $v_5$  was searched before node  $v_4$  because  $|V^{(5)}| = 16 > 8 = |V^{(4)}|$ . The mixed template node  $v_6$  was not searched. After searching  $v_5$ , the network activated the node  $v_7$  which possessed the smallest superset template. A comparison of rows 8 and 9 in column 7 shows how the superset template of  $v_7$  was recoded to match the input pattern. Node  $v_7$  was searched before node  $v_6$  because the ratio  $|I| |V^{(7)}|^{-1} = 17/21$  was larger than  $|V^{(6)} \cap I| |V^{(6)}|^{-1} = 14/18$ .

The eight input patterns of Figure 12 were chosen to illustrate a search followed by coding of an uncommitted node. The last input pattern  $I$  in Figure 12 was the same as the last input pattern in Figure 11. In Figure 12, however, there were no superset templates corresponding to input pattern  $I$ . Consequently  $I$  was coded by a previously uncommitted node  $v_8$  on trial 8. In particular, on trial 8, the network first searched the nodes with subset templates in the order  $v_2 \rightarrow v_1$ . Then the mixed template nodes were searched in the order  $v_4 \rightarrow v_6 \rightarrow v_5 \rightarrow v_7$ . The mixed template node  $v_3$  was not searched because its template badly mismatched the input pattern  $I$ . Instead, the uncommitted node  $v_8$  was activated and learned a template that matched the input pattern.

If parameter  $\beta$  is not small enough to satisfy inequality (18), then mixed templates or superset templates may be searched before subset templates. The order of search when  $\beta$  violates (18) is characterized in Carpenter and Grossberg (1985b). In all cases, direct access of a perfectly coded pattern is achieved.

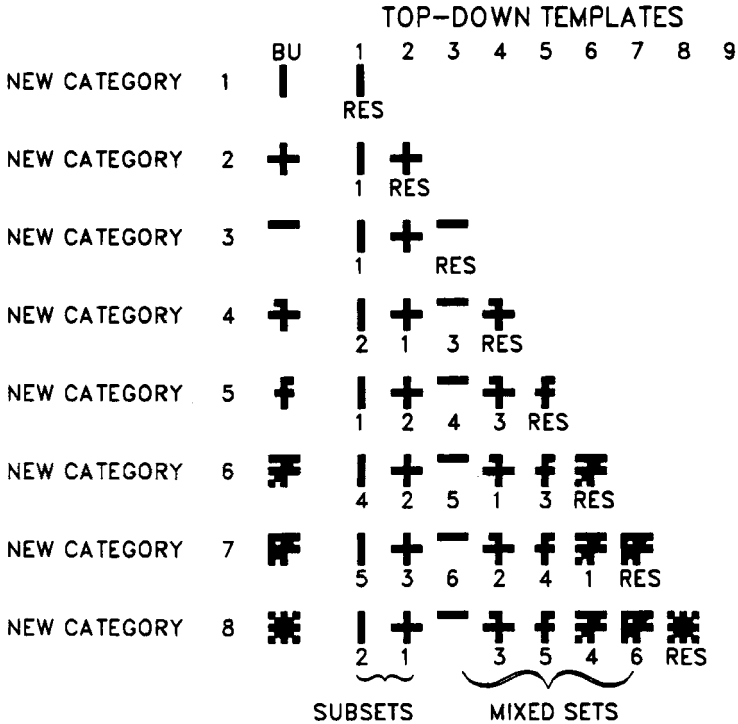


**Figure 11.** Computer simulation to illustrate order of search: On trial 9, the input pattern first searches subset templates, next searched some, but not all, mixed templates, and finally recodes the smallest superset template. A smaller choice of vigilance parameter could have terminated the search at a subset template or mixed set template node.

#### 14. The Nature of Categorical Invariance During Learning

The preceding discussion casts new light on the issue of how invariant properties of a category can persist even while new learning takes place. Two main cases need to be differentiated. In the first case, a novel input pattern is coded by a node whose bottom-up filter and top-down template have previously undergone learning. In the second case, a novel input pattern is coded by a previously unchosen node. Our remarks herein will focus on the first case.

In this case, presentation of the novel input pattern does not immediately change the number of categories that are coded by the network, nor the set of nodes which code these categories in STM at  $F_2$ . Output signals from  $F_2$  generate the network's observable responses. Hence, in this case, the novel pattern is assimilated into the



**Figure 12.** Computer simulation to illustrate order of search: On trial 8, the input pattern first searches subset templates and then searches some, but no all, mixed templates before choosing an uncommitted node, whose template learns the input pattern.

previously established set of categorical alternatives and observable responses. At least two different types of learning can accompany such an assimilation process: learning that is external to the categorical recognition process and learning that is internal to this process.

As an example of external learning, suppose that the novel input is associated with a different reinforcement schedule than previous inputs in the same category. New learning between the category in  $F_2$  and reinforcement mechanisms may alter the network's response to *all* the inputs in the category. Thus the very fact of membership in the same category may force forgetting of old external contingencies as new category exemplars are associated with new external contingencies.

As an example of internal learning, we consider the following facts. Even if a novel input pattern is coded by an "old"  $F_2$  node, this input pattern may alter the bottom-up filter and top-down template corresponding to that node. In so doing, the novel input pattern may alter the categorical boundaries of the network as a whole. Input patterns which were coded by prescribed nodes on previous trials may no longer be coded by the

same nodes when they are presented later on. Thus, even if the number of categories and their pathways to overt responses do not change, the categorical invariants may change.

The 2/3 Rule implies, however, that the filters and templates of a category are subsets of all the input patterns that are coded by that category. Adding a new input pattern to a category through learning can only refine further the filters and templates of the category. Thus, after a template becomes a subset of an input pattern by coding that pattern, the template remains a subset of the input pattern for all future time, no matter how many times the template is refined as other input patterns join the same category. As a template becomes progressively finer, the mismatch between the template and the largest input patterns coded by its category becomes progressively greater. If this mismatch becomes too great, then some of these large input patterns may eventually be recoded. For example, in Figure 10b, pattern B is coded by node  $v_1$  on trial 2, and no new categories are established. Later, however, when pattern A is next presented on trial 4, it can no longer adequately match the template from node  $v_1$ , as it did after trial 1. Hence pattern A establishes a new category.

Two main conclusions follow from these considerations. 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. For example, two different templates may overlap an input pattern to  $F_1$  at the same set of feature detectors, yet the network could reset the  $F_2$  node of one template yet not reset the  $F_2$  node of the other template. The degree of mismatch of template and input *as a whole* determines whether recoding 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, and hence base the code on sets of critical feature groupings. This article shows how these two opposing tendencies can be resolved, leading to dynamic equilibration, or self-stabilization, of recognition categories in response to a prescribed input environment.

The next section describes how a sufficiently large mismatch between an input pattern and a template can lead to STM reset, while a sufficiently good match can terminate the search and enable learning to occur.

## 15. Vigilance, Orienting, and Reset

We now show how matching within the attentional subsystem at  $F_1$  determines whether or not the orienting subsystem will be activated, thereby leading to reset of the attentional subsystem at  $F_2$ . The discussion can be broken into three parts:

### A. Distinguishing Active Mismatch from Passive Inactivity

A severe mismatch at  $F_1$  activates the orienting subsystem A. In the worst possible case of mismatch, none of the  $F_1$  nodes can satisfy the 2/3 Rule, and thus no supraliminal activation of  $F_1$  can occur. Thus in the worst case of mismatch, wherein  $F_1$  becomes totally inactive, the orienting subsystem must surely be engaged.

On the other hand,  $F_1$  may be inactive simply because no inputs whatsoever are being processed. In this case, activation of the orienting subsystem is not desired. How does the network compute the difference between active mismatch and passive inactivity at  $F_1$ ?

This question led Grossberg (1980) to assume that the bottom-up input source activates two parallel channels (Figure 5a). The attentional subsystem receives a specific input pattern at  $F_1$ . The orienting subsystem receives convergent inputs at A from all the active input pathways. Thus the orienting subsystem can be activated only when  $F_1$  is actively processing bottom-up inputs.

### B. Competition between the Attentional and Orienting Subsystems

How, then, is a bottom-up input prevented from resetting its own  $F_2$  code? What mechanism prevents the activation of  $A$  by the bottom-up input from *always* resetting the STM representation at  $F_2$ ? Clearly inhibitory pathways must exist from  $F_1$  to  $A$  (Figure 5a). When  $F_1$  is sufficiently active, it prevents the bottom-up input to  $A$  from generating a reset signal to  $F_2$ . When activity at  $F_1$  is attenuated due to mismatch, the orienting subsystem  $A$  is able to reset  $F_2$  (Figure 5b,c,d). In this way, the orienting subsystem can distinguish between active mismatch and passive inactivity at  $F_1$ .

Within this general framework, we now show how a finer analysis of network dynamics, with particular emphasis on the 2/3 Rule, leads to a vigilance mechanism capable of regulating how coarse the learned categories will be.

### C. Collapse of Bottom-Up Activation due to Template Mismatch

Suppose that a bottom-up input pattern has activated  $F_1$  and blocked activation of  $A$  (Figure 5a). Suppose, moreover, that  $F_1$  activates an  $F_2$  node which reads-out a template that badly mismatches the bottom-up input at  $F_1$  (Figure 5b). Due to the 2/3 Rule, many of the  $F_1$  nodes which were activated by the bottom-up input alone are suppressed by the top-down template. Suppose that this mismatch event causes a large collapse in the total activity across  $F_1$ , and thus a large reduction in the total inhibition which  $F_1$  delivers to  $A$ . If this reduction is sufficiently large, then the excitatory bottom-up input to  $A$  may succeed in generating a nonspecific reset signal from  $A$  to  $F_2$  (Figure 5c).

In order to characterize when a reset signal will occur, we make the following natural assumptions. Suppose that an input pattern  $I$  sends positive signals to  $|I|$  nodes of  $F_1$ . Since every active input pathway projects to  $A$ ,  $I$  generates a total input to  $A$  that is proportional to  $|I|$ . We suppose that  $A$  reacts linearly to the total input  $\gamma|I|$ . We also assume that each active  $F_1$  node generates an inhibitory signal of fixed size to  $A$ . Since every active  $F_1$  node projects to  $A$ , the total inhibitory input  $\delta|X|$  from  $F_1$  to  $A$  is proportional to the number  $|X|$  of active  $F_1$  nodes. When  $\gamma|I| > \delta|X|$ ,  $A$  receives a net excitatory signal and generates a nonspecific reset signal to  $F_2$  (Figure 5c).

In response to a bottom-up input pattern  $I$  of size  $|I|$ , as in Figure 5a, the total inhibitory input from  $F_1$  to  $A$  equals  $\delta|I|$ , so the net input to  $A$  equals  $(\gamma - \delta)|I|$ . In order to prevent  $A$  from firing in this case (Figure 5a), we assume that  $\delta \geq \gamma$ . We call

$$\rho = \frac{\gamma}{\delta} \quad (32)$$

the *vigilance parameter* of the orienting subsystem. The constraints  $\delta \geq \gamma \geq 0$  are equivalent to  $0 \leq \rho \leq 1$ . The size of  $\rho$  determines the proportion of the input pattern which must be matched in order to prevent reset.

When both a bottom-up input  $I$  and a top-down template  $V^{(j)}$  are simultaneously active (Figure 5b), the 2/3 Rule implies that the total inhibitory signal from  $F_1$  to  $A$  equals  $\delta|V^{(j)} \cap I|$ . In this case, the orienting subsystem is activated only if

$$\gamma|I| > \delta|V^{(j)} \cap I|; \quad (33)$$

that is, if

$$\frac{|V^{(j)} \cap I|}{|I|} < \rho. \quad (34)$$

The function which determines whether or not  $F_2$  will be reset in response to an input pattern  $I$  is called the Reset Function. Inequality (34) shows that the Reset Function should be defined as follows.



*Reset Function*

$$R_j = \frac{|V^{(j)} \cap I|}{|I|}. \quad (35)$$

The Reset Function  $R_j$  and the Order Function

$$T_j = \frac{\alpha |V^{(j)} \cap I|}{\beta + |V^{(j)}|} \quad (10)$$

determine how the search will proceed.

This line of argument can be intuitively recapitulated as follows. Due to the 2/3 Rule, a bad mismatch at  $F_1$  causes a large collapse of total  $F_1$  activity, which leads to activation of  $A$ . In order for this to happen, the system must maintain a measure of the prior level of total  $F_1$  activity and compare this criterion level with the collapsed level of total  $F_1$  activity. The criterion level is computed by summing bottom-up inputs at  $A$ . This sum can provide a criterion because it is proportional to the initial activation of  $F_1$  by the bottom-up input, and yet it remains unchanged as the matching process unfolds in real-time.

Figure 13 summarizes the total network architecture. It includes the modulatory processes, such as attentional gain control, which regulate matching within  $F_1$ , as well as the modulatory processes, such as orienting arousal, which regulate reset within  $F_2$ . Figure 13 also includes an attentional gain control process at  $F_2$ . Such a process enables offset of the input pattern to terminate all STM activity within the attentional subsystem in preparation for the next input pattern. In this example, STM storage can persist after the input pattern terminates only if an internally generated or intermodality input source maintains the activity of the attentional gain control system.

## 16. Distinguishing Signal from Noise in Patterns of Variable Complexity: Weighing the Evidence

A variety of important properties follow from the conception outlined in Section 15 of how the orienting system is engaged by mismatch within the attentional subsystem. These properties all address the fundamental issue of how a system can distinguish between signal and noise as it processes inputs of variable complexity.

We now indicate how the network automatically rescales its noise criterion as the complexity of the input pattern varies. In particular, even with fixed parameters, the network can tolerate larger mismatches in response to larger input patterns. Suppose, for example, that the network processes two input patterns at different times. One input pattern  $I^{(1)}$  activates just a few  $F_1$  feature detectors, whereas the other input pattern  $I^{(2)}$  activates many  $F_1$  feature detectors; that is,

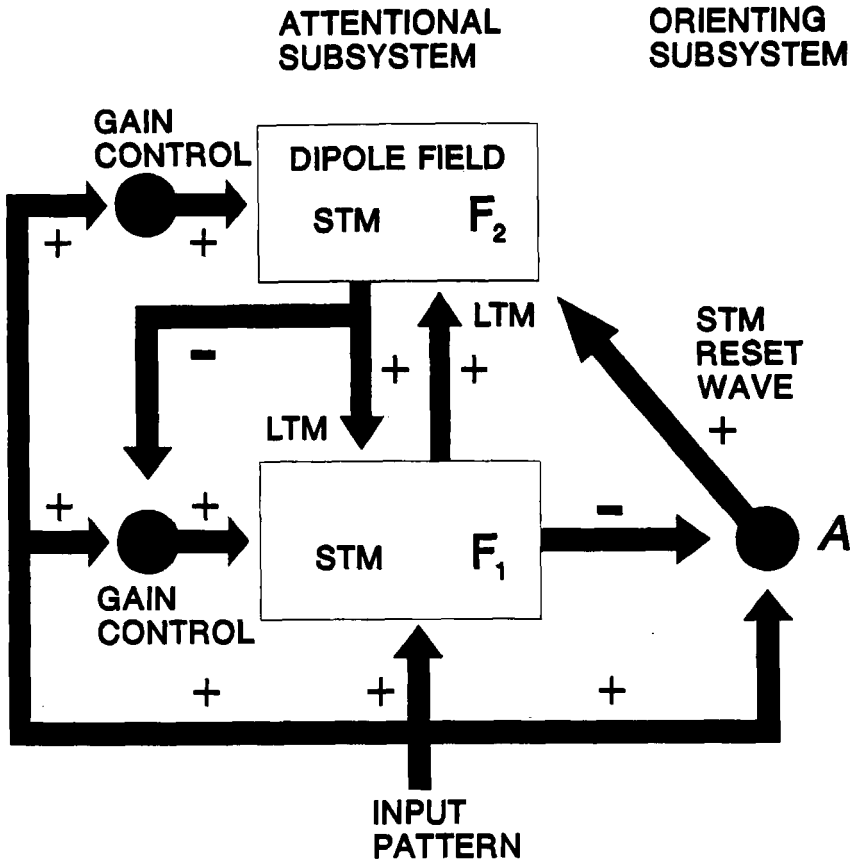
$$|I^{(1)}| < |I^{(2)}|. \quad (36)$$

Suppose, moreover, that  $I^{(1)}$  activates the  $F_2$  node  $v_1$ ,  $I^{(2)}$  activates the  $F_2$  node  $v_2$ , and that

$$|V^{(1)} \cap I^{(1)}| = |V^{(2)} \cap I^{(2)}|. \quad (37)$$

In other words, both input patterns overlap their templates by the same amount. Due to (36), however,

$$R_1 = \frac{|V^{(1)} \cap I^{(1)}|}{|I^{(1)}|} > \frac{|V^{(2)} \cap I^{(2)}|}{|I^{(2)}|} = R_2. \quad (38)$$



**Figure 13.** Anatomy of the attentional-orienting system: This figure describes all the interactions of the model without regard to which components are active at any given time.

By inequalities (34) and (38), the network is more likely to reset  $v_2$  in response to  $I^{(2)}$  than it is to reset  $v_1$  in response to  $I^{(1)}$ . Thus a fixed amount of match with a large input pattern provides less evidence for coding than the same amount of match with a small input pattern. If (37) holds, then the larger pattern  $I^{(2)}$  disagrees with the template at more features than does the smaller pattern  $I^{(1)}$ . Hence, by (38),  $v_2$  may be reset whereas  $v_1$  may not be reset; this will, in fact, be the case when  $\rho$  lies between  $R_1$  and  $R_2$ .

The rescaling property shows that the network processes input patterns as a whole. The functional units of the network are activation patterns across a field of feature detectors, rather than individual activations of feature detectors.

If the network does not reset  $v_1$  in response to  $I^{(1)}$ , then the template of  $v_1$  is refined to equal the intersection  $V^{(1)} \cap I^{(1)}$ . In other words, given that the network accepts the evidence that  $I^{(1)}$  should be coded by  $v_1$ , it then suppresses as noise the features at which  $I^{(1)}$  disagrees with  $V^{(1)}$ , both in STM and in LTM.

Using this property, the network can also distinguish finer differences between small input patterns than between large input patterns. Suppose that the amount of mismatch between a small input pattern  $I^{(1)}$  and its template  $V^{(1)}$  equals the amount of mismatch between a large input pattern  $I^{(2)}$  and its template  $V^{(2)}$ ; that is,

$$|I^{(1)} - |V^{(1)} \cap I^{(1)}|| = |I^{(2)} - |V^{(2)} \cap I^{(2)}||. \tag{39}$$

By (36) and (39),

$$R_1 = \frac{|V^{(1)} \cap I^{(1)}|}{|I^{(1)}|} < \frac{|V^{(2)} \cap I^{(2)}|}{|I^{(2)}|} = R_2. \tag{40}$$

Thus  $v_1$  is more likely to be reset by  $I^{(1)}$  than is  $v_2$  to be reset by  $I^{(2)}$ . This shows that a fixed amount of mismatch offers more evidence for reset when the input pattern is simple than when it is complex. Otherwise expressed, since the network is reset by smaller mismatches when processing smaller input patterns, it automatically makes finer distinctions between smaller input patterns than between larger input patterns.

The simulation in Figure 14 illustrates how the network automatically rescales its matching criterion. On the first four presentations, the patterns are presented in the order ABAB. By trial 2, coding is complete. Pattern A directly accesses node  $v_1$  on trial 3, and pattern B directly accesses node  $v_2$  on trial 4. Thus patterns A and B are coded within different categories. On trials 5–8, patterns C and D are presented in the order CDCD. Patterns C and D are constructed from patterns A and B, respectively, by adding identical upper halves to A and B. Thus, pattern C differs from pattern D at the same locations where pattern A differs from pattern B. However, because patterns C and D represent many more active features than patterns A and B, the difference between C and D is treated as noise, whereas the difference between A and B is considered significant. In particular, both patterns C and D are coded within the same category on trials 7 and 8.

The network's different categorization of patterns A and B vs. patterns C and D can be understood as follows. The core issue is: why on trial 2 does B reject the node  $v_1$  which has coded A, whereas D on trial 6 accepts the node  $v_3$  which has coded C? This occurs despite the fact that the mismatch between B and  $V^{(1)}$  equals the mismatch between D and  $V^{(3)}$ :

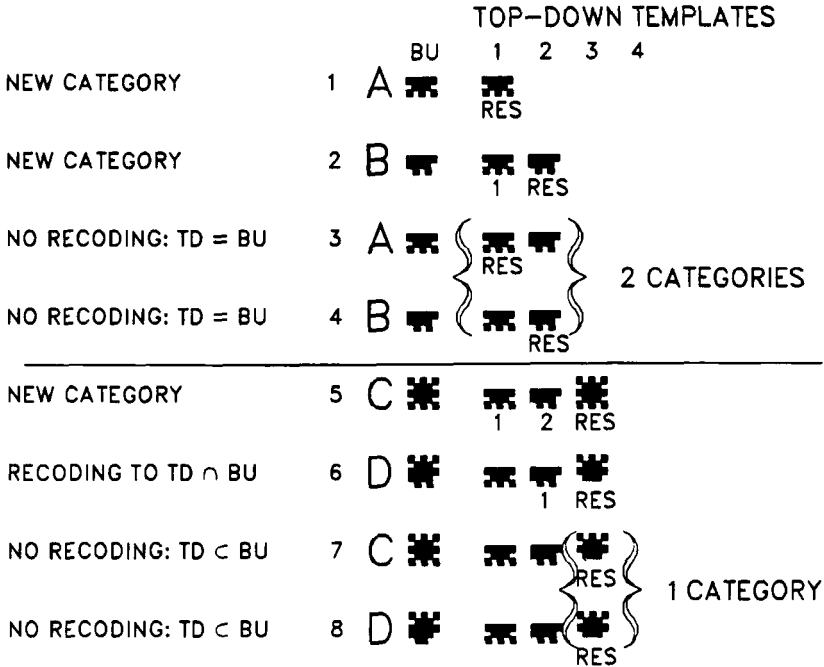
$$|B| - |V^{(1)} \cap B| = 3 = |D| - |V^{(3)} \cap D|, \tag{41}$$

as in equation (39). The reason can be seen by comparing the relevant reset functions:

$$R_{1B} = \frac{|V^{(1)} \cap B|}{|B|} = \frac{8}{11} \tag{42}$$

and

$$R_{3D} = \frac{|V^{(3)} \cap D|}{|D|} = \frac{14}{17}. \tag{43}$$



**Figure 14.** Distinguishing noise from patterns of inputs of variable complexity: Input patterns A and B are coded by the distinct category nodes  $v_1$  and  $v_2$ , respectively. Input patterns C and D include A and B as subsets, but also possess identical subpatterns of additional features. Due to this additional pattern complexity, C and D are coded by the same category node  $v_3$ . At this vigilance level ( $\rho = .8$ ), the network treats the difference between C and D as noise, and suppresses the discordant elements in the  $v_3$  template. By contrast, it treats the difference between A and B as informative, and codes the difference in the  $v_1$  and  $v_2$  templates, respectively.

In this simulation, the vigilance parameter  $\rho = .8$ . Thus

$$R_{1B} < \rho < R_{3D}. \tag{44}$$

By (34), pattern B resets  $v_1$  but D does not reset  $v_3$ . Consequently, B is coded by a different category than A, whereas D is coded by the same category as C.

**17. Vigilance Level Tunes Categorical Coarseness: Environmental Feedback**

The previous section showed how, given each fixed vigilance level, the network automatically rescales its sensitivity to patterns of variable complexity. The present section shows that changes in the vigilance level can regulate the coarseness of the categories that are learned in response to a fixed sequence of input patterns.

A low vigilance level leads to learning of coarse categories, whereas a high vigilance level leads to learning of fine categories. Suppose, for example, that a low vigilance level has led to a learned grouping of inputs which need to be distinguished for successful adaptation to a prescribed input environment. Suppose, moreover, that a punishing event occurs as a consequence of this erroneous grouping. Such a punishing event may have multiple effects on the organism. In addition to its negative reinforcing effects, we suppose that it also has a direct cognitive effect; namely, it increases attentive sensitivity to the environment. Such an increase in sensitivity is modeled within the network by an increase in the vigilance parameter,  $\rho$ . Increasing this single parameter enables the network to discriminate patterns which previously were lumped together. Once these patterns are coded by different categories in  $F_2$ , the different categories can be associated with different behavioral responses.

In this way, environmental feedback such as a punishing event can act as a "teacher" for a self-organizing recognition system. This teaching function does not take the form of an algorithm or any other type of pattern-specific information. Rather, it sets a single nonspecific parameter whose interaction with the internal organization of the network enables the network to parse more finely whatever input patterns happen to occur. The vigilance parameter will be increased, for example, if all the signals from the input pattern to  $A$  are nonspecifically amplified, so that parameter  $\gamma$  increases. A nonspecific decrease in the size of signals  $\delta$  from  $F_1$  to  $A$  will also increase  $\rho$ . Alternatively, reinforcement-activated nonspecific excitatory input to  $A$  can also facilitate mismatch-mediated activation of  $A$ . The process whereby the level of vigilance is monitored is one of the three types of nonspecific arousal that exist within the network.

Figure 15 describes a series of simulations in which four input patterns— $A$ ,  $B$ ,  $C$ ,  $D$ —are coded by a network with 4 nodes in  $F_2$ . In this simulation,  $A \subset B \subset C \subset D$ . The different parts of the figure show how categorical learning changes with changes of  $\rho$ . The simulation shows that any consecutive pair of patterns— $(A, B)$ ,  $(B, C)$ ,  $(C, D)$ —can be coded in the same category at different vigilance levels. When  $\rho = .8$  (Figure 15a), 4 categories are learned:  $(A)(B)(C)(D)$ . When  $\rho = .7$  (Figure 15b), 3 categories are learned:  $(A)(B)(C,D)$ . When  $\rho = .6$  (Figure 15c), 3 different categories are learned:  $(A)(B,C)(D)$ . When  $\rho = .5$  (Figure 15d), 2 categories are learned:  $(A,B)(C,D)$ . When  $\rho = .3$  (Figure 15e), 2 different categories are learned:  $(A,B,C)(D)$ . When  $\rho = .2$  (Figure 15f), all the patterns are lumped together into a single category.

## 18. Universal Recognition Design Across Modalities

The properties that we have demonstrated using illustrative simulations generalize to the coding of arbitrary sequences of input patterns. The ability to group arbitrary inputs is needed, we suggest, because the same mechanisms of grouping are used across modalities. Each modality, such as speech and vision, undergoes multiple stages of preprocessing through which different invariant properties of its environmental inputs are abstracted. These abstract representations then feed, as input patterns, into an attentional-orienting system. We suggest that the attentional-orienting system obeys the same processing rules across modalities. In this sense, the attentional-orienting system realizes a universal processing design.

In order to illustrate how such a network codifies a more complex series of patterns, we show in Figure 16 the first 20 trials of a simulation using alphabet letters as input patterns. In Figure 16a, the vigilance parameter  $\rho = .5$ . In Figure 16b,  $\rho = .8$ . Three properties are notable in these simulations. First, choosing a different vigilance parameter can determine different coding histories, such that higher vigilance induces coding into finer categories. Second, the network modifies its search order on each trial to reflect the cumulative effects of prior learning, and bypasses the orienting system to directly access categories after learning has taken place. Third, the templates of coarser categories tend to be more abstract because they must approximately match a larger number of input pattern exemplars.

$$A \subset B \subset C \subset D$$

$$| \quad + \quad * \quad \times$$

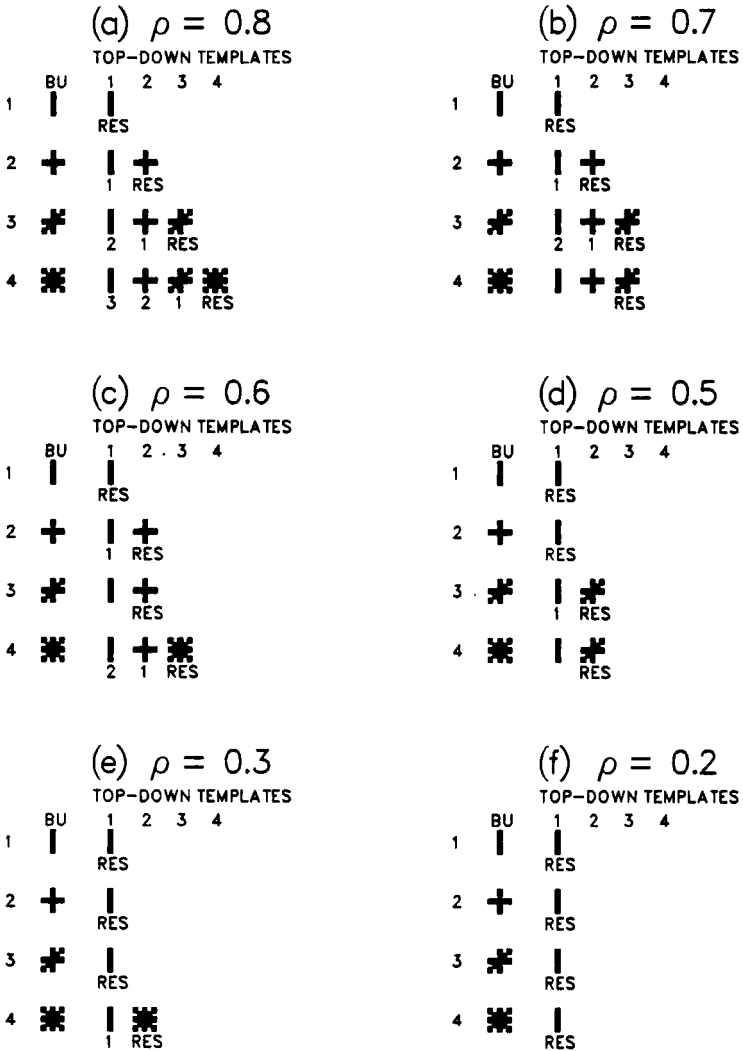


Figure 15. Influence of vigilance level on categorical groupings: As the vigilance parameter  $\rho$  decreases, the number of categories progressively decreases.

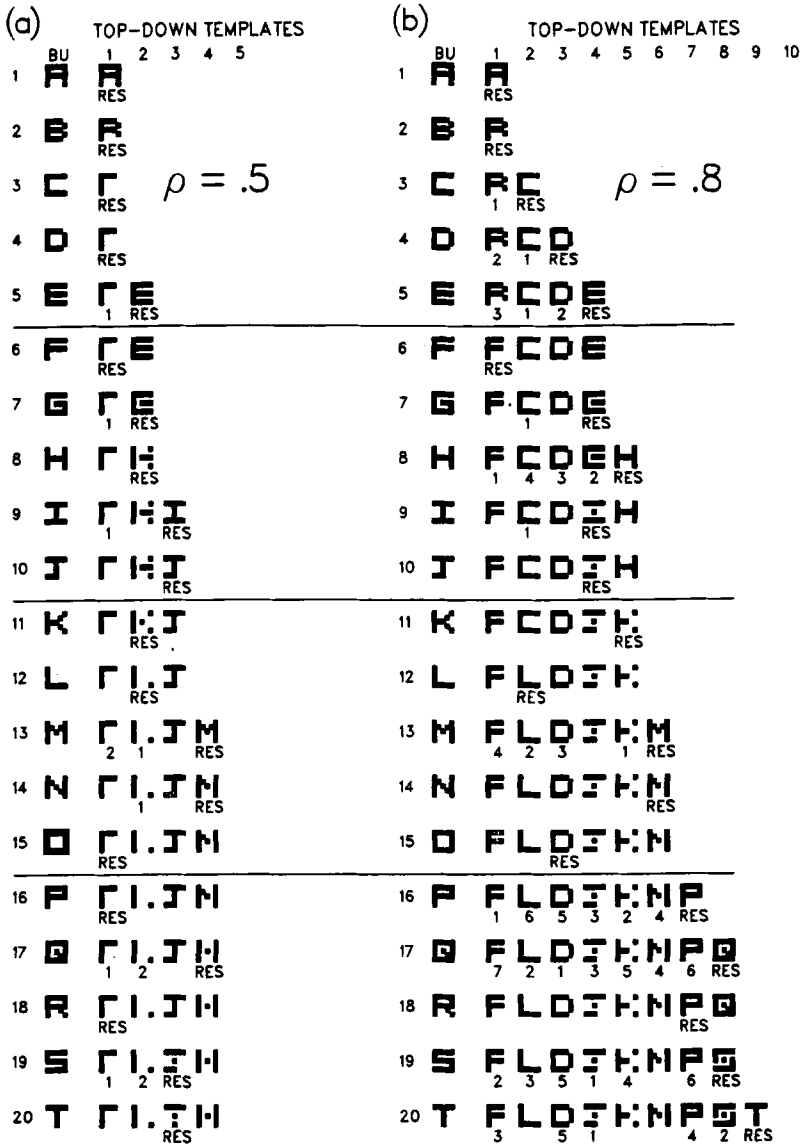


Figure 16. Alphabet learning: Different vigilance levels cause different numbers of letter categories to form.

Given  $\rho = .5$ , the network groups the 26 letter patterns into 8 stable categories within 3 presentations. In this simulation,  $F_2$  contains 15 nodes. Thus 7 nodes remain uncoded because the network self-stabilizes its learning after satisfying criteria of vigilance and global code self-consistency. Given  $\rho = .8$  and 15  $F_2$  nodes, the network groups 25 of the 26 letters into 15 stable categories within 3 presentations. The 26th letter is rejected by the network in order to self-stabilize its learning while satisfying its criteria of vigilance and global code self-consistency. These simulations show that the network's use of processing resources depends upon an evolving dynamical organization with globally context-sensitive properties. This class of networks is capable of organizing arbitrary sequences of arbitrarily complex input patterns into stable categories subject to the constraints of vigilance, global code self-consistency, and number of nodes in  $F_1$  and  $F_2$ . If slow learning rather than fast learning rates are used (Section 9), then the categorical code may be learned more slowly but it still enjoys the critical properties just listed.

### 19. Interdisciplinary Relationships: Word Recognition, Evoked Potentials, and Medial Temporal Amnesia

In this article, we have described the formal properties of a neural network which is capable of self-stabilizing its learning of recognition categories. The theory which this network develops arose from an analysis of several types of data, and is currently being refined through its use in explaining other types of data.

For example, the adaptive resonance theory anticipated the discovery of the processing negativity evoked potential and has successfully predicted several important properties of the processing negativity, mismatch negativity, and P300 evoked potentials. A review of these applications is found in Grossberg (1984a). This article is contained in a book (Karrer, Cohen, and Tueting, 1984) which includes detailed descriptions of relevant evoked potential data. The attentional-orienting network exhibits properties that are homologous to those of evoked potentials. In particular, the process whereby a top-down attentional prime is matched against a bottom-up input pattern at  $F_1$  may be compared with data about the *processing negativity* evoked potential. The process whereby the orienting subsystem is activated at  $A$  when a mismatch occurs may be compared with data about the *mismatch negativity* evoked potential. The process whereby STM is reset at  $F_2$  in response to an unexpected event may be compared with data about the *P300* evoked potential.

The bottom-up and top-down interactions within the attentional subsystem have also been used to explain and predict data about word recognition and recall in normal subjects (Grossberg, 1984b, 1985a; Grossberg and Stone, 1985). In these data analyses, concepts such as attentional gain control and attentional priming, which we have here related to code stabilization via the 2/3 Rule (Section 12), have enabled us to clarify and modify empirical models of "automatic activation" and "conscious attention" (Neely, 1977; Posner and Snyder, 1975a, 1975b).

Certain abnormal learning and recognition phenomena are strikingly similar to properties of a damaged attentional-orienting system. In considering this comparison, it is necessary to keep in mind that the attentional-orienting system is only one component in a larger neural theory of learning and memory. In particular, we do not herein extend this comparison to consider theoretical circuits for learned cognitive-motivational interactions, for serially ordered language utterances, or for sensory-motor coordination. Despite these limitations, it is of interest that injury to the orienting subsystem generates a type of amnesia that is reminiscent of amnesia in human patients, such as H.M., who have suffered injury to their medial temporal brain structures (Lynch, McGaugh, and Weinberger, 1984; Squire and Butters, 1984). In making this comparison, we will focus on issues relating to retrograde and anterograde amnesia, memory consolidation, impaired reactions to novel events, and differences between priming and recognition capabilities.



Suppose that the orienting subsystem ceases to function. Then the network cannot generate a search for new recognition categories. Consequently it cannot build up new recognition codes that would require a depthful search. On the other hand, well-established recognition categories can be directly accessed. Since they do not require intervention of the orienting subsystem, recognition codes which were established before the orienting subsystem failed are still accessible. Codes which were partially learned when the orienting subsystem failed may suffer variable degrees of impairment. Thus, failure of the orienting subsystem generates an amnesic syndrome with temporally limited retrograde amnesia and a temporally prolonged anterograde amnesia.

This amnesic syndrome is, in some respects, consistent with the following statement of Squire and Cohen (1984). "The medial temporal region establishes a relationship with distributed memory storage sites in neocortex and perhaps elsewhere; it then maintains the coherence of these ensembles until, as a result of consolidation, they can be maintained and can support retrieval on their own . . . the amnesic deficit is due to impaired consolidation" (p.45). In a normal attentional-orienting system, memory consolidation occurs as the system progresses from searching the attentional subsystem via the orienting subsystem to directly accessing its learned codes without engaging the orienting subsystem. During this consolidation process, the orienting subsystem is disengaged as unfamiliar environmental events gain familiarity by building learned recognition categories. The amnesic syndrome of the attentional-orienting subsystem is thus due to "impaired consolidation," in agreement with Squire and Cohen (1984). However, the orienting subsystem does not "maintain the coherence of these ensembles." Rather, when these ensembles become coherent and globally self-consistent, they disengage the orienting subsystem.

The role played by the orienting subsystem in driving a search for a globally self-consistent code coexists with its equally important role in enabling the network to react to the mismatches generated by unexpected and/or unfamiliar events. This latter role is the basis for calling this system the orienting subsystem (Grossberg, 1982, 1984a). The theory thus shows how memory consolidation and novelty detection can be mediated by the same structure, which is suggested to be a medial temporal brain structure such as hippocampus. This interpretation is consistent with data concerning the inability of hippocampotomized rats to orient to novel cues (O'Keefe and Nadel, 1978) and with the progressive reduction in novelty-related hippocampal potentials as learning proceeds (Deadwyler, West, and Lynch, 1979; Deadwyler, West, and Robinson, 1981). In summary, ablation of the orienting subsystem, and by interpretation medial temporal brain regions such as hippocampus, can interfere both with reactions to novel cues and with memory consolidation.

The attentional-orienting subsystem clarifies how normal priming and abnormal recognition can coexist in amnesia. In brief, the attentional priming mechanism may be intact even if the orienting subsystem is not working. An attentional prime can improve recognition by facilitating direct access to the correct learned category. These properties are consistent with data showing effective priming in amnesic patients (Cohen, 1984; Graf, Squire, and Mandler, 1984; Mattis and Kovner, 1984; Warrington and Weiskrantz, 1970, 1974).

The dynamics of the attentional-orienting system also shed new light on concepts about the properties of multiple memory systems (Lynch, McGaugh, and Weinberger, 1984; Squire and Butters, 1984). These memory systems have been given different names by different authors. Ryle (1949) distinguished "knowing that" from "knowing how"; Bruner (1969) discussed "memory with record" and "memory without record"; Mishkin (1982) analysed "memories" and "habits"; Squire and Cohen (1984) contrasted "declarative memory" and "procedural memory." The attentional-orienting system may be classified, at least qualitatively, as a "declarative memory" system because it governs "the storage of or access to memory ordinarily acquired during the learning experience" (Squire and Cohen, 1984, p.39). Recent theoretical progress has enabled such a learned

recognition system to be clearly distinguished, on the level of neural mechanism, from the learning systems which govern the acquisition of sensory-motor coordinations and plans (Grossberg, 1985a, 1985b; Grossberg and Kuperstein, 1985). These sensory-motor learning circuits provide examples of "procedural memory" systems.

When analysed on the level of mechanism, however, different types of memory systems cannot be neatly separated. For example, Squire and Cohen (1984) assume that attentional priming mechanisms form part of a procedural memory system *because* they are effective in amnesics whose recognition memory is impaired. In an attentional-orienting system, priming mechanisms form part of the attentional subsystem. The attentional subsystem, however, governs "the storage of or access to memory ordinarily acquired during the learning experience" (Squire and Cohen, 1984, p.39). Hence, by this criterion, attentional priming mechanisms should be included in a declarative memory system, not a procedural memory system. This difficulty reflects the general proposition that "procedures" cannot be separated from the contents, or "facts," that they manipulate, either in recognition systems or sensory-motor systems. This proposition is necessitated by the fact that the contents are learned, and thus the procedures must be defined interactively with respect to the evolving contents in order to be effective. In a sensory-motor system, the contents may not be "facts" that represent ordinary events. They may represent different types of information, such as terminal motor maps, or short term memory patterns of temporal order information over item representations (Cohen and Grossberg, 1985; Grossberg and Kuperstein, 1985). These contents define the "procedures" which govern how the sensory-motor systems will operate.

Another example of this interdependence can be seen in the attentional-orienting system. This system, on the level of mechanism, exhibits both "procedural" and "declarative" elements. Moreover, it is a defect of its procedures that leads to amnesia for its facts. The "procedures" of the attentional-orienting system are the search routines that are mediated by the orienting subsystem. The orienting subsystem cannot search except through its interactions with the attentional subsystem, in keeping with the goal of the search to preserve old "facts" while learning new "facts" within the attentional subsystem. Thus there can be no search programs—no independently definable procedures—within the orienting subsystem because the global organization of the codes being searched changes during learning.

In summary, the processing terms which have been chosen to emphasize the separateness of multiple memory systems—such as procedures and facts—become less clear-cut on the mechanistic level. Both types of process seem to exist in each memory system. This observation does not deny the basic fact that different memory systems react to environmental inputs in different ways, so that a patient may be able to learn a sensory-motor skill without being able to recognize a person's face. However, it does clarify how an amnesic can use familiar visual recognition codes as the inputs which trigger new learning within a sensory-motor system, without also generating new visual recognition codes within the very object recognition system which processes the visual signals. In other words, these results suggest how H.M. may use familiar "facts" to generate novel "procedures" without also learning to recognize the unfamiliar "facts" that are perceptually grouped in new ways during the "procedures."

We conclude with a prediction. If the data about evoked potentials and medial temporal amnesics both reflect a common level of neural processing, then the mismatch negativity and P300 evoked potential of medial temporal amnesic patients should be much more impaired than their processing negativity evoked potentials during attentional priming experiments, with the processing negativity tested in a match situation and the mismatch negativity and P300 tested in a mismatch situation.

## APPENDIX NETWORK EQUATIONS

### STM Equations

The STM activity of any node  $v_k$  in  $F_1$  or  $F_2$  obeys a membrane equation of the form

$$\frac{d}{dt}x_k = -Ax_k + (B - Cx_k)J_k^+ - Dx_kJ_k^-, \quad (A1)$$

where  $J_k^+$  and  $J_k^-$  are the total excitatory input and total inhibitory input, respectively, to  $v_k$  and  $A, B, C, D$  are nonnegative parameters. If  $C > 0$ , then the STM activity  $x_k(t)$  remains within the finite interval  $[0, BC^{-1}]$  no matter how large the inputs  $J_k^+$  and  $J_k^-$  are chosen.

We denote nodes in  $F_1$  by  $v_i$ , where  $i = 1, 2, \dots, M$ . We denote nodes in  $F_2$  by  $v_j$ , where  $j = M + 1, M + 2, \dots, N$ . Thus by (A1),

$$\frac{d}{dt}x_i = -A_1x_i + (B_1 - C_1x_i)J_i^+ - D_1x_iJ_i^- \quad (A2)$$

and

$$\frac{d}{dt}x_j = -A_2x_j + (B_2 - C_2x_j)J_j^+ - D_2x_jJ_j^-. \quad (A3)$$

The input  $J_i^+$  is a sum of the bottom-up input  $I_i$  and the top-down template

$$V_i = \sum_j f(x_j)z_{ji}, \quad (A4)$$

that is,

$$J_i^+ = I_i + V_i, \quad (A5)$$

where  $f(x_j)$  is the signal generated by activity  $x_j$  of  $v_j$ , and  $z_{ji}$  is the LTM trace in the pathway from  $v_j$  to  $v_i$ .

The inhibitory input  $J_i^-$  controls the attentional gain:

$$J_i^- = F \sum_j f(x_j). \quad (A6)$$

Thus  $J_i^- = 0$  if and only if  $F_2$  is inactive (Figures 6 and 7).

The inputs and parameters of STM activities in  $F_2$  were chosen so that the  $F_2$  node which received the largest input from  $F_1$  wins the competition for STM activity. Theorems in Ellias and Grossberg (1975), Grossberg (1973), and Grossberg and Levine (1975) show how these parameters can be chosen. The inputs  $J_j^+$  and  $J_j^-$  have the following form.

Input  $J_j^+$  adds a positive feedback signal  $g(x_j)$  from  $v_j$  to itself to the bottom-up adaptive filter input

$$T_j = \sum_i h(x_i)z_{ij}, \quad (A7)$$

that is,

$$J_j^+ = g(x_j) + T_j, \quad (A8)$$

where  $h(x_i)$  is the signal emitted by  $v_i$  and  $z_{ij}$  is the LTM trace in the pathway from  $v_i$  to  $v_j$ . Input  $J_j^-$  adds up negative feedback signals  $g(x_k)$  from all the other nodes in  $F_2$ :

$$J_j^- = \sum_{k \neq j} g(x_k). \quad (\text{A9})$$

Such a network behaves approximately like a binary switching circuit:

$$x_j = \begin{cases} G & \text{if } T_j > \max(T_k : k \neq j) \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A10})$$

### LTM Equations

The LTM trace of the bottom-up pathway from  $v_i$  to  $v_j$  obeys a learning equation of the form

$$\frac{d}{dt} z_{ij} = f(x_j) [-H_{ij} z_{ij} + K h(x_i)]. \quad (\text{A11})$$

In (A11), term  $f(x_j)$  is a postsynaptic sampling, or learning, signal because  $f(x_j) = 0$  implies  $\frac{d}{dt} z_{ij} = 0$ . Term  $f(x_j)$  is also the output signal of  $v_j$  to pathways from  $v_j$  to  $F_1$ , as in (A4).

The LTM trace of the top-down pathway from  $v_j$  to  $v_i$  also obeys a learning equation of the form

$$\frac{d}{dt} z_{ji} = f(x_j) [-H_{ji} z_{ji} + K h(x_i)]. \quad (\text{A12})$$

In the present simulations, the simplest choice of  $H_{ji}$  was made for the top-down LTM traces:

$$H_{ji} = H = \text{constant.} \quad (\text{A13})$$

A more complex choice of  $H_{ji}$  was made for the bottom-up LTM traces. This was done to directly generate the Weber Law Rule of Section 8 via the bottom-up LTM process itself. The Weber Law Rule can also be generated indirectly by exploiting a Weber Law property of competitive STM interactions across  $F_1$ . Such an indirect instantiation of the Weber Law Rule enjoys several advantages and will be developed elsewhere. In particular, it would enable us to also choose  $H_{ij} = H = \text{constant}$ . Instead, we allowed the bottom-up LTM traces at each node  $v_j$  to compete among themselves for synaptic sites. Malsburg and Willshaw (1981) have used a related idea in their model of retinotectal development. In the present usage, it was essential to choose a shunting competition to generate the Weber Law Rule, unlike the Malsburg and Willshaw usage. Thus we let

$$H_{ij} = Lh(x_i) + \sum_{k \neq i} h(x_k). \quad (\text{A14})$$

A physical interpretation of this choice can be seen by rewriting (A11) in the form

$$\frac{d}{dt} z_{ij} = f(x_j) [(K - Lz_{ij})h(x_i) - z_{ij} \sum_{k \neq i} h(x_k)]. \quad (\text{A15})$$

By (A15), when the postsynaptic signal  $f(x_j)$  is positive, a positive presynaptic signal  $h(x_i)$  commits receptor sites to the LTM process  $z_{ij}$  at a rate  $(K - Lz_{ij})h(x_i)f(x_j)$ . Simultaneously, signals  $h(x_k)$ ,  $k \neq i$ , which reach  $v_j$  at different regions of the  $v_j$

membrane compete for sites which are already committed to  $z_{ij}$  via the mass action competitive terms  $-z_{ij}f(x_j)h(x_k)$ . When  $z_{ij}$  equilibrates to these competing signals,

$$z_{ij} = \frac{K h(x_i)}{(L-1)h(x_i) + \sum_k h(x_k)}. \quad (A16)$$

The signal function  $h(w)$  was chosen to rise quickly from 0 to 1 at a threshold activity level  $w_0$ . Thus if  $v_i$  is a suprathreshold node in  $F_1$ , (A16) approximates

$$z_{ij} \cong \frac{K}{(L-1) + |X|} \quad (A17)$$

where  $|X|$  is the number of active nodes in  $F_1$ . Thus  $z_{ij}$  obeys a Weber Law Rule if  $L > 1$ . By comparison with (2),  $\alpha = K$  and  $\beta = L - 1$ .

### STM Reset System

The simplest possible mismatch-mediated activation of  $A$  and STM reset of  $F_2$  by  $A$  were implemented in the simulations. As outlined in Section 15, each active input pathway sends an excitatory signal of size  $\gamma$  to  $A$ . Potentials  $x_i$  of  $F_1$  which exceed a signal threshold  $T$  generate an inhibitory signal of size  $-\delta$  to  $A$ . Population  $A$ , in turn, generates a nonspecific reset wave to  $F_2$  whenever

$$\gamma |I| - \delta |X| > 0, \quad (A18)$$

where  $I$  is the current input pattern and  $|X|$  is the number of nodes across  $F_1$  such that  $x_i > T$ . The nonspecific reset wave shuts off the active  $F_2$  node until the input pattern  $I$  shuts off. Thus (A10) must be modified to shut off all  $F_2$  nodes which have been reset by  $A$  during the presentation of  $I$ .

## REFERENCES

- Basar, E., Flohr, H., Haken, H., and Mandell, A.J. (Eds.), **Synergetics of the brain**. New York: Springer-Verlag, 1983.
- Bruner, J.S., Modalities of memory. In G.A. Talland and N.C. Waugh (Eds.), **The pathology of memory**. New York: Academic Press, 1969.
- Carpenter, G.A. and Grossberg, S., Neural dynamics of adaptive pattern recognition: Priming, search, attention, and category formation. *Society for Neuroscience Abstracts*, **11**, 1985 (a).
- Carpenter, G.A. and Grossberg, S., Self-organization of neural recognition categories. In preparation, 1985 (b).
- Cohen, M.A. and Grossberg, S., Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology*, in press, 1985.
- Cohen, N.J., Preserved learning capacity in amnesia: Evidence for multiple memory systems. In L. Squire and N. Butters (Eds.), **The neuropsychology of memory**. New York: Guilford Press, 1984, pp.83-103.
- Deadwyler, S.A., West, M.O., and Lynch, G., Activity of dentate granule cells during learning: Differentiation of perforant path inputs. *Brain Research*, 1979, **169**, 29-43.
- Deadwyler, S.A., West, M.O., and Robinson, J.H., Entorhinal and septal inputs differentially control sensory-evoked responses in the rat dentate gyrus. *Science*, 1981, **211**, 1181-1183.
- Ellias, S.A. and Grossberg, S., Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics*, 1975, **20**, 69-98.
- Graf, P., Squire, L.R., and Mandler, G., The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1984, **10**, 164-178.
- Grossberg, S., Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 1973, **52**, 217-257.
- Grossberg, S., Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 1976, **23**, 121-134 (a).
- Grossberg, S., Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 1976, **23**, 187-202 (b).
- Grossberg, S., 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. New York: Academic Press, 1978, pp.233-374.
- Grossberg, S., How does a brain build a cognitive code? *Psychological Review*, 1980, **87**, 1-51.
- Grossberg, S., Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, 1982, **89**, 529-572.
- Grossberg, S., The quantized geometry of visual space: The coherent computation of depth, form, and lightness. *Behavioral and Brain Sciences*, 1983, **6**, 625-692.
- Grossberg, S., 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**. New York: New York Academy of Sciences, 1984 (a).

- Grossberg, S., Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory*, 1984, **7**, 263-283 (b).
- Grossberg, S., 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, 1985 (a).
- Grossberg, S., The role of learning in sensory-motor control. *Behavioral and Brain Sciences*, in press, 1985 (b).
- Grossberg, S. and Kuperstein, M., **Adaptive neural dynamics of sensory-motor control: Ballistic eye movements**. Amsterdam: North-Holland, 1985.
- Grossberg, S. and Levine, D.S., Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology*, 1975, **53**, 341-380.
- Grossberg, S. and Mingolla, E., Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 1985, **92**, 173-211 (a).
- Grossberg, S. and Mingolla, E., Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. Submitted for publication, 1985 (b).
- Grossberg, S. and Stone, G.O., Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, in press, 1985.
- Karrer, R., Cohen, J., and Tueting, P. (Eds.), **Brain and information: Event related potentials**. New York: New York Academy of Sciences, 1984.
- Lynch, G., McGaugh, J.L., and Weinberger, N.M. (Eds.), **Neurobiology of learning and memory**. New York: Guilford Press, 1984.
- Malsburg, C. von der and Willshaw, D.J., Differential equations for the development of topological nerve fibre projections. In S.Grossberg (Ed.), **Mathematical psychology and psychophysiology**. Providence, RI: American Mathematical Society, 1981.
- Mattis, S. and Kovner, R., Amnesia is as amnesia does: Toward another definition of the anterograde amnesias. In L. Squire and N. Butters (Eds.), **Neuropsychology of memory**. New York: Guilford Press, 1984.
- Mishkin, M., A memory system in the monkey. *Philosophical Transactions of the Royal Society of London*, 1982, **B298**, 85-95.
- Neely, T.H., Semantic priming and retrieval from lexical memory: The roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, 1977, **106**, 226-254.
- O'Keefe, J. and Nadel, L., **The hippocampus as a cognitive map**. Oxford: Oxford University Press, 1978.
- Posner, M.I. and Snyder, C.R.R., Attention and cognitive control. In R.L. Solso (Ed.), **Information processing and cognition: The Loyola symposium**. Hillsdale, NJ: Erlbaum, 1975 (a).
- Posner, M.I. and Snyder, C.R.R., Facilitation and inhibition in the processing of signals. In P.M.A. Rabbitt and S. Dornic (Eds.), **Attention and performance V**. New York: Academic Press, 1975 (b).
- Ryle, G., **The concept of mind**. San Francisco: Hutchinson, 1949.
- Squire, L.R. and Butters, N. (Eds.), **Neuropsychology of memory**. New York: Guilford Press, 1984.
- Squire, L.R. and Cohen, N.J., Human memory and amnesia. In G. Lynch, J. McGaugh, and N.M. Weinberger (Eds.), **Neurobiology of learning and memory**. New York: Guilford Press, 1984, pp.3-64.

- Warrington, E.K. and Weiskrantz, L., The amnesic syndrome: Consolidation or retrieval? *Nature*, 1970, **228**, 628-630.
- Warrington, E.K. and Weiskrantz, L., The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia*, 1974, **12**, 419-428.