

Masking fields: a massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data

Michael A. Cohen and Stephen Grossberg

A massively parallel neural network architecture, called a masking field, is characterized through systematic computer simulations. A masking field is a multiple-scale self-similar automatically gain-controlled cooperative-competitive feedback network F_2 . Network F_2 receives input patterns from an adaptive filter $F_1 \rightarrow F_2$ that is activated by a prior processing level F_1 . Such a network F_2 behaves like a content-addressable memory. It activates compressed recognition codes that are predictive with respect to the activation patterns flickering across the feature detectors of F_1 and competitively inhibits, or masks, codes which are unpredictable with respect to the F_1 patterns. In particular, a masking field can simultaneously detect multiple groupings within its input patterns and assign activation weights to the codes for these groupings which are predictive with respect to the contextual information embedded within the patterns and the prior learning of the system. A masking field automatically rescales its sensitivity as the overall size of an input pattern changes, yet also remains sensitive to the microstructure within each input pattern. In this way, a masking field can more strongly activate a code for the whole F_1 pattern than for its salient parts, yet amplifies the code for a pattern part when it becomes a pattern whole in a new input context. A masking field can also be primed by inputs from F_1 : it can activate codes which represent predictions of how the F_1 pattern may evolve in the subsequent time interval. Network F_2 can also exhibit an adaptive sharpening property: repetition of a familiar F_1 pattern can tune the adaptive filter to elicit a more focal spatial activation of its F_2 recognition code than does an unfamiliar input pattern. The F_2 recognition code also becomes less distributed when an input pattern contains more contextual information on which to base an unambiguous prediction of which the F_1 pattern is being processed. Thus a masking field suggests a solution of the credit assignment problem by embodying a real-time code for the predictive evidence contained within its input patterns. Such capabilities are useful in speech recognition, visual object recognition, and cognitive information processing. An absolutely stable design for a masking field is disclosed through an analysis of the computer simulations. This design suggests how associative mechanisms, cooperative-competitive interactions, and modulatory gating signals can be joined together to regulate the learning of compressed recognition codes. Data about the neural substrates of learning and memory are compared to these mechanisms.

I. Introduction: Context-Sensitive Grouping in Recognition Processes

One fundamental problem area in perception, cognition, and artificial intelligence concerns the characterization of the functional units into which perceptual and cognitive mechanisms group the patterned information that they process. A core issue concerns the context-sensitivity of these functional units or the manner in which a grouping into functional units can depend on the spatiotemporal patterning of all the

signals being processed. Another core issue concerns the adaptive tuning of recognition mechanisms and the manner in which such tuning can alter the groupings which emerge within a context containing familiar elements. Adaptive tuning of recognition processes is one of the mechanisms whereby representations become compressed, chunked, or unitized into coherent recognition codes through experience.

The present paper describes the further development of a real-time neural network model, called a masking field, which was introduced in Grossberg.¹ A masking field is a multiple-scale self-similar automatically gain-controlled cooperative-competitive feedback network. This type of network acts like a content-addressable memory whose properties are useful for understanding how a large class of compressed recognition code is established during real-time speech recognition, visual object recognition, and cognitive information processing. The analyses of cooperative-

The authors are with Boston University, Center for Adaptive Systems, Boston, Massachusetts 02215.

Received 12 May 1986.

0003-6935/87/101866-26\$02.00/0.

© 1987 Optical Society of America.

competitive content-addressable memories which led to the masking field concept are found in Grossberg² and Grossberg and Levine.³ Since its introduction in 1978, the masking field model has played an important role in development of the adaptive resonance theory which was introduced in Grossberg^{4,5} and which has since undergone extensive development and application.⁶⁻²¹

A related family of cooperative-competitive content-addressable networks, which also grew out of the Grossberg² analysis, was identified through the discovery of a global Liapunov function in Cohen and Grossberg.²² The Liapunov function described in Hopfield²³ is a special case of the Cohen-Grossberg function, a case that was explicitly noted in Cohen and Grossberg (p. 819).²² The Cohen-Grossberg Liapunov function requires symmetry of connections between pairs of network nodes. Symmetry of connections is strongly violated in a masking field. This symmetry violation is a consequence of the network's self-similar design. In particular, the masking fields analyzed herein obey the differential equations

$$\frac{d}{dt}x_i^{(j)} = -Ax_i^{(j)} + (B - x_i^{(j)}) \left\{ \sum_{j \in J} I_j \left[\frac{1}{|J|} (1 - p_{ij}) + r_{ji}^{(j)} p_{ij} \right] z_{ji}^{(j)} + D|J|f(x_i^{(j)}) \right\} - F(x_i^{(j)} + C) \frac{\sum_{m,K} g(x_m^{(K)})|K|(1 + |K \cap J|)}{\sum_{m,K} |K|(1 + |K \cap J|)}, \quad (1)$$

$$\frac{d}{dt}z_{ji}^{(j)} = \epsilon f(x_i^{(j)}) (-z_{ji}^{(j)} + LI_j), \quad (2)$$

where the variables $x_i^{(j)}$ are activations, or short-term memory (STM) traces, of F_2 nodes, and the variables $z_{ji}^{(j)}$ are adaptive weights, or long-term memory (LTM) traces, of the pathways within the $F_1 \rightarrow F_2$ adaptive filter. These equations are derived in the Appendix. The competitive interaction coefficients in Eq. (1) from an F_2 node $v_m^{(K)}$ to an F_2 node $v_i^{(j)}$, namely,

$$\frac{|K|(1 + |K \cap J|)}{\sum_{m,L} |L|(1 + |L \cap J|)}, \quad (3)$$

are asymmetric functions of K and J .

Despite the asymmetry of the coefficients in (3), the masking field Eq. (1) can be written in Cohen-Grossberg form using a simple change of variables. This fact clarifies why the networks considered herein always approach an equilibrium point. Grossberg²⁴ describes how Eq. (1) as well as a number of other well-studied content-addressable network memory models can be written in this form. To achieve a parametric understanding of masking fields, we have developed these networks through the use of systematic computer simulations. The present paper reports computer simulations carried out to design masking fields capa-

ble of robustly solving the following type of adaptive coding problem.

II. Detecting and Encoding Multiple Groupings in Short-Term Memory Using a Masking Field

A masking field is capable of simultaneously detecting multiple groupings within its input patterns and automatically assigning activation weights to the codes for these groupings which are predictive with respect to the context of the patterns and the prior learning of the system.

For example, a word such as *Myself* is used by a fluent speaker as a unitized verbal chunk. In different verbal contexts, however, the components *My*, *Self*, and *Elf* of *Myself* are all words in their own right. Moreover, although an utterance which ended at *My* generates one grouping of the speech flow, an utterance which goes on to include the entire word *Myself* supplants this encoding with one appropriate to the longer word. Thus to understand how context-sensitive content-addressable language units are perceived by a fluent speaker, one must analyze how all possible groupings of the speech flow are analyzed through time and how certain groupings can be chosen in one context without preventing other groupings from being chosen in a different context.

The same considerations hold when words such as *Myself* are presented visually rather than auditorily. Then the problem becomes one of visual object recognition and of figure-ground segmentation. The problem exists also on a finer level of visual or auditory processing, since letters such as *E* contain as parts letters such as *L* and *F*. The masking field design is capable of sensing multiple-pattern groupings, which subtend multiple spatial scales, and assigns each of these groupings a proper activation weight in its STM representation of these groupings.

III. Developmental Rules Imply Cognitive Rules as Emergent Properties of Neural Network Interactions

It has been shown how a masking field network F_2 can arise through simple rules of neuronal growth^{1,11,13,15} for the connections from its input source F_1 and among its own cells or nodes. These rules include a random growth of connections along spatial gradients from F_1 to F_2 , activity-dependent self-similar cell growth within F_2 , and intercellular interactions among F_2 cells which compete for conserved synaptic sites [Fig. 1(a)]. Although these growth rules are of interest at the present time primarily in applications of masking fields to cognitive psychology and developmental neurobiology, they may at a future time suggest procedures for realizing a masking field in hardware. In addition, these growth processes illustrate how simple rules of neuronal development can give rise to a system whose parallel interactions act as if it obeys complex rules of context-sensitive cognitive coding. Because these growth rules can be obeyed by any number of network levels, masking fields can be linked into a coding hierarchy F_1

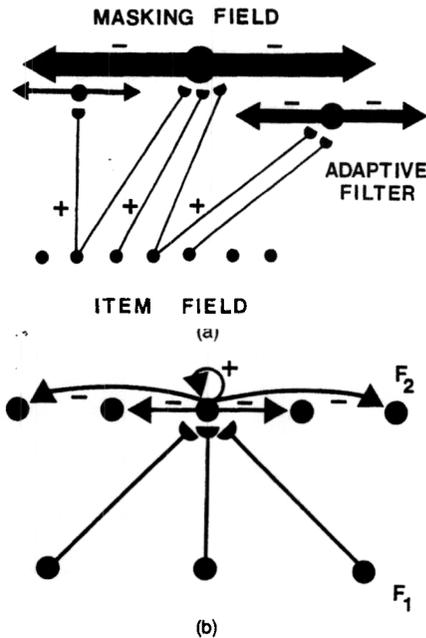


Fig. 1. Masking field interactions: (a) cells from an item field F_1 grow randomly to a masking field F_2 along positionally sensitive gradients. The nodes in the masking field grow so that larger item groupings, up to some optimal size, can activate nodes with broader and stronger inhibitory interactions. Thus the $F_1 \rightarrow F_2$ connections and the $F_2 \leftrightarrow F_2$ interactions exhibit properties of self-similarity. (b) The interactions within a masking field F_2 include positive feedback from a node to itself and negative feedback from a node to its neighbors. Long-term memory traces at the ends of $F_1 \rightarrow F_2$ pathways (designated by hemidisks) adaptively tune the filter defined by these pathways to amplify the F_2 reaction to item groupings which have previously succeeded in activating their target F_2 nodes.

$\rightarrow F_2 \rightarrow F_3 \rightarrow \dots F_n$ whose successive levels are able to detect and manipulate ever more abstract recognition codes and hypotheses about the input patterns received by F_1 .

A masking field network F_2 selects its compressed representations by performing a new type of multiple scale analysis of the activity patterns which reach it from its input level F_1 . This analysis enhances correct groupings and competitively inhibits, or masks, inappropriate groupings in STM. Otherwise expressed, the masking field network does not confuse wholes with their parts, yet—despite this fact—it enables familiar parts to emerge as wholes in their own right in an appropriate input context, just as the words My and Self may be processed as wholes if they are presented separately or as parts within Myself when presented together (Sec. II).

The spatial pattern of enhanced STM activities across F_2 embodies a hypothesis, or compressed content-addressable code, which represents the input stream. As will be described in greater detail, this code can predict, or anticipate, subsequent events by assigning activities to groupings which have not yet fully occurred based on the available evidence. Thus the network acts like a real-time prediction, or evidence gathering, machine. No serial programs, cognitive rule structures, or teachers exist within the mask-

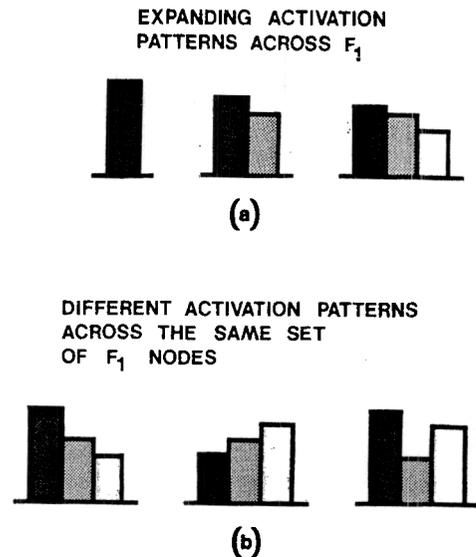


Fig. 2. Two types of masking field sensitivity: (a) A masking field F_2 can automatically rescale its sensitivity to differentially react as the F_1 activity pattern expands through time to activate more F_1 cells. It hereby acts like a multiple spatial frequency filter. (b) A masking field can differentially react to different F_1 activity patterns which activate the same set of F_1 nodes. By (a) and (b), F_2 acts like a spatial pattern discriminator which can compensate for changes in overall spatial scale without losing its sensitivity to pattern changes at the finest spatial scale.

ing field network to accomplish these properties. Instead, the model nodes, or neurons, obey membrane equations undergoing shunting (mass action) on-center off-surround (cooperative-competitive) recurrent (feedback) interactions [Fig. 1(b)]. The STM activation code of a masking field is an automatic emergent property of these interactions.

IV. Sensitivity to Multiple Pattern Scales and to Intrascale Microstructure

The multiple-scale analysis that is performed by a masking field is sensitive to two different types of pattern change.

A. Sensitivity to Multiple Pattern Scales

As a word like Myself is processed, a subword such as My occurs before the entire word Myself is experienced. Figure 2(a) schematizes this type of informational change. As the word is presented, it activates an increasing number of F_1 nodes, or feature detectors, through time. As increasing numbers of F_1 nodes are activated, earlier STM activations within F_1 may be modified as they are supplemented by later STM activations. After Myself is fully stored within F_1 , parts such as My, Self, and Elf are still present within the whole. However, the masking field F_2 can automatically rescale its initial response to My as the remainder of Myself is presented. In this way, the masking field can favor the whole word Myself rather than its parts, such as My, Self, or Elf, even though My may have been favored before Self also occurred.

A masking field's ability to favor a representation of a whole word rather than its parts derives from its sensitivity to the overall scale of each of the groupings which it can detect. The subtlety of this automatic scaling property is revealed through the following issue. If a masking field favors a whole pattern rather than its parts, why does not the field continue to favor the same whole pattern code when only part of the pattern is presented? How does the field sensitively respond to the part as a new content-addressable whole in its own right? Otherwise expressed, if larger pattern codes are favored when the larger patterns actually occur, how can smaller pattern codes be favored when smaller patterns occur?

B. Sensitivity to Internal Pattern Microstructure

The second type of masking field sensitivity is illustrated by the two words Left and Felt. This comparison is merely illustrative. It does not attempt to characterize the many subtle context-sensitive alterations that occur in evolving sound patterns or visually detected reading patterns. The words Left and Felt illustrate the issue that the same set of item representations within F_1 may be activated by different item orderings. To distinguish two such activity patterns across F_1 , sensitivity within F_2 to different spatial scales of F_1 is insufficient because both lists may activate the same spatial scale of F_1 . Instead, sensitivity to different STM patterns which excite the same set of items is required [Fig. 2(b)].

The automatic rescaling and microstructure detection properties follow from the manner in which nonlinear feedback interactions among F_2 nodes automatically reorganize the sizes of the inputs received at F_2 by F_1 . This type of nonlinear feedback is absent from many alternative grouping algorithms, such as the Hough transform.^{25,26} In recent contributions to developing the Hough transform, a central problem is to discover how to use negative votes to cancel off-peak positive votes in parameter space.²⁷ A related type of problem is solved by a masking field. However, a masking field replaces algorithms for positive and negative voting with a real-time network undergoing positive and negative feedback interactions. The key insights of the present paper concern how to combine the design of nonlinear feedback within F_2 with the proper type of nonlinear learning in the $F_1 \rightarrow F_2$ adaptive filter to generate stable learning of unitized groupings with environmentally predictive properties.

V. Hypothesis Formation, Anticipation, Evidence, and Prediction

The dynamics of a masking field express in an abstract language a number of important properties of cognitive information processing, no less than of perceptual grouping. Consider for definiteness a masking field F_2 that is capable of simultaneously discriminating more than one grouping within a list of events that activates F_1 . For example, a masking field F_2 might respond to the F_1 representation of the word Myself by strongly activating an F_2 population that is

sensitive to the whole word and weakly activating F_2 populations that are sensitive to the word's most salient parts. More generally, it might react to a pair of events A and B by representing the events singly and as a unitized configuration. In such a representation, the total STM pattern across F_2 represents the F_1 STM pattern. The relative sizes of the F_2 's STM activities weight the relative importance of the unitized groupings which are coded by the respective F_2 cell populations.

The suprathreshold STM activities across F_2 are approximately normalized, or conserved, due to the fact that its feedback interactions obey a type of shunting cooperative-competitive law which is capable of automatic gain control [Fig. 1(b)]. The STM activities across F_2 thus function like a type of real-time probabilistic logic, or hypothesis-testing algorithm, or model of the evidence which F_2 has about the pattern across F_1 . The self-normalizing properties of such cooperative-competitive feedback networks also invite comparisons with the classical formalisms of statistical mechanics and quantum mechanics.^{1,13}

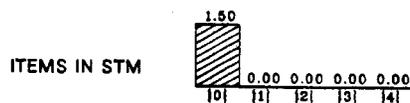
A masking field also possesses a predictive, anticipatory, or priming capability. In response to a single item across F_1 , the F_2 population which is most vigorously activated may code that item. In addition, less vigorous activations may arise at those F_2 populations which represent the most salient larger groupings of which the item forms a part. Such a masking field can predictively prime the masking field to anticipate the larger groupings of which the item may form a part during the next time interval. As more items are stored by F_1 through time, F_2 's uncertainty concerning the information represented at F_1 may decrease due to the emergence of a more predictive overall F_1 pattern. As F_2 's uncertainty decreases, the spatial distribution of STM activity across F_2 becomes more focused, or spatially localized, and includes fewer predictive groupings. This type of spatial sharpening measures the degree of informational uncertainty within the F_2 code about the activation pattern at F_1 .

VI. Computer Simulations of Global Choices and Multiple Groupings without Learning

The masking field design is described mathematically in the Appendix. This description is self-contained and may be read at any point during the subsequent exposition. The grouping properties of a masking field are illustrated in Secs. VI-X through computer simulations. These simulations are then used to motivate a refinement of masking field design in Secs. XI-XIV. This refined design leads to a number of predictions about mechanisms of neural learning.

Figures 3-5 depict the simplest type of grouping by a masking field. In this example, each distinct STM activity pattern across F_1 activates a unique node, or population, for STM storage within F_2 . In other words, such a masking field globally groups an activity pattern across F_1 into an STM choice within F_2 . Distinct choices are made in response to F_1 patterns which

ITEM FIELD (F_1)



MASKING FIELD (F_2)

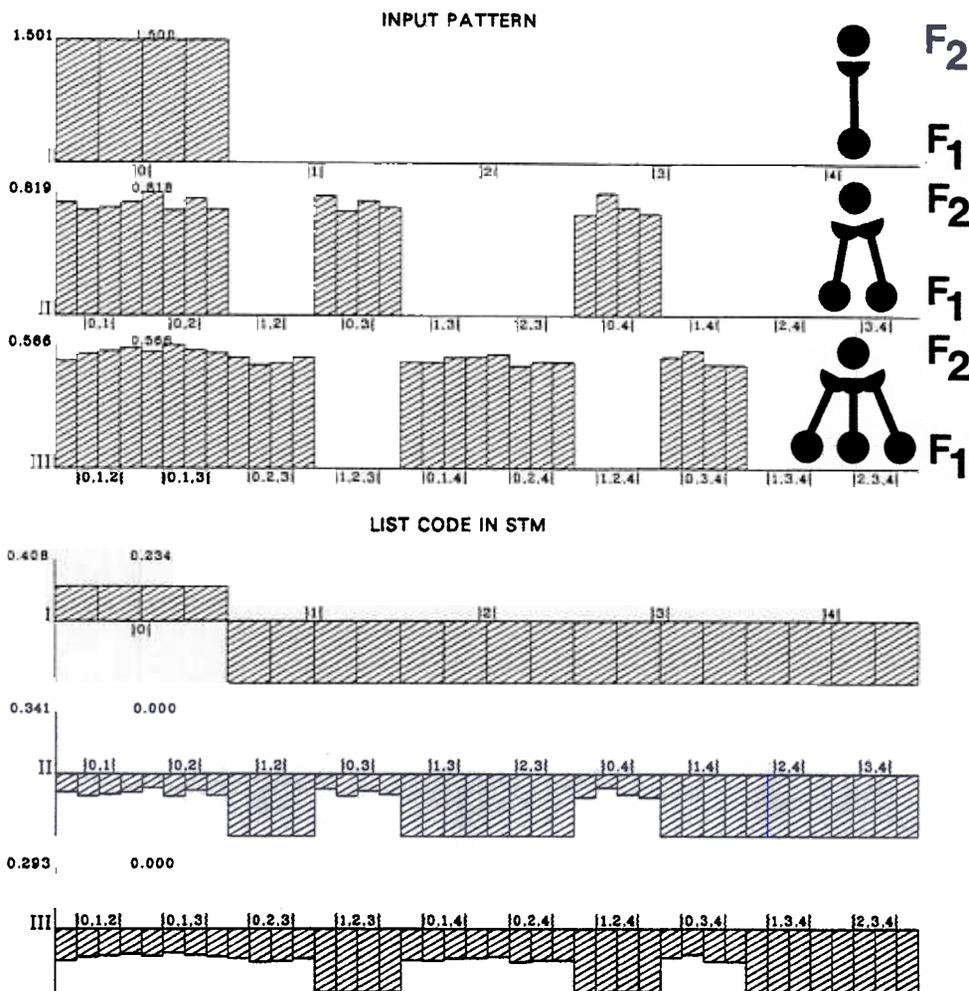


Fig. 3. List coding of a single item: Network F_1 encodes in short-term memory a spatial pattern of activation over item representations. In this figure, the single item $\{0\}$ is activated. Network F_2 encodes in STM the pattern of sublist chunks that are activated by F_1 . The first three rows depict the inputs from F_1 to F_2 . They are broadly distributed across F_2 . The list code in STM depicts the STM response to these inputs. Only the $\{0\}$ cells in F_2 are stored in STM despite the broad distribution of inputs.

vary in overall scale as well as in their microstructure, thereby demonstrating the properties summarized in Sec. IV. The same numerical parameters were used in all these simulations to demonstrate that a single masking field can generate all the properties being claimed. Sensitivity analyses were also carried out to determine the robustness of the design, but these will not be reported here. In this series of simulations, no learning was allowed to occur within the long-term memory (LTM) traces, or adaptive weights, that multiply the signals in the $F_1 \rightarrow F_2$ pathways (see Appendix).

In Figs. 6 and 7, a fixed but different set of parameters was used to illustrate how a masking field can generate STM representations which encode multiple groupings, including predictive groupings, of activity patterns across F_1 . In these STM representations, the masking field is maximally sensitive to the total STM pattern across F_1 , but it also generates partial activations to salient subpatterns (parts) and superpatterns (predictions) of this pattern. As in Figs. 3-5, the simulations described in Figs. 6 and 7 do not allow the LTM traces in the $F_1 \rightarrow F_2$ pathways to change due to learning. The computer simulation results in Figs. 3-7

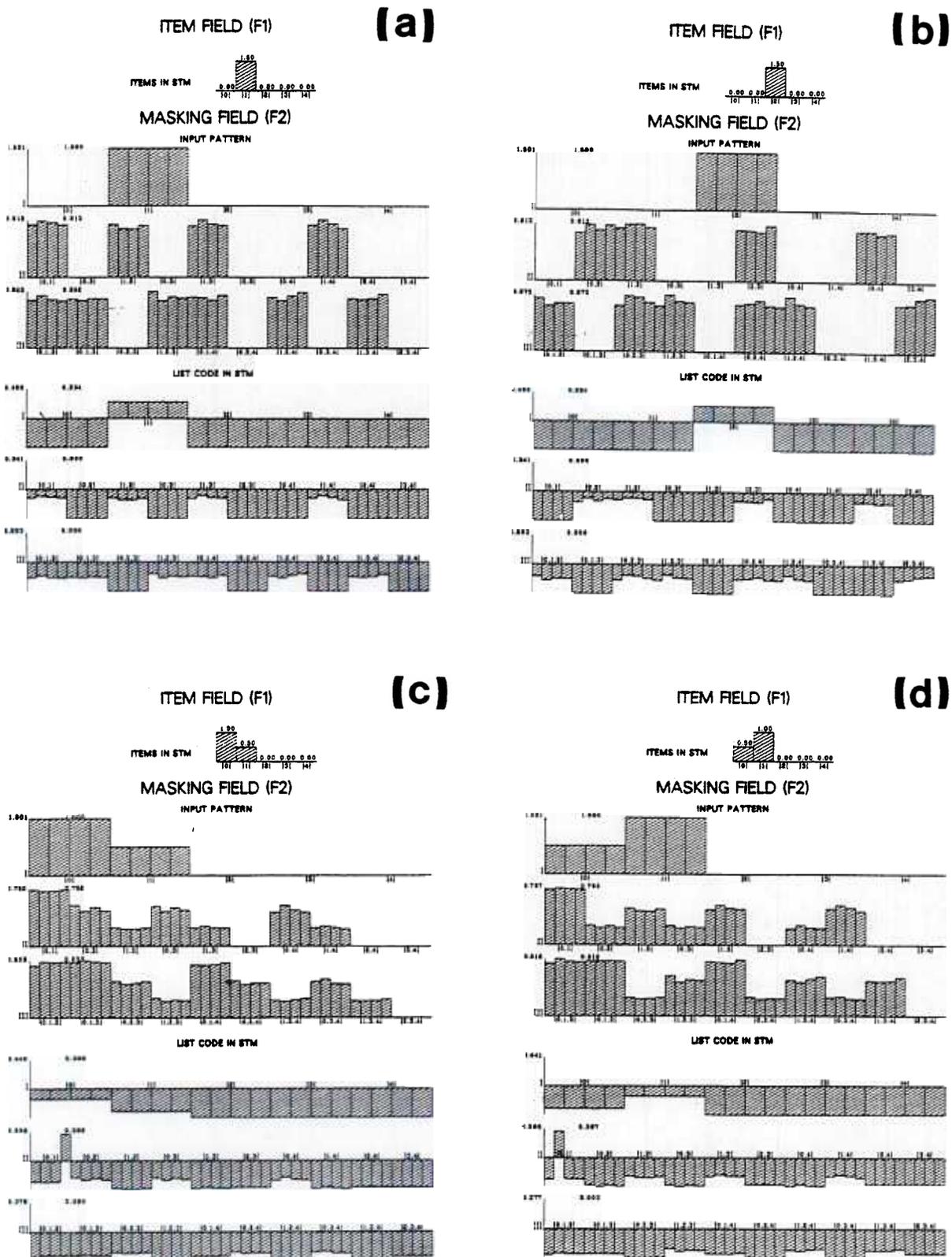


Fig. 4. (a) List coding of a single item: In response to item [1] in F_1 , the masking field in F_2 chooses the {1} cells in response to a broad distribution of inputs. Thus the list code in STM responds selectively to individual items in F_1 . The same thing is true in the next figure. (b) Here item [2] chooses the subset of list nodes [2] for storage in STM. (c) List coding of an STM primacy gradient across two items: A primacy gradient in STM across two items of F_1 generates a broader input pattern to F_2 . The list code in STM no longer responds at either the {0} or {1} cells. Instead a choice occurs among the set of possible {0,1} cells. Comparison with Fig. 3 shows that F_2 can update its internal representation in a context-sensitive way. (d) List coding of an STM recency gradient across two items: A recency gradient in STM occurs across the same two items of F_2 rather than a primacy gradient. Again the {0} and {1} cells are suppressed. A different choice among the {0,1} cells occurs from that in response to the primacy gradient of the preceding figure. Thus F_2 can distinguish different temporal orderings of the same items.

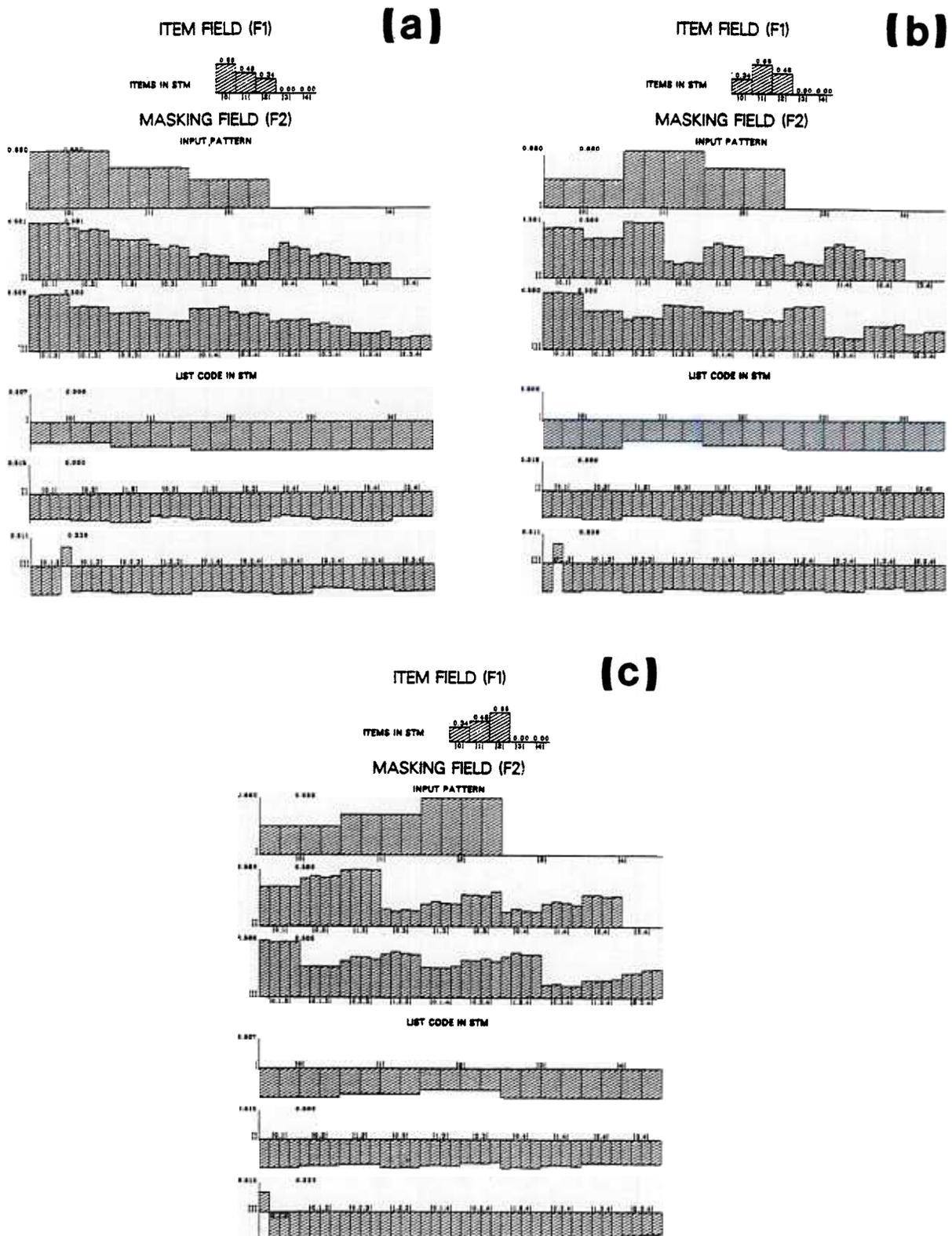


Fig. 5. (a) List coding of an STM primacy gradient across three items: In this figure, a primacy gradient in STM occurs across three items of F_1 . The input pattern to F_2 is even broader than before. However, the STM response of F_2 retains its selectivity. Network F_2 suppresses all $\{0\}$, $\{1\}$, $\{2\}$, $\{0,1\}$, $\{0,2\}$, \dots cells and chooses for STM storage a population from among the $\{0,1,2\}$ cells. (b) List codings of different temporal orderings across three items: In this and the next figure, different temporal orderings of the same three items generate selective STM responses among the $\{0,1,2\}$ cells. Thus, as future items activate an updated STM item code across F_1 , the STM list coding within F_2 is also updated in a context-sensitive way. (c) See legend for (b).

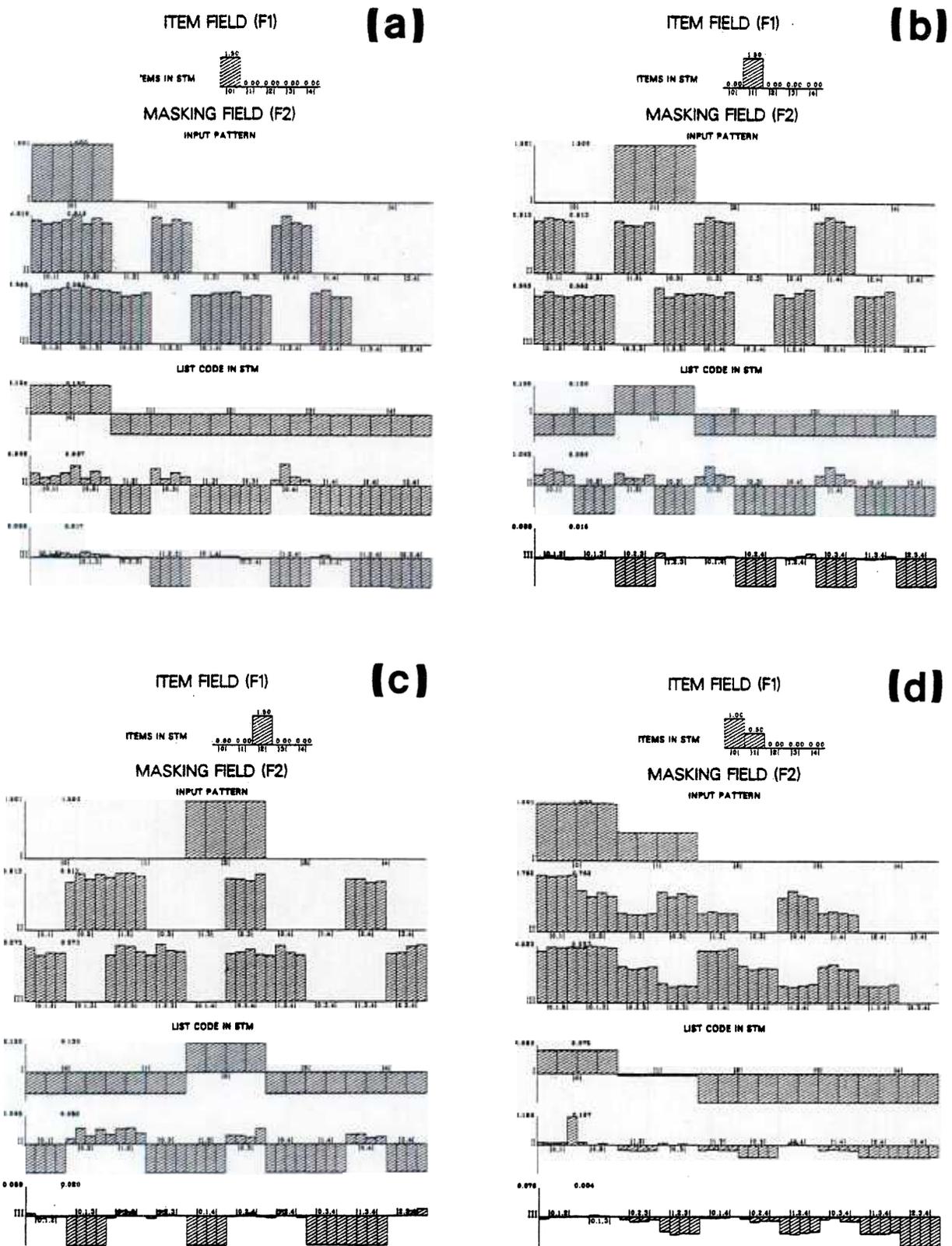


Fig. 6. (a) The correct list code {0} is preferred in STM, but predictive list codes which include {0} as a part are also activated with lesser STM weights. The prediction gets less activation if {0} forms a smaller part of it. (b) The correct list code {1} is preferred in STM, but the predictive list codes which include {1} as a part are also activated with lesser STM weights. (c) The list code in response to item {1} also generates an appropriate reaction. (d) A list code of type {0,1} is maximally activated, but part codes {0} and predictive codes which include {0,1} as a part are also activated with lesser STM weights.

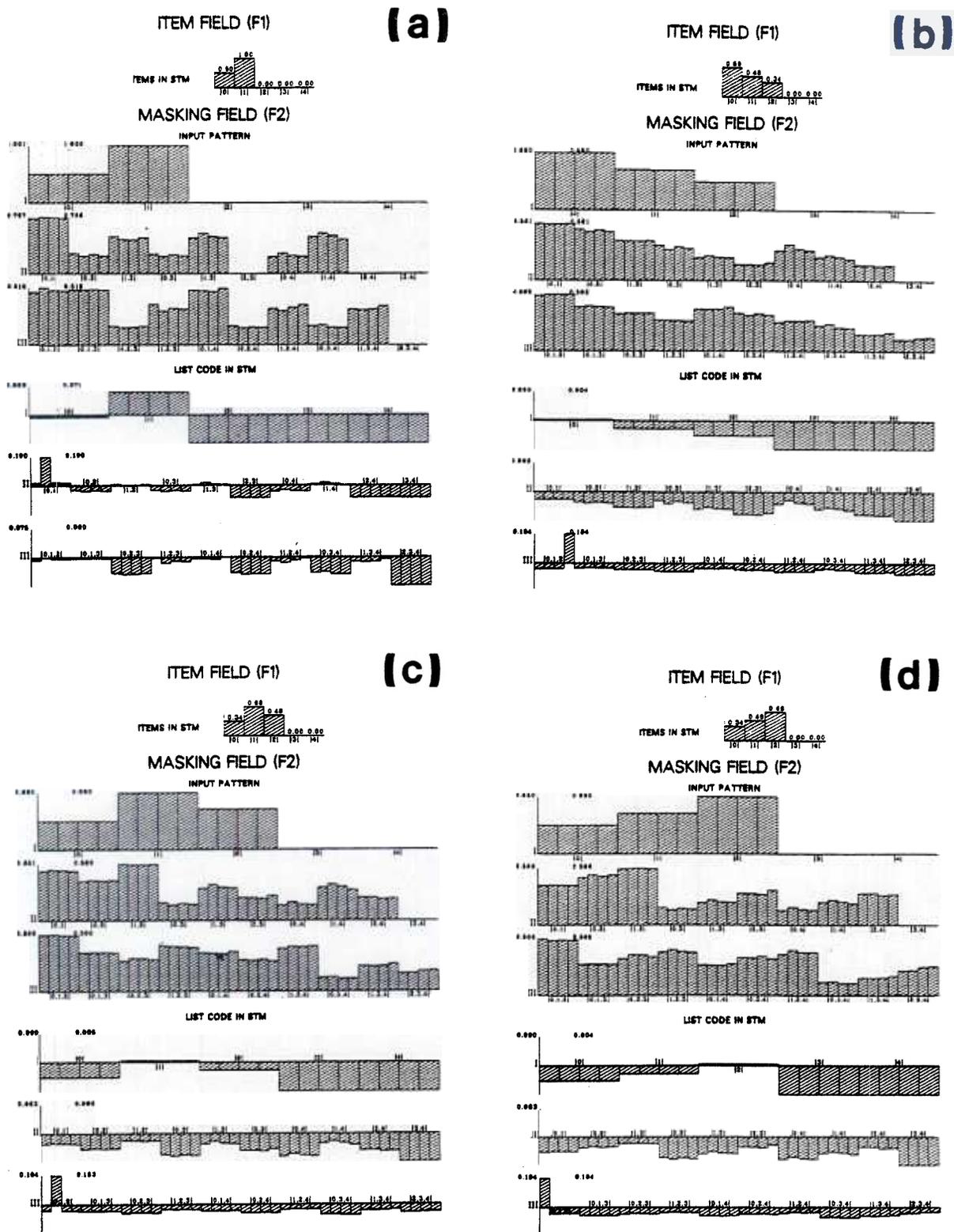


Fig. 7. (a) Different list code of type $\{0,1\}$ is maximally activated, but part codes $\{1\}$, are also activated with lesser STM weight. Due to the random growth of $F_1 \rightarrow F_2$ pathways, no predictive list codes are activated (to three significant digits). (b)–(d) When the STM pattern across F_1 includes three items, the list code in STM strongly activates an appropriate list code. Part groupings are suppressed due to the high level of predictiveness of this list code. Comparison of Fig. 6(a), 6(d), and 7(b) shows that as the item code across F_1 becomes more constraining, the list code representation becomes less distributed across F_2 .

provide the foundation for analyzing how learning within $F_1 \rightarrow F_2$ pathways interacts with the cooperative-competitive interactions within F_2 . Before proceeding to this analysis, we summarize the results due to the cooperative-competitive interactions in the absence of learning.

In these simulations, the level F_1 is called the item level, and the level F_2 is called the list level. These abstract terms are consistent with the dynamic properties of the two levels F_1 and F_2 and avoid pitfalls of alternative nomenclatures—such as letter level and word level—which do not adequately deal with the context-sensitivity of code reorganizations that occur during perceptual and cognitive processing. These nomenclature issues are discussed more fully in Grossberg^{20,28,29} and Grossberg and Stone.^{20,21}

In Fig. 3, a single item in F_1 is active. This item broadcasts positive inputs to a large number of nodes in F_2 . The input sizes over the target F_2 nodes are depicted by the heights of the bars in the three rows labeled Input Pattern. Each row lists all F_2 nodes which receive the same number of pathways from F_1 . The first row consists of F_2 nodes which receive one pathway, the second row consists of F_2 nodes which receive two pathways, and the third row consists of F_2 nodes which receive three pathways. In row 1, each F_2 node in the set labeled $\{i\}$ receives a pathway from the F_1 item node labeled $\{i\}$, $i = 0, 1, 2, \dots, 4$. Note that four F_2 nodes receive inputs from the $\{0\}$ F_1 node. In row 2, all F_2 nodes labeled $\{0,1\}$ receive pathways from the F_1 nodes $\{0\}$ and $\{1\}$. In row 3, all F_2 nodes labeled $\{0,1,2\}$ receive pathways from the F_1 nodes $\{0\}$, $\{1\}$, and $\{2\}$. The mathematical rules whereby these connections and input sizes are established are described in the Appendix.

The inputs to all the F_2 nodes which receive pathways from the F_1 node $\{0\}$ are positive. There are forty-four such nodes in Fig. 3. Despite this fact, the only F_2 nodes capable of becoming persistently active in STM are the nodes which receive pathways only from the active item node $\{0\}$. These are the F_2 nodes labeled $\{0\}$. The STM activities of all other F_2 nodes are quickly inhibited by the competitive feedback interactions within F_2 [Fig. 1(b)], despite the fact that many of these F_2 nodes also receive large excitatory inputs from F_1 . The equilibrium STM activities of the F_2 nodes are listed in three rows under the heading List Code in STM. These are the activities which the nodes store in STM after the network equilibrates to the entire input pattern. Figure 3 thus illustrates how F_2 can transform a widespread input pattern into a focal, and appropriate, STM activation.

Figures 4 and 5 further illustrate this property. Each panel in these figures represents the network response to a different input pattern. The panels are reduced relative to the scale of Fig. 3 to present a larger number of simulations. In Fig. 4(a) and (b), a different item at F_1 is activated. Each item generates a widespread input pattern to F_2 . Each input pattern is contrast-enhanced into a focal STM activation. This STM activation is restricted to the F_2 nodes which

receive pathways from only the active item node.

A comparison of Figs. 3, 4(a), and 4(c) illustrates the self-scaling property of masking field dynamics. Suppose that the list of items $\{0\}$, $\{1\}$ is successively received by F_1 . The list as a whole generates a different spatial pattern across F_1 [Fig. 4(c)] than does its first item (Fig. 3) or its second item [Fig. 4(a)] taken in isolation. The list as a whole also activates even more nodes than does either item taken separately, eighty-two nodes in all. Despite this fact, only a single F_2 node's activity is stored in STM. This F_2 node is, moreover, an appropriate node because it is one of the $\{0,1\}$ nodes that receive pathways only from the F_1 items $\{0\}$ and $\{1\}$. This comparison thus illustrates the ability of F_2 nodes which are activated by larger numbers of F_1 nodes to mask the activity of F_2 nodes which are activated by smaller subsets of F_1 nodes. This is a key property in the F_2 's functioning as a content-addressable memory.

A comparison of Figs. 4(c) and (d) illustrates the ability of F_2 to distinguish item patterns with different microstructures. In both of these figures, the same set of F_1 items— $\{0\}$ and $\{1\}$ —is activated, but a different spatial pattern of activity exists across the items. The spatial pattern in Fig. 4(c) may represent the list of items $\{0, 1\}$, whereas the spatial pattern in Fig. 4(d) may represent the list of items $\{0,1\}$. The simulations show that F_2 is sensitive to the item pattern as a whole, because F_2 can generate different STM responses to these patterns even though they activate the same unordered set of F_1 nodes. In particular, in Figs. 4(c) and (d), different F_2 nodes become active within the set of F_2 nodes which receives pathways only from items $\{0\}$ and $\{1\}$.

A comparison of Figs. 3, 4, and 5(a) illustrates a more demanding variant of these F_2 properties. As an ordered list of items $\{0\}$, $\{1\}$, $\{2\}$ is successively stored by F_1 , all the items become active at F_1 as the spatial patterns in Figs. 3, 4(c), and 5(a) evolve through time. The stored STM pattern in Fig. 5(a) is, however, restricted to a single F_2 node, which is one of the nodes receiving pathways only from items $\{0\}$, $\{1\}$, and $\{2\}$. Thus F_2 selects a content-addressable representation of the whole pattern at F_1 rather than of its constituent items.

A comparison of Figs. 5(a)–(c) makes the same point as the comparison of Figs. 4(c) and (d) but in a more demanding variation. In each of the panels in Fig. 5, the same unordered set of items— $\{0\}$, $\{1\}$, and $\{2\}$ —is active across F_1 . The different spatial patterns across F_1 represent different orderings of these items: $\{0,1,2\}$, $\{1,2,0\}$, and $\{2,1,0\}$, respectively. In each figure, a different F_2 node is activated. The active F_2 node is, moreover, one of the nodes that receives pathways only from the item nodes $\{0\}$, $\{1\}$, and $\{2\}$. Thus the content-addressable F_2 code is sensitive to the microstructure of the F_1 activity patterns.

Figures 6 and 7 describe the reactions of a masking field whose parameters are chosen to enable multiple groupings of F_1 patterns to be coded in STM at F_2 . Multiple groupings can emerge when the competitive interactions across F_2 are uniformly weakened. The

same input patterns were used as in Figs. 3–5. Comparison of Figs. 6(a), 6(d), and 7(b) shows how the automatic scaling property enables F_2 to update its STM representations based on all the groupings which it can detect as the F_1 activity pattern expands. In Fig. 6(a), item {0} most strongly activates the {0} nodes of F_2 but also weakly activates predictive F_2 nodes that represent groupings which include {0}. The F_2 nodes which receive an item pathway only from {0} have a maximal activity of 0.163. The F_2 nodes which receive two item pathways, including a pathway from {0}, have a maximal activity of 0.07. The F_2 nodes which receive three item pathways, including a pathway from {0}, have a maximal activity of 0.007. These activity weights characterize the degree of predictive evidence which the masking field possesses that each grouping is reflected in the input pattern.

In Fig. 6(d), the {0,1} spatial pattern across F_1 most strongly activates a node within the {0,1} subfield of F_2 but also weakly activates other nodes of F_2 which receive inputs from {0}. The activity levels are 0.246 and 0.04, respectively. Thus the multiple-scale self-similar interactions cause a reversal in activation strength when item {1} follows item {0} at F_1 : whereas the F_2 code for {0} is strong and for {0,1} is weak in response to item {0} at F_1 , the F_2 code for {0} is weak and for {0,1} is strong in response to the list {0}, {1} of items at F_1 . In Fig. 7(b), the {0,1,2} spatial pattern across F_1 most strongly activates a node within the {0,1,2} subfield of F_2 (with activity 0.184) but also weakly activates the {0} subfield of F_2 (with activity 0.004). The STM activity pattern across F_2 becomes more focused from Fig. 6(a) to 6(d) to 7(b) as increasing contextual information across F_1 reduces F_2 's predictive uncertainty.

VII. Adaptive Discovery of Segmentation Rules: The Adaptive Sharpening Property

The following criterion was applied to test the adequacy of associative learning laws for adaptive tuning of STM groupings across F_2 :

Adaptive sharpening property: Suppose that an arbitrary unfamiliar input pattern to F_1 generates an STM representation across F_2 . The LTM law for the adaptive weights within the $F_1 \rightarrow F_2$ pathways must learn from this $F_1 - F_2$ pairing so that, after learning occurs, the same input pattern to F_1 generates a spatially sharpened, or contrast-enhanced, STM pattern across F_2 .

In particular, if F_2 makes a choice in STM, as in Figs. 3–5, learning which satisfies the adaptive sharpening property acts to confirm this choice. More generally, the adaptive sharpening property prevents learning in the pathways which adaptively filter signals between F_1 and F_2 from destroying the good prewired properties of the masking field. Learning can accentuate the initial decisions due to interactions of the adaptive filter with the masking field but cannot upset this balance due merely to repeated presentations of the same F_1 pattern.

The adaptive sharpening property is not trivially satisfied by all the associative learning laws that one might reasonably wish to consider. This is because F_2 automatically reorganizes its STM reactions based on the global patterning of the inputs received by all its nodes (Sec. IV). A single LTM law, used in all the $F_1 \rightarrow F_2$ pathways, must be able to react to all possible combinations of activity patterns across F_1 and F_2 with adaptive sharpening and not a destruction of the global balance between $F_1 \rightarrow F_2$ inputs and $F_2 \leftrightarrow F_2$ interactions.

After such a LTM law is characterized, the adaptive sharpening property which it guarantees provides a foundation for studying how segmentation of an F_1 pattern into multiple groupings can be influenced by learning. For example, if a prescribed pattern across F_1 is repeatedly presented, this pattern becomes familiar by tuning the adaptive filter to code preferentially its most salient groupings in STM at F_2 . If a novel superset pattern at F_1 is then presented, that is, a pattern which includes the familiar pattern as a subpattern, the subset pattern groupings of the familiar pattern can coherently break away from the complementary superset groupings. The superset pattern can consequently be represented by an STM pattern of resonant parts, or structural groupings, across F_2 . In other words, prior adaptive tuning can enable a novel F_1 pattern to generate a directly accessed STM reaction across F_2 which segments the F_1 pattern into a distributed code of familiar groupings.

A related implication of the adaptive sharpening property is that a repeated presentation of a superset grouping may gradually mask otherwise possible subset groupings, unless the subset patterns are also frequently presented in their own right to F_1 . In intuitive terminology, a coherent set of familiar parts may come to represent the whole, or a more global segmentation may come to represent the whole, depending on the statistics of the input time series. Interactions between an adaptive filter and a masking field can hereby dynamically organize incoming input patterns into structural relationships which are learned from the statistics of a unique input environment, rather than trying to outguess the environment using prewired segmentation rules that are bound to fail in most environments.

VIII. Functional Unit of Associative Learning is a Spatial Pattern: A Nonlinear Non-Hebbian Learning Law

In our computer simulations, we demand a strict version of the adaptive sharpening property to direct our investigation of STM and LTM interactions. Given all the STM groupings in Figs. 3–7, we demanded that adaptive sharpening transform these groupings into STM choices at F_2 in response to repeated presentation of individual activation patterns at F_1 . In particular, we demanded that adaptive sharpening choose that F_2 population which was maximally favored by F_2 in response to that F_1 pattern before learning began.

As in Eq. (2), an LTM law which satisfies this property has the form

$$\frac{d}{dt}z_{ji} = \epsilon f(x_i)[-z_{ji} + LI_j]. \quad (4)$$

In Eq. (4), z_{ji} is the LTM trace in the pathway from the j th node v_j in F_1 to the i th node v_i in F_2 , I_j is the input from v_j , x_i is the STM activity of v_i , $f(x_i)$ is a nonlinear sampling signal that is activated by sufficiently large values of x_i , and ϵ and L are constants. Such a law was introduced into the associative learning literature in Grossberg^{30,31} and has since been used in many models.³²⁻³⁶ In particular, it was the associative law that was chosen to introduce the adaptive resonance theory,^{4,5} and it has played an important role in the complete numerical and mathematical characterization of an adaptive resonance circuit.⁷ Recent neurophysiological experiments about cortical and hippocampal processing have, moreover, supported this associative rule both qualitatively and quantitatively.³⁷⁻⁴¹

One reason such a law cannot be taken for granted is that it violates the Hebbian associative postulate⁴² that is the basis for many current learning models. On p. 64 of his classic book, Hebb proposed his famous Hebb postulate: "When the axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process takes place in one or both cells such that A 's efficiency, as one of the cells firing B is increased." The development of neural network models of conditioning since Hebb's work is discussed in detail by Levine.⁴³ The learning rule in Eq. (4) is called an associative rule, whereby LTM efficacy changes as a function of a time average of correlated presynaptic cell activities. Associative rules are often called Hebbian rules to honor the pioneering work of Hebb.⁴² This convention has, we believe, caused a great deal of confusion in the conditioning literature because different associative rules can support qualitatively different types of learning property.

The Hebb postulate seems plausible if one assumes that the unit of associative learning is a single cell's activity whose correlation with another cell's activity can increase the LTM strength of a pathway between the cells. A different associative rule is needed, however, if one agrees that the unit of associative learning is a spatial pattern of activity across a network of cells, as is required by Figs. 3-7. Then the correlation of a spatial pattern across F_1 with a cell's activity in F_2 enables the LTM traces in the set of pathways from F_1 to the active F_2 cell to encode the entire spatial pattern of activity into LTM. In this situation, an associative rule is needed which can encode both increases and decreases of LTM strength as a function of the pairing of cell activities, because an inactive cell v_j at F_1 should cause z_{ji} to approach zero when correlated with an active cell v_i at F_2 . Thus a change in the functional unit of learning from a single cell to a spatial pattern across a network of cells necessitates an associative rule that violates the Hebb postulate.

Another nonclassical property of the learning law (1) is that the sampling signal $f(x_i)$ is a nonlinear function of x_i ; in particular, $f(x_i)$ is a non-negative function which grows faster than linearly, for example, quadrat-

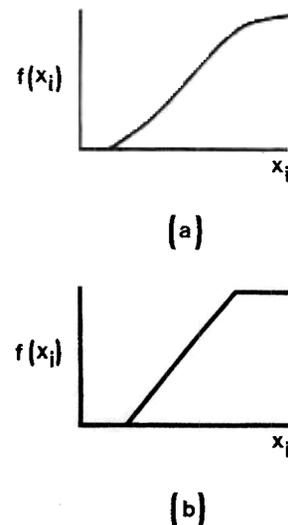


Fig. 8. Sampling functions: (a) a faster-than-linear increase of $f(x_i)$ with x_i over a large domain of x_i activities or (b) a threshold-linear increase of $f(x_i)$ over a large domain of x_i activities illustrates two possible types of sampling function.

ically or in a threshold linear fashion (Fig. 8), as x_i increases above zero. In Eq. (4), the sampling signal $f(x_i)$ multiplies the constant ϵ to determine a state-dependent learning rate $\epsilon f(x_i)$. Due to the faster-than-linear growth of $f(x_i)$ with x_i , small values of x_i determine a much slower learning rate in z_{ji} than do large values of x_i . Consequently, F_2 cells which acquire an initial STM advantage can greatly amplify that advantage by speeding up the learning of their contiguous LTM traces. In contrast, F_2 cells whose activities remain below zero cannot trigger any learning in their contiguous LTM traces because $f(x_i) = 0$ if $x_i \leq 0$. This property justifies calling $f(x_i)$ a sampling signal.

Speaking intuitively, the state-dependent learning rate $\epsilon f(x_i)$ in Eq. (4) says that learning can occur only at LTM traces z_{ji} whose target activities x_i are chosen by the cooperative-competitive decision-making machinery of F_2 . All LTM traces z_{ji} whose F_1 items receive positive inputs I_j can influence F_2 's decision making by multiplying these inputs on their way to F_2 . In contrast, a much smaller number of LTM traces can learn from the decision-making process due to the property that F_2 chooses a compressed recognition code which is much less distributed than the input patterns which it receives. In this sense, LTM readout through the adaptive $F_1 \rightarrow F_2$ filter and LTM readin by the associative law (4) are at least partly dissociated due to intervention of the cooperative-competitive interactions within F_2 . Such dissociation of signaling and learning by code compression mechanisms is not easily accomplished using an autoassociator. It was, in part, to avoid the attendant limitations, notably the instabilities, of learning by an autoassociator that hierarchical neural networks with multiple network levels began to be designed in the late 1960s. These developments included the introduction of hierarchies of nonlinear

avalanche-type circuits for spatiotemporal pattern learning, of prewired and adaptive pattern recognition circuits, including competitive learning and adaptive resonance circuits, and of circuits for the active regulation of information processing by reinforcement and homeostatic mechanisms to attentively direct information processing toward the realization of desired goals. Several of these early contributions are described in Grossberg.¹³ More recent contributions built on this foundation are brought together in Grossberg.^{14,15}

IX. Computer Simulations: Adaptive Sharpening of Multiple Groupings

In this section, we illustrate the adaptive sharpening property by showing how the multiple groupings depicted in Figs. 6 and 7 are adaptively transformed into STM choices when the learning law (1) is used in the $F_1 \rightarrow F_2$ adaptive filter. We have demonstrated in addition that the learning law (1) confirms all the STM choices described in Figs. 3–5. These simulations are not, however, displayed herein.

All the figures in this section describe the equilibrium STM choice that is generated by F_2 when the learning process approaches a limit in response to sustained presentation of each input pattern. The fact that the system always approaches equilibrium STM and LTM values is in itself a fundamental property, since feedback interactions between STM (fast) and LTM (slow) processes can easily lead to sustained oscillations, such as traveling waves, bursts, or even chaotic oscillations.^{44–48} In some physical systems, complex oscillations are functionally desirable. In the present applications, they are not. Theorems which guarantee a global approach to equilibrium in related cooperative–competitive feedback networks are found in Cohen and Grossberg,²² Elias and Grossberg,⁴⁹ Grossberg and Levine,³ and Grossberg.^{2,13}

To supplement the global theorem about masking fields,²⁴ we have studied the approach within F_2 to STM and LTM limits using a variety of techniques. The simplest technique uses a singular approximation to the full dynamic system. In the full dynamic system, STM reacts to an input pattern more quickly than does the slower LTM learning process. In a singular approximation, it is assumed that LTM does not change at all until the STM activities have almost reached an equilibrium value. Then the LTM learning process is switched on, and both STM and LTM interact until they conjointly approach equilibrium. Using such a singular approximation, a much faster LTM learning rate [namely, a larger ϵ in Eq. (4)] can be used without significantly changing the equilibrium STM and LTM patterns that are found using the full system. A computer simulation of a singular system can thus be done much more quickly than a simulation in which the full system is integrated with a small ϵ until it reaches equilibrium. Carpenter^{44,50,51} and Fenichel⁵² have proved theorems which describe conditions in which solutions of a nonlinear dynamic system with fast and slow processes lie close to solutions of a singular approximation to the full dynamic system.

Once we confirmed the adaptive sharpening property using a singular approximation, we did simulations with the full system using several different choices of the learning rate parameter ϵ in Eq. (4). Our goal was to understand how fast the learning rate could be before it might disrupt the adaptive sharpening process. More generally, we wanted to understand whether LTM changes must necessarily occur more slowly than STM changes to achieve basic functional properties such as adaptive sharpening.

Figures 9 and 10 describe the equilibrium patterns in a singular system, all of whose parameters, except the learning rate ϵ , are the same as in the simulations of Figs. 6 and 7. In Figs. 6 and 7, the learning rate $\epsilon = 0$. In Figs. 9 and 10, ϵ was set equal to zero until the STM traces across F_2 were close to equilibrium. Then we switched ϵ to equal 1 to allow the full system to approach equilibrium.

Comparison of Figs. 9 and 10 with Fig. 6 and 7 shows that the adaptive sharpening property is obtained. Comparison of the input patterns to F_2 nodes without learning and after learning shows how LTM changes in the $F_1 \rightarrow F_2$ pathways alter the total inputs to the F_2 nodes and thereby bias the competitive feedback process within F_2 to make global choices in STM.

Having achieved the adaptive sharpening property in a singular system, we demonstrated the property, without a change of parameters other than ϵ , in the full system. In one successful series of full system simulations, the choice $\epsilon = 0.01$ was made. In all these simulations, the decay rate of STM activities across F_2 , in the absence of internal feedback signals, was chosen equal to 1. Thus the adaptive sharpening property was confirmed in the full system using plausible relative rates of STM and LTM change. Figure 11 depicts a computer simulation of how the LTM values in a subset of $F_1 \rightarrow F_2$ pathways changed through time due to learning. The simulations show that the present masking field and associative learning laws are sufficient to generate all the properties that we have claimed.

Despite these successful results, a finer study of the transient behavior of the full system, before equilibrium was reached, raised a number of issues which have led us to propose a refinement of masking field design which promises to generate even stronger properties.

X. Transient STM Surge Precedes Competitive Contrast Enhancement

Two major phases in F_2 's reaction to an input pattern at F_1 can be identified. In Phase 1, the input pattern starts to deliver signals to F_2 nodes via the $F_1 \rightarrow F_2$ pathways, and many F_2 nodes thereby start to become activated. As these nodes become activated, they begin to generate feedback signals, notably competitive signals, to other F_2 nodes [Fig. 1(b)]. The balance between excitatory and inhibitory signals to each node quickly contrast enhances the input pattern from F_1 and generates the more focal STM reactions at F_2 which are depicted in Figs. 3–7. In the absence of

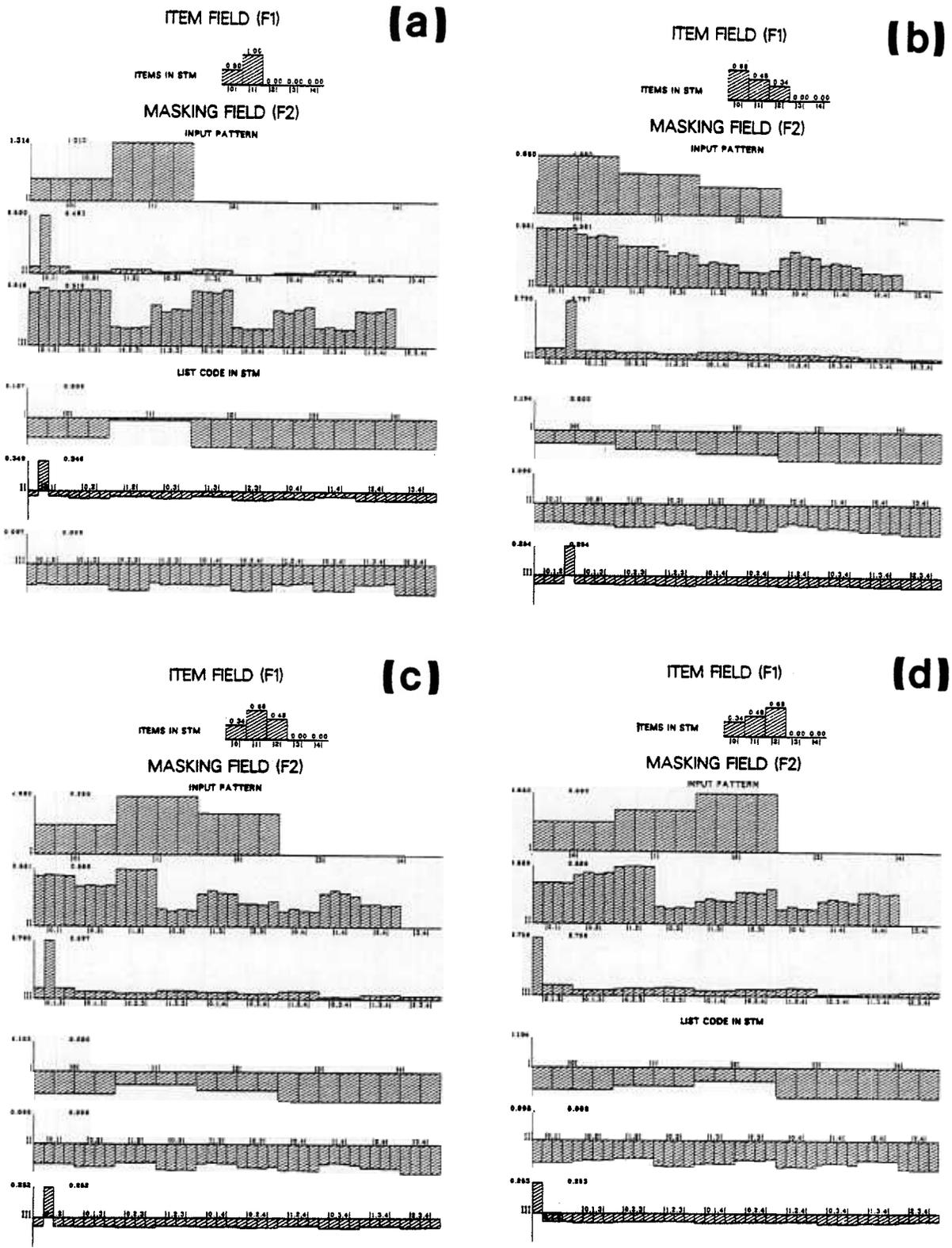


Fig. 10. Adaptive sharpening in response to the input patterns of Fig. 7.

additional learning, reset, or habituated mechanisms, these focal STM reactions are stored by the balance of inputs and feedback signals within F_2 . Phase 2 consists in the contrast enhancement and storage of these

STM patterns. In the language of the Hough transform, the positive and negative votes cast by the masking field cancel both off-peaks and false peaks caused by the adaptive filter.

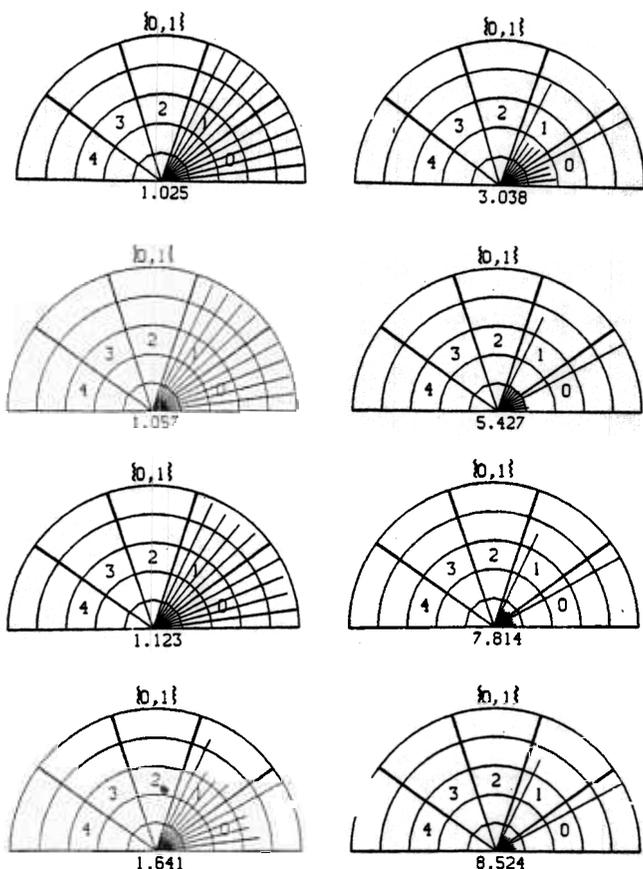


Fig. 11. Changes in LTM strength through time due to learning. Lengths of the spokes are proportional to the sizes of the corresponding LTM traces. LTM traces at increasing times are plotted down column 1 and then column 2. The numbers under each figure designate the size of the maximal LTM trace in that figure.

Figure 12 summarizes a computer simulation of the transition from Phase 1 to Phase 2. The parameters are the same as those in Fig. 4(c). Each successive picture depicts the STM activities of F_2 nodes at a later time after the onset of the input pattern to F_1 .

In summary, after an input pattern activates F_1 , there is a massive but transient activity burst across F_2 which is quickly sculpted by F_2 's feedback interactions. The key question is: How quickly, relative to the learning rate?

XI. Spurious Learning of the Transient Surge

The following problem can arise if the learning rate is too fast. Suppose that ϵ in Eq. (4) is chosen so large that significant learning can occur during Phase 1. Then many F_2 nodes v_j can sample the F_1 activity pattern because their learning rates $\epsilon f(x_i)$ are large during Phase 1. In contrast, if ϵ is small, insignificant learning occurs during Phase 1 because the duration of Phase 1 is not long enough to integrate a large LTM change at rate $\epsilon f(x_i)$. During Phase 2, only those F_2 nodes which are selected by the internal feedback interactions within F_2 can sample the input pattern and

thereby tune their LTM traces, because $f(x_i) = 0$ at all other F_2 nodes.

In summary, if the learning rate is fast relative to the duration of Phase 1, learning is not controlled by the masking field's grouping process. Moreover, such spurious learning can interfere with the masking field's ability to select a predictive grouping during Phase 2.

Figure 13 describes computer simulations which illustrate how a change in the learning parameter ϵ can alter the equilibrium grouping that is finally learned. Choosing ϵ too large can also cause violations of the adaptive sharpening property. Figure 13(a) repeats Fig. 4(c) to facilitate comparison of the no-learning case with several learned groupings. In Fig. 13(b)–(d), ϵ was chosen equal to 1, 0.1, and 0.01, respectively. When $\epsilon = 1$, F_2 chose the {0} nodes. When $\epsilon = 0.1$, F_2 selected both {0} and {0,1} nodes. When $\epsilon = 0.01$, F_2 chose the correct {0,1} node. In all cases, the learned F_2 grouping exhibited a form of adaptive sharpening. In Fig. 13(b), however, the chosen F_2 nodes do not code information about item {1} at all.

The reason for this bias toward {0} nodes at fast learning rates can be traced to properties of the Phase 1 surge. In Fig. 11, an initial advantage of {0} nodes above {0,1} nodes can be seen before the self-scaling feedback interactions within F_2 reverse this advantage.

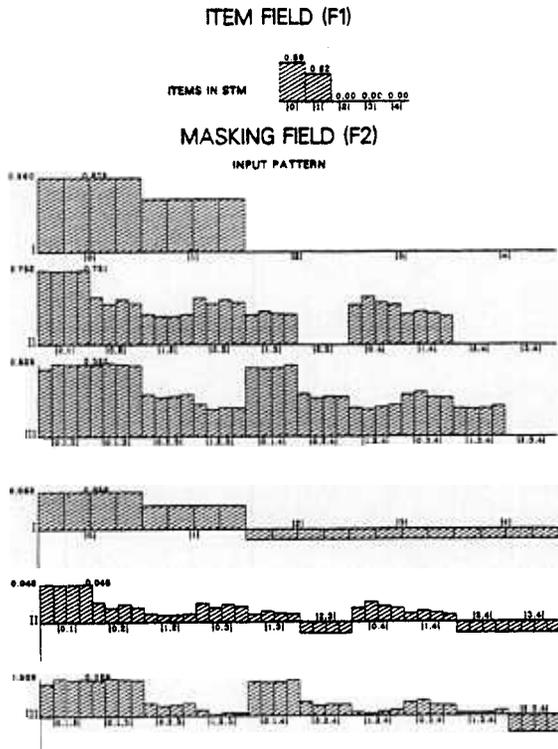
These results illustrate that, in the masking heretofore described, there exists a trade-off between the rate of cooperative-competitive decision making by F_2 and the rate of learning by the $F_1 \rightarrow F_2$ adaptive filter. Learning must be sufficiently slow relative to the decision-making process to avoid spurious learning of transient decisions. The results also show, however, that a proper scaling of rates, with LTM ~ 100 times slower than STM, can avoid this sampling problem. On the other hand, these simulations also call attention to the following design problem, should one wish to be freed from concerns about the proper scaling of slow LTM rates against fast STM rates.

XII. Structurally Stabilized Learning

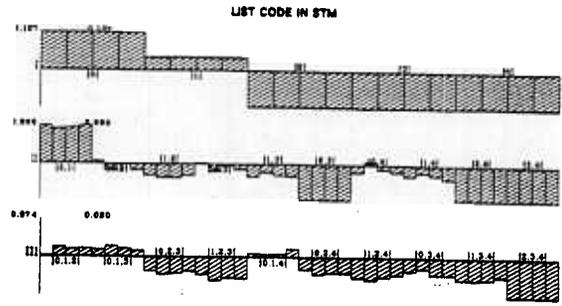
The design problem that is raised by the simulations in Figs. 12 and 13 can be stated as follows:

Structurally stabilized masking field: Does there exist a modification of masking field design which overcomes the Phase 1 surge (Fig. 12) and the fast learning (Fig. 13) problems, given essentially any choice of STM and LTM rate parameters?

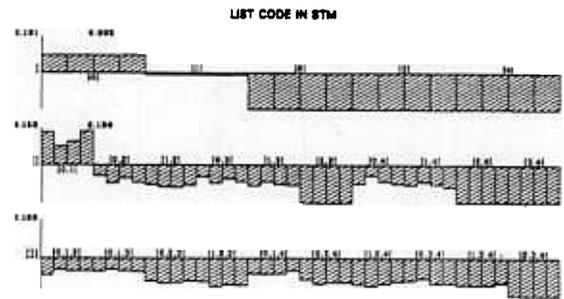
We now suggest a modification of a masking field's internal connections which can substantially reduce the Phase 1 surge. We also suggest a modification of a masking field's internal connections which enables it to learn in a way that is insensitive to whatever residual surge may still occur. We hereby overcome a problem that may arise due to improperly chosen rates by modifying the system's interactive structure to work well given a more careless choice of rates. Otherwise expressed, a structurally stabilized masking field is more fault-tolerant of a poor choice of processing rates.



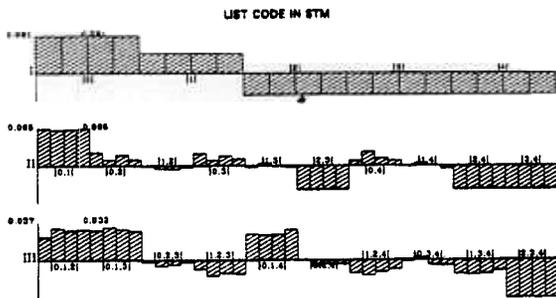
(a)



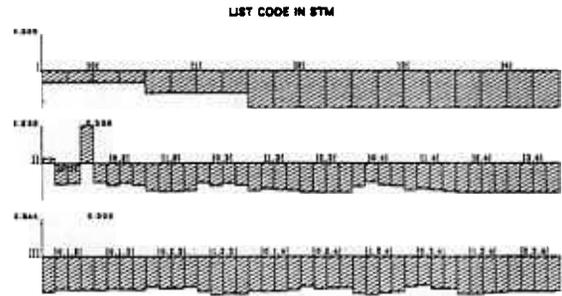
(c)



(d)



(b)



(e)

Fig. 12. Transition through time from a widespread Phase 1 burst of activation across F_2 to a more focal Phase 2 activation that is under inhibitory control. Successive list codes in columns 1 and 2 are evaluated at logarithmically increasing times.

XIII. Feedforward and Feedback Sharing of Internal Feedback Pathways

The Phase 1 surge is due to the fact that all $F_1 \rightarrow F_2$ inputs are excitatory. We propose that, before these inputs can influence their target cells in F_2 , they activate internal feedback pathways within F_2 which bal-

ance the excitatory signals with inhibitory signals (Fig. 14). Inhibitory signals will, therefore, be registered at the same moment that excitatory signals are registered. There does not exist a time interval during which excitatory inputs can activate a Phase 1 burst that is not controlled by inhibitory signals.

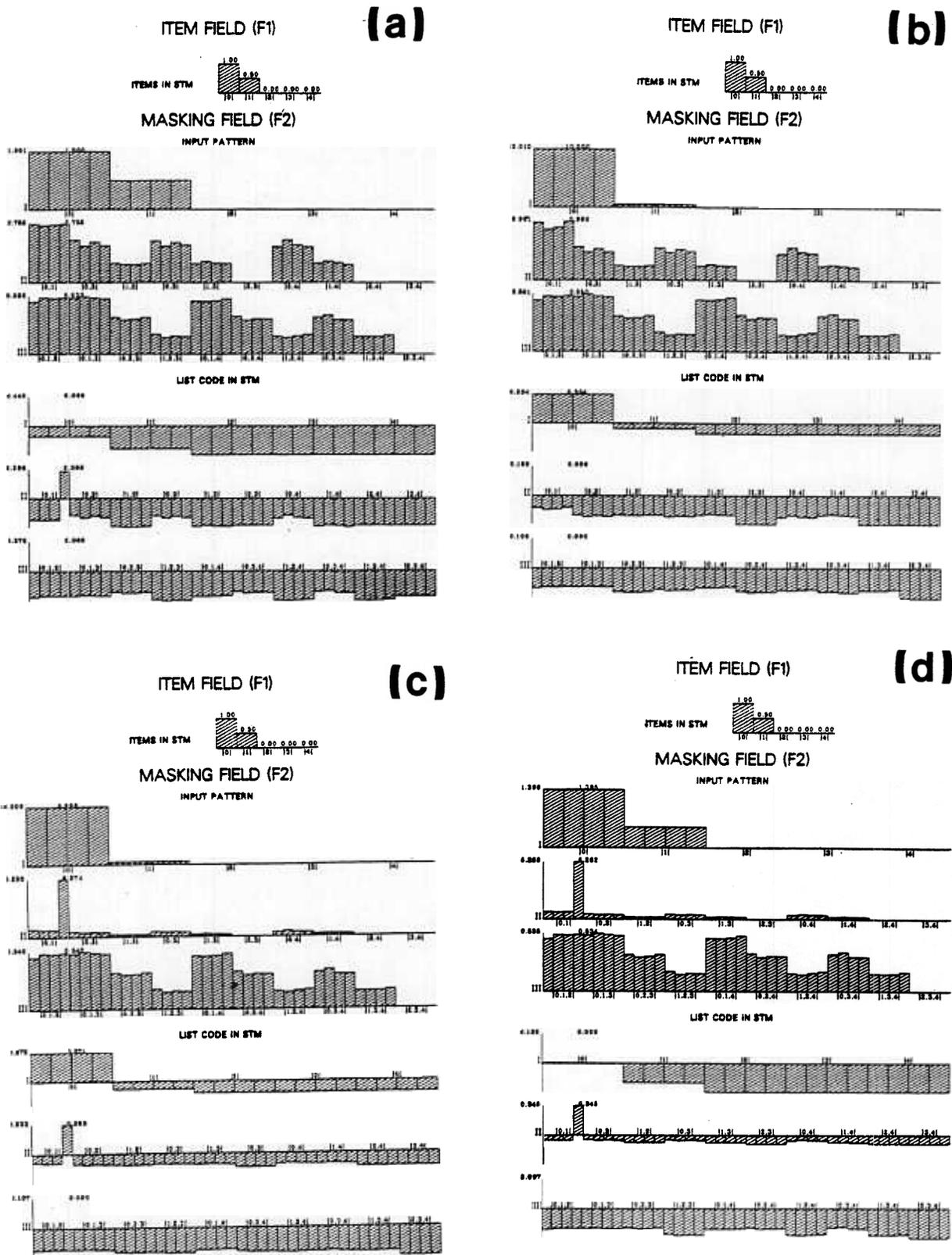


Fig. 13. Comparison of the list code stored in STM at F_2 in a no-learning case (a) with the list code that is stored after learning with (b) $\epsilon = 1$, (c) $\epsilon = 0.1$, and (d) $\epsilon = 0.01$. The learning rates $\epsilon = 1$ and $\epsilon = 0.1$ are both too fast to achieve the adaptive sharpening property because the LTM traces can learn significantly during the Phase 1 burst.

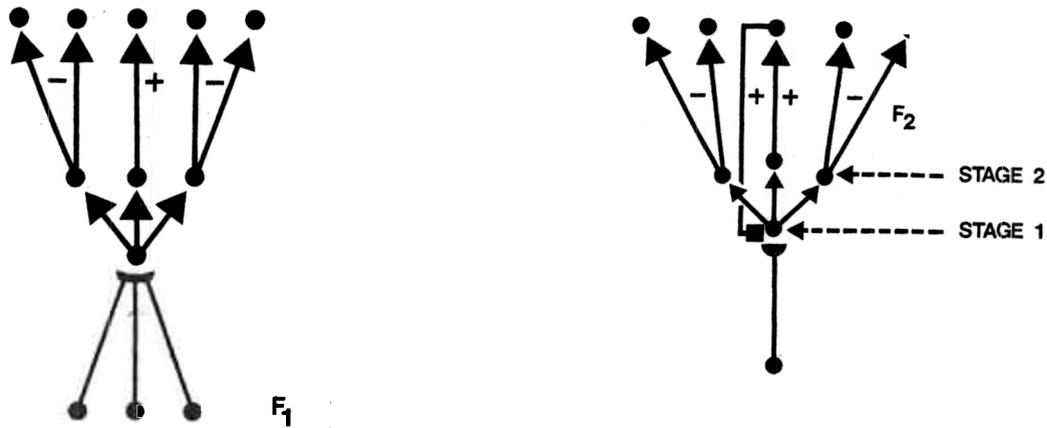


Fig. 14. Phase 1 burst can be partially eliminated by causing each $F_1 \rightarrow F_2$ input to activate both excitatory and inhibitory pathways leading to target F_2 nodes.

How should these excitatory and inhibitory signals be chosen? In particular, how can they be chosen so that they do not upset the feedback interactions that are the basis of a masking field's grouping properties? A simple answer is available: Let the feedforward inputs from the adaptive filter use the same interneurons, or internal feedback cells, that are used to define the masking field (Fig. 15). Such a design was first described by Grossberg.⁵ Then the uncontrolled Phase 1 burst is prevented by a structural mechanism which immediately begins the grouping process when it receives an input burst.

As Fig. 15 shows, the masking field is now broken into three internal stages. Stage 1 receives the excitatory inputs from F_1 . Stage 2 contains the internal pathways which distribute excitatory and inhibitory signals across the masking field. Stage 3 contains the target cells of these internal pathways. These target cells always receive a mixture of excitatory and inhibitory signals. They are never exposed to an uncontrolled Phase 1 burst. The Stage 3 cells give rise to topographic positive feedback pathways to their Stage 1 source cells. These positive feedback pathways close the feedback loops within the masking field. Using these stages, the internal feedback interactions of the masking field remain unchanged, yet the F_1 inputs engage these interactions before they influence Stage 3 cells.

The architecture in Fig. 15 prevents a totally uncontrolled Phase 1 burst from occurring. On the other hand, the internal feedback within the masking field does not instantaneously select an equilibrium grouping. Rapidly cycling feedback signals within the masking field select such a grouping. It remains to say how the LTM traces within the $F_1 \rightarrow F_2$ pathways can be buffered against learning activity patterns that are far from equilibrium.

XIV. Internal Feedback as a Sampling Signal

The main problem to be overcome is clearly illustrated in Fig. 15. Although the Stage 3 cells receive a mixture of excitatory and inhibitory signals, the Stage

1 cells receive only excitatory signals. Moreover, the $F_1 \rightarrow F_2$ pathways about the Stage 1 cells. What prevents the LTM traces within the endings of these pathways from being activated by sampling signals from the Stage 1 cells?

We hypothesize that the sampling signal which activates an LTM trace is not derived from a Stage 1 cell. Rather, the sampling signal is activated by feedback from a Stage 3 cell (Fig. 15). Many Stage 3 cells will be immediately inhibited by Stage 2 interneurons when an input pattern turns on. Use of Stage 3 cells as a source of sampling signals enables masking field interactions to restrict learning from its very first moments of interaction, because many Stage 1 cells which are initially activated by F_1 inputs correspond to Stage 3 cells which are never activated during the ensuing grouping process. To instantiate this constraint, we simply replace Eq. (4) by equation

$$\frac{d}{dt}z_{ji} = ef(x_i^{(3)})(-z_{ji} + LI_j), \quad (5)$$

where $x_i^{(3)}$ is the activity of the i th cell population in Stage 3 of the masking field.

The concept that internal feedback signals generate LTM sampling signals was introduced in Grossberg.⁵³ We now believe that it may be a design principle which is widely used in the brain, whether the feedback signal is intercellularly generated, as in Fig. 15, or intracellularly generated by a network of biochemical feedback interactions. Computer simulations which illustrate how such a feedback signal regulates learning are described in a related type of circuit for combining cooperative-competitive and associative mechanisms in Grossberg and Schmajuk.¹⁹ Some of the properties

which can be used to experimentally test for this design are now summarized.

XV. Dissociation of LTM Readin and Readout: Feedback as a Neural Modulator

Readout of LTM occurs when an LTM trace multiplicatively gates a signal on its way from F_1 to F_2 (Appendix). In the masking fields which we have simulated, LTM is read out into the same F_2 cells which enable the LTM traces to sample, or read in, new LTM values [Fig. 1(b)]. The design in Fig. 15 structurally dissociates the processes of LTM readout and LTM readin by enabling some Stage 1 cells to become activated without triggering any learning, no matter how fast the learning rate is chosen.

The feedback signals from Stage 3 to Stage 1 do not, however, act only as sampling signals. They must also activate their target Stage 1 cells to close the internal nonlinear feedback loops which enable the masking field to select its compressed recognition code for storage in STM. If the feedback signals can activate Stage 1 cells, how can the LTM traces which abut Stage 1 cells tell the difference between the activation of Stage 1 cells by inputs from F_1 and activation of Stage 1 cells by feedback signals from Stage 3? If such a distinction cannot be made, a functional dissociation of LTM readout and LTM readin cannot be achieved.

There exist two types of solution to the dissociation problem: a dynamical solution and a structural solution, which can be instantiated either chemically or electrically. In the dynamical solution, the LTM traces continue to use Stage 1 cells as sampling signals, but the threshold for activating the sampling signal $f(x_j)$ is chosen high. It is assumed that Stage 1 cells can only be activated enough to exceed the sampling threshold when their direct activation by inputs from F_1 is supplemented by large positive feedback signals from Stage 3 cells. Although such a mechanism may be adequate to solve simple learning problems, it is inadequate in a complex learning system. For example, in a masking field, if the sampling threshold is chosen too small, the Phase 1 surge can be learned. If the sampling threshold is chosen too large, many groupings which should induce adaptive tuning will fail to do so. We have performed many computer simulations which support our contention that such a design is not robust.

In contrast, a structural solution to the problem is manifestly robust. In one such structural solution, the feedback signal is delivered via a different chemical transmitter than the chemical transmitter which gates signals from F_1 to F_2 and regulates learned LTM changes in $F_1 \rightarrow F_2$ pathways. Term $f(x_j^{(3)})$ in Eq. (5) can then be realized by a modulatory action of the feedback transmitter on the feedforward transmitter. A modulatory action of catecholaminergic transmitters on learning by cholinergic transmitters has been reported in neural data (e.g., Friedhoff^{54,55}) and has also been postulated in neural models of classical and instrumental conditioning.^{13,14}

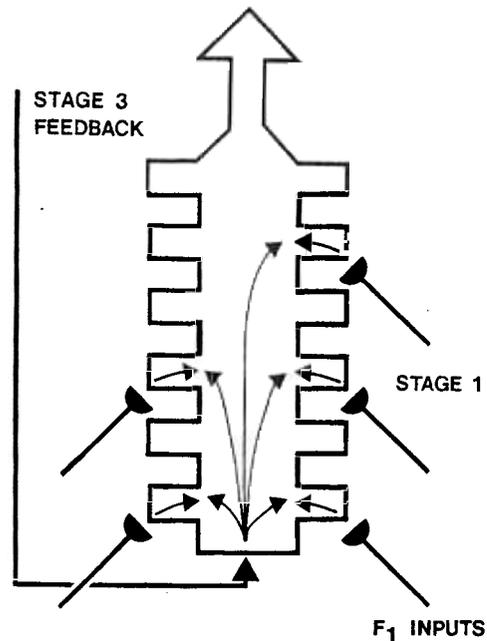


Fig. 16. Stage 1 cells receive inputs from F_1 on branches, or dendrites, whose local activations summate to generate the total output signal to Stage 2 cells. Stage 3 feedback signals cause a massive global activation of Stage 1 cells which triggers the biophysical events that enable LTM traces (hemidisks) in active F_1 pathways to learn.

The use of two transmitters enables both transmitter systems to electrically activate Stage 1 cells, yet also enables LTM traces abutting Stage 1 cells to distinguish between feedback signals from Stage 3 and their aggregate effects on Stage 1 cells. In one microscopic realization of such a dual-transmitter system, either transmitter can cause macromolecular changes in the cell membranes of Stage 1 cells which enable electrical activation to occur, but only their conjoint action can cause those macromolecular changes which enable the learning process to unfold. Data concerning associative learning in invertebrates implicate a Ca^{2+} -dependent membrane current which is activated only when pairs of critical events occur together.⁵⁶⁻⁵⁸ A catecholaminergic transmitter may, moreover, participate in the activation of this Ca^{2+} current.⁵⁷ The feedback signal from Stage 3 to Stage 1 plays an analogous formal role in the circuit depicted in Fig. 15. The suggestion that associative learning may depend on a Ca^{2+} current was made in Grossberg,^{30,59} based on the fragmentary biochemical evidence then available, to explain how a learning equation such as Eq. (4) could be physically realized.

Another structural solution of the problem can also be envisaged. In this solution, each $F_1 \rightarrow F_2$ pathway causes a local change in its target cell membranes at Stage 1 (Fig. 16). These local membrane channels cause local changes in potential which are summated by the Stage 1 cells before these cells activate Stage 2 cells. Feedback signals from Stage 3 cells cause global action potentials throughout the Stage 1 cells. These global action potentials activate membrane channels

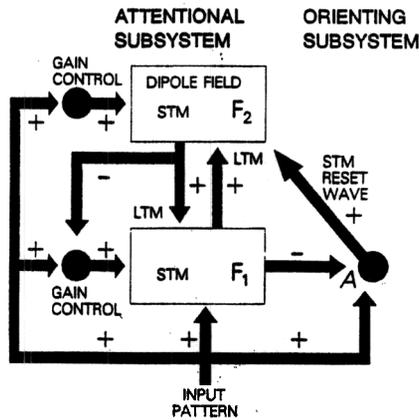


Fig. 17. Adaptive resonance theory architecture: Two successive stages F_1 and F_2 of the code learning, or attentional, subsystem are depicted. In addition to a bottom-up adaptive filter $F_1 \rightarrow F_2$, a top-down adaptive filter $F_2 \rightarrow F_1$ learns templates, or critical feature patterns, which are matched against input patterns at F_1 . This matching process protects already learned codes against unstable recoding by activating the orienting subsystem A. A reset wave from A quickly resets the list code at F_2 before it can be erroneously associated with the current activity pattern at F_1 and initiates a self-adjusting search for a better list code. The gain control channels enable F_1 to distinguish between bottom-up input patterns and top-down templates during the matching process and enable stored STM activity to decay when gain control is shut off.

which cannot be activated merely by local signals from F_1 . These membrane channels enable learning to occur within the abutting LTM traces. This possibility was used in Grossberg⁵³ to discuss classical conditioning within the hippocampus and in Grossberg¹ to discuss possible sites of neocortical conditioning. It is a structural rather than a dynamic scheme because all feedback signals are assumed to trigger the global change which enables learning to occur, not only feedback signals which can summate sufficiently with feed-forward signals. Such a structural scheme could also be used to trigger a Ca^{2+} -dependent current when the cell is globally activated. This type of structural scheme is used to interpret the systems simulated by Grossberg and Schmajuk.¹⁹

XVI. Concluding Remarks: Self-Stabilization of Learning Within ART Circuits

Subsequent work on masking fields will proceed in several directions. In addition to quantitative analyses of the structurally stabilized masking field architecture summarized in Fig. 15, each masking field design will be embedded within the total architecture which defines an adaptive resonance theory (ART) module (Fig. 17). Such an ART architecture is capable of self-organizing and self-stabilizing its recognition codes in response to arbitrary orderings of arbitrarily many and arbitrarily complex input patterns. Carpenter and Grossberg⁷⁻¹⁰ have numerically and mathematically characterized ART architectures which use a masking field F_2 that always makes a global choice. In this special case, they have rigorously proved that the learned recognition code self-stabilizes in response to arbitrary orderings of arbitrarily

many arbitrarily chosen binary input patterns.⁷ Using masking fields capable of coding multiple groupings, the design of masking field hierarchies $F_1 \leftrightarrow F_2 \leftrightarrow \dots \leftrightarrow F_n$ becomes possible. Such hierarchies show promise of being able to self-organize highly abstract grouping, hypothesis testing, and logical operations. The design of masking field hierarchies can now be pursued by combining the results of Carpenter and Grossberg with the results described herein.

The work of Michael Cohen was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-86-C-0037) and the National Science Foundation (NSF IST-84-17756). The work of Stephen Grossberg was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-86-C-0037, the Army Research Office (ARO DAAG-29-85-K-0095), and the National Science Foundation (NSF IST-84-17756).

We wish to thank Cynthia Suchta and Carol Yanakakis for their valuable assistance in preparing the manuscript and illustrations.

Appendix: Mathematical Description of a Masking Field

A. Shunting On-Center Off-Surround Networks

The cell populations v_i of a masking field have potentials $x_i(t)$, or STM activities, which obey the membrane equations of neurophysiology, namely,

$$C \frac{\partial V}{\partial t} = (V^+ - V)g^+ + (V^- - V)g^- + (V^p - V)g^p. \quad (\text{A1})$$

In Eq. (A1), $V(t)$ is a variable voltage, C is a constant capacitance, the constants V^+ , V^- , and V^p are excitatory, inhibitory, and passive saturation points, respectively; and the terms g^+ , g^- , and g^p are conductances which can vary through time as a function of input signals. Due to the multiplicative relationship between conductances and voltages in Eq. (A1), a membrane equation is also said to describe a shunting interaction.

In a masking field, the cells are linked together via recurrent, or feedback, on-center off-surround interactions [Fig. 1(b)]. The properties of a masking field are thus part of the general theory of shunting recurrent on-center off-surround networks. Grossberg^{15,60} reviews general properties of this class of network.

The masking field equations are most simply built up in stages. Rewrite Eq. (A1) for the potential $x_i(t)$ in the form

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)P_i - (x_i + C)Q_i, \quad (\text{A2})$$

where 0 is the passive equilibrium point, $B(> 0)$ is the excitatory saturation point, and $-C(\leq 0)$ is the inhibitory saturation point. Term P_i is the total excitatory input, and term Q_i is the total inhibitory input to v_i . Potential $x_i(t)$ can vary between B and $-C$ in Eq. (A2) as the inputs P_i and Q_i fluctuate through time. The multiplication of P_i and Q_i by terms which include x_i endow the circuit with properties of automatic gain control.

The excitatory input P_i is a sum of two components: the total input from the item field plus a positive feedback signal from v_i to itself [Fig. 1(b)]. Thus P_i can be written in the form

$$P_i = \sum_{j \in J} I_j p_{ji} z_{ji} + Df(x_i). \quad (\text{A3})$$

In Eq. (A3), term I_j is the output from the item node $\{j\}$, p_{ji} is the connection strength of the pathway from v_j in F_1 to v_i in F_2 , and z_{ji} is the LTM trace, or adaptive weight, within this pathway. Term $Df(x_i)$ describes the positive feedback signal from v_i to itself. This feedback signal enables v_i to store activities in STM after the inputs I_j terminate.

The inhibitory input Q_i in Eq. (A2) is a sum of feedback signals $g(x_m)$ from other populations v_m in the masking field. Thus Q_i can be written as

$$Q_i = \sum_{m \in I} g(x_m) E_{mi}. \quad (\text{A4})$$

B. Mass Action Interaction Rules

The notation in Eqs. (A2)–(A4) is now refined to express the fact that the cells in different subfields of a masking field possess different parameters. To express the fact that an F_2 population receives F_1 pathways only from a prescribed (unordered) set J of items, let $x_i^{(j)}$ denote the STM activity of an F_2 population $v_i^{(j)}$ which receives input pathways only from the set J of F_1 items. Any number of different populations $v_i^{(j)}$ in F_2 may correspond to each fixed set J of F_1 items. Equation (A2) is replaced by the equation

$$\frac{d}{dt} x_i^{(j)} = -Ax_i^{(j)} + (B - x_i^{(j)})P_i^{(j)} - (x_i^{(j)} + C)Q_i^{(j)}, \quad (\text{A5})$$

which holds for all unordered sets J of F_1 items that can selectively send pathways to nodes in F_2 .

Equation (A3) for the excitatory input P_i is replaced by

$$P_i^{(j)} = \sum_{j \in J} I_j p_{ji}^{(j)} z_{ji}^{(j)} + D_{|J|} f(x_i^{(j)}). \quad (\text{A6})$$

In term $D_{|J|}$, notation $|J|$ denotes the size of set J . Thus $D_{|J|}$ depends on the size of set J but not on the items in set J . Thus the excitatory feedback coefficient $D_{|J|}$ is sensitive to the spatial scale of the population $v_i^{(j)}$.

Equation (A4) for the inhibitory input Q_i is refined in several stages. Function $Q_i^{(j)}$ obeys an equation of the form

$$Q_i^{(j)} = \sum_{m \in K} g(x_m^{(K)}) E_{KJ}, \quad (\text{A7})$$

where coefficient E_{KJ} determines the strength of the inhibitory feedback pathway from $v_m^{(K)}$ to $v_i^{(j)}$. This path strength depends only on the unordered sets K and J of items to which $v_m^{(K)}$ and $v_i^{(j)}$ respond. Coefficient E_{KJ} expresses the randomness of the self-similar growth process between populations in F_2 (Ref. 11) as follows:

Mass action interactions:

$$E_{KJ} = F_{|J|} G_{|K|} H_{|K \cap J|}. \quad (\text{A8})$$

By Eq. (A8), E_{KJ} is a product of three factors. Each factor depends only on the size of an unordered set of items. These unordered sets are set K , set J , and their intersection $K \cap J$. Equation (A8) can be explained by assuming that the inhibitory interaction strength from $v_m^{(K)}$ to $v_i^{(j)}$ is the result of an interaction of three independent random factors. The net strength E_{KJ} can thus arise from a statistically independent interaction between growth factors that depend on the sizes of K , J , and their overlap. By putting together all these constraints, we find the following:

Masking field equations:

$$\frac{d}{dt} x_i^{(j)} = -Ax_i^{(j)} + (B - x_i^{(j)}) \left[\sum_{j \in J} I_j p_{ji}^{(j)} z_{ji}^{(j)} + D_{|J|} f(x_i^{(j)}) \right] - (x_i^{(j)} + C) \sum_{m \in K} g(x_m^{(K)}) F_{|J|} G_{|K|} H_{|K \cap J|}. \quad (\text{A9})$$

We now define how the coefficients $D_{|J|}$, $F_{|J|}$, $G_{|K|}$, and $H_{|K \cap J|}$ depend on the unordered sets K and J ; how the positive and negative feedback functions $f(w)$ and $g(w)$ depend on their activities w ; how the path strengths $p_{ji}^{(j)}$ from F_1 to F_2 express a random growth rule; and how numerical parameters were chosen.

C. Self-Similar Growth Within List Nodes

The coefficient $D_{|J|}$ determines how the positive feedback from a node to itself varies with the node's self-similar scale. We assume that $D_{|J|}$ increases with scale, thereby enabling nodes corresponding to longer sublists to gain a competitive advantage in STM, other things being equal. The simplest choice is made in our simulations, namely,

$$D_{|J|} = D|J|, \quad (\text{A10})$$

where D is a positive constant. This rule is consistent with the possibility that, as an F_2 cell (population) grows in response to high levels of F_1 input, it also produces more excitatory synaptic sites for its own axon collaterals.

D. Conservation of Synaptic Sites

The dependence of the intermodal connection strengths $p_{ji}^{(j)}$, $F_{|J|}$, $G_{|K|}$, and $H_{|K \cap J|}$ on the sets K and J is now described. The total connection strength to each population $v_i^{(j)}$ from all cells in F_1 and the total inhibitory connection strength to each population $v_i^{(j)}$ from all cells in F_2 are both chosen to be independent of K and J . This property is compatible with the interpretation that the size of each cell (population) is scaled to the total strength of its input pathways. If more pathways input to such a cell, each input's effect is diluted more due to the larger size of the cell. The property of matching cell (population) volume to the total number of input pathways is called conservation of synaptic sites.

Conservation of synaptic sites enables the network to overcome the following problem. Due to the randomness of the growth rules, there may exist different

numbers of cells in each of the F_2 masking subfields. As these F_2 cells compete for STM activity, the competitive balance could be biased by accidents of random growth. A mechanism is needed to control the proliferation of random connections. Conservation of synaptic sites is one effective mechanism. A masking field embodies a new functional role for such a growth rule. Thus we impose the following constraints:

Synaptic conservation rule:

Let

$$\sum_{j \in \mathcal{J}} p_{ji}^{(j)} = \text{constant} = 1 \quad (\text{A11})$$

$$\sum_{m,k} F_{|J|} G_{|K|} H_{|K \cap J|} = \text{constant} = F. \quad (\text{A12})$$

By Eq. (A12)

$$F_{|J|} = \frac{F}{\sum_{m,k} G_{|K|} H_{|K \cap J|}}. \quad (\text{A13})$$

We also did simulations in which the coefficients $p_{ji}^{(j)}$ are replaced by coefficients $p_{ji}^{(j)} [1 + \alpha \sum_k p_{jk}^{(j)}]^{-1}$ which obey a Weber law rule and found similar results within a reasonable parameter range.

E. Random Growth from Item Nodes to List Nodes

The connections $p_{ji}^{(j)}$ from F_1 to F_2 are chosen to satisfy the conservation law (A11) as well as a random growth law.

Random normalized growth rule:

Let

$$p_{ji}^{(j)} = \frac{1}{|\mathcal{J}|} (1 - p_{|J|}) + r_{ji}^{(j)} p_{|J|}. \quad (\text{A14})$$

The fluctuation coefficient $p_{|J|}$ in Eq. (A14) determines how random the growth is from F_1 to F_2 . If $p_{|J|} = 0$, growth is deterministic (but spatially distributed) because $p_{ji}^{(j)} = 1/|\mathcal{J}|$. In this limiting case, all connection strengths from item nodes in F_1 to a fixed list node in F_2 are equal and vary inversely with the number $|\mathcal{J}|$ of item nodes that contact the list node. If $0 < p_{|J|} \leq 1$, the coefficients $r_{ji}^{(j)}$ in Eq. (A14) influence the connection strengths $p_{ji}^{(j)}$. The numbers $\{r_{ji}^{(j)} : j \in \mathcal{J}\}$ are chosen pseudo-randomly. They are uniformly distributed between 0 and 1 so that

$$\sum_{j \in \mathcal{J}} r_{ji}^{(j)} = 1. \quad (\text{A15})$$

Equations (A14) and (A15) together imply the conservation rule (11).

It remains to say how the fluctuation coefficients $p_{|J|}$ depend on the set size $|\mathcal{J}|$. We choose these coefficients to keep the statistical variability of the connection strengths independent of $|\mathcal{J}|$. In other words, we choose $p_{|J|}$ so that the standard deviation of $\{p_{ji}^{(j)} : j \in \mathcal{J}\}$ divided by the mean of $\{p_{ji}^{(j)} : j \in \mathcal{J}\}$ is independent of $|\mathcal{J}|$. This is accomplished as follows.

To produce a pseudo-random sequence of numbers $\{r_{ji}^{(j)} : j \in \mathcal{J}\}$ distributed uniformly over the simplex

$$S_n = \{(y_1, y_2, \dots, y_{n+1}) : y_j \geq 0, \sum_{j=1}^{n+1} y_j = 1\}, \quad (\text{A16})$$

we proceed as follows. By a standard algorithm⁶¹ we obtain a vector of numbers $w = (w_1, w_2, \dots, w_n)$ uniformly distributed over the n -cube $I_n = \times_{j=1}^n [0, 1]$. Rearrange the numbers in w in order of increasing size to produce a new vector $w' = (w'_1, w'_2, \dots, w'_n)$ so that $w'_1 \leq w'_2 \leq \dots \leq w'_n$. The map $w \rightarrow w'$ from I_n into itself is determined by a permutation σ of the indices $\{1, 2, \dots, n\}$ so that $w'_i = w_{\sigma(i)}$. Each permutation σ can transform a different subset of I_n into vectors with increasing entries. Thus I_n can be decomposed into sets D_σ so that a single permutation σ can map all $w \in D_\sigma$ into $w' \in I_n$. Hence the map $w \rightarrow w'$ transforms uniformly distributed vectors in I_n onto uniformly distributed vectors in I_n with elements in increasing order.

We next map vectors w' in I_n with elements in increasing order onto vectors y in S_{n+1} via the one-to-one linear transformation $y_1 = w'_1$, $y_2 = w'_2 - w'_1, \dots, y_n = w'_n - w'_{n-1}$, and $y_{n+1} = 1 - w'_n$. Since this linear transformation maps equal volumes onto equal surface areas, the vectors y are uniformly distributed on the simplex S_{n+1} .

The coefficient of variation of $\{p_{ji}^{(j)} : j \in \mathcal{J}\}$ is made independent of $|\mathcal{J}|$ (> 1) as follows. By the above construction, the marginal distribution $r_{ji}^{(j)}$ in Eq. (A14) is distributed with density function $(|\mathcal{J}| - 1)(1 - x)^{|\mathcal{J}| - 2}$. The mean of this distribution is $1/|\mathcal{J}|$, and its standard deviation is

$$\frac{1}{|\mathcal{J}|} \sqrt{\frac{|\mathcal{J}| - 1}{|\mathcal{J}| + 1}}.$$

Thus the mean of $p_{ji}^{(j)}$ is also $1/|\mathcal{J}|$, and its standard deviation is

$$p_{|J|} \frac{1}{|\mathcal{J}|} \sqrt{\frac{|\mathcal{J}| - 1}{|\mathcal{J}| + 1}}. \quad (\text{A17})$$

The coefficient of variation of $p_{ji}^{(j)}$ is its standard deviation divided by its mean, which we set equal to a constant p independent of $|\mathcal{J}|$. Thus we chose

$$p_{|J|} = p \sqrt{\frac{|\mathcal{J}| + 1}{|\mathcal{J}| - 1}}. \quad (\text{A18})$$

In the simulations reported herein, $p = 1/10\sqrt{3}$.

F. Self-Similar Competitive Growth Between List Nodes

Coefficient $F_{|J|}$ in Eq. (A9) describes the total number of inhibitory synaptic sites within a population $v_i^{(j)}$. By Eq. (A13), this quantity is chosen to keep the number of synaptic sites constant across all the cells. Small random variations could also be allowed, but we have absorbed all the effects of randomness into the coefficients $p_{ji}^{(j)}$ in Eq. (A14) for simplicity.

Coefficient $G_{|K|}$ in Eq. (A9) measures the total number of inhibitory connections, or axons, emitted by each population $v_m^{(K)}$ to all other F_2 populations. Due to self-similar growth, $G_{|K|}$ increases with $|K|$. In our simulations, we make the simplest choice.

Self-similar axon generation:
Let

$$G_{|K|} = |K|. \quad (\text{A19})$$

Thus $G_{|K|} = 0$ if $|K| = 0$.

Coefficient $H_{|K \cap J|}$ in Eq. (A9) describes how well growing axons from a population $v_m^{(K)}$ can compete for synaptic sites at a population $v_i^{(J)}$. In particular, coefficient $G_{|K|}$ describes the number of emitted axons, whereas coefficient $H_{|K \cap J|}$ measures the fraction of these axons that can reach $v_i^{(J)}$ and compete for synaptic space there. Due to self-similar growth,¹¹ $H_{|K \cap J|}$ increases with $|K \cap J|$. Consequently, if either set K or J increases, $H_{|K \cap J|}$ also increases, other things being equal. Given fixed sizes of K and J , $H_{|K \cap J|}$ increases as the overlap, or intersection, of the sets increases. In other words, list nodes become list nodes due to the random growth of connections from item nodes. Two list nodes, therefore, tend to be closer in F_2 if they receive more input pathways from the same item nodes in F_1 . If a pair of list nodes in F_2 is closer, their axons can more easily contact each other, other things being equal. In the simulations, we choose $H_{|K \cap J|}$ as follows. Let

$$H_{|K \cap J|} = 1 + |K \cap J|. \quad (\text{A20})$$

By Eq. (A20), $H_{|K \cap J|}$ increases linearly with $|K \cap J|$. Because $H_{|K \cap J|}$ is always positive, when $H_{|K \cap J|}$ multiplies $G_{|K|}$ in Eq. (A9), every population $v_m^{(K)}$ can send weak long-range inhibitory pathways across the whole of F_2 , but these pathways tend to arborize with greater density at populations $v_i^{(J)}$, which receive inputs from the same F_1 nodes. Equations (A13), (A19), and (A20) imply that

$$F_{|J|} = \frac{F}{\sum_{m,K} |K| (1 + |K \cap J|)}. \quad (\text{A21})$$

G. Contrast Enhancement by Sigmoid Signal Functions

The positive and negative feedback signals $f(x_i^{(J)})$ and $g(x_m^{(K)})$ in Eq. (A9) enable the network to contrast enhance its input patterns before storing them in STM. To achieve this property, we choose both $f(w)$ and $g(w)$ to be sigmoid, or S-shaped, functions of the activity level w .^{2,13} In particular, we let

$$f(w) = \frac{([w]^+)^2}{f_0 + ([w]^+)^2}, \quad (\text{A22})$$

$$g(w) = \frac{([w]^+)^2}{g_0 + ([w]^+)^2}. \quad (\text{A23})$$

The notation $[w]^+$ in Eqs. (A22) and (A23) stands for $\max(w, 0)$. Thus $f(w)$ and $g(w)$ do not generate feedback signals if w is smaller than the signal threshold zero. As w increases above zero, both $f(w)$ and $g(w)$ grow quadratically with w until they begin to saturate at their maximum value 1.

H. Associative Learning

The associative law that we have used is that described in Eq. (2).

Associative learning law:

$$\frac{d}{dt} z_{ji}^{(J)} = \epsilon f(x_i^{(J)}) (-z_{ji}^{(J)} + LI_j). \quad (\text{A24})$$

In Eq. (24), the sampling signal $f(x_i^{(J)})$ is assumed to equal the positive feedback signal in Eq. (A9) and is thus a sigmoid function (A22) of activity $x_i^{(J)}$. The parameter ϵ determines the learning rate, and the parameter L is a constant that multiplies the input I_j from node v_j in F_1 .

The learning law contains term I_j rather than term $I_j p_{ji}^{(J)}$ as in Eq. (A9) due to the following interpretation. Term $z_{ji}^{(J)}$ in Eq. (A9) is the LTM density, or LTM strength per unit cross-sectional area, in the pathways from v_j in F_1 to v_i in F_2 . Term $p_{ji}^{(J)}$ describes the total cross-sectional area of these pathways. The input term I_j is broadcast along all these pathways, where it influences the LTM densities as in Eq. (A24). The total signal that is read out from these pathways into v_i equals the readout of all the LTM densities $z_{ji}^{(J)}$ by I_j , summed across all the pathways. This sum equals $I_j p_{ji}^{(J)} z_{ji}^{(J)}$, as in Eq. (A9).

All the above constraints can be summarized in the following system of equations.

Adaptively filtered masking field

$$\begin{aligned} \frac{d}{dt} x_i^{(J)} = & -Ax_i^{(J)} + (B - x_i^{(J)}) \left\{ \sum_{j \in J} I_j \left[\frac{1}{|J|} (1 - p_{|J|}) \right. \right. \\ & \left. \left. + r_{ji}^{(J)} p_{|J|} \right] z_{ji}^{(J)} + D |J| f(x_i^{(J)}) \right\} \\ & - F(x_i^{(J)} + C) \frac{\sum_{m,K} g(x_m^{(K)}) |K| (1 + |K \cup J|)}{\sum_{m,K} |K| (1 + |K \cap J|)} \end{aligned} \quad (\text{A25})$$

and

$$\frac{d}{dt} z_{ji}^{(J)} = \epsilon f(x_i^{(J)}) (-z_{ji}^{(J)} + LI_j),$$

where f and g are sigmoid signal functions. All the intelligence of a masking field is embodied in the emergent properties which arise from the parallel interactions defined by these equations.

I. Parameters

The following parameter choices were made: $A = 1$, $B = 1$, $D = 4$, $L = 10$, $f_0 = 1$, $g_0 = 16$. In all runs $CF = 1088$. Additional parameters are listed by figure. Unless otherwise noted, the system has run to near equilibrium value.

Figure 3: $\epsilon = 0$, $C = 1$, $F = 1088$, $I_0 = 1.5$.

Figure 4(a): same as Fig. 3 except $I_1 = 1.5$; Fig. 4(b): $I_2 = 1.5$; Fig. 4(c): $I_0 = 1$, $I_1 = 0.5$; Fig. 4(d): $I_0 = 0.5$, $I_1 = 1$.

Figure 5(a): $I_0 = 0.68$, $I_1 = 0.48$, $I_2 = 0.34$; Fig. 5(b): $I_0 = 0.34$, $I_1 = 0.68$, $I_2 = 0.48$; Fig. 5(c): $I_0 = 0.34$, $I_1 = 0.48$, $I_2 = 0.68$.

Figure 6: $\epsilon = 0$, $C = 0.125$, $F = 8704$.

Figure 6(a): $I_0 = 1.5$; Fig. 6(b): $I_1 = 1.5$, Fig. 6(c): $I_2 = 1.5$, Fig. 6(d): $I_0 = 1.0$, $I_1 = 0.5$.

Figure 7(a): $I_0 = 0.5$, $I_1 = 1.0$; Fig. 7(b): $I_0 = 0.68$, $I_3 = 0.48$, $I_2 = 0.34$; Fig. 7(c): $I_0 = 0.34$, $I_3 = 0.68$, $I_2 = 0.48$; Fig. 7(d): $I_0 = 0.34$, $I_1 = 0.48$, $I_2 = 0.68$.

Figure 9: Simulation is run at $\epsilon = 0$ until no single step or the size of any component of the derivative is $>1.0 \times 10^{-4}$. ϵ is then set equal to 1, and simulation proceeds to equilibrium parameters; $C = 0.125$, $F = 8704$.

Figure 9(a): $I_0 = 1.5$; Fig. 9(b): $I_1 = 1.5$; Fig. 9(c): $I_2 = 1.5$; Fig. 9(d): $I_0 = 1.0$, $I_1 = 0.5$.

Figure 10: Same parameters and conditions as in Fig. 9 except where noted.

Figure 10(a): $I_0 = 0.5$, $I_1 = 1.0$; Fig. 10(b): $I_0 = 0.68$, $I_1 = 0.48$, $I_2 = 0.34$; Fig. 10(c): $I_0 = 0.34$, $I_1 = 0.68$, $I_2 = 0.48$; Fig. 10(d): $I_0 = 0.34$, $I_1 = 0.48$, $I_2 = 0.68$.

Figure 11: $\epsilon = 0.1$, $C = 0.125$, $F = 8704$. Figures are output of $\{0,1\}$ long-term memory traces at times 1, 2, 4, 8, 16, 32, 64, 96.

Figure 12: $\epsilon = 0$, $C = 1$, $F = 1088$, $t = 0.1$, $t = 0.2$, $t = 0.4$, $t = 0.8$, $t = 1.6$.

Figure 13(a): $\epsilon = 0$, $C = 0.125$, $F = 8704$, $I_0 = 1$, $I_1 = 0.5$; Fig. 13(b): $\epsilon = 1$; Fig. 13(c): $\epsilon = 0.1$; Fig. 13(d): $\epsilon = 0.01$.

References

1. S. Grossberg, "A Theory of Human Memory: Self-Organization and Performance of Sensory-Motor Codes, Maps, and Plans, in *Progress in Theoretical Biology*, Vol. 5, R. Rosen and F. Snell, Eds. (Academic, New York, 1978).
2. S. Grossberg, "Contour Enhancement, Short Term Memory, and Constancies in Reverberating Neural Networks," *Studies in Applied Math.* 52, 217 (1973).
3. S. Grossberg, and D. S. Levine, "Some Developmental and Attentional Biases in the Contrast Enhancement and Short Term Memory of Recurrent Neural Networks," *J. Theoret. Biol.* 53, 341 (1975).
4. S. Grossberg, "Adaptive Pattern Classification and Universal Recoding, I: Parallel Development and Coding of Neural Feature Detectors," *Biol. Cybernet.* 23, 121 (1976).
5. S. Grossberg, "Adaptive Pattern Classification and Universal Recoding, II: Feedback, Expectation, Olfaction, and Illusions," *Biol. Cybernet.* 23, 187 (1976).
6. J.-P. Banquet and S. Grossberg, "Probing Cognitive Processes Through the Structure of Event-Related Potentials During Learning: An Experimental and Theoretical Analysis," *Applied Optics*, in press (1987).
7. G. A. Carpenter and S. Grossberg, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine," *Comput. Vision Graphics Image Process.* 37, 54 (1987).
8. G. A. Carpenter and S. Grossberg, "Neural Dynamics of Category Learning and Recognition: Attention, Memory Consolidation, and Amnesia," in *Brain Structure, Learning, and Memory*, J. Davis, R. Newburgh, and E. Wegman, Eds., AAAS Symposium Series (1987).
9. G. A. Carpenter and S. Grossberg, "Neural Dynamics of Category Learning and Recognition: Structural Invariants, Reinforcement, and Evoked Potentials," in *Pattern Recognition and Concepts in Animals, People, and Machines*, M. L. Commons, S. M. Kosslyn, and R. J. Herrnstein, Eds. (Erlbaum, Hillsdale, NJ, 1987).
10. G. A. Carpenter and S. Grossberg, "ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, in press (1987).
11. M. A. Cohen and S. Grossberg, "Neural Dynamics of Speech and Language Coding: Developmental Programs, Perceptual Grouping, and Competition for Short Term Memory," *Human Neurobiol.* 5, 1 (1986).
12. M. A. Cohen and S. Grossberg, "Unitized Recognition Codes for Parts and Wholes: The Unique Cue in Configurational Discriminations," in *Pattern Recognition and Concepts in Animals, People, and Machines*, M. L. Commons, S. M. Kosslyn, and R. J. Herrnstein, Eds. (Erlbaum, Hillsdale, NJ, 1987).
13. S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control* (Reidel, Boston, 1982).
14. S. Grossberg, Ed., *The Adaptive Brain, I: Cognition, Learning, Reinforcement, and Rhythm* (Elsevier/North-Holland, Amsterdam, 1987).
15. S. Grossberg, Ed., *The Adaptive Brain, II: Vision, Speech, Language, and Motor Control* (Elsevier/North-Holland, Amsterdam, 1987).
16. S. Grossberg and W. Gutowski, "Neural Dynamics of Decision Making under Risk: Affective Balance and Cognitive-Emotional Interactions," *Psychol. Rev.*, in press (1987).
17. S. Grossberg and M. Kuperstein, *Neural Dynamics of Adaptive Sensory-Motor Control: Ballistic Eye Movements* (Elsevier/North-Holland, Amsterdam, 1986).
18. S. Grossberg, and D. S. Levine, "Neural Dynamics of Attentionally Modulated Pavlovian Conditioning: Blocking, Inter-stimulus Interval, and Secondary Reinforcement," submitted for publication (1987).
19. S. Grossberg and N. A. Schmajuk, "Neural Dynamics of Attentionally Modulated Pavlovian Conditioning: Conditioned Reinforcement, Inhibition, and Opponent Processing," submitted for publication (1987).
20. S. Grossberg and G. O. Stone, "Neural Dynamics of Word Recognition and Recall: Attentional Priming, Learning, and Resonance," *Psychol. Rev.* 93, 46 (1986).
21. S. Grossberg and G. O. Stone, "Neural Dynamics of Attention Switching and Temporal Order Information in Short Term Memory," *Memory Cognition*, 14, 451 (1986).
22. M. A. Cohen and S. Grossberg, "Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks," *IEEE Trans. Syst. Man Cybernet.* SMC-13, 815 (1983).
23. J. J. Hopfield, "Neurons with Graded Response have Collective Computational Properties Like Those of Two-State Neurons," *Proc. Nat. Acad. Sci. USA* 81, 3088 (1984).
24. S. Grossberg, "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," *Applied Optics*, in press (1987).
25. R. O. Duda and P. E. Hart, "Use of the Hough Transform to Detect Lines and Curves in Pictures," *Commun. Assoc. Comput. Mach.* 15, 11 (1972).
26. P. V. C. Hough, "Method and Means for Recognizing Complex Patterns," U.S. Patent 3,069,654 (18 Dec. 1962).
27. C. M. Brown, "Inherent Bias and Noise in the Hough Transform," *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-5, 493 (1983).
28. S. Grossberg, "Unitization, Automaticity, Temporal Order, and Word Recognition," *Cognition Brain Theory*, 7, 263 (1984).
29. S. Grossberg, "The Adaptive Self-organization of Serial Order in Behavior: Speech, Language, and Motor Control," in *Pattern Recognition by Humans and Machines, Vol. 1: Speech Production*, E. C. Schwab and H. C. Nusbaum, Eds. (Academic, New York, 1986).
30. S. Grossberg, "Some Physiological and Biochemical Consequences of Psychological Postulates," *Proc. Natl. Acad. Sci. USA* 60, 758 (1968).
31. S. Grossberg, "On Learning and Energy-Entropy Dependence in Recurrent and Nonrecurrent Signed Networks," *J. Stat. Phys.* 1, 319 (1969).
32. S. Amari, "Competitive and Cooperative Aspects in Dynamics of Neural Excitation and Self-Organization," in *Competition and Cooperation in Neural Networks*, S. Amari and M. Arbib, Eds. (Springer-Verlag, New York, 1982).

- (Springer-Verlag, New York, 1982).
33. S. Amari and A. Takeuchi, "Mathematical Theory on Formation of Category Detecting Nerve Cells," *Biol. Cybern.* **29**, 127 (1978).
 34. T. Kohonen, "A Simple Paradigm for the Self-Organized Formation of Structural Feature Maps," in *Competition and Cooperation in Neural Networks* S. Amari and M. Arbib, Eds. (Springer-Verlag, New York, 1982).
 35. T. Kohonen, "Representation of Information in Spatial Maps which are Produced by Self-organization," in *Synergetics of the Brain*, E. Basar, H. Flohr, H. Haken, and A. J. Mandell, Eds. (Springer-Verlag, New York, 1983).
 36. D. E. Rumelhart and D. Zipser, "Feature Discovery by Competitive Learning," *Cognitive Sci.* **9**, 75 (1985).
 37. W. B. Levy, "Associative Changes at the Synapse: LTP in the Hippocampus," in *Synaptic Modification, Neuron Selectivity and Nervous System Organization*, W. B. Levy, J. Anderson, and S. Lehmkuhle, Eds. (Erlbaum, Hillsdale, NJ, 1985) pp. 5-33.
 38. W. B. Levy, S. E. Brassel, and S. D. Moore, "Partial Quantification of the Associative Synaptic Learning Rule of the Dentate Gyrus," *Neuroscience* **8**, 799 (1983).
 39. W. B. Levy and N. L. Desmond, "The Rules of Elemental Synaptic Plasticity," in *Synaptic Modification, Neuron Selectivity and Nervous System Organization*, W. B. Levy, J. Anderson, and S. Lehmkuhle, Eds. (Erlbaum, Hillsdale, NJ, 1985), pp. 105-121.
 40. J. P. Rauschecker and W. Singer, "Changes in the Circuitry of the Kitten's Visual Cortex are Gated by Postsynaptic Activity," *Nature London* **280**, 58 (1979).
 41. W. Singer, "Neuronal Activity as a Shaping Factor in the Self-Organization of Neuron Assemblies, in *Synergetics of the Brain*, E. Basar, H. Flohr, H. Haken, and A. J. Mandell, Eds. (Springer-Verlag, New York, 1983).
 42. D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
 43. D. S. Levine, "Neural Population Modeling and Psychology: A Review," *Math. Biosci.* **66**, 1 (1983).
 44. G. A. Carpenter, "Bursting Phenomena in Excitable Membranes," *SIAM J. Appl. Math.* **36**, 334 (1979).
 45. G. A. Carpenter, "Normal and Abnormal Signal Patterns in Nerve Cells," *SIAM-AMS Proc.* **13**, 49 (1981).
 46. G. A. Carpenter and S. Grossberg, "A Neural Theory of Circadian Rhythms: The Gated Pacemaker," *Biol. Cybern.* **48**, 35 (1983).
 47. G. A. Carpenter and S. Grossberg, "A Neural Theory of Circadian Rhythms: Aschoff's Rule in Diurnal and Nocturnal Mammals," *Am. J. Physiol.* **247**, R1067 (1984).
 48. G. A. Carpenter and S. Grossberg, "A Neural Theory of Circadian Rhythms: Split Rhythms, After-Effects, and Motivational Interactions," *J. Theor. Biol.* **113**, 163 (1985).
 49. S. A. Ellias and S. Grossberg, "Pattern Formation, Contrast Control, and Oscillations in the Short Term Memory of Shunting On-Center Off-Surround Networks," *Biol. Cybernet.* **20**, 69 (1975).
 50. G. A. Carpenter, "A Geometric Approach to Singular Perturbation Problems with Applications to Nerve Impulse Equations," *J. Differential Equations* **23**, 335 (1977).
 51. G. A. Carpenter, "Periodic Solutions of Nerve Impulse Equations," *J. Math. Anal. Appl.* **58**, 152 (1977).
 52. N. Fenichel, "Geometric Singular Perturbation Theory for Ordinary Differential Equations," *J. Differential Equations* **31**, 53 (1979).
 53. S. Grossberg, "A Neural Model of Attention, Reinforcement, and Discrimination Learning," *In. Rev. Neurobiol.* **18**, 263 (1975).
 54. A. J. Friedhoff, Ed., *Catecholamines and Behavior, Vol. I: Basic Neurobiology*. (Plenum, New York, 1975).
 55. A. J. Friedhoff, Ed., *Catecholamines and Behavior, Vol. II: Neuropsychopharmacology*. (Plenum, New York, 1975).
 56. D. L. Alkon, "Voltage-Dependent Calcium and Potassium Ion Conductances: A Contingency Mechanism for an Associative Learning Model," *Science* **205**, 810 (1979).
 57. D. L. Alkon, "Calcium-Mediated Reduction of Ionic Currents: A Biophysical Memory Trace," *Science* **226**, 1037 (1984).
 58. D. L. Alkon, "Changes of Membrane Currents During Learning," *J. Exp. Biol.* **112**, 95 (1984).
 59. S. Grossberg, "On the Production and Release of Chemical Transmitters and Related Topics in Cellular Control," *J. Theor. Biol.* **22**, 325 (1969).
 60. S. Grossberg, Ed., "Adaptive Resonance in Development, Perception, and Cognition," in *Mathematical Psychology and Psychophysiology* (American Mathematical Society, Providence, RI, 1981).
 61. D. E. Knuth, *Seminumerical Algorithms: The Art of Computer Programming, Vol. 2* (Addison-Wesley, Reading, MA, 1981).