"Proceedings of the 3rd Army Conference on Applied Mathematics and Computing" 1985

CATEGORY LEARNING AND ADAPTIVE PATTERN RECOGNITION: A NEURAL NETWORK MODEL

Gail A. Carpenter†
Department of Mathematics, Northeastern University
Boston, Massachusetts 02115
and
Center for Adaptive Systems, Boston University
Boston, Massachusetts 02215
AND
Stephen Grossberg‡
Center for Adaptive Systems, Boston University
Boston, Massachusetts 02215

ABSTRACT. A theory is presented 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

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

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

Acknowledgements: We wish to thank Cynthia Suchta for her valuable assistance in the preparation of the manuscript.

presentations. The interactions generate emergent rules such as a Weber Law Rule, a 2/3 Rule, and an Associative Decay Rule. No serial programs or algorithmic rule structures are used.

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* [1]. A theory of how recognition categories can self-organize is outlined in this report, which summarizes the model's design and mathematical analysis, developed in other articles [2–4]. In those articles, the *adaptive resonance theory* is also related 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. 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 [5–7]. 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 [8,9].

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 an *attentional subsystem*. Pattern I is transformed into a pattern X of activation across the nodes of F_1 (Figure 1). 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 2a). 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. 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

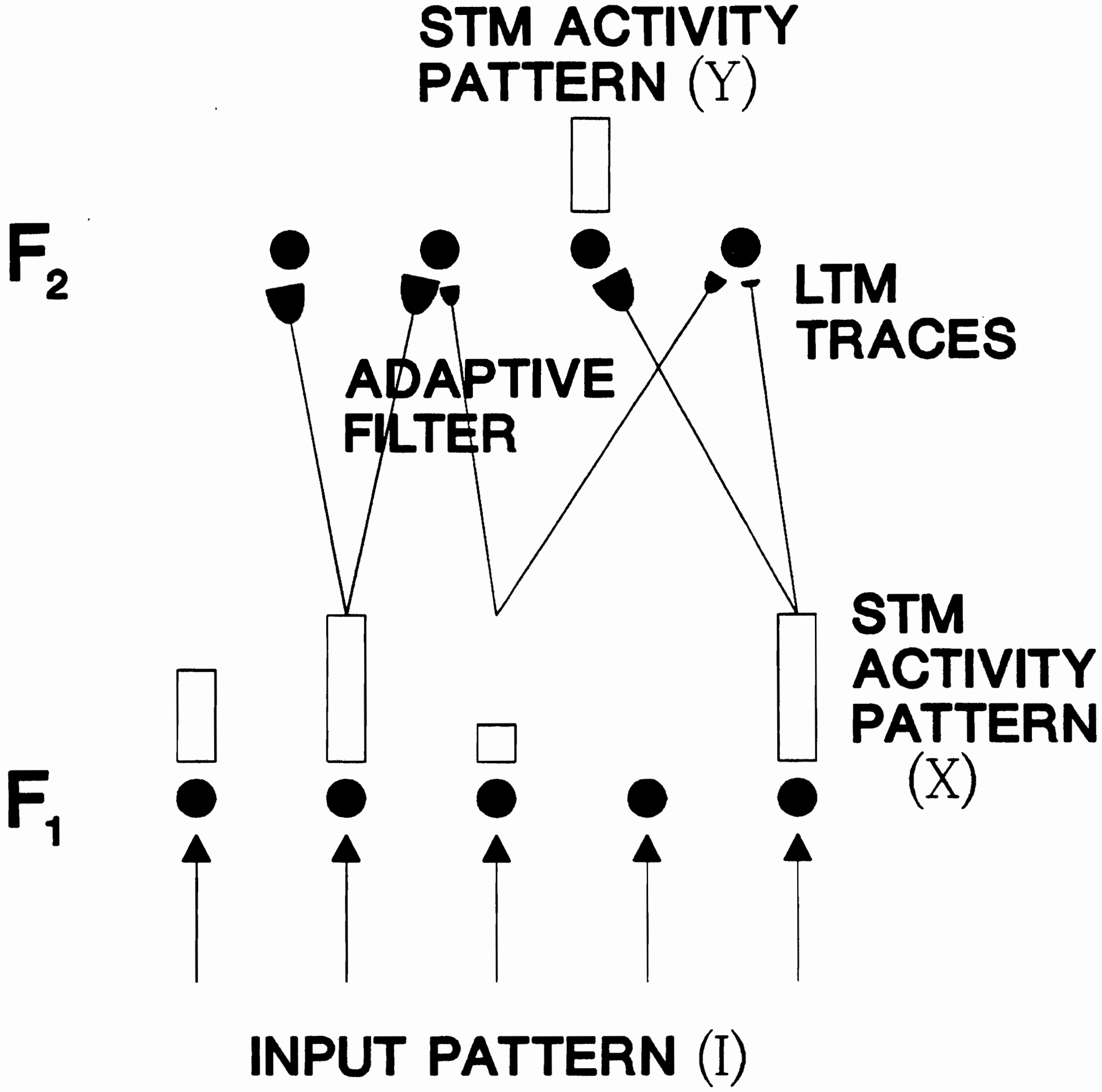

Figure 1. Stages of bottom-up activation: The input pattern I generates a pattern of STM activation X across F_1. Sufficiently active F_1 nodes emit bottom-up signals to F_2. This signal pattern S is gated by long term memory (LTM) traces within the $F_1 \rightarrow F_2$ pathways. The LTM-gated signals 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.

groupings. of an input pattern [6.10]. When F_2 is designed to make a choice in STM. it selects that global grouping of the input pattern which is preferred by the adaptive filter. This process automatically enables the network to partition all the input patterns which are received by F_1 into disjoint sets of recognition categories, each corresponding to a particular node in F_2.

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

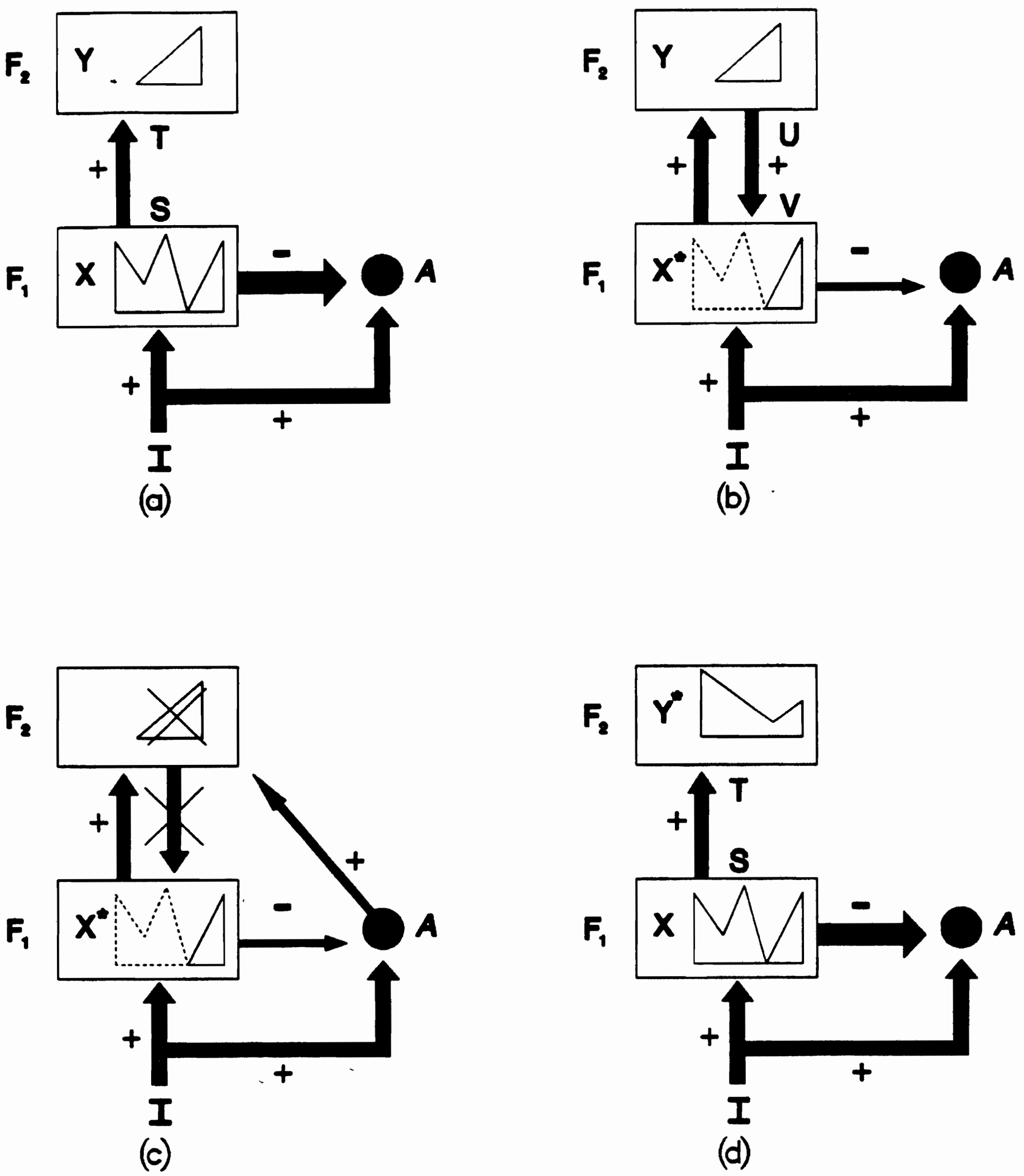

Figure 2. Search for a correct F_2 code: (a) The input pattern I generates the specific STM activity pattern X at F_1 as it nonspecifically activates A. Pattern X both inhibits A and generates the output signal pattern S. Signal pattern S is transformed into the input pattern T, which activates the STM 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.

The bottom-up STM transformation I→X→S→ T→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→T (Figure 2a) 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 7. It was the instability of bottom-up adaptive coding that led Grossberg [11,12] 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 [4,6,7]. These mechanisms have now been developed into a rigorously characterized learning system whose properties have been quantitatively analysed [2,3]. 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.

Figure 3 summarizes the total network architecture. It includes modulatory processes, such as attentional gain control, which regulate matching within F_1, as well as modulatory processes, such as orienting arousal, which regulate reset within F_2. Figure 3 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.

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→X→ S→T→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→T can change. As soon as the bottom-up STM transformation X→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 $F_2 \to F_1$.

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* (Figure 2b).

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.

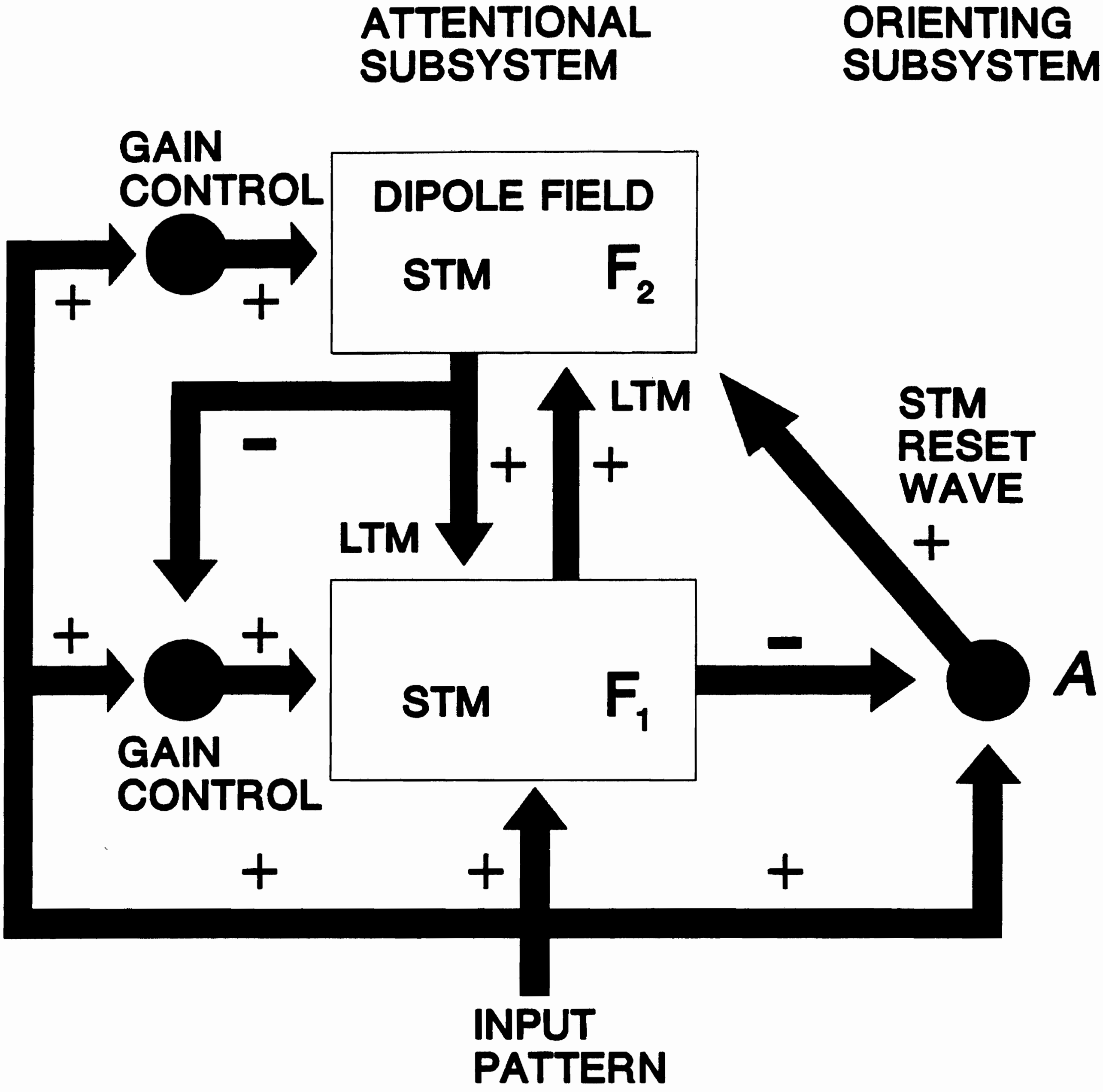

Figure 3. Anatomy of the attentional-orienting system: This figure describes all the interactions of the model without regard to which components are active at any given time.

The entire activation sequence

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

takes place very quickly relative to the rate with which the LTM traces in either the bottom-up adaptive filter S→T or the top-down adaptive filter U→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 works together with an *orienting subsystem* 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 2 depicts a typical interaction between the attentional subsystem and the orienting subsystem.In Figure 2a, 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 2b, 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 2c 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* [4,13]. 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 2c, 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 2d). 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 can be based upon on-center off-surround interactions among cells obeying membrane equations [4.14].

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 2). 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.

Figure 4 depicts one possible scheme 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. Figure 4d shows, 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.

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 2). 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 4a, 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 4b, 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 signalling, the attentional gain control signal is inhibited by the top-down channel (Figure 4c). 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 2a and 2b).

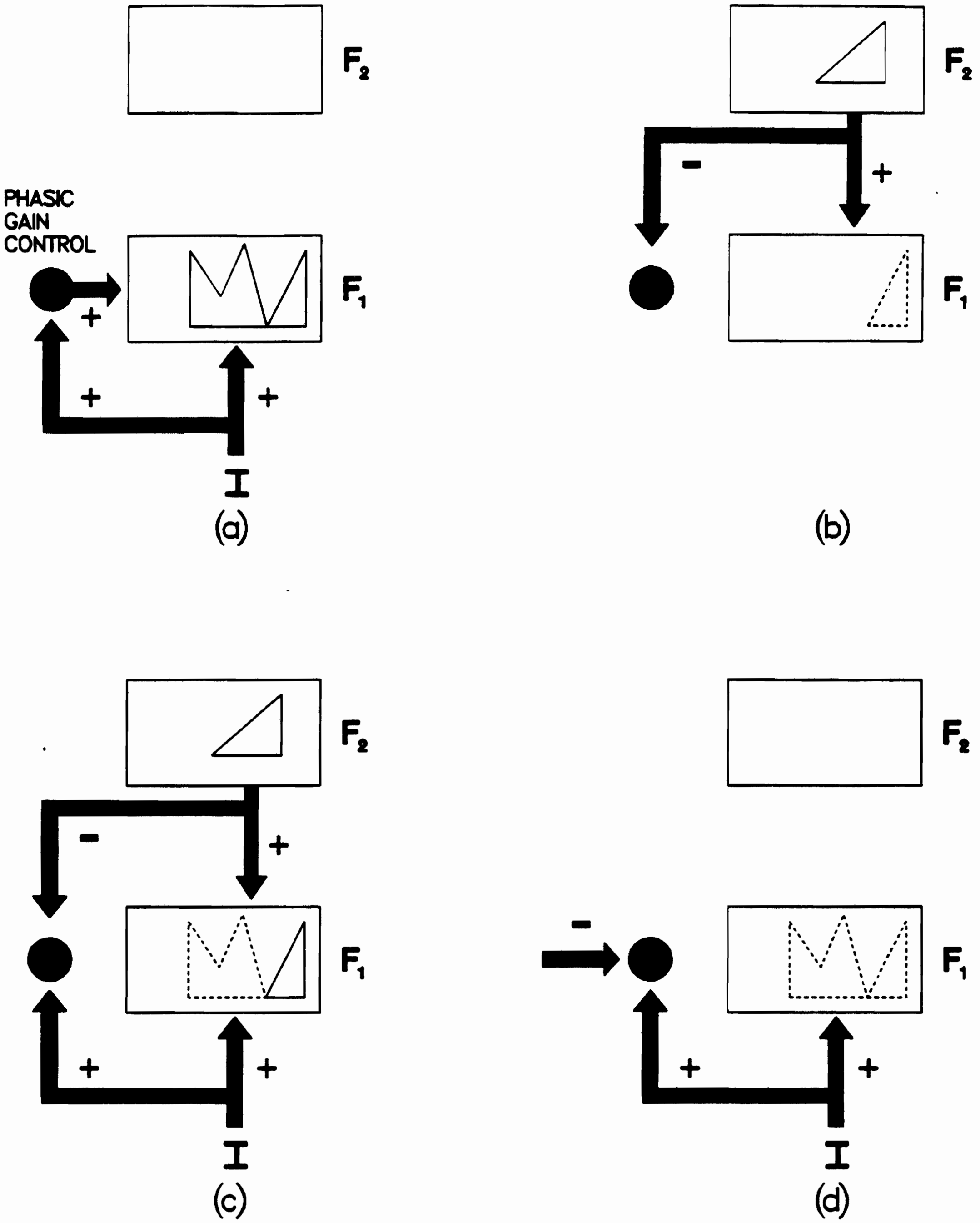

Figure 4. Matching by 2/3 Rule: (a) In this example. nonspecific attentional gain control signals are phasically activated by the bottom-up input. In this network, the bottom-up input arouses two different nonspecific channels: the attentional gain control channel and the orienting subsystem. Only F_1 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.

7. Example of Code Instability. We now illustrate the importance of the 2/3 Rule 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 [11]. 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 3 and 4c) 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 5a summarizes a computer simulation of unstable code learning. Figure 5b summarizes a computer simulation that illustrates how reinstatement of the 2/3 Rule can stabilize code learning.

The first column of Figure 5a 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

$$D \subset C \subset A. \tag{2}$$

$$B \subset A, \tag{3}$$

$$B \bigcap C = \phi. \tag{4}$$

$$\mid D \mid < \mid B \mid < \mid C \mid . \tag{5}$$

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 5a and 5b, the input patterns were periodically presented in the order ABCAD.

Each of the Top-Down Template columns in Figure 5 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 5a, 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.

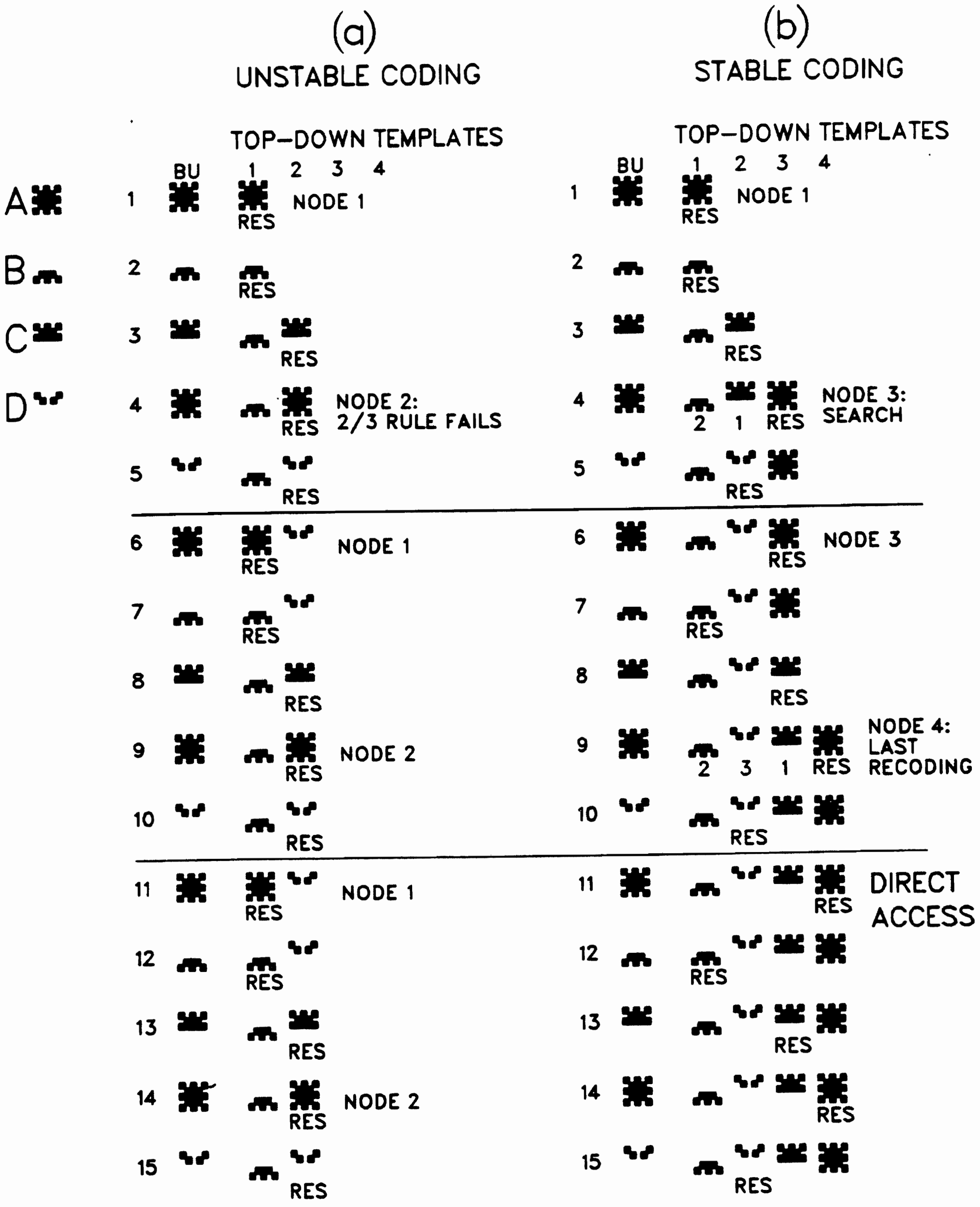

Figure 5. Stabilization of categorical learning by the 2/3 Rule: In both (a) and (b), four input patterns A, B, C, and D are presented repeatedly in the list order ABCAD. In (a), the 2/3 Rule is violated because the top-down inhibitory gain control mechanism be weak (Figure 4c). Pattern A is periodically 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.

In Figure 5b. 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 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 5b, 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, 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 5a and 5b. 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 5b. 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.

8. 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 $\mathcal{A}$. 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 [4] to assume that the bottom-up input source activates two parallel channels (Figure 2a). The attentional subsystem receives a specific input pattern at F_1. The orienting subsystem receives convergent inputs at $\mathcal{A}$ 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 2a). 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 2b,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 2a). 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 2b). 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 2c).

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 $\mid I \mid$ nodes of F_1. Since every active input pathway projects to A, I generates a total input to A that is proportional to $\mid I \mid$. We suppose that A reacts linearly to the total input $\gamma \mid I \mid$. 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 $\delta \mid X \mid$ from F_1 to A is proportional to the number $\mid X \mid$ of active F_1 nodes. When $\gamma \mid I \mid > \delta \mid X \mid$, A receives a net excitatory signal and generates a nonspecific reset signal to F_2 (Figure 2c).

In response to a bottom-up input pattern I of size $\mid I \mid$, as in Figure 2a, the total inhibitory input from F_1 to A equals $\delta \mid I \mid$, so the net input to A equals $(\gamma - \delta) \mid I \mid$. In order to prevent A from firing in this case (Figure 2a), we assume that $\delta \geq \gamma$. We call

$$\rho = \frac{\gamma}{\delta} \tag{6}$$

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 ρ 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 2b), the 2/3 Rule implies that the total inhibitory signal from F_1 to A equals $\delta \mid V^{(j)} \cap I \mid$. In this case, the orienting subsystem is activated only if

$$\gamma \mid I \mid > \delta \mid V^{(j)} \cap I \mid; \tag{7}$$

that is, if

$$\frac{\mid V^{(j)} \cap I \mid}{\mid I \mid} < \rho. \tag{8}$$

In order to illustrate how the network codifies a series of patterns, we show in Figure 6 the first 20 trials of a simulation using alphabet letters as input patterns. In Figure 6a,

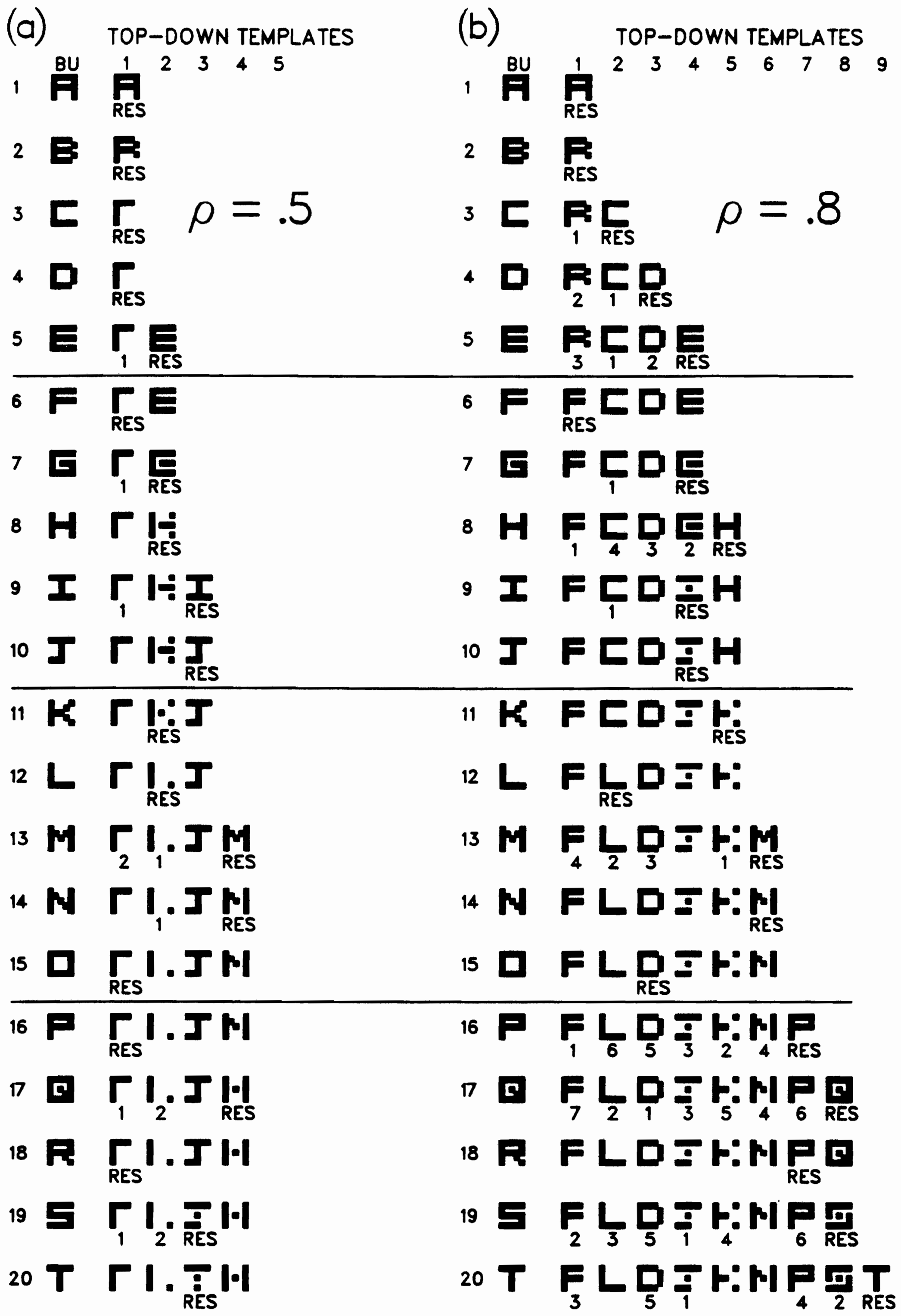

Figure 6. Alphabet learning: Different vigilance levels cause different numbers of letter categories to form.

the vigilance parameter $\rho = .5$. In Figure 6b, $\rho = .8$. 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.

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.

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

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

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 $i = 1.2.\ldots, M$. We denote nodes in F_2 by v_j, where $j = M+1, M+2, \ldots, N$. Thus by (A1),

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

and

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

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

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

that is,

$$J_i^+ = I_i + V_i. \qquad (A5)$$

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:

$$J_i^- = F\sum_j f(x_j). \tag{A6}$$

Thus $J_i^- = 0$ if and only if F_2 is inactive (Figure 4).

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 show how these parameters can be chosen [15–17]. 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

$$T_j = \sum_i h(x_i)z_{ij}. \tag{A7}$$

that is,

$$J_j^+ = g(x_j) + T_j, \tag{A8}$$

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:

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

Such a network behaves approximately like a binary switching circuit:

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

LTM Equations

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

$$\frac{d}{dt} z_{ij} = f(x_j)[-H_{ij}z_{ij} + Kh(x_i)]. \tag{A11}$$

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

$$\frac{d}{dt} z_{ji} = f(x_j)[-H_{ji}z_{ji} + Kh(x_i)]. \tag{A12}$$

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

$$H_{ji} = H = \text{constant}. \tag{A13}$$

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 [2] 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. In particular, it would enable us to also choose $H_{ji} = H =$ constant. Instead, we allowed the bottom-up LTM traces at each node v_j to compete among themselves for synaptic sites. Malsburg and Willshaw [18] 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

$$H_{ji} = Lh(x_i) + \sum_{k \neq i} h(x_k). \tag{A14}$$

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

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

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,

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

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

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

where $\mid X \mid$ is the number of active nodes in F_1. Term z_{ij} obeys a Weber Law Rule if $L > 1$.

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 8, each active input pathway sends an excitatory signal of size γ 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

$$\gamma \mid I \mid - \delta \mid X \mid > 0, \tag{A18}$$

where I is the current input pattern and $\mid X \mid$ 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

[1] Basar, E., Flohr, H., Haken, H., and Mandell, A.J. (Eds.), **Synergetics of the brain.** New York: Springer-Verlag, 1983.

[2] Carpenter, G.A. and Grossberg, S., Neural dynamics of category learning and recognition: Attention, memory consolidation, and amnesia. In J. Davis, R. Newburgh, and E. Wegman (Eds.), **Brain structure, learning, and memory**. AAAS Symposium Series, 1985.

[3] Carpenter, G.A. and Grossberg, S., Self-organization of neural recognition categories. In preparation, 1985.

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

[5] 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.

[6] 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.), **Perception of speech and visual form: Theoretical issues, models, and research.** New York: Academic Press, 1985.

[7] Grossberg, S. and Stone, G.O., Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, in press, 1985.

[8] Grossberg, S. and Mingolla, E., Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 1985, **92**, 173–211.

[9] Grossberg, S. and Mingolla, E., Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, in press, 1985.

[10] 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.

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

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

[13] 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.

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

[15] 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.

[16] Grossberg, S., Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 1973, **52**, 217–257.

[17] 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.

[18] 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.