

# Neural Dynamics of Word Recognition and Recall: Attentional Priming, Learning, and Resonance

Stephen Grossberg and Gregory Stone  
Center for Adaptive Systems  
Boston University

Data and models about recognition and recall of words and nonwords are unified using a real-time network processing theory. Lexical decision and word frequency effect data are analyzed in terms of theoretical concepts that have unified data about development of circular reactions, imitation of novel sounds, the matching of phonetic to articulatory requirements, serial and paired associate verbal learning, free recall, unitization, categorical perception, selective adaptation, auditory contrast, and word superiority effects. The theory, called adaptive resonance theory, arose from an analysis of how a language system self-organizes in real time in response to its complex input environment. Such an approach emphasizes the moment-by-moment dynamical interactions that control language development, learning, and stability. Properties of language performance emerge from an analysis of the system constraints that govern stable language learning. Concepts such as logogens, verification, automatic activation, interactive activation, limited-capacity processing, conscious attention, serial search, processing stages, speed-accuracy trade-off, situational frequency, familiarity, and encoding specificity are revised and developed using this analysis. Concepts such as adaptive resonance, resonant equilibration of short-term memory, bottom-up adaptive filtering, top-down adaptive template matching, competitive masking field, unitized list representation, temporal order information over item representations, attentional priming, attentional gain control, and list-item error trade-off are applied.

## 1. Role of Learning in Word Recognition

An explosive outpouring of data during the past two decades has documented many aspects of how humans process language in response to visual and auditory cues. With the data have arisen a number of conceptual frameworks and models aimed at integrating data from individual experimental paradigms and suggesting new experiments within these paradigms. A complex patchwork of experiments and models has thus far organized the data into a loose confederation of relatively isolated data domains. The time is ripe for a synthesis.

A parallel line of theoretical development over the past two decades has begun to achieve such a synthesis (Grossberg, 1980, 1982b, 1982c, 1986b, 1986c; Grossberg & Kuperstein, 1985; Grossberg & Mingolla, 1985). In this article the theory's principles and mechanisms are used to explain some challenging data about letter and word recognition and recall. Alternative models of letter and word recognition and recall are also reviewed and discussed. The theory's principles and mechanisms arose from an analysis of how behaving individuals can adapt in real

time to environments whose rules can change unpredictably. Only a few principles and real-time mechanisms are needed to unify a large data base. We believe that the unifying power of the theory is due to the fact that principles of adaptation—such as the laws regulating development, learning, and unitization—are fundamental in determining the design of behavioral mechanisms. This perspective suggests that the lack of a unifying account of the data base is due not to insufficient data quality or quantity but to the use of conceptual paradigms that do not sufficiently tap the principles of adaptation that govern behavioral designs. Such adaptive principles are often called principles of self-organization in theoretical biology and physics (Basar, Flohr, Haken, & Mandell, 1983).

Many of the information processing models that have been suggested during the past two decades have ignored principles of self-organization. In models in which learning was included, the learning rules and the information processing rules were usually introduced as independent hypotheses. We suggest that the linkage between learning and information processing is more intimate than these models suggest. A growing appreciation of this close linkage is suggested by experiments that demonstrate that five or six presentations of a pseudoword can endow it with many of the identification properties of a high-frequency word (Salasoo, Shiffrin, & Feustel, 1985). Such an intimate linkage was also evident in classical paradigms such as serial verbal learning (Underwood, 1966; Young, 1968), wherein the functional units, or chunks, governing a subject's performance can change in a context-sensitive way from trial to trial. The great successes of the 1970s in exploring information processing paradigms made it possible to temporarily ignore vexing issues concerning dynamically changing functional units, but the price paid

---

Stephen Grossberg's research was supported in part by Air Force Office of Scientific Research Grant No. AFOSR 82-0148, National Science Foundation Grant No. NSF IST-8417756, and Office of Naval Research Contract ONR N00014-83-K0337. Gregory Stone's research was supported in part by Office of Naval Research Contract ONR N00014-83-K0337.

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

Correspondence concerning this article should be addressed to Stephen Grossberg, Room 244, Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston, Massachusetts 02215.

has been the fragmentation of explanatory concepts concerning this important data base.

To deeply understand word recognition and recall data, one needs to analyze the computational units that subserve speech, language, and visual processing. One needs to consider how these computational units acquire behavioral meaning by reacting to behavioral inputs and generating behavioral outputs, how a particular choice of computational units determines a model's processing stages and interactions between stages, and how such concepts are supported or challenged by functionally related data other than those under particular scrutiny. All the while, one needs to explicate all the hidden processing assumptions that go into a model and to test their plausibility and ability to arise through self-organization in a stable fashion.

The organization of the present article reflects the multifaceted nature of its task. Sections 2-5 review some of the main models and empirical concepts that have been used to explain word recognition data. The logogen model, the verification model, and the Posner and Snyder model are reviewed using concepts such as automatic activation, limited capacity attention, serial search, and interactive activation. Experimental evidence is cited that suggests the need for multiple processing stages in whatever model is chosen. Internal and predictive limitations of these models are noted to prepare for a resolution of these difficulties using our theoretical framework.

Section 6 begins the exposition of adaptive resonance theory with a discussion of the relations among the theory's computational units, its processing stages, and its mechanisms. These are network mechanisms whose interactions define the theory's processing stages and give rise to its computational units. The computational units are spatial patterns of short-term memory activation and of long-term memory strength. The computational properties of these units are emergent, or collective, properties of the network interactions. These interactions do not include serial programs, algorithms, or cognitive rule structures. Instead, the network as a whole acts as if intelligence is programmed into it.

The network's emergent computational properties are not adequately described by the familiar metaphors of symbol manipulation and number crunching. The spatial pattern units are concrete yet indivisible entities that are capable of coding highly abstract context-sensitive information. Breaking such a pattern down into its constituent parts destroys the pattern's contextual information and its behavioral meaning.

Many possible theories use spatial patterns as their computational units. Section 7 describes the main mechanisms that set apart the present theory from possible alternatives. Just a few mechanisms are needed, despite the theory's broad predictive range. Section 8 shows how these mechanisms can be used to clarify fundamental concepts found in other models, such as attention, limited capacity, and processing stage. Section 9 goes on to show how the distinction between attentional priming and attentional gain control can further clarify concepts of automatic spreading activation and conscious attention.

On the basis of these general ideas, we describe the theory's hierarchical organization in Section 10. This discussion indicates how the theory's computational units and stages differ in critical ways from those used in other theories. We also clarify how word recognition and recall are related to other types of speech and

language processes, notably to processes governing auditory feature detection, analysis by synthesis, matching phonetic to articulatory requirements, imitation of novel sounds, word superiority effects, temporal-order information over item representations in short-term (or working) memory, and list chunking. The several design principles that are used to determine the architecture of each processing stage are also reviewed.

With both mechanistic and organizational processing concepts in hand, we turn in Section 11 to a detailed analysis of the lexical decision data of Schvaneveldt and McDonald (1981). We show how these data are clarified by a property of the theory called the *list-item error trade-off*. This new trade-off is closely related to the *speed-accuracy trade-off*. The analysis considers the moment-by-moment influence of related, unrelated, and neutral word primes on subsequent recognition of word and nonword targets under mask and no-mask presentation conditions. The analysis clarifies how unconscious processes can strongly influence the course of recognition (Marcel, 1980; MacKay, 1973).

Sections 12-16 illustrate how the theory can explain a different type of word recognition and recall data. We analyze data concerning the recognition of prior occurrence, familiarity, situational frequency, and encoding specificity, such as that of Mandler, Tulving, Underwood, and their collaborators. We note problems within the empirical models that have arisen from these data and indicate how our theory overcomes these problems. This synthesis of several domains of recognition and recall data illustrates the explanatory power of processing concepts that have been derived from an analysis of language self-organization.

We now turn to a description of several of the most important word recognition models and representative data upon which they were built. We recognize that many variations are possible within each framework. Our critique identifies core problems with certain *types* of models. We acknowledge that variations on these model types may be devised to partially overcome these core problems. Our critique suggests that any improved model must deal with these problems in a principled and uncontrived fashion. At the present time, these models do not seem to contain any principles that naturally overcome the problems. Then we show how our theory deals with these problems in a principled way.

## 2. Logogens and Embedding Fields

Many recent experiments on word recognition have been influenced either directly or indirectly by the seminal work of Morton (1969, 1970). The functional unit of Morton's model is called the *logogen*:

The logogen is a device which accepts information from the sensory analysis mechanisms concerning the properties of linguistic stimuli and from context-producing mechanisms. When the logogen has accumulated more than a certain amount of information, a response (in the present case the response of a single word) is made available. (Morton, 1969, p. 165)

The logogen model can be instantiated as a real-time network. In such a network, a combination of visual and auditory feature detectors and semantically related contextual information can input to network nodes that represent the logogens. When the total input at a logogen node exceeds a threshold value, an output

from the node is triggered. A node's threshold defines the "level of evidence" that its logogen requires before it can generate outputs that influence other logogens or postlexical processes.

The logogen model is one of a family of network models that arose in psychology and neurobiology during the 1960s. In the domain of visual neurophysiology, the classical Hartline-Ratcliff model of lateral inhibition in the retina of the horseshoe crab (Ratcliff, 1965) obeys the same formal rules. The Hartline-Ratcliff model also compares activation due to input increments and decrements to output thresholds at network nodes. Although the interpretation of nodal activations in such models may differ depending on the processing levels that they contain, all of the models are, from a formal perspective, continuous versions of the influential binary-threshold McCullough-Pitts networks that were introduced in the 1940s (Pitts & McCullough, 1947).

The logogen model differs, however, from the Hartline-Ratcliff model in yet another, equally important way. Semantically related logogens are assumed to interact more strongly than semantically unrelated logogens. Thus a logogen is tacitly the outcome of a learning process. Within Morton's theory, familiarization with known words is conceptualized as a process whereby output thresholds for existing logogens are lowered by repeated presentations. Lowering a threshold facilitates selection of a logogen. However, the logogen's internal organization does not change as a function of experience. The learning processes whereby internal representations for words are organized and maintained are not rendered explicit or used to derive the model. We show that mechanisms whereby internal representations, such as logogens, are organized and maintained have properties that also help to explain familiarization effects. Moreover, changes in output threshold are not among these mechanisms. Thus an analysis of how logogens become logogens leads to different explanations of the types of data that the logogen theory was constructed to explain.

At the same time that Morton was developing his network model for word recognition, Grossberg was developing network models in which learning issues were central. These networks, called *embedding fields*, were introduced to explain data about human learning, such as serial verbal learning and paired associate learning (Grossberg, 1969d; Grossberg & Pepe, 1971); performance speed-up and perceptual masking due to the interaction between learning, lateral inhibition, and thresholds (Grossberg, 1969b); and unitization of hierarchically organized sensory-motor plans and synergies (Grossberg, 1969c, 1970b). Grossberg also showed that the Hartline-Ratcliff network could be derived as a special case of an embedding field in which no learning occurs (Grossberg, 1969b) and that the simplest circuits for learning and performing sequential motor acts—the so-called avalanche circuits—were similar to the command cells of invertebrates (Grossberg, 1970b, 1982c). Such derivations suggested that formal similarities exist between the processing laws utilized by vertebrates and invertebrates yet that these laws are organized into different circuits across species to solve different classes of environmental problems. Thus it was clear by the late 1960s that an analysis of learning could lead to conclusions about performance that cut across species and psychological-neural boundaries.

The original embedding field theory also had a limited domain of applicability beyond which its predictive power failed. The

nature of this failure helped to extend the theory by suggesting other principles and mechanisms of self-organization that are used to generate complex behaviors (Grossberg, 1982c). We indicate in this article how offspring of the original logogen networks can be unified by offspring of the original embedding field networks.

### 3. Verification by Serial Search

The incompleteness of the logogen model was suggested by several types of evidence. Meyer, Schvaneveldt, and Ruddy (1974) showed that in a lexical decision experiment, stimulus quality and semantic context interact. Becker and Killion (1977) studied the interaction between stimulus quality and word frequency. This interaction, although in the direction predicted by the logogen model, did not reach statistical significance, and a power calculation indicated that an interaction as small as 25 ms could have been detected. Under the assumptions of the Sternberg (1969) additive factor method, these data suggested that stimulus quality and semantic context affect at least one common stage of information processing, whereas stimulus quality and word frequency influence different stages of information processing. In particular, semantic context can influence the relatively early stages of processing at which stimulus quality is important, and word frequency influences a later processing stage. Because both semantic context and word frequency effects arise through learning, these data also indicate that learning can go on at multiple stages during language learning experiences.

The logogen model did not specify two levels at which learned factors can operate, although considered as a framework rather than a definite model, any number of levels could in principle be contemplated. Instead, within the logogen model, semantic context acts by supplying activation to target logogens from semantically related logogens. This type of context effect is equivalent to lowering the output threshold, or criterion, of the target logogens, which is the mechanism used to account for faster recognition of high-frequency words. In a lexical decision task, such a lowering of criterion should induce bias toward word responses. Schvaneveldt and McDonald (1981) summarized data of Antos (1979), Lapinski and Tweedy (1976), McDonald (1977), McDonald and Schvaneveldt (1978), and Schuberth (1978) that disconfirmed this expectation. Semantic context thus does not merely induce a criterion shift. Somehow, context facilitates the processing of information in the stimulus.

The verification model was introduced to compensate for this weakness of the logogen model (Becker, 1976; Becker & Killion, 1977; Becker, Schvaneveldt, & Gomez, 1973; Paap, Newsome, McDonald, & Schvaneveldt, 1982; Schvaneveldt, Meyer, & Becker, 1976). To deal with the two stages at which learned factors can operate, the verification model

assumed that the feature analysis and feature increment process is an indeterminate one: a process that results in numerous word detectors exceeding criterion. . . . The function of this type of processing is to delineate a subset of lexical memory items that are consistent with the primitive visual feature information. . . . It is assumed that the verification process selects a word from the sensory set and uses an abstract representation stored with the word to 'predict' the relational features contained in visual memory. If the predicted relational features match those found in visual memory, then the word that generated the predictions

is recognized. If the predictions fail to match the stimulus, another word is sampled from the sensory set to be compared against the stimulus. Thus verification is an interactive *comparison process that operates on one word at a time* [italics ours]. . . . The operation of semantic context in the verification model is as follows: When a context word is recognized, a semantic priming process, similar to that suggested by Morton (1969), activates word detectors that are semantically related to the prime word . . . it is assumed that the semantically defined set of words is sampled by the verification process during the time that the sensory feature set is being defined. . . . Thus, if the stimulus presented following the prime word DOCTOR is semantically related, then that stimulus would be recognized by the successful verification of a word selected from the semantic set . . . the effect of a semantic context is to bypass the visual feature analyzer component of the model. If the new stimulus is not related to the context word, then the semantic set would be *exhaustively sampled* [italics ours], and verification would proceed to sample the sensory set . . . for a word in context, the only effects of intensity would be those localized in the peripheral visual system . . . the interaction of intensity with context derives from the effect of intensity on a process that is necessary to recognize a word out of context but that is bypassed for a word in context. (Becker & Killion, 1977, pp. 395-396)

Adding the verification process introduced a second stage at which learning could occur into the word recognition literature. This second stage operates according to processing assumptions that deserve further analysis. The assumption that verification operates on one word at a time creates problems of implementation *if serial mechanisms are used*. In a serial processor, order information is typically stored via some type of buffer. The order in which items are arranged in the buffer slots determines the order of search. In the verification model, two types of buffers would be needed. One would order items in terms of decreasing semantic relatedness, the other in terms of decreasing word frequency.

To determine decreasing word frequency, pairwise comparisons of all word frequencies in the sensory set or some equivalent procedure would have to occur before any item could be stored in the corresponding buffer. The system would have to be able to compare arbitrary pairs of items, store the results of these arbitrary comparisons, and then move and store the data repeatedly until a winner could be found. Only then could an item be stored in the buffer. After the first item is stored, the entire process would be repeated on the remaining items. In short, ordering items according to word frequency in a serial buffer implies the existence of a complex presorting device in which arbitrary pairs of items can be accessed and compared using a procedure whose total number of operations increases exponentially with candidate set size. The same type of presorting process would be needed to order items in the semantic set according to decreasing semantic relatedness. It is left unclear in the verification model how to index word relatedness or word frequency in a way that could support such comparisons.

Moreover, because the sensory set cannot be searched before the semantic set is fully sampled, one must assume that matches based on word frequency do not even begin until every item of the semantic set is serially eliminated. Thus the demands of the verification model on a serial buffer are much more severe than the demands placed on a serial buffer that attempts merely to store items in their presentation order. Even such classical serial buffer concepts do not fare well when they are analyzed from the perspective of self-organization (Grossberg, 1978b). We have

not succeeded in finding any mechanism that could self-organize these processing requirements for a serial buffer. Of course, one might argue that a concept of verification can be salvaged by characterizing suitable parallel mechanisms. Such a characterization is a primary task of this article. Once such a characterization is articulated, however, a finer processing language replaces concepts such as verification.

Coltheart, Davelaar, Jonasson, and Besner (1977) obtained evidence against serial search in word recognition using a lexical decision experiment. They varied the number ( $N$ ) of English words that can be produced by changing just one letter in the target letter string. They found that reaction times to respond "word" were independent of  $N$ , but reaction times to respond "nonword" increased with  $N$ . Coltheart et al. used this type of evidence to argue against serial search and for some type of parallel access to lexical information. The relevance of the Coltheart et al. data to serial search concepts can be understood by considering several cases.

Consider a serial search that uses the complete lexicon as its search set. Then both "word" and "nonword" responses should depend on  $N$  in a similar way. If the time to make each match were independent of  $N$ , then both sets of reaction times should be independent of  $N$ . If the time to make each match depended on  $N$ , then both sets of reaction times should depend on  $N$ . In the verification model, by contrast, a serial search would be restricted to the visual set, the size of which increases with  $N$ . In this situation, because times to respond "nonword" would require an exhaustive search of the visual set, increasing set size should increase reaction times. Times to respond "word" would not require an exhaustive search, but they too should increase with  $N$ , although less rapidly. Thus the verification model is consistent with the Coltheart et al. experiment if the latter was not powerful enough to detect the smaller effect of  $N$  on "word" response reaction times. On the other hand, this explanation becomes less plausible if serial presorting operations increase exponentially with  $N$ . To reconcile this implication of a serial presorting mechanism with the Coltheart et al. data seems to require that the sum total of an exponentially growing number of presorting operations takes much less time than does a linearly increasing number of buffer matching operations. Thus a specification of real-time serial mechanisms to instantiate the verification model raises concerns not only about a serial model's ability to self-organize, but also about the plausibility of the model's explanations of certain data.

Even if a parallel presorting mechanism is assumed, a mechanistic analysis raises a serious problem concerning the assumption that "the interaction of intensity with context derives from the effect of intensity on a process that is necessary to recognize a word out of context but that is bypassed for a word in context" (Becker & Killion, 1977, p. 396). The model asserts that the time to match the information in the sensory buffer with items in the semantic set does not depend on the quality of the sensory information. The model also asserts that the activation of the visual feature analyzers does depend on the quality of information in the sensory buffer. It is unclear what kind of matching mechanism could be insensitive to the quality of the information being matched, especially when the quality of this information does effect other processing channels. Thus, despite the great heuristic value of the verification model, attempts to embody the model

using real-time mechanisms provide converging evidence against certain versions of the verification process.

#### 4. Automatic Activation and Limited-Capacity Attention

The verification model explicitly recognizes that at least two processes are required to discuss data about word recognition. Other two-process theories have also been useful in the analysis of word recognition data. The popular model of Posner and Snyder (1975a) posits one process whereby a stimulus rapidly and automatically activates its logogen, which in turn rapidly activates a set of semantically related logogens. The second process is realized by a limited-capacity attentional mechanism. It is slower acting, cannot operate without conscious attention, and inhibits the retrieval of information stored in logogens on which attention is not focused. If a stimulus activates an unattended logogen, that logogen can build up activation but cannot be read out for postlexical processing or response selection. Before output can occur from an active logogen, a limited-capacity attentional process must be shifted to that logogen.

This two-process theory successfully explains various data about word recognition (Neely, 1977; Posner & Snyder, 1975b). It also, however, raises serious questions. For example, suppose that attention happens to be misdirected away from the semantic set that is activated by a stimulus word. How does the processing system then redirect attention to this semantic set? Unless attention is randomly redirected, which is manifestly false, or an undefined agent redirects attention toward the semantic set, signals from the logogens in the semantic set to the attentional mechanism are the only agents whereby attention can be redirected to the semantic set. As Neely (1977) noted,

the conscious-attention mechanism does not inhibit the build-up of activation in unattended logogens, but rather inhibits the readout of information from unattended logogens . . . before the information stored at the unattended logogen can be analyzed in preparation for response initiation, the conscious-attention readout mechanism must be 'shifted' to that logogen. (p. 228)

The Posner-Snyder model is silent about how logogens may draw attention to themselves without outputting to postlexical processes.

The key feature of the Posner-Snyder model lies in the fact that attention can eventually get redirected, that competitive or limited-capacity processes are often involved in its redirection and sharpening, and that these processes occur after the initial wave of logogen activation. At first glance, it may appear that the incompleteness of the Posner-Snyder model can be resolved if an attention shift can be a consequence of logogen output as well as a cause of logogen output, as in the attention theory of Grossberg (1975). Further consideration shows, however, that the problem cannot be resolved just by assuming that active logogens can draw attention to themselves. In a spreading activation model, the activation across logogens can be of unlimited capacity. How an unlimited capacity output from logogens to the attentional mechanism can be transformed into a limited capacity output from logogens to other postlexical processes is not explained within the Posner-Snyder model.

We trace this problem within the Posner-Snyder model to its very foundations, in particular, its choice of spreading activation as a computational unit. By design, spreading activation of their type does not have limited capacity. Even if each logogen were to activate connected logogens by only a fraction of its current activation, then the total logogen activation could still grow with the number of activated logogens and hence would not be of limited capacity. By contrast, any model whose lexical processing takes place among logogens must generate a limited capacity logogen output to postlexical processes. In the Posner-Snyder framework, an external mechanism that interacts with spreading activation is needed to generate a limited capacity among the logogens, because of the lack of an internal mechanism to restrict the capacity.

In our theory, the computational unit is no longer a spreading activation among network nodes. It is a spatial pattern of activity that is processed as a whole across a field of network nodes. Such a spatial pattern has limited capacity whether or not attention is directed toward it. Thus our computational unit per se does not require a two-process model. This change of unit does, however, necessitate a far-reaching revision in how one thinks about basic notions such as attention and capacity (Grossberg, 1984).

The nature of these changes is hinted at by a comparison of the verification model with the Posner-Snyder model. Both models have posited a second process subsequent to logogen activation to deal with word recognition data. In the light of our analysis of the verification model, we can now ask: How does the rapid activation of logogens initiate competitive interactions, notably interactions due to read-out of a top-down expectancy that is matched against sensory data? How does the notion of verification relate to the notion of attention? Can conscious attention be an *outcome* of verification, rather than its cause?

#### 5. Interactive Activation and Parallel Access

The previous analysis suggests some of the problems faced by a verification process that utilizes a serial search mechanism. To overcome such difficulties, McClelland and Rumelhart described a two-level model in which serial search is replaced by parallel access (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). The theory that we now apply, which was introduced in Grossberg (1978a), also has this "interactive activation" characteristic. We do not use the McClelland and Rumelhart formulation for several reasons. We have elsewhere argued that their model cannot arise in a stable manner as a result of a self-organization process and that both its nodes and its interactions are incompatible with some word recognition data that our theory has successfully predicted (Grossberg, 1984, 1985). Within the general framework of real-time network models exist many different possibilities. The present theory is one of many "interactive activation," or real-time network, theories. It happens to be one that is capable of stable self-organization and able to explain a larger data base than the McClelland and Rumelhart version. Rumelhart and Zipser (1985) have recently considered how learning mechanisms may be appended to the interactive activation framework. The relation of these results to those in the Grossberg (1976a) study and related learning models are described in the Grossberg (1986a) study.

## 6. The View From Adaptive Resonance Theory

We explain word recognition data using the theory of human memory that was developed by Grossberg (1978a). This theory is more complex than many psychological models because it analyzes how auditory, visual, cognitive, and motor representations can develop and be learned in real time within individual learning subjects. The simultaneous consideration of several modalities by the theory enables one to discover many more design constraints than can be discovered by consideration of any one factor alone. For example, interactions between auditory and motor factors impose constraints on how internal language representations are initiated, imitated, chunked, recognized, and used to generate motor commands for recall. Interactions between visual, auditory, and motor factors impose constraints on how visual symbols for language units, such as letters, numbers, and words, can be recognized through learned verbally mediated language representations, which in turn can generate motor commands for recall.

A central issue in each of these intermodal interactions concerns the manner in which development, learning, and unitization are stabilized via the action of feedback loops. Some of these feedback loops are closed internally, via bottom-up and top-down signal exchanges. Others are closed externally, via the registration of environmentally mediated sensory feedback signals. These cyclic organizations, rather than one-way traffic between processing stages, define the computational units that have behavioral meaning. Piaget (1963) and Neisser (1967) both emphasized the importance of the cyclic organization of perceptual and cognitive units—Piaget through his theory of circular reactions and Neisser through his theory of perceptual cycles. Circular reactions are, in fact, dynamically characterized within the theory and play an important role in initiating the self-organization of its cyclic memory structures. The term *adaptive resonance* was introduced by Grossberg (1976b) to describe this cyclic aspect of the theory. An adaptive resonance is a fully elaborated recognition event within a feedback network capable of dynamically buffering, or self-stabilizing, its learned codes and expectancies against recoding by irrelevant cues.

The final results of the Grossberg (1978a) analysis were a macrotheory and a microtheory. These two aspects of the theory coexist in a mutually supportive relation. The macrotheory consists of several design principles, dynamical laws, and macrocircuits whose macrostages compute functionally characterized properties. The microtheory describes the processes that generate the properties of the various macrostages. Unlike many artificial intelligence models, the macrotheory and the microtheory cannot easily be dissociated. This is because the critical properties at the macrostages are interactive, or collective, properties of the microtheory's processes. Even the apparently local concept of feature detector is the net effect of widespread interactions within a microtheory network.

The microtheory thus does not freely invent properties at each macrostage. Each process of the microtheory generates a formal syndrome of interactive properties in response to prescribed experimental and system constraints. The internal structuring of these syndromes defines the macrotheory properties and is the source of the theory's predictive force. The macrotheory's general principles and laws severely constrain the types of microprocesses

that are allowed to occur at any macrostage. Only a few principles and laws are used in the entire theory, despite its broad scope. For example, every stage of the theory is a mixed cooperative-competitive network, and every interstage signal process is an adaptive filter. Furthermore, the same mechanisms are used to generate chunking and temporal order properties of both language and motor control processes.

That feedback cycles define the basic building blocks of the theory leads to a sobering conclusion. Such feedback cycles must be built up out of nonlinear mechanisms, because linear mechanisms have been proven to be unstable (Grossberg, 1973, 1983). Thus a certain amount of mathematics is needed to achieve a deep understanding of behavioral self-organization. The human mind does not easily grasp nonlinear interactions among thousands or millions of units without mathematical tools. Fortunately, once one has identified good nonlinear laws, a mathematical theory with data-predictive properties can be developed.

## 7. Elements of the Microtheory: Tuning, Categories, Matching, and Resonance

In this section we describe some of the interactions of short-term memory (STM) processes at successive network stages with bottom-up and top-down long-term memory (LTM) processes between these stages. We denote the  $i$ th stage in such a network hierarchy by  $F_i$ .

Suppose that a pattern  $X$  of STM activities is present at a given time across the nodes of  $F_i$ . Each sufficiently large STM activity can generate excitatory signals that are transmitted by the pathways from its node to target nodes within  $F_{i+1}$ . When a signal from a node in  $F_i$  is carried along a pathway to  $F_{i+1}$ , the signal is multiplied, or *gated*, by the pathway's LTM trace. The LTM gated signal (signal times LTM trace), not the signal alone, reaches the target node. Each target node sums up all of its LTM gated signals. In this way, a pattern  $X$  of STM activities across  $F_i$  elicits a pattern  $S$  of output signals from  $F_i$ . Pattern  $S$  in turn generates a pattern  $T$  of LTM gated and summed input signals to  $F_{i+1}$ . This transformation from  $S$  to  $T$  is called an *adaptive filter*.

The input pattern  $T$  to  $F_{i+1}$  is itself quickly transformed by a cooperative-competitive interaction among the nodes of  $F_{i+1}$ . In the simplest example of this process, these interactions choose the node that received the largest input (Grossberg, 1976a, 1982c). The chosen node is the only one that can store activity in STM. In other words, the chosen node "wins" the competition for STM activity. The choice transformation executes the most severe type of contrast enhancement by converting the input pattern  $T$ , in which many signals can be positive, into a pattern  $Y$  of nodal activities in which at most one activity is positive. In more realistic cooperative-competitive interaction schemes, the contrast enhancing transformation from  $T$  to  $Y$  is more subtle than a simple choice because it is designed to properly weight many possible groupings of an input pattern. Such multiple grouping networks are generically called *masking fields* (see Section 10). In every case, the transformed activity pattern  $Y$ , not the input pattern  $T$ , is the one stored in STM at  $F_{i+1}$  (see Figure 1a). In every case, the transformation of  $T$  into  $Y$  is nonlinear.

Only nodes of  $F_{i+1}$  that support stored activity in STM can elicit new learning at their contiguous LTM traces. Thus, whereas all the LTM traces in the adaptive filter, and thus all learned past

experiences of the network, are used to determine recognition via the transformation  $X \rightarrow S \rightarrow T \rightarrow Y$ , only those LTM traces whose target activities in  $F_{i+1}$  survive the competitive-cooperative struggle for stored STM activity can learn in response to the activity pattern  $X$ . The fact that only the transformed STM patterns  $Y$ , rather than the less focused input pattern  $T$ , can influence LTM in the network helps to explain why recognition and attention play an important role in learning ( Craik & Lockhart, 1972; Craik & Tulving, 1975).

This type of feedback interaction between associative LTM mechanisms and cooperative-competitive STM mechanisms has many useful properties (Grossberg, 1976a, 1978a). It generalizes

the Bayesian tendency to minimize risk in a noisy environment. It spontaneously tends to form learned categories  $Y$  in response to the activity patterns  $X$ . Novel activity patterns  $X$  can be unambiguously classified into a category on their first presentation if they are weighted averages of experienced patterns that all lie within a single category. Input patterns that are closely related to previously classified patterns can elicit faster processing than can input patterns that are not. The rate of processing is also sensitive to the number and dissimilarity of experienced patterns that fall within a category. For example, a larger number of dissimilar patterns can cause slower processing, other things being equal. Learning by each LTM trace is sensitive to the entire activity pattern  $X$  that is active at that time, as well as to all prior learning at all LTM traces. This property follows from the fact that the input pattern  $T$  engages the cooperative-competitive STM interactions across  $F_{i+1}$  to generate  $Y$ , before  $Y$  in turn regulates how each LTM trace will change in response to  $X$ . Thus both learning and recognition are highly context sensitive. The learning and recognition capabilities of the choice model are mathematically characterized in Carpenter and Grossberg (1985) and Grossberg (1976a). The properties of the masking field model are described in Cohen and Grossberg (in press) and Grossberg (1978a, 1985).

The bottom-up STM transformation  $X \rightarrow S \rightarrow T \rightarrow Y$  is not the only signal process that occurs between  $F_i$  and  $F_{i+1}$ . In the absence of top-down processing, the LTM traces within the learned map  $S \rightarrow T$  can respond to a sequence of input patterns by being ceaselessly recoded by inappropriate events (Grossberg, 1976a). In other words, the learning within the bottom-up code can become temporally unstable in that individual events are never eventually encoded by a single category as they are presented on successive trials. Carpenter and Grossberg (1985) described an infinite class of examples in which such ceaseless recoding occurs. Information processing models could not articulate this basic stability problem because they did not use learning in an essential way. Simulation studies of conditioning in response to impoverished input environments also failed to deal with the problem. In Grossberg (1976b), it was shown that properly designed mechanisms for top-down signaling from  $F_{i+1}$  to  $F_i$  and for matching within  $F_i$  can stabilize learning within the bottom-up code against recoding by inappropriate events. Because information processing models could not achieve this insight about the functional role of top-down processing, the constraints that follow from this insight were not available to help choose among the many possible embodiments of top-down signaling and matching. The aspects of this top-down scheme that we need are reviewed below.

The STM transformation  $X \rightarrow S \rightarrow T \rightarrow Y$  takes place very quickly. By "very quickly" we mean more quickly than the rate at which the LTM traces in the adaptive filter  $S \rightarrow T$  can change. As soon as the bottom-up transformation  $X \rightarrow Y$  takes place, the activities  $Y$  in  $F_{i+1}$  elicit top-down excitatory signals  $U$  back to  $F_i$  (see Figure 2a). The rules whereby top-down signals are generated are the same as the rules by which bottom-up signals are generated. Signal thresholds allow only sufficiently large activities in  $Y$  to elicit the signals  $U$  in pathways from  $F_{i+1}$  to  $F_i$ . The signals  $U$  are gated by LTM traces in these pathways. The LTM gated signals excite their target nodes in  $F_i$ . These LTM gated signals summate at the target nodes to form the total input

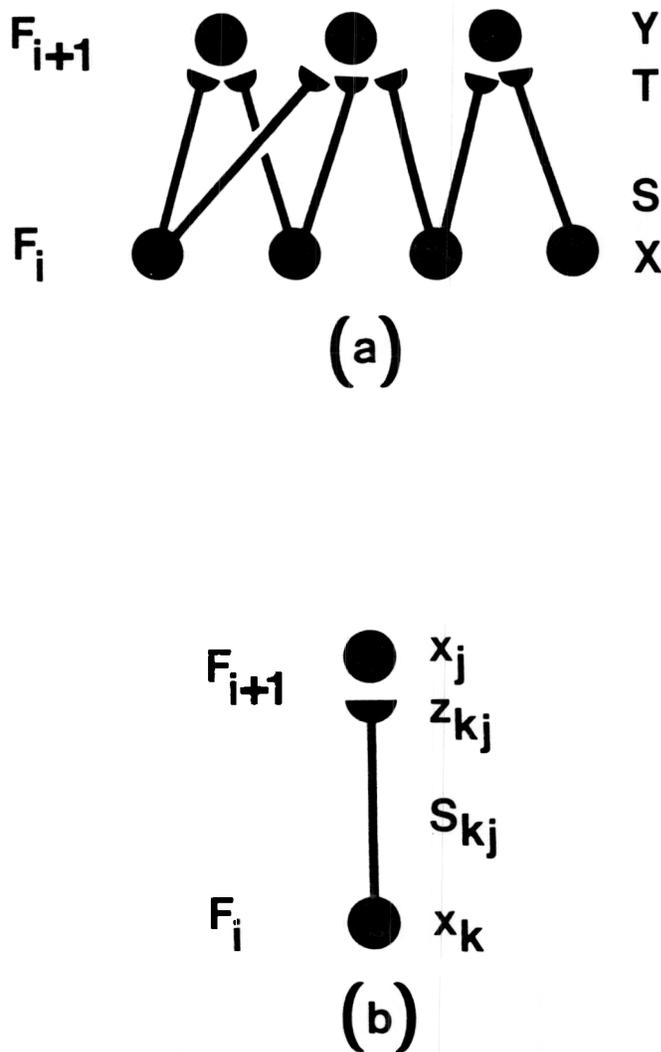


Figure 1. Bottom-up interaction of short-term memory and long-term memory between network levels: (a) An activity pattern  $X$  at level  $F_i$  gives rise to a pattern of output signals  $S$ . Pattern  $S$  is multiplicatively gated by long-term memory traces. These gated signals summate to form the input pattern  $T$  to level  $F_{i+1}$ . Level  $F_{i+1}$  contrast enhances  $T$  before storing the contrast-enhanced activity pattern  $Y$  in short-term memory. (b) Each activity  $x_k$  in  $F_i$  gives rise to a signal  $S_{kj}$  (possibly zero), which is gated by a long-term memory trace  $z_{kj}$  before the gated signal activates  $x_j$  in  $F_{i+1}$ .

signals from  $F_{i+1}$ . In this way, the pattern  $U$  of output signals from  $F_{i+1}$  generates a pattern  $V$  of input signals to  $F_i$ . The map  $U \rightarrow V$  is said to define a top-down template, or learned expectation,  $V$ . Note that  $V$  is not defined exclusively in terms of LTM traces. It is a combination of LTM traces and the STM activities across  $F_{i+1}$ .

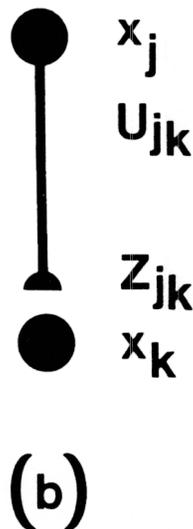
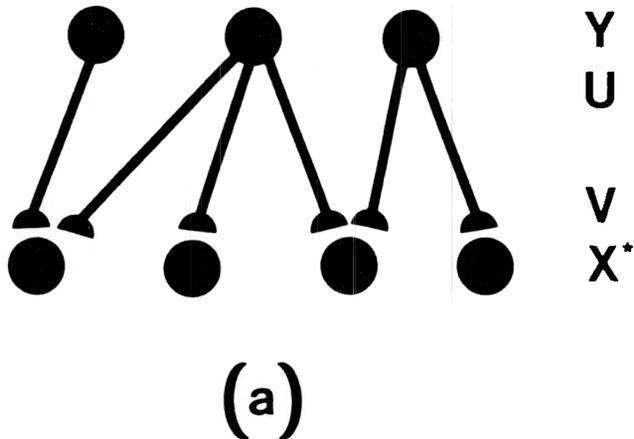


Figure 2. Top-down interaction of short-term memory and long-term memory between network levels: (a) An activity pattern  $Y$  at level  $F_{i+1}$  gives rise to a pattern of output signals  $U$ . Pattern  $U$  is multiplicatively gated by long-term memory traces. These gated signals summate to form the input pattern  $V$  to level  $F_i$ . Level  $F_i$  matches the bottom-up input pattern  $I$  with  $V$  to generate a new activity pattern  $X^*$  across  $F_i$ . (b) Each activity  $x_j$  in  $F_{i+1}$  gives rise to a signal  $U_{jk}$  (possibly zero), which is gated by a long-term memory trace  $Z_{jk}$  before the gated signal activates  $x_k$  in  $F_i$ .

Two sources of input now perturb  $F_i$ : the bottom-up input pattern  $I$  that gives rise to the original activity pattern  $X$  and the top-down template  $V$  that results from activating  $X$ . At this point, descriptive language breaks down unless it is supported by precise mathematical concepts. For once the feedback loop  $X \rightarrow S \rightarrow T \rightarrow Y \rightarrow U \rightarrow V \rightarrow X^*$  closes, it continues to reverberate quickly between  $F_i$  and  $F_{i+1}$  until its activity patterns equilibrate. Notations such as  $X$  and  $Y$  are inadequate to describe this equilibration process because the activity pattern  $X^*$  across  $F_i$  that is induced by  $I$  and  $V$  taken together may not be the same activity pattern  $X$  that was induced by  $I$  alone. This composite activity pattern quickly activates a new template, which quickly activates a new activity pattern, which quickly activates a new template, ad infinitum, as the system equilibrates. This conceptual problem is naturally handled by the formalism of nonlinear systems of differential equations, as is the quantitative analysis of equilibration and its relation to learning, recognition, and recall. To sidestep these technical difficulties, we complete our intuitive discussion using suggestive but incompletely specified terminology.

Because of the design of its cooperative-competitive interactions,  $F_i$  is capable of matching a bottom-up input pattern  $I$  with a top-down template  $V$  (Grossberg, 1976b, 1983). The functional units that are matched or mismatched at  $F_i$  are whole input patterns, which are spatially distributed across the nodes of  $F_i$  rather than the inputs to each node. This choice of functional units is not an independent hypothesis. It is forced by two sets of mathematical results that follow from more basic theoretical hypotheses. The first set of mathematical results proves that the functional unit of associative learning and LTM in such networks is a spatial pattern of LTM traces (Grossberg, 1969a, 1982c). The second set of mathematical results proves that the functional unit of matching and STM is a spatial pattern of STM traces (Grossberg, 1970a, 1982c).

We illustrate what is meant by saying that the functional units are spatial patterns of STM and LTM traces, rather than individual traces, with the following properties: A bottom-up input and a top-down template signal to a node can be equal, yet part of mismatched bottom-up and top-down patterns. A bottom-up input and top-down template signal to a node can be different, yet part of perfectly matched bottom-up and top-down patterns. The relative sizes of the traces in an STM pattern or an LTM pattern determine the "information" carried by the pattern. The scaling parameter, which multiplies these relative sizes into their actual sizes, is an "energy" variable that determines properties such as how quickly this "information" is processed, rather than what the information is.

In the special case wherein a top-down template  $V$  perfectly matches a bottom-up input pattern  $I$ , the relative sizes of the activities that compose  $X$  can be preserved after  $V$  is registered, whereas the absolute size of each activity is proportionally increased. Thus a perfect top-down match does not disrupt ongoing bottom-up pattern processing. Rather it facilitates such processing by amplifying its energy without causing pattern distortions. This energy amplification due to pattern matching is one of the properties used to generate adaptive resonances.

Pattern matching within such a network thus does not just compare bottom-up and top-down inputs at each node, as in Euclidean matching algorithms. Instead, the network senses the

degree of match between subpatterns of the total bottom-up and top-down input. As Figure 3 shows, approximate matches between subpatterns of  $I$  and  $V$  tend to enhance the corresponding activities of  $X$ , whereas serious mismatches between  $I$  and  $V$  tend to suppress the corresponding activities of  $X$ . The effect of an approximate match is to deform  $X$  so that it forms a compromise between  $I$  and  $V$ . When  $X$  is deformed, so are all the patterns that form the map  $X \rightarrow S \rightarrow T \rightarrow Y$ . The net effect of all these shifts is to represent the same input pattern  $I$  by a new activity pattern  $Y$  whose template  $V$  provides the best match to  $I$ . As this best match equilibrates, the system enters a state of energy amplification, or resonance. The resonant state persists long enough for the slowly varying LTM traces of the bottom-up adaptive filter  $S \rightarrow T$  and the top-down template  $U \rightarrow V$  to adjust their values to the resonating STM patterns. Only LTM traces lying in pathways that carry signals to or away from active nodes in  $F_{i+1}$  can learn after resonance has been reached. The computational unit of the feedback network  $F_i \leftrightarrow F_{i+1}$  is called an *adaptive resonance* because resonance triggers learning.

When learning does occur, its form is quite simple. Denote by  $z_{kj}$  the LTM traces in the pathways from node  $v_k$  in  $F_i$  to node  $v_j$  in  $F_{i+1}$  (see Figure 1b). We restrict attention to the LTM traces  $z_{kj}$  in the pathways that lead to a fixed node  $v_j$  of  $F_{i+1}$ . Whenever the STM trace  $x_j$  at  $v_j$  remains suprathreshold for a long enough time, each  $z_{kj}$  can gradually become proportional to the signal  $S_{kj}$  emitted by  $v_k$  into its pathway. Two important properties are implicit in this statement. First, the rate with which  $z_{kj}$  changes increases with the size of  $x_j$ . Second, while  $x_j$  is active, the LTM vector  $z_j \equiv (z_{1j}, z_{2j}, \dots, z_{nj})$  of the LTM traces leading to  $v_j$  gradually becomes parallel to the signal vector  $S_j = (S_{1j}, S_{2j}, \dots, S_{nj})$  emitted by  $F_i$  to  $v_j$ .

The following learning law (Grossberg, 1969a, 1976a) is the simplest differential equation for the rate of change  $(d/dt)z_{kj}$  of the LTM trace  $z_{kj}$  that rigorously captures these properties:

$$\frac{d}{dt}z_{kj} = \alpha x_j (-z_{kj} + \beta S_{kj}). \quad (1)$$

In Equation 1,  $\alpha$  and  $\beta$  are positive constants. At times when  $x_j = 0$ , it follows that  $(d/dt)z_{kj} = 0$  so that no learning occurs. Positive values of  $x_j$  induce positive learning rates. Increasing  $x_j$  increases the learning rate. The rate constant  $\alpha$  is sufficiently small that momentary activations of  $x_j$  cause insignificant changes in  $z_{kj}$ . If  $x_j$  remains positive for a sufficiently long time—that is, if  $x_j$  is stored in STM at  $F_{i+1}$ —then  $z_{kj}$  approaches the value  $\beta S_{kj}$ , which is proportional to the signal  $S_{kj}$ . For simplicity of exposition, we suppose that  $\beta = 1$ . Then Equation 1 shows that the LTM vector  $z_j$  approaches the signal vector  $S_j$  at a rate that increases with the STM activity  $x_j$ .

The total input  $T_j$  to  $v_j$  is defined by

$$T_j = \sum_{k=1}^n S_{kj} z_{kj}, \quad (2)$$

which can also be written as the dot product  $T_j = S_j \cdot z_j$  of the vectors  $S_j$  and  $z_j$ . As  $z_j$  approaches  $S_j$  because of learning,  $T_j$  becomes larger. This property indicates how a fixed signal vector  $S_j$  can generate an amplified and faster reaction of  $x_j$  due to prior

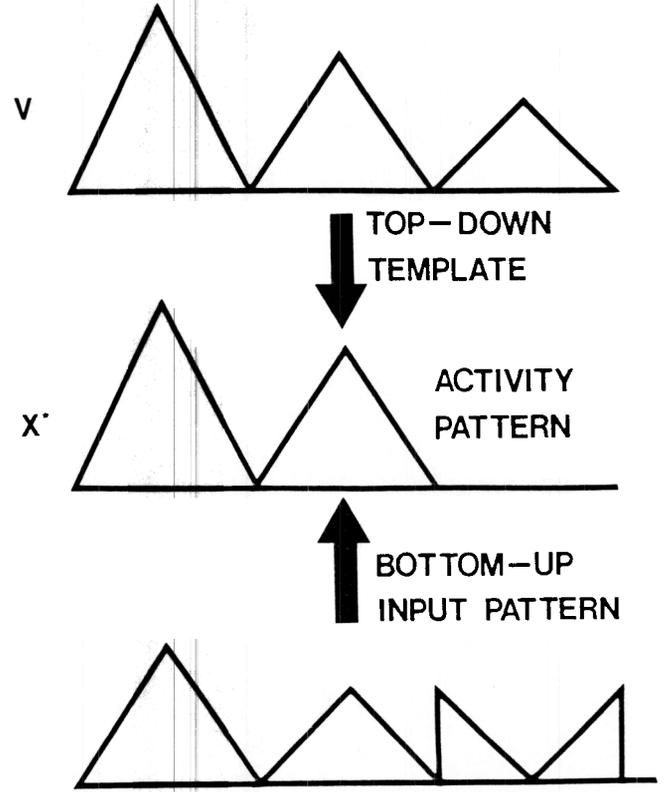


Figure 3. Matching of a bottom-up input pattern with a top-down template pattern: Regions of approximate match lead to an amplification of the corresponding activities. Regions of serious mismatch lead to a suppression of the corresponding activities. ( $I$  = input pattern;  $V$  = template pattern;  $X^*$  = activity pattern due to conjoint action of  $I$  and  $V$ .)

learning by  $z_j$ . This type of performance speed-up does not require a larger total size

$$\sum_{k=1}^n z_{kj} \quad (3)$$

of the LTM traces abutting  $v_j$ . Rather, it requires a repatterning of the LTM traces within the vector  $z_j$ .

The signal patterns  $S_j$  that succeed in activating  $v_j$  can differ on successive learning trials. The LTM vector  $z_j$  thus encodes a weighted time average of all the signal patterns  $S_j$  that  $v_j$  has actually sampled. The weight of a particular pattern  $S_j$  in this average increases with the intensity and duration of  $x_j$ 's suprathreshold activity during the learning episodes when  $S_j$  was active. Even though  $v_j$  may have intensely sampled a particular  $S_j$  on a previous learning trial, the LTM vector  $z_j$  will not generally equal  $S_j$  on the current trial. The LTM vector  $z_j$  provides a statistical measure of all the patterns  $S_j$  that ever activated  $v_j$ . If many signal vectors  $S_j$  succeed in activating  $v_j$ , then  $z_j$  will be different from any one of the vectors  $S_j$ , since it computes an average of all the vectors. Thus the dot product  $T_j = S_j \cdot z_j$  of any one of these vectors  $S_j$  with the average encoded by  $z_j$  may become smaller as the number of exemplars  $S_k$  within the category corresponding to  $v_j$  increases. The amount of practice on a single pattern (i.e., familiarity) and category variance thus tend to have opposite effects on the reaction rate of  $F_{i+1}$ .

A similar type of learning rule governs the LTM traces  $Z_{jk}$  in the top-down pathways from nodes  $v_j$  in  $F_{i+1}$  to nodes  $v_k$  in  $F_i$  (see Figure 2b). Again  $x_j$  must be suprathreshold before  $Z_{jk}$  can learn. At least one fundamental difference exists, however, between bottom-up learning and top-down learning. During bottom-up learning, each LTM vector  $z_j$  determines its own total input  $T_j$  to  $v_j$ . These inputs trigger a cooperative-competitive struggle among the STM activities  $x_j$  of  $F_{i+1}$ . Consider, for example, the case in which the node  $v_j$  that receives the largest input  $T_j$  wins the STM competition. Then each input  $T_j$  acts to improve the competitive advantage of its own STM trace  $x_j$ .

The same is not true during top-down learning. Once the bottom-up filter determines which STM activities  $x_j$  in  $F_{i+1}$  will be stored, all of the suprathreshold  $x_j$  values generate signal vectors  $U_j = (U_{j1}, U_{j2}, \dots, U_{jm})$  to  $F_i$ . All of the signal vectors  $U_j$  simultaneously read out their top-down LTM patterns  $Z_j = (Z_{j1}, Z_{j2}, \dots, Z_{jm})$  to form a single top-down template  $V = (V_1, V_2, \dots, V_m)$ , where

$$V_k = \sum_{j=1}^m U_{jk} Z_{jk}. \quad (4)$$

Thus, whereas bottom-up filtering separates the effects of the LTM vectors  $z_j$  in order to drive the STM competition within  $F_{i+1}$ , the survivors of this competition pool their top-down LTM vectors to form a consensus in the form of a composite template to  $F_i$ . This template, as a whole, is then matched against the bottom-up input pattern to  $F_i$ . If the LTM pattern  $Z_j$  of one node  $v_j$  in  $F_{i+1}$  matches the input pattern much better than the other LTM patterns that form the template consensus, then this node will win a larger portion of the total STM activity across  $F_{i+1}$  as the resonance equilibrates. In other words, the template presents a consensus to  $F_i$  so that the matching process within  $F_i$  can rearrange the competitive STM balance across  $F_{i+1}$  and thereby better match the bottom-up data at  $F_i$ . Computer simulations that illustrate these competitive and matching properties during pattern recognition and category formation via adaptive resonance are described by Carpenter and Grossberg (1985).

## 8. Counting Stages: Resonant Equilibration as Verification and Attention

Many of the types of properties that the Verification model and the Posner-Snyder model addressed are mechanized and extended by using the parallel process of resonant equilibration. For example, other things being equal, high-frequency words are more parallel to the LTM vectors  $z_j$  and  $Z_j$  that they persistently activate than low-frequency words are to their preferred LTM vectors. This property helps to explain how a verification-type process can occur without invoking serial search. To understand why, suppose that a low-frequency word activates an output signal pattern  $S$  from  $F_i$  to  $F_{i+1}$  that does not correspond to any high-frequency word. Before  $S$  can influence  $F_{i+1}$  via  $T$ , it is gated by the bottom-up LTM traces. Because of the relatively large effect of high-frequency words on LTM tuning, the largest inputs  $T_j$  to  $F_{i+1}$  may initially bias the STM reaction of  $F_{i+1}$  toward a high-frequency interpretation of the low-frequency word. Consequently, the fastest and largest top-down signals in the template

$V$  may initially tend to code high-frequency words. The template  $V$  plays the role of the verification signal. Thus high-frequency word components can tend to make a faster and larger contribution to the early phase of "verification," even though the process that instantiates the template is a parallel read-out rather than serial search.

After the matching process between bottom-up and top-down templates is initiated at  $F_i$ , it can change the bottom-up inputs  $T_j$  in favor of the low-frequency word. That is, although high-frequency components across  $F_{i+1}$  may be favored initially, once "verification" begins at  $F_i$ , lower frequency components across  $F_{i+1}$  whose top-down signals better match the sensory data at  $F_i$  are amplified in STM as the high-frequency components are attenuated. No serial search is needed for this to occur. This capability is already implied by the conjoint action of rules whereby (a) the bottom-up filter biases the STM competition across  $F_{i+1}$  and (b) the top-down template synthesizes a consensus for matching across  $F_i$ . We maintain the intuition of the verification model—that many alternative interpretations are tested until the correct one is found—but the serial search mechanism is replaced by a parallel zooming-in process whereby top-down consensus becomes better matched to the sensory data as the bottom-up filter chooses a better STM representation. No serial mechanisms are needed to accomplish this process of matching, deformation, and equilibration.

After the resonance quickly equilibrates in this way, the final resonant STM pattern is the one to which attention is paid. As in the Posner-Snyder model, the initial phase of activating the "logogens" across  $F_{i+1}$  uses excitatory  $F_i \rightarrow F_{i+1}$  signals. The subsequent stages of competition for STM storage across  $F_{i+1}$  and matching across  $F_i$  both use lateral inhibitory signals, which is also consistent with the Posner-Snyder model. Attention does not, however, switch to the logogens before inhibition can act. Inhibition can help to draw attention to its final locus by focusing the process of resonant equilibration. Also, in these processes, the mechanism of inhibition does not primarily function to impose a limited capacity. This fact reflects a major difference of our theory from the Posner-Snyder and verification models. The computational unit in our networks is not a logogen that can be excited or inhibited by other network nodes. The computational unit is a distributed pattern of activity across a field of network nodes. Inhibition may help to cause an amplified STM reaction to a matched bottom-up input pattern and top-down template pattern, as well as to suppress the STM response to mismatched input patterns. These facilitative and suppressive reactions can occur without changing the set of nodes that the input patterns activate or the total activities of the input patterns. Thus the mechanism of inhibition can cause amplification or suppression, not as a function of network capacity, but rather as a function of pattern match or mismatch.

Another reflection of the difference between computational units in the theories is that inhibition also acts before signals ever leave  $F_i$  for  $F_{i+1}$ . The same inhibitory interactions within  $F_i$  that are used for top-down template matching at  $F_i$  are also used to accurately register the bottom-up input at  $F_i$  without noise or saturation before  $F_i$  ever elicits its excitatory signals to  $F_{i+1}$ . The fact that this early phase of inhibitory interaction prior to the activation of the "logogens" at  $F_{i+1}$  does not primarily function to limit capacity reflects the deeper property that the

computational unit of the network is not a logogen at all (Grossberg, 1980, 1984).

Perhaps the strongest departure from both the Posner-Snyder model and the verification model concerns the concept of processing stages. Even if we restrict our consideration to two levels  $F_i$  and  $F_{i+1}$ , these two levels do not correspond to two processing stages of the sort that these other models have described. In particular, these stages are not separable from one another. Rather, each feeds into the next and thereby alters its own processing due to feedback. The feedback continues to cycle between the stages as it drives the approach to resonance and maintains the resonance after it has been attained. We believe that it is the absence of this type of insight in the Posner-Snyder model that led to its problematic argument about how inhibition and attention are related (see Section 4). We see no way to fully understand these basic intuitions except through the use of nonlinear systems of differential equations. Once this is accepted, serial stage models and binary computer models of human information processing seem much less appealing than they once did. In fact, the computational units of our macrotheory are not even necessarily determined within individual macrotheory levels. Rather, computational units such as adaptive resonances emerge from interactions using microtheory mechanisms that are distributed across several macrotheory levels.

### 9. Attentional Gain Control Versus Attentional Priming: The 2/3 Rule

The present theory makes another distinction that is not mechanistically elaborated within the Posner-Snyder theory. As Neely (1977) noted, experiments like those of Posner and Snyder (1975b) and Neely (1976) "confounded the facilitatory effects of conscious attention with the facilitatory effects of automatic spreading activation" (p. 231). The Posner-Snyder model accounts for these two types of facilitation by positing a separate process for each. In the present theory, at the level of mechanism, the two types of facilitation both share "automatic" properties. They are distinguished by factors other than their automaticity. A central distinction in the development of the present theory is the difference between attentional gain control and attentional priming (Grossberg, 1975, 1982b).

The need for distinct mechanisms of attentional gain control and attentional priming can be appreciated by considering Figure 4. In Figure 4a, a learned top-down template from  $F_{i+1}$  to  $F_i$  is activated before a bottom-up input pattern activates  $F_i$ . The level  $F_i$  is then primed, or ready, to receive a bottom-up input that may or may not match the active template. The template represents the input that the network expects to receive. The template plays the role of an expectancy.

Level  $F_i$  can be primed to receive a bottom-up input without necessarily eliciting suprathreshold output signals in response to the priming expectancy. If this were not possible, then every priming event would lead to suprathreshold reactions. Such a property would prevent anticipation of a future event. On the other hand, certain top-down expectancies can lead to suprathreshold consequences, much like we can at will experience internal conversations and musical events, as well as other fantasy activities. Thus there exists a difference between the read-out of a top-down expectancy, which is a mechanism of attentional

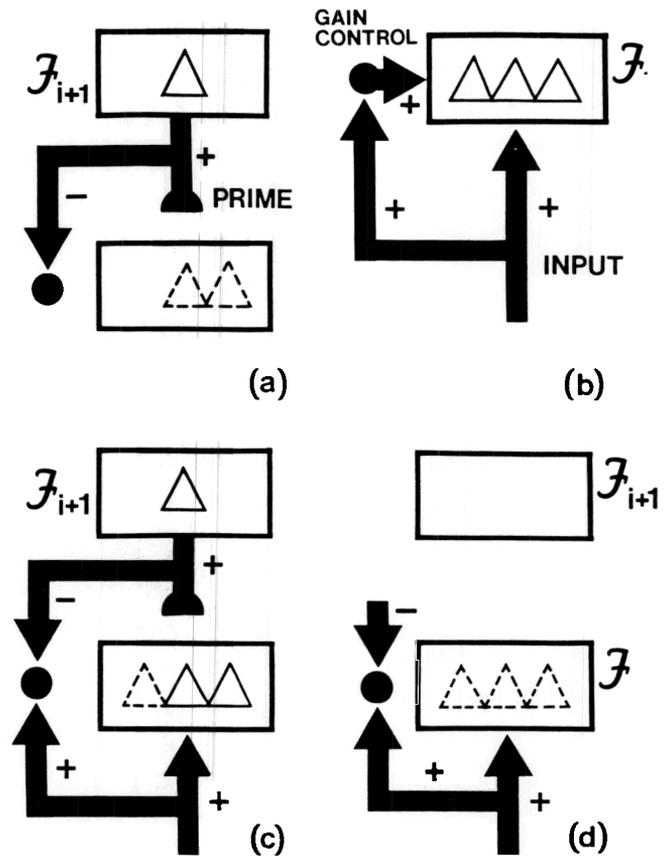


Figure 4. Interaction of attentional priming with attentional gain control: (a) A supraliminal activity pattern within  $F_{i+1}$  causes a subliminal response within  $F_i$ . (b) A bottom-up input pattern can instate a supraliminal activity pattern within  $F_i$  by engaging the attentional gain control channel. (c) During bottom-up and top-down matching, only cells at  $F_i$  that receive convergent bottom-up and top-down signals can generate supraliminal reactions. (d) An inhibitory attentional gain control signal from a competing source can block supraliminal reaction to a bottom-up input.

priming, and the translation of this operation into suprathreshold signals due to attentional gain control.

The distinction between attentional priming and attentional gain control can be sharpened by considering the opposite situation, in which a bottom-up input pattern  $I$  activates  $F_i$  before a top-down expectancy from  $F_{i+1}$  can do so. As discussed earlier, we want the input pattern  $I$  to generate a suprathreshold activity pattern  $X$  across  $F_i$  so that the sequence of transformations  $X \rightarrow S \rightarrow T \rightarrow Y \rightarrow U \rightarrow V \rightarrow X^*$  can occur. How does  $F_i$  know that it should generate a suprathreshold reaction to the bottom-up input pattern but not to the top-down input pattern? In both cases, an input pattern stimulates the cells of  $F_i$ . Some other mechanism must exist that distinguishes between bottom-up and top-down inputs. We use the mechanism of attentional gain control.

These considerations suggest that at least two of the three signal sources to  $F_i$  must be simultaneously active in order for some  $F_i$  cells to become supraliminally active. Carpenter and Grossberg (1985, 1986) called this constraint the  $\frac{2}{3}$  Rule. These three signal

sources are (a) the bottom-up input channel that delivers specific input patterns to  $F_i$ ; (b) the top-down template channel that delivers specific expectancies, or priming signals, to  $F_i$ ; and (c) the attentional gain control channel that nonspecifically modulates the sensitivity of  $F_i$ . Figure 4 illustrates one realization of the  $\frac{2}{3}$  Rule.

As shown in Figure 4a, supraliminally active cells within  $F_{i+1}$  read out a specific top-down expectancy to  $F_i$  along excitatory (+) and conditionable pathways. The active  $F_{i+1}$  cells also read out inhibitory (-) signals that converge in a nonspecific fashion on the cells that regulate attentional gain control. Because all the cells in  $F_i$  receive inputs from at most one channel, they cannot generate supraliminal activations. By contrast, as shown in Figure 4b, the bottom-up input channel instates a specific input pattern at  $F_i$  and excites the attentional gain control channel, which nonspecifically sensitizes all the cells of  $F_i$ . (Alternatively, the attentional gain control channel may remain endogenously, or tonically, active.) Those cells at which a bottom-up input and an attentional gain control signal converge can generate supraliminal activations. Cells that receive no bottom-up input cannot generate supraliminal activations, because they receive inputs only from a single signal source. As shown in Figure 4c, a bottom-up input pattern and a top-down template pattern are simultaneously active. The top-down signal source shuts off the nonspecific gain control channel. However, if the bottom-up input pattern and the top-down template pattern are not too different, then some cells in  $F_i$  will receive inputs from both signal channels. By the  $\frac{2}{3}$  Rule, these cells can become supraliminally active. Cells that receive inputs from only the bottom-up input or the top-down template, but not both, cannot become supraliminally active. Nor can cells become active that receive no inputs whatsoever.

Thus, in addition to suggesting how  $F_i$  can respond supraliminally to bottom-up inputs and subliminally to top-down templates, the  $\frac{2}{3}$  Rule suggests a rule for matching bottom-up and top-down patterns at  $F_i$ . Carpenter and Grossberg (1985) showed, moreover, that use of the  $\frac{2}{3}$  Rule for matching is necessary to prevent ceaseless recoding of the learned categories that form in response to sequences of bottom-up input patterns.

The  $\frac{2}{3}$  Rule leads to other useful conclusions as well. For example, a supraliminal reaction to bottom-up inputs does not always occur, especially when attention is focused on a different processing modality. This can be understood as a consequence of intermodal competition between attentional gain control signals. As Figure 4d shows, the attentional gain control channel is inhibited by such a competitive signal; hence, only a subliminal reaction to the bottom-up input occurs. In a similar way, top-bottom signals do not always generate subliminal reactions. We can willfully generate conscious internal fantasy activities from within. Supraliminal reactions in response to top-down signals can occur if the attentional gain control channel can be activated by an "act of will."

To ground these concepts in a broader framework, we make the following observations. The  $\frac{2}{3}$  Rule was first noticed in a study of reinforcement and attention (Grossberg, 1975; reprinted in Grossberg, 1982c, p. 290). In that context, the two specific channels carried internal drive inputs and conditioned reinforcer signals, respectively. The third channel carried nonspecific arousal signals. The  $\frac{2}{3}$  Rule in that situation suggests how incentive mo-

tivational signals can be generated by pairwise combinations of these signal channels. In that application, the  $\frac{2}{3}$  Rule suggested how attention is modulated by the processing of emotion-related signals. More generally, sensitization of specific processing channels by a nonspecific channel is a theme that occurs even in the modeling of invertebrate motor control (Grossberg, 1978a; reprinted in Grossberg, 1982c, pp. 517-531). Thus the concepts of attentional gain control and attentional priming reflect design constraints that seem to be used in several modalities as well as across species.

The attentional priming and attentional gain control processes can be used to clarify and modify the "automatic activation" and "conscious attention" constructs found elsewhere in the literature. We discuss these constructs to form a bridge to other models. Our discussion simultaneously refines the mechanistic concepts that we introduced above and shows that these mechanistic concepts do not match well with descriptive ideas such as "automatic activation" and "conscious attention." (Quotation marks emphasize this mismatch.) For present purposes, we suppose that a supraliminal activity pattern across  $F_i$  that can survive matching against a top-down expectancy from  $F_{i+1}$  becomes "conscious." In other words, activity across  $F_i$  can become conscious if it persists long enough to undergo resonant equilibration. Activity within  $F_{i+1}$  never becomes conscious, whether or not it is supraliminal. The discussion is broken up into four parts: (a) *top-down subliminal control*—"automatic" attentional priming in the absence of attentional gain control; (b) *top-down supraliminal control*—"automatic" attentional priming plus "willed" excitatory attentional gain control; (c) *bottom-up supraliminal control*—"automatic" content-addressable input activation plus "automatic" excitatory attentional gain control; (d) *bottom-up subliminal control*—"automatic" content-addressable input activation plus "automatic" or "willed" inhibitory attentional gain control. These properties are explained mechanistically as follows.

*Top-down subliminal control.* A top-down expectancy from  $F_{i+1}$  to  $F_i$  has a direct excitatory effect on  $F_i$  (Figure 4a). The  $F_{i+1}$  cells do not control a nonspecific excitatory signal capable of sensitizing all the cells of  $F_i$  to its inputs. Such a nonspecific arousal signal is said to lower the quenching threshold (QT) of  $F_i$  (Grossberg, 1973, 1980), which in the absence of inputs is chosen to be large. The QT is a parameter that STM activities must exceed to elicit a suprathreshold reaction. Thus  $F_i$  cannot generate a suprathreshold output in response to a top-down expectancy alone. The top-down expectancy can prime  $F_i$  but cannot release the priming pattern.

*Top-down supraliminal control.* An "act of will" can activate the attentional gain control channel. This act generates a nonspecific arousal signal that sensitizes all the cells of  $F_i$  to whatever inputs happen to be delivered at that time. This type of willed control does not deliver a specific pattern of information to be processed. Rather it exerts a nonobligatory type of sensitivity modulation on the entire processing channel.

This way of mechanizing the distinction between attentional priming and attentional gain control is a special case of a general design principle in neural information processing, namely, the *factorization of pattern and energy*, or *factorization of information and arousal* (Grossberg, 1982c), which is also mentioned in Section 7. Another example of the dissociation between information and arousal occurs during willed motor acts. When one looks at

an object, one can decide to reach it with one's left hand or right hand or not at all. Learned target position maps are automatically read out by the eye-head system to the hand-arm systems corresponding to both hands. These maps remain subliminally active at their target cells until an "act of will" generates nonspecific arousal to the desired hand-arm system. Suprathreshold read-out of the corresponding target position map is thereby effected. In this example, the top-down expectancy is a target position map that encodes the desired terminal positions of the hand-arm system with respect to the body.

*Bottom-up supraliminal control.* When a bottom-up input pattern activates  $F_i$ , it has two simultaneous effects on  $F_i$  (Figure 4b). The obvious effect is the delivery of the input pattern directly to  $F_i$ . As this is happening, the input also activates a nonspecific channel. This nonspecific channel controls the nonspecific arousal, or attentional gain control, that sensitizes  $F_i$  by lowering its quenching threshold. Thus the bottom-up pathway "automatically" instates its input and "automatically" activates the attentional gain control system. It is misleading to suggest that "automatic spreading activation" and "conscious attention" are two independent stages of information processing, because the same mechanisms that give rise to "conscious attention" are often "automatically" activated in parallel with the mechanisms that subserve "automatic spreading activation."

*Bottom-up subliminal control.* The automatic activation of attentional gain control by a bottom-up input pattern can be prevented from generating a suprathreshold response at  $F_i$  (Figure 4d). If a given processing channel is already active, then its attentional gain control mechanism can competitively inhibit the automatic activation of the gain control mechanisms within other channels. A large-scale competition exists between the gain control sources of different processing channels in addition to the small-scale cooperative-competitive interactions within each processing channel that regulate its STM matching and contrast enhancement properties.

Bottom-up subliminal control shows that inhibition is mobilized not only by "conscious attention" processes, as the Posner-Snyder model suggests. Attentional gain control signals elicited by bottom-up inputs in one channel can cause "automatic" attentional inhibition in a different processing channel. Moreover, "automatic" excitatory attentional gain control signals in a given channel cause "automatic" inhibitory signals in that channel by rendering suprathreshold the small-scale competitions that regulate STM matching and contrast enhancement within the channel. "Conscious attention" can be the outcome, rather than the cause, of this inhibitory process.

Neely (1977) designed and performed a remarkable set of experiments to unconfound the facilitatory effects of conscious attention and the facilitatory effects of automatic spreading activation. The present theory also recognizes at least two different types of facilitatory effects, attentional gain control and attentional priming, but does not attribute these properties to "conscious" and "automatic" processes in the same manner as the Posner-Snyder (1975a) theory. Despite this fact, the present theory can explain the Neely (1977) data. A more serious test of the theory concerns its ability to explain transient versus equilibration effects, such as effects of inconsistent versus consistent primes and of mask versus no-mask manipulations. Another serious test concerns the theory's ability to predict different out-

comes of recognition and recall tests. These applications of the theory are presented after the next section.

Before turning to these applications, we outline a theoretical macrocircuit that embodies the theory's view of how learning, recognition, and recall take place in real time. The macrocircuit embeds our analysis of word recognition and recall in a broader theory. This theory clarifies the types of information that are coded by the functional units at different processing stages and locates the stages subserving word recognition and recall processes. Such a theory is necessary to understand how the functional units arise and how they give rise to observable behaviors. The reader can skim this section on a first reading.

## 10. A Macrocircuit for the Self-Organization of Recognition and Recall

Figure 5 depicts a macrocircuit that is capable of self-organizing many recognition and recall properties of visual, verbal, and motor lists. The boxes  $A_i$  are macrostages for the elaboration of audition, speech, and language. The boxes  $M_j$  are macrostages for the elaboration of language-related motor representations. The box  $V^*$  designates a source of preprocessed visual signals.

At an early stage of development, the environmentally activated auditory patterns at stage  $A_1$  start to tune the long-term memory (LTM) traces within the pathways of the adaptive filter from  $A_1$  to  $A_2$ , and thus to alter the patterning of short-term memory (STM) auditory "feature detector" activation across  $A_2$ . After this tuning process begins, endogenous activations of the motor command stage  $M_1$  can elicit simple verbalizations (babbling) whose environmental feedback can also tune the  $A_1 \rightarrow A_2$  adaptive filter. The learning within the feedback pathway  $M_1 \rightarrow A_2 \rightarrow A_1$  helps to tune auditory sensitivities to articulatory requirements. This process is consistent with the motor theory of speech perception (Cooper, 1979; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978; Mann & Repp, 1981; Repp & Mann, 1981; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

Just as the auditory patterns across  $A_1$  tune the  $A_1 \rightarrow A_2$  adaptive filter, the endogenously activated motor command patterns across  $M_1$  tune the  $M_1 \rightarrow M_2$  adaptive filter. The activation patterns across  $M_2$  encode the endogenously activated motor commands across  $M_1$  using the same mechanisms by which the activation patterns across  $A_2$  encode the exogenously activated auditory patterns across  $A_1$ .

The flow of adaptive signaling is not just bottom-up from  $A_1$  to  $A_2$  and from  $M_1$  to  $M_2$ . Top-down conditionable signals from  $A_2$  to  $A_1$  and from  $M_2$  to  $M_1$  are also hypothesized to exist. These top-down signal patterns represent learned expectancies, or templates. Their most important role is to stabilize the learning that goes on within the adaptive filters  $A_1 \rightarrow A_2$  and  $M_1 \rightarrow M_2$ . In so doing, these top-down signal patterns also constitute the read out of optimal templates in response to ambiguous or novel bottom-up signals. These optimal templates predict the patterns that the system expects to find at  $A_1$  or  $M_1$  on the basis of past experience. The predicted and actual patterns merge at  $A_1$  and  $M_1$  to form completed composite patterns that are a mixture of actual and expected information.

Auditory and motor features are linked via an associative map from  $A_2$  to  $M_2$ . When  $M_1$  is endogenously activated, it activates

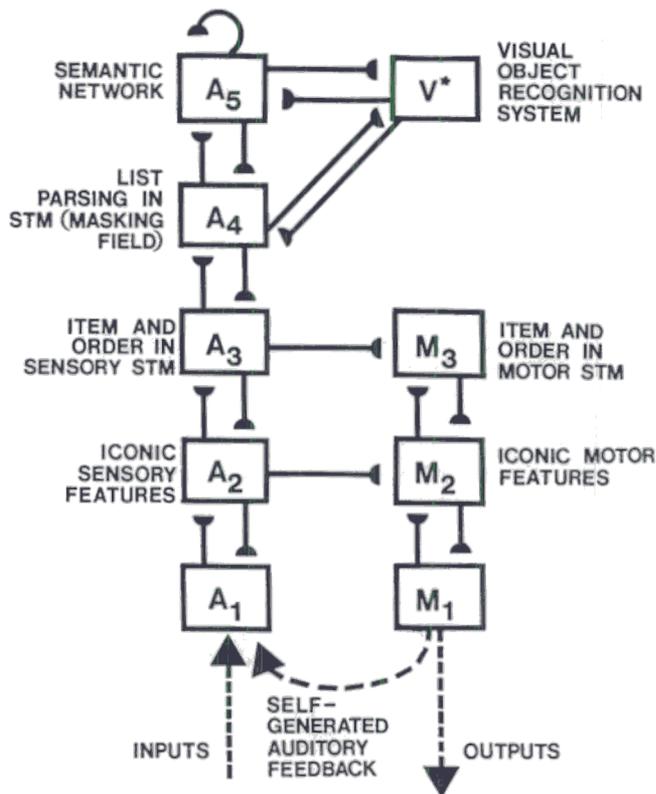


Figure 5. A macrocircuit governing self-organization of recognition and recall processes: The text explains how auditorily mediated language processes (the  $A_i$ ), visual recognition processes ( $V^*$ ), and motor control processes (the  $M_i$ ) interact internally via conditionable pathways (black lines) and externally via environmental feedback (dotted lines) to self-organize the various processes which occur at the different network stages.

a motor representation at  $M_2$  via its adaptive filter  $M_1 \rightarrow M_2$ , as well as an auditory representation at  $A_2$  via environmental feedback  $M_1 \rightarrow A_1$  and the adaptive filter  $A_1 \rightarrow A_2$ . Because  $A_2$  and  $M_2$  are then simultaneously active, the associative map  $A_2 \rightarrow M_2$  can be learned. This map also links auditory and articulatory features.

The associative map  $A_2 \rightarrow M_2$  enables the imitation of novel sounds—in particular, of non-self-generated sounds—to get underway. It does so by analyzing a novel sound via the bottom-up auditory filter  $A_1 \rightarrow A_2$ , mapping the activation patterns of auditory features detectors into activation patterns of motor feature detectors via the associative map  $A_2 \rightarrow M_2$ , and then synthesizing the motor feature pattern into a net motor command at  $M_1$  via the top-down motor template  $M_2 \rightarrow M_1$ . The motor command, or synergy, that is synthesized in this way generates a sound that is closer to the novel sound than are any of the sounds currently coded by the system. The properties whereby the learned map  $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$  enables imitation of novel sounds to occur are in agreement with the analysis-by-synthesis approach to speech recognition (Halle & Stevens, 1962; Stevens, 1972; Stevens & Halle, 1964).

The environmental feedback from  $M_1$  to  $A_1$  followed by the learned map  $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$  defines a closed feedback loop, or "circular reaction" (Piaget, 1963). Thus the present the-

ory's explication of the developmental concept of circular reaction helps to clarify the speech performance concepts of motor theory and analysis-by-synthesis in the course of suggesting how an individual can begin to imitate non-self-generated speech sounds.

The stages  $A_2$  and  $M_2$  can each process just one spatial pattern of auditory or motor features at a time. Thus  $A_2$  can process an auditory "feature code" that is derived from a narrow time slice of a speech spectrogram, and  $M_2$  can control a simple motor synergy of synchronously coordinated muscle contractions. These properties are consequences of the fact that spatial patterns, or distributed patterns of activity across a field of network nodes, are the computational units in embedding field networks. This computational unit is a mathematical consequence of the associative learning laws that govern these networks (Grossberg, 1969a, 1982c). This fact is not intuitively obvious and was considered surprising when first discovered. The later stages  $A_i$  and  $M_i$  in Figure 5 are all devoted to building up recognition and recall representations for temporal groupings, or lists, of spatial pattern building blocks. These higher stages embody solutions to aspects of the fundamental problem of self-organizing serial order in behavior (Lashley, 1951).

A spatial pattern of activation across  $A_2$  encodes the relative importance of each "feature detector" of  $A_2$  in representing the auditory pattern that is momentarily activating  $A_1$ . To encode temporal lists of auditory patterns, one needs first to simultaneously encode a sequence of spatial patterns across  $A_2$ 's auditory feature detectors. The following way to accomplish this also addresses the vexing problem that individual speech sounds and thus their spatial patterns across  $A_2$  can be altered by the temporal context of other speech sounds in which they are embedded.

In addition to activating the associative map from  $A_2$  to  $M_2$ , each spatial pattern across  $A_2$  also activates an adaptive filter from  $A_2$  to  $A_3$ . Although all the adaptive filters of the theory obey the same laws, each filter learns different information depending on its location in the network. Because the  $A_2 \rightarrow A_3$  filter is activated by feature patterns across  $A_2$ , it builds up learned representations, or chunks, of these feature patterns. Each such representation is called an *item representation* within the theory.

It is important to realize that all new learning about item representations is encoded within the LTM traces of the  $A_2 \rightarrow A_3$  adaptive filter. Although each item representation is expressed as a pattern of activation across  $A_3$ , the learning of these item representations does not take place within  $A_3$ . This flexible relation between learning and activation is needed so that temporal codes for lists can be learned and performed. For example, whereas the spatial patterns across  $A_2$  can rapidly decay via a type of iconic memory (Sperling, 1960), the item representations across  $A_3$  are stored in short-term memory (STM), also called *working memory* (Cermak & Craik, 1979). As a sequence of sound patterns activates  $A_1$ , a succession of item representations is stored in STM across  $A_3$ . The spatial pattern of STM activity across  $A_3$  represents temporal order information across the item representations of  $A_3$ .

This temporal order information cannot be laid down arbitrarily without causing temporally unstable LTM recordings to occur in the adaptive filter from  $A_3$  to  $A_4$ . Laws for regulating temporal order information in STM have been derived from the LTM *invariance principle*. This principle shows how to alter the STM activities of previous items in response to the presentation

of new items so that the repatterning of STM activities that is caused by the new items does not inadvertently obliterate the LTM codes for old item groupings. These STM patterns of temporal order information have been used, for example, to explain and predict data showing primacy and recency gradients during free-recall experiments (Grossberg, 1978b). Computer simulations that illustrate how these temporal order patterns evolve through time are described by Grossberg and Stone (1985).

The concept of temporal order information across item representations is necessary but not sufficient to explain how lists of items can be learned and performed. One also needs to consider the analogous bottom-up filtering process from  $M_2$  to  $M_3$  that builds up unitized representations of motor items (synergies), the top-down learned templates from  $M_3$  to  $M_2$ , and the associative map  $A_3 \rightarrow M_3$  that is learned from sensory items to motor items. In particular, suppose that analysis by synthesis (the map  $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$ ) has elicited a novel pattern of sensory features across  $A_2$  and of motor features across  $M_2$ . These feature patterns can then generate unitized item representations at  $A_3$  and  $M_3$  even though the network never endogenously activated these patterns during its babbling phase. A map  $A_3 \rightarrow M_3$  between these unitized item representations can then be learned. Using these building blocks, we can now show how a unitized representation of an entire list can be learned and performed.

When the network processes a verbal list, it establishes an STM pattern of temporal order information across the item representations of  $A_3$ . Because every sublist of a list is also a list, the adaptive filter from  $A_3$  to  $A_4$  simultaneously "looks at" all the sublist groupings to which it is sensitive as a list is presented through time. The cooperative-competitive interaction across  $A_4$  then determines which of these sublist representations will be stored in STM at  $A_4$ .

To enable  $A_4$  to store maximally predictive sublist chunks, the interactions within  $A_4$  simultaneously solve several problems. A core problem is called the *temporal chunking problem*. Consider the problem of unitizing an internal representation for an unfamiliar list of familiar items, for example, a novel word composed of familiar phonemes (auditory) or letters (visual). The most familiar groupings of the list are the items themselves. To even know what the novel list is, all of its individual items must first be presented. All of these items are more familiar than the list itself. What mechanisms prevent item familiarity from forcing the list always to be processed as a sequence of individual items, rather than eventually as a whole? How does a not-yet-established word representation overcome the salience of well-established phoneme or syllable representations? How does unitization of unfamiliar lists of familiar items even get started? If the temporal chunking problem is not solved, then internal representations of lists with more than one item can never be learned.

The cooperative-competitive design of  $A_4$  that solves the temporal chunking problem is called a *masking field*. One property of this design is that longer lists, up to some maximal length, can selectively activate populations in  $A_4$  that have a prewired competitive advantage over shorter sublists in the struggle for STM storage. Simple growth rules are sufficient to achieve this competitive STM advantage of longer sublists. Such a competitive advantage enables a masking field to exploit the fact that longer sublists, other things being equal, are better predictors of subsequent events than are shorter sublists, because they embody a

more unique temporal context. A masking field's preferential STM response to longer sublists leads in turn to preferential LTM chunking, or representation, of longer sublists using the LTM law given by Equation 1. As an important side benefit, the competitive advantage of longer unfamiliar sublists enables them to compete effectively for STM activity with shorter familiar sublists, thereby providing a solution to the temporal chunking problem. The postulate that longer sublists, up to some maximum length, have a competitive STM advantage led to the prediction of a word length effect by Grossberg (1978a, Section 41). A word length effect was reported in the word superiority experiments of Samuel, van Santen, and Johnston (1982, 1983), with longer words producing greater word superiority. Grossberg (1984, 1985) analyzed these and related data from the perspective of masking field design. Computer simulations that illustrate how a masking field can group temporal order information over item representations into sublist chunks are described by Cohen and Grossberg (in press).

The word length property is only one of several differences between properties of stages  $A_3$  and  $A_4$  and those of the stages in alternative theories. Instead of letting  $A_3$  and  $A_4$  represent letters and words, as in the McClelland and Rumelhart (1981) theory,  $A_3$  and  $A_4$  represent items (more precisely, temporal order and item information in STM) and lists (more precisely, sublist parsings in STM), respectively. These properties do not require that individual nodes exist for all items and lists. Learning enables distributed item and list representations to be formed over a network substrate whose rules do not change through time. All familiar letters possess both item and list representations, not just letters such as *A* and *I* that are also words. This property helps to explain Wheeler's (1970) data showing that letters such as *A* and *I* are not recognized more easily than letters such as *E* and *F*. In contrast, the McClelland and Rumelhart (1981) model postulates a letter level and a word level instead of an item level and a list level. By their formulation, letters such as *A* and *I* must be represented on both the letter level and the word level, whereas letters such as *E* and *F* are represented only on the letter level. This choice of levels leads to both conceptual and data-related difficulties with the McClelland and Rumelhart (1981) model, including a difficulty in explaining the Wheeler (1970) data without being forced into further paradoxes (Grossberg, 1984). More generally, any model whose nodes represent letters and words, and only these units, faces the problem of describing what the model nodes represented before a particular letter or word entered the subject's lexicon or what happens to these nodes when such a verbal unit is forgotten. This issue hints at more serious problems concerning such a model's inability to self-organize. Such concerns are dealt with by a theory whose levels can learn to encode abstract item and list representations on a substrate of previously uncommitted nodes.

These abstract item and list processing units of  $A_3$  and  $A_4$  play an important role in the theory's explanation of how unfamiliar and unitized words are recalled. For example, suppose that a list has just been represented in STM across the item representations of  $A_3$ . Before the items in the list can be rehearsed, the entire list begins to tune the  $A_3 \rightarrow A_4$  adaptive filter. The LTM traces within this adaptive filter learn to encode temporal order information in LTM. After this learning has occurred, the tuned filter activates unitized sublist representations across  $A_4$ . These sublist

representations contribute to the recognition of words but cannot by themselves elicit recall. This raises the issue of how short novel lists of familiar items can be recalled even before they are unitized. The fact that a verbal unit can have both an item representation and a list representation now becomes crucial.

Recall of a short novel list of familiar items is triggered by a nonspecific rehearsal wave to  $A_3$ . This wave opens an output gate that enables output signals of active items to be emitted from  $A_3$  to  $M_3$ . As each item is read out, it activates a negative feedback loop to itself that selectively inhibits its item representation, thereby enabling the next item representation to be read out. Each item representation is recalled via the learned  $A_3 \rightarrow M_3 \rightarrow M_2 \rightarrow M_1$  sensory-motor map.

The recall is an immediate recall from STM, or working memory, of a list of unitized item representations. It is a type of "controlled" process. It is not "automatic" recall out of LTM. For a unitized list chunk in  $A_4$  to learn how to read out its list of motor commands from LTM, the chunk must remain active long enough during the learning process to sample pathways to all of these motor commands. We next briefly sketch the simplest version of how learning and recall can occur in the correct order. It should be realized, however, that mechanisms that sometimes control recall in the correct order can also generate recall in an incorrect order. In fact, these mechanisms provide an explanation of the bowed and skewed serial position curve of serial verbal learning as well as related breakdowns of temporal order information in LTM. We do not consider why these STM and LTM temporal order mechanisms cannot always encode veridical STM and LTM order information. See Grossberg (1982a, 1985) for recent discussions of this issue.

In the simplest example of how temporal order information across item representations is encoded and read out of LTM, the top-down template from  $A_4$  to  $A_3$  learns this information while the adaptive filter from  $A_3$  to  $A_4$  is being tuned. Thus the learning of temporal order information is part of the learning of an adaptive resonance. Later activation of a list chunk in  $A_4$  can read this LTM temporal order information into an STM pattern of order information across the item representations of  $A_3$ . Activation of the rehearsal wave at this time enables the list to be read out of STM. In sum, recall can occur via the learned  $A_4 \rightarrow A_3 \rightarrow M_3 \rightarrow M_2 \rightarrow M_1$  sensory-motor map.

All the stages  $A_1$ ,  $A_2$ ,  $A_3$ , and  $A_4$  are sensitive to phonetic information to different degrees. The next stage  $A_5$  can group the list representations of  $A_4$  into list representations that exceed the maximal list length that can be represented within  $A_4$  due to the finite STM capacity of  $A_3$ . In other words, the list representations of  $A_4$  spontaneously parse a list into the most predictive sublist grouping that  $A_4$ 's prior experience permits, and  $A_5$  groups together the parsed components via associative mechanisms. Associative bonds also exist among the chunks of stage  $A_5$ . The learned groupings from  $A_5$  to  $A_4$  can bind together multisyllable words as well as supraword verbal units. The learned interactions within  $A_5$  tend to associate verbal units that are highly correlated in the language. Because the verbal units that are capable of activating  $A_5$  are already of a rather high order,  $A_5$ 's associations encode semantic information in addition to other highly correlated properties of these verbal units.

The visual stage  $V^*$  is not broken down in the present analysis because its several processing stages, such as boundary formation,

featural filling in, binocular matching, and object recognition (Carpenter & Grossberg, 1985; Cohen & Grossberg, 1984a, 1984b; Grossberg & Mingolla, 1985, in press), go beyond the scope of this article. The stages within  $V^*$  that are used for visual object recognition (Carpenter & Grossberg, 1985, 1986) as distinct from visual form perception (Grossberg & Mingolla, 1985, in press) also use bottom-up adaptive filters and top-down learned expectancies. This is so because the problem of stabilizing a self-organizing code in a complex input environment imposes similar general design constraints on all sensory modalities (Grossberg, 1980). Not all of these visual processing stages input to the language system. We assume that associative maps from the object recognition stages in  $V^*$  to  $A_4$  or  $A_5$  can lead to phonetic and semantic recognition as well as to motor recall of a visually presented letter or word via the sensory-motor paths previously described. Associative maps from  $A_4$  or  $A_5$  to  $V^*$  can in turn match the correct visual template of a word, such as NURSE, against a phonetically similar target nonword, such as NERSE.

### 1. The Schvaneveldt-McDonald Lexical Decision Experiments: Template Feedback and List-Item Error Trade-Off

Adaptive resonance theory predicts ordinal relations between accuracy and reaction time. To illustrate these properties, we use the theory to analyze the lexical decision experiments of Schvaneveldt and McDonald (1981), which included both reaction time and tachistoscopic conditions. These experiments used three types of primes (semantically related, neutral, and semantically unrelated) and two types of targets (normal words and altered words). The neutral primes were used to establish a baseline against which effects of related primes and unrelated primes could be evaluated. The altered words were formed by replacing one interior letter of each word with a different letter to form a nonword (e.g., TIGAR from TIGER). Assignment of nonwords to the related or unrelated prime condition was based on the relation between the prime and the word from which the nonword was constructed. Subjects responded "word" or "nonword" manually by pressing a left-hand key or a right-hand key. Schvaneveldt and McDonald's (1981) procedure was as follows:

Each trial consisted of two events in the reaction time paradigm and three events in the tachistoscopic paradigm. In either case the first event was always the priming signal, which consisted of a string of x's or a valid English word. If the prime was neutral (x's), it was the same length . . . as the related word prime for the target on that trial. The prime remained on for 750 msec and was followed by a blank interval of 500 msec. No response to the prime was required, and subjects were only told that the first event was to prepare them to respond to the target. The second event, or target, appeared in the same location on the screen and was either an English word or an altered word, as defined by the task requirements.

In the reaction time experiments, the target remained visible until the subject responded. The instructions were that the subject was to respond as rapidly and accurately as possible.

In the tachistoscopic experiments, the target was displayed for approximately 33.3 msec and was followed by a masking pattern consisting of a string of number signs (#). Subjects were instructed to make as few errors as possible, and speed was not encouraged. The interval between target and mask, or interstimulus interval (ISI), was adjusted at the end of each

block of trials in order to maintain an error rate of approximately .250. (p. 678)

The results of these manipulations are summarized in Figure 6. A systematic explanation of these data follows, comparing every pair of data points within each graph. Because of the qualitative nature of the explanation, only the relative values of compared data points, not their absolute values, can be derived from such an analysis.

The main point of this analysis is to compare and contrast the interactions between a visual list level and an auditory list level. The visual list level exists in the visual subsystem  $V^*$ , which projects to the list levels  $A_4$  and/or  $A_5$  in the auditory system. Reciprocal connections from  $\{A_4, A_5\}$  to  $V^*$  are also assumed to exist and act as feedback templates. Under conditions of auditory presentation, an analogous analysis could be given of reciprocal interactions between  $A_3$  and  $A_4$ .

A critical issue in our analysis concerns the criteria used by subjects to select a response. Our microtheory implies that an individual letter or word can be completely identified within its modality when the corresponding resonance equilibrates (see Section 7). Many conditions of lexical decision experiments do not enable a fully blown resonance to evolve. Thus subjects are forced to use incomplete processing measures. Within our theory, the size and speed of an initial burst of activation at the appropriate list level correlates well with subject performance. The characteristics of such a burst depend on factors such as whether a mask is imposed and whether priming events, among other factors, lead to a match or mismatch situation. Our explanation therefore emphasizes the context-dependent nature of subject response criteria. An analysis of these transient dynamical events provides new insights into speed-accuracy trade-off (Pachella, 1974; Pew, 1969) and statistical decisionlike performance under uncertainty (Green & Swets, 1966).

We first consider the tachistoscopic condition (see Figure 6a), whose analysis is simpler than that of the reaction time condition. The lower curve in the tachistoscopic condition describes the error rate when word targets (W) followed related (R), neutral (N), or unrelated (U) primes. Unrelated primes caused more errors than did neutral primes, whereas related primes caused fewer errors than did neutral primes. The upper curve describes the error rate when nonword targets (Nw) followed R, N, or U primes. A nonword target is constructed from a word target that is in the designated relation R, N, or U to a prime by changing one of the word target's interior letters. With nonword targets, the reverse effect occurred: Related primes caused more errors than did neutral primes, and unrelated primes caused fewer errors than did neutral primes.

These curves are consistent with two hypotheses:

1. The bottom-up pathways from  $V^*$  to  $\{A_4, A_5\}$  are capable of activating the auditory list representations, but the action of the visual mask 33.3 ms after the onset of the target obliterates target-induced item activation and prevents top-down template signals  $\{A_4, A_5\} \rightarrow V^*$  from causing resonant sharpening and equilibration. Thus, in this case transient activations are the only dynamic events on which the model can base a decision.
2. On the average, R, N, and U primes cause equal amounts of interference in the bottom-up registration of word and nonword targets at the item level of  $V^*$ . This hypothesis is compatible

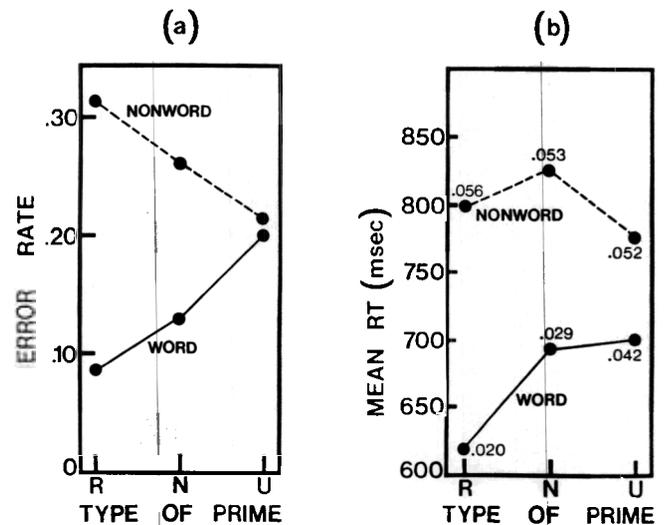


Figure 6. Results from the Schvaneveldt and McDonald (1981) lexical decision experiments: (a) A tachistoscopic experiment with a backward pattern mask. Error rates in response to word (W) and nonword (Nw) targets that follow related word (R), neutral (N), or unrelated word (U) primes. (b) A reaction-time experiment without a backward pattern mask. Reaction times in response to W and Nw targets that follow R, N, or U primes. (From "Semantic Context and the Encoding of Words: Evidence for Two Modes of Stimulus Analysis" by R. W. Schvaneveldt and J. E. McDonald, 1981, *Journal of Experimental Psychology: Human Perception and Performance*, pp. 680, 683. Copyright 1981 by the American Psychological Association. Reprinted by permission.)

with the R, N, and U categories' being defined by semantic relatedness to the targets and not by similarity of visual features.

Independent experimental evidence for Hypothesis 2 was provided by the gap detection experiment of Schvaneveldt and McDonald (1981). This experiment differed from the lexical decision experiments only in that the altered-word foils were constructed by introducing a gap in the letter that had been replaced in the lexical decision study. The task was to indicate whether or not a gap had been present. Detecting these gaps should not have required semantic information. In the tachistoscopic condition of this experiment, word targets (without gaps) were recognized with equal error rates after R, N, or U word primes. This result would not be expected if the different prime categories had caused unequal amounts of interference with the visual registration of the targets. The following discussion compares priming effects first on words, then on nonwords. After that, we consider word-nonword comparisons to describe effects of response bias.

#### Word Target Comparisons—Tachistoscopic

Each tachistoscopic condition is identified by prime type followed by target type: For example, R/Nw denotes an R prime followed by an Nw target. The following discussion suggests why nonword responses to word targets increase in the order from R to N to U primes.

*R/W-N/W.* According to Hypothesis 2, the target word receives approximately equal interference on the item level from the prior R and N primes because relatedness is defined seman-

tically, not in terms of shared item features. By contrast, the R prime strongly activates list nodes of the word target because of their semantic relatedness. When the word target occurs after an R prime, its input to its list representation augments the prior activation that has lingered since the R prime presentation. The N prime does not significantly activate the list representations of target words. Hence the R prime facilitates recognition more than does the N prime.

*N/W-U/W.* The N prime and the R prime cause equal amounts of interference on the item level. The N prime does not significantly activate the list representations of any target words. The U prime does not strongly activate list nodes of the word target. In fact, the U prime can activate list nodes that competitively inhibit the list nodes of the word target, because of the recurrent lateral inhibition between list representations that exists within a masking field (see Section 10). Hence a target word is recognized more easily after an N prime than after a U prime.

### *Nonword Target Comparisons—Tachistoscopic*

The following discussion suggests why word responses to nonword targets decrease in the order from R to N to U primes.

*R/Nw-N/Nw.* According to Hypothesis 2, the target word receives approximately equal interference on the item level from the prior R and N primes. The N prime does not significantly activate the list level. The R prime does. Moreover, the R prime significantly activates the same list representation that the subsequent target word activates. Thus the nonword target is more often misidentified as a word after an R prime than after an N prime.

*N/Nw-U/Nw.* The N prime does not significantly activate the list level. The U prime significantly activates word representations on the list level that inhibit the word representation that is activated by the nonword target. The nonword target can significantly activate its word representation on the list level. Template feedback cannot act to correct this misidentification. Inhibition from the U prime can. Thus a nonword target is misidentified as a word more frequently after an N prime than after a U prime.

We now compare the transient bursts caused by word and nonword targets after the same type of prime. We show that both bursts are similar if the mask acts sufficiently quickly after target onset. The sizes of these bursts increase across the prime conditions from R to N to U. Thus the different error rates in response to word and nonword targets can be ascribed to response biases. For example, a subject who demands a fully blown resonance in order to respond "word" will be biased to respond "nonword" in all conditions.

*R/W-R/Nw.* Both word and nonword targets receive equal amounts of interference on the item level because of the R prime. Both word and nonword targets similarly activate the list representation of the word. This list representation received significant activation from the prior R prime. The target mask prevents template feedback from correcting activation of a list representation of a word by a nonword. Consequently, both word and nonword list representations generate similar activation bursts, both of which are amplified by the prior R prime.

*U/W-U/Nw.* According to Hypothesis 2, both the word target

and the nonword target receive equal amounts of interference from the prior U word prime activating their item representations within  $V^*$ . Because of the brief presentation of word and nonword targets, both targets activate similar list nodes in  $A_4$  and  $A_5$ . In particular, letters that are interior to words or word syllables are less important in the adaptive filtering of items into lists than are letters at the ends of words or syllables (Grossberg, 1978a, 1985). This property is due to the primacy and recency gradients of temporal order information that form in the pattern of STM activation across active item representations. The rapid onset of the target mask prevents the template feedback from lists to items from correcting the misclassification of a nonword target as a word. Consequently, both word and nonword list representations generate similar activation bursts, both of which are attenuated by the prior U prime.

Schvaneveldt and McDonald (1981) obtained a different pattern of results with the same design when no mask was presented and reaction times were recorded. We now trace the differences between the data of the reaction time condition (Figure 6b) and of the tachistoscopic condition (Figure 6a) to the action of the feedback template from list to item representations. To emphasize these differences, we compare pairs of data points in the reaction time condition with the corresponding pairs of data points in the tachistoscopic condition. Several of our explanations depend on a trade-off that exists in these networks between an initial tendency to misidentify a word at the list level and the ability of this initial tendency to generate an error-correcting mismatch at the item level before false recognition can occur. We call this trade-off the *list-item error trade-off*.

Two changes in the reaction time data are of particular interest. First, there was no increase in reaction time on word trials due to a U prime relative to an N prime, although a U prime increased error rate relative to an N prime in the tachistoscopic experiment. Second, an R prime decreased reaction time relative to an N prime whereas an R prime increased error rate relative to an N prime in the tachistoscopic experiment. We will analyze the difference between the two types of experiments by making pairwise comparisons of the data points.

### *Nonword Comparisons—Reaction Time*

Each reaction time condition is again identified by prime type followed by target type such that unmasked targets are indicated by an apostrophe: For example,  $N/W'$  denotes an N prime followed by an unmasked W target.

*R/Nw'-U/Nw' versus R/Nw-U/Nw.* These comparisons analyze the large difference between the error rates for  $R/Nw$  and  $U/Nw$  in the tachistoscopic condition and the insignificant difference between both the error rates and the reaction times of  $R/Nw'$  and  $U/Nw'$  in the reaction time condition.

Compare  $R/Nw$  with  $R/Nw'$ . In  $R/Nw$ , both the R prime and the nonword target activate the list representation of the word from which the nonword was derived. Hence the nonword target generates a relatively large number of word misidentifications.

The processing in  $R/Nw'$  starts out just as it does in  $R/Nw$ . In  $R/Nw'$ , however, the conjoint activation of the word's list representation by both the R prime and the nonword target generates large template feedback from the word list representation to the

nonword item representation. Thus the very factor that causes many false word identifications of the nonword target in R/Nw leads to a relatively strong feedback signal with which to disconfirm this misidentification in R/Nw'. The mismatch between the nonword's item representation and the word's item representation causes a significant collapse in the item-to-list signals that were supporting the word's list representation. Thus the number of word misidentifications of the nonword target in R/Nw' is reduced relative to R/Nw.

Moreover, the fact that the word representation is still active because of the R prime when the nonword target is presented speeds up as well as amplifies the read-out of the word template, thereby causing a speed-up in reaction time.

Compare R/Nw' with U/Nw'. In contrast with R/Nw', in U/Nw' the U prime inhibits the list representation of the word that the nonword target activates. Thus the net activation of the word's list representation by a nonword target is weaker in U/Nw' than it is in R/Nw'. Consequently, the word template that is read out by the list representation in U/Nw' is weaker than that read out in R/Nw'. The U/Nw' template is therefore less effective than the R/Nw' template in mismatching the nonword item representation.

Thus a trade-off (list-item error trade-off) exists between the initial tendency toward word misidentification at the list level and the ability of this initial tendency to trigger template feedback that can correct this tendency before erroneous recognition can occur. The two factors—degree of incorrect initial list activation and degree of item mismatch—tend to cancel each other out.

This trade-off holds for both activity levels and rates of activation. A large prior activation of the word's list representation by the R prime helps to cause a rapid read-out of the word template. This rapid reaction elicits a strong item mismatch capable of undercutting the already large initial activation of the word's list representation. The greater speed is hereby compensated for by the larger activation that must be undercut. Thus in both error rate and reaction time, R/Nw' and U/Nw' are similar despite the large difference between R/Nw and U/Nw.

*N/Nw-U/Nw' versus N/Nw-U/Nw.* The main points of interest concern why in the reaction time paradigm U/Nw' is reliably faster than is N/Nw' even though their error rates are comparable, whereas in the tachistoscopic paradigm, the error rate in U/Nw is reliably less than that of N/Nw. Our task is to show how the feedback template in the N/Nw'-U/Nw' comparison alters the N/Nw-U/Nw dynamics of the tachistoscopic case. Once again, a list-item error trade-off between the amount and timing of list representation activation and its effects on the amount and timing of item mismatch forms the core of our analysis. Our explanation, moreover, differs from the hypothesis that Schvaneveldt and McDonald (1981) derived from these data.

The N prime does not activate the list representations nearly as much as a word prime does. Hence a nonword target can modestly activate its word representation without major interference or augmentation from the prior N prime. The word's list representation then reads out a template whose ability to mismatch the nonword's item representation depends on how strongly the nonword was able to activate the word's list representation. Thus the size of the initial tendency to misidentify the nonword target covaries with the ability of the feedback template to correct this error.

This balance between initial list activation and subsequent item mismatch also occurs in the U prime condition. In this condition, however, the U prime inhibits the word list representation that the nonword target activates. This list representation is thus activated by the nonword less after a U prime than after an N prime. The weaker tendency to misidentify the nonword as a word after a U prime leads to a weaker template read-out and a weaker item mismatch with the nonword's item representation. The similar error rates in the N and U nonword conditions can thus be traced to the list-item error trade-off.

Why then is the reaction time in the U prime condition reliably faster ( $36 \pm 12$  ms,  $p < .01$ ) than in the N prime condition? We suggest that a major factor is the following one. In the U prime condition, the nonword target causes a relatively small initial activation of the word level because of the prior occurrence of the U prime. This relatively small initial activation tends to cause a relatively weak item mismatch that can only cause the initial list activation to become even smaller. We suggest that the absence of a large rate or amount of activation within the list level at any time provides a relatively rapid cue that a nonword has occurred.

By contrast, within the N prime condition, the nonword target can cause a relatively large initial activation of the list level. Although the large template read-out that is caused by this activation can compensate for it, via list-item error trade-off, more time is required to inhibit this initial activity surge than is required to register the absence of such a surge in the U prime condition. Thus the reaction time tends to be longer in the N prime condition than in the U prime condition. This explanation assumes that subjects in the reaction time condition tend to respond to the network's equilibrated activities rather than to momentary activity surges. Subjects who do respond to initial surges might respond faster and be more prone to make word misidentifications.

Schvaneveldt and McDonald (1981) discuss these data in terms of the hypothesis that "the priming event can lead to general activation of linguistic information-processing mechanisms" (p. 681) in response to an R prime or U prime but not an N prime. This hypothesis is used to explain why the reaction times to nonword targets after R or U primes are faster than those after an N prime. This explanation does not seem to suffice to explain these data, at least not using a serial search model of verification. This is because the N prime should not generate any semantic set. Even if R and U primes can speed up search through their semantic sets, such a search would presumably take longer than no search at all.

Our questioning how a nonspecific activating mechanism could modulate serial search does not mean that we deny the existence of nonspecific activating mechanisms. Within the context of adaptive resonance theory, a level  $F_i$  can also nonspecifically activate a level  $F_{i+1}$  in addition to specifically activating  $F_{i+1}$ . This nonspecific activation lowers the gain of  $F_{i+1}$  and thereby enables the  $F_i \rightarrow F_{i+1}$  signals to supraliminally activate the STM traces of  $F_{i+1}$ . As summarized in Section 9, the need for such nonspecific gain control can be seen by considering how a higher stage  $F_{i+2}$  reads top-down templates into  $F_{i+1}$ . The theory suggests that  $F_{i+2}$  can actively read such template signals into  $F_{i+1}$  without supraliminally activating  $F_{i+1}$ . These templates subliminally prepare  $F_{i+1}$  for supraliminal bottom-up activation from  $F_i$ . When  $F_i$  does send bottom-up signals to  $F_{i+1}$ , it "opens

the  $F_{i+1}$  gate"—that is, nonspecifically alters the gain—to enable the bottom-up signals from  $F_i$  and the top-down template from  $F_{i+2}$  to begin to supraliminally match or mismatch, as the case might be.

A role for bottom-up nonspecific gain control within our theory would require that R and U primes are more vigorously processed by  $V^*$  than are N primes and hence can elicit larger nonspecific signals to  $\{A_4, A_5\}$ . This property does not occur in the outputs of the  $V^*$  item level, because all the R, N, and U primes are constructed from equal numbers of letters. The property can, however, occur in the outputs of the  $V^*$  list level, because a familiar word prime of a fixed length can generate greater activation within a masking field than can an unfamiliar prime with the same number of letters (Section 10). Thus, the Schvaneveldt and McDonald (1981) hypothesis can be restated as the suggestion that  $V^*$  contains a list representational stage that obeys the laws of a masking field. Then the present analysis still holds if lists in  $V^*$  project to lists in  $A_4$  and/or  $A_5$ .

*R/Nw'-N/Nw' versus R/Nw-N/Nw*. This comparison is already implicit in the  $R/Nw'-U/Nw'$  versus  $R/Nw-U/Nw$  and  $N/Nw'-U/Nw'$  versus  $N/Nw-U/Nw$  comparisons. It is included to emphasize the importance of the list-item error trade-off.

In the tachistoscopic experiments,  $R/Nw$  has a higher error rate than condition  $N/Nw$  due to significant conjoint activation of word list representations by the R prime and the subsequent nonword target. In  $R/Nw'$ , by contrast, the large initial activation of the nonword target due to the prior R prime elicits a faster and stronger template read-out and item mismatch than in the N prime condition. The faster acting template in the R prime case than in the N prime case leads to a faster reaction time in  $R/Nw'$  than in  $N/Nw'$ . The stronger template read-out and item mismatch in the R prime case than in the N prime case compensates for the larger initial activation via the list-item error trade-off, thereby reducing the relative error rates of  $R/Nw'$  to  $N/Nw'$  compared with the error rates of  $R/Nw$  to  $N/Nw$  by causing a significant collapse in the incorrect initial activation of word list representations.

Our theory overcomes the objection we made to a serial search version of verification in the following way. At the moment when an  $Nw'$  target occurs, a word representation is already significantly active in case  $R/Nw'$  but not case  $N/Nw'$ . Even if the word representation were capable of generating suprathreshold top-down signals to the item level at this time, these signals would not elicit suprathreshold item activation until the target occurred (see Figure 4a). In any case, the  $Nw'$  target causes the list representation of the corresponding word to exceed its output threshold faster when it follows an R prime than when it follows an N prime, because the N prime does not significantly activate this list representation. Thus list-item error trade-off begins to act sooner and more vigorously in the  $R/Nw'$  case than in the  $N/Nw'$  case.

### Word Comparisons—Reaction Time

*N/w'-U/w' versus N/W-U/W versus N/Nw'-U/Nw'*. The main points of interest are that in the reaction time experiments, the reaction times to word targets after N and U primes were not significantly different (the U prime condition was slightly slower), although the error rate in the U prime condition was

larger (.042) than in the N prime condition (.029), although this trend was not significant. A similar pattern of error rates was found in the tachistoscopic experiments ( $N/W-U/W$ ). By contrast, the reaction time to a nonword target after a U prime was significantly less than the reaction time to a nonword target after an N prime ( $N/Nw'-U/Nw'$ ). Why does letting a nonword-activated template act preserve relative error rates while producing approximately equal reaction times ( $N/W'-U/W'$  versus  $N/W-U/W$ ), whereas letting a word-activated template act does not produce approximately equal reaction times?

When a word target occurs after an N prime, it can activate its representation on the list level without experiencing inhibition or facilitation due to the prior N prime. In the reaction time experiments, the word target can also cause read-out of a template capable of matching its item representation. By contrast, when a word target occurs after a U prime, its list representation is still experiencing residual inhibition because of the prior U prime. However, the word target remains on long enough for its list representation to be activated and to read out a template capable of matching its item representation. Matching the item representation can in turn further amplify input to the list representation. In all, despite a slower initial start in being activated, if the word target stays on long enough, it generates equal levels of equilibrated list activation after both an N prime and a U prime. The initial difference in list activation levels, which is so important in the tachistoscopic experiments ( $N/W-U/W$ ), becomes less so in the reaction time experiments ( $N/W'-U/W'$ ) because of the similar course of list-item equilibration after the initial activation difference is overcome. Nonetheless, a slightly longer reaction time can be caused by the U prime's inhibitory effect on the initial course of processing.

This explanation makes important use of the different effects of template-mediated item matching or mismatching on subsequent list activation. An item match due to a word target tends to strengthen the word list representation. An item mismatch due to a nonword target tends to weaken the word list representation according to list-item error trade-off. These different consequences of word targets and nonword targets suffice to explain the reaction time differences between  $N/W'-U/W'$  and  $N/Nw'-U/Nw'$ .

It remains to say why the error rates are not the same in  $N/W'-U/W'$ , whereas the reaction times are approximately the same. The upper limit on reaction time differences is set by the equal equilibration of list activation due to word targets that follow N primes and U primes. Even if subjects respond with a statistical distribution of reaction times that is concentrated throughout the time interval until equilibration is finalized, the difference in mean reaction times should not significantly exceed the brief interval needed to offset U prime inhibition. Nonetheless, if some subjects do respond at times before equilibration occurs and use the level of list activation to determine word or nonword responses, then the initial U prime inhibition can cause a significant increase in nonword responses to the word target. This tendency should correlate with shorter reaction times.

*R/W'-N/W' versus R/W-N/W*. The  $R/W$  error rate is less than the  $N/W$  error rate because of the priming of the correct word list representation after an R prime but not after an N prime. The same is true for the relative sizes of the  $R/W'$  and  $N/W'$  error rates, although this trend was not significant. In

R/W' and N/W', a word target that follows an R prime can more quickly and strongly read out its template than can a word target that follows an N prime. These templates tend to match and amplify the word item representation. Thus R/W' possesses a faster reaction time and a lower error rate than does condition N/W'.

In summary, these lexical decision data can be qualitatively explained using the following properties:

1. R, N, and U primes all generate comparable levels of interference to later targets at the item level.

2. R primes subliminally activate semantically related word list representations via recurrent conditioned excitatory pathways within the list levels. U primes inhibit semantically unrelated word list representations via recurrent unconditioned inhibitory pathways within the list levels. N primes neither activate nor inhibit word list representations by a significant amount.

These priming properties assume that the list representations of words with more than one letter can inhibit the list representations of their constituent letters, including the list representations of letters that form the N prime. This property, which is needed to learn selective word list representations, is achieved by designing the list levels as masking fields (see Section 10).

3. A larger activation of a word list representation due to a word target causes faster and stronger item matching and activity amplification at the item and list levels.

4. A larger activation of a word list representation due to a nonword target causes faster and stronger item mismatch and activity suppression at the item and list levels. This compensatory property is called the list-item error trade-off.

This property assumes that mismatch of a single letter in a word at the item level can cause a significant collapse in the activation of the word's list representation. Thus for our explanation to hold, activation of a list level representation must be selectively sensitive to word length in a manner that is consistent with the properties of a masking field (Cohen & Grossberg, in press).

## 12. Word Frequency Effects in Recognition and Recall

Recognition events that occur during lexical decision experiments are often analyzed as a world unto themselves. Relations with other sorts of word recognition and recall phenomena are often neither noted nor used to provide additional constraints upon understanding of word recognition phenomena. The remaining sections of this article relate lexical decision data to another sort of word recognition and to recall.

A number of experiments have demonstrated that word frequency manipulations can have different effects on recall than on recognition of prior occurrence. A unified explanation of these effects is suggested by our theory and leads to interesting comparisons with previous explanations. Our explanation invokes unitization and interlist associative reactions in a basic way. Thus whereas our explanation of lexical decision data did not require an analysis of LTM changes, our analysis of memory of previous occurrence does. Some of the main experimental phenomena are now summarized.

In lexical decision experiments, different effects of word frequency occur with and without the use of a backward pattern mask. Under conditions of backward masking, word frequency

typically does not have a significant effect on accuracy of word recognition, although insignificant improvements in recognition have been noted as a function of increasing frequency (Manelis, 1977; Paap & Newsome, 1980). By contrast, if a backward mask is not used, then high-frequency words are consistently classified faster than low-frequency words (Landauer & Freedman, 1968; Rubenstein, Garfield, & Millikan, 1970; Scarborough, Cortese, & Scarborough, 1977). This difference has been used to support the verification model hypothesis that word frequency does not influence word encoding but does influence the later stage of word verification.

A paradoxical pattern of data emerges when influences of word frequency on recognition and recall are contrasted. This data pattern, which is often called the *word frequency effect*, states that high-frequency words are recalled better than low-frequency words, but low-frequency words are recognized better than high-frequency words (Glanzer & Bowles, 1976; Gorman, 1961; Schulman, 1967; Underwood & Freund, 1970). To explain this effect, it is necessary to carefully define the relevant experimental procedures.

Underwood and Freund (1970) used a two-alternative forced-choice recognition procedure. In Stage 1 of their experiment, subjects studied a list of 50 low-frequency (L) words or a list of 50 high-frequency (H) words. In Stage 2, subjects were shown pairs of words. One word in each pair was chosen from the list of study words. The other word in each pair was chosen from a list of either H words or from a list of L words. Thus subjects fell into one of four categories—L-L, L-H, H-L, H-H—in which the first letter designated whether the study list was composed of L or H words and the second letter designated whether the distractor word in each pair was chosen from an L or H list. The subjects were instructed to identify the word in each pair that they had seen on the study trial.

The results of Underwood and Freund (1970) are summarized in Figure 7. The main word frequency effect compares H-H with L-L; that is, when studied L words were paired with unstudied L words, recognition was better than when studied H words were paired with unstudied H words. This effect reversed in the H-L and L-H conditions. Studied H words in the H-L condition were recognized with fewer errors than studied L words in either the L-H or L-L conditions. To understand these results, one needs to consider the effect of word frequency on the study trial as well as the effect of word frequency differences on the test trial. Underwood and Freund (1970) offered an interesting explanation of their results. Our explanation is compared and contrasted with theirs below.

This experiment raises fundamental questions about the processes that lead to judgments of recognition. Unlike a judgment between word and nonword, all items in the Underwood and Freund (1970) experiment are words. The task is to judge which of these words have recently been seen. This type of recognition has been called different names by different authors. Concepts such as "the judgment of previous occurrence" (Mandler, 1980; Mandler, Pearlstone, & Koopmans, 1969), "familiarity" (Juola, Fischler, Wood, & Atkinson, 1971; Kintsch, 1967; Mandler, 1980), "situational frequency" (Underwood & Freund, 1970), and "encoding specificity" (Tulving, 1976; Tulving & Thomson, 1973) have been used to distinguish this type of recognition from other types, such as the word-nonword recognition of lexical

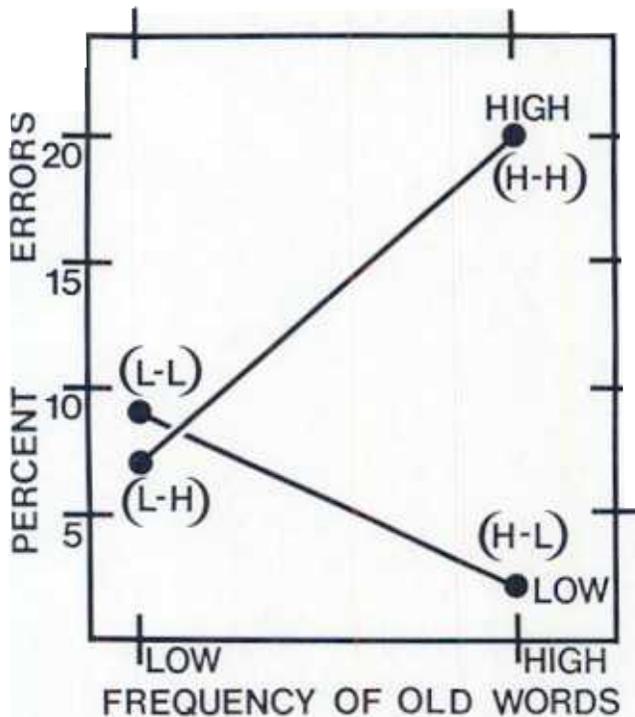


Figure 7. Results from the Underwood and Freund (1970) experiment concerning word frequency influences on the recognition of previous occurrence: Forced-choice recognition of (old-new) word pairs leads to more errors for old high-frequency words paired with new high-frequency words (H-H) than for old low-frequency words paired with new low-frequency words (L-L). In contrast, more errors occur for old low-frequency words paired with new high-frequency words (L-H) than for the converse (H-L). (From "Word Frequency and Short Term Recognition Memory" by B. J. Underwood and J. S. Freund, 1970, *American Journal of Psychology*, 83, p. 347. Copyright 1970 by the American Journal of Psychology. Reprinted by permission.)

decision tasks, or recognition of the individual items in a list. The Underwood and Freund (1970) experiment underscored the difficulty of making such a concept precise by showing that recent presentation can interact in a complex fashion with word frequency. In other words, these data show that subjects may confuse the internal recognition indices that are due to recent presentation with internal recognition indices that are due to the cumulative effects of many past presentations. This is not surprising when one considers that it is the long-term cumulative effects of many "recent presentations" that yield the internal representations that subserve word frequency effects. How one should proceed from such general observations is, however, far from obvious, as we shall see by reviewing two of the leading models of this type of recognition.

### 13. Analysis of the Underwood and Freund Theory

Underwood and Freund's (1970) explanation of the data in Figure 7 is a prototype of later explanations of the word frequency effect. A measure of "situational frequency" is assigned to each item. A study trial is assumed to increase the situational frequency of a studied word from 0 to 1. A second study trial is

assumed to increase the situational frequency to 2. However, discriminability of an item increases as a slower-than-linear function of situational frequency, so that two study trials have little more effect than one study trial on discriminability. It is also assumed that a subject chooses that item in a pair of items with the higher discriminability.

A critical assumption that differentiates the effects of H and L words concerns the role of implicit associational responses (IARs). If a studied item elicits an IAR, then the IAR also acquires a frequency of 1. It becomes an "old" item even if it does not explicitly appear in the study list. It is also assumed that H words have more IARs than do L words and that the IARs of H words tend to be other H words.

Consequently, H-L recognition is best. H words directly receive an increment of 1 due to study. They may also indirectly achieve a greater discriminability than 1 by being the IARs of other studied H words. By contrast, the unstudied L words are unlikely to be the IARs of studied H words, so that the frequency differences between H and L words in H-L pairs is maximal. By contrast, in the H-H situation, many of the unstudied IARs of studied H words may be the new words in the H-H pairs. Furthermore, studied H words that are IARs of other studied H words derive little extra advantage from this fact. Hence many of the H-H pairs tend to have similar discriminability values, so that many errors occur. L-H should produce fewer errors than H-H, because the studied L words acquire a large frequency 1 without increasing the frequency of the unstudied H words. L-L should also produce fewer errors than H-H, for a similar reason.

The other comparisons follow less easily from this analysis. The advantage of H-L over L-L can be attributed only to the slight benefit received by studied H words that are IARs of other studied words. In the data, the error difference between H-L and L-L is at least a third of the difference between H-L and H-H. If studied words derive little benefit from being IARs of other studied H words, then this difference should be small. If, however, studied H words derive a great deal of benefit from being IARs of other studied H words, then H-H should not produce nearly so many errors. However, this difficulty would be less pronounced if L words have more L word IARs than do H words.

A further difficulty concerns the model's implications concerning the real-time events that translate situational frequencies into decisions. The situational frequency changes may be likened to changes in LTM. The theory does not, however, explain how a new H item that was an IAR of a studied H item translates its LTM situational frequency value of 1 into a decision on a test trial. In particular, suppose that the frequency change between a studied H word and an IAR H word were due to a change in the LTM strengths within the conditionable pathways between internal representations of these words. Then these LTM traces would have no effect on the activation of the IAR H word on a test trial unless the studied H word with which it is paired is the word that activated it as an IAR on the study trial. Thus, something more than the strengthening of interword associative linkages is needed to explicate the situational frequency concept. Suppose, for example, that all the activated H words, both studied words and IAR words, form new LTM linkages with internal representations of contextual cues, notably with visual repre-

sentations of the experimental context. Then these contextual cues could activate the list representations of the studied words and the IAR H words on recognition trials. Such a contextual contribution could save the formal properties of the situational frequency concept but still could not explain the relatively large error difference between H-L and L-L. Despite these uncertain and incomplete aspects, the Underwood and Freund (1970) model provided a seminal framework for later models of the word frequency effect.

#### 14. Analysis of the Mandler Theory

Mandler (1980) and Mandler, Goodman, and Wilkes-Gibbs (1982) further developed the theory and the data of these recognition and recall differences. In the Mandler et al. (1982) experiments, a lexical decision task provided the occasion for studying H and L words. The subjects' task, as usual, was to identify words and nonwords. They were not told that they would be asked to remember these stimuli. Half the subjects were then given the same words and asked to define them. The remaining subjects were not. Then all subjects were asked to return in 24 h. At that time, half of the subjects in each group underwent a recognition test in which old and new words were intermixed. More L words were recognized than H words. After the recognition test, these subjects were also given a recall test. Recall was better for H words than for L words. The other half of the subjects were tested for recall before recognition. A prior recognition test was found to increase recall significantly, but it also increased errors due to distractors from the recognition test.

Mandler et al. (1982) also did an analysis aimed at replicating the major results of Underwood and Freund (1970), although they did not use a two-alternative forced-choice paradigm. This analysis was restricted to data from the definition task group, because this task provided a learning experience analogous to that in the Underwood and Freund (1970) study. Mandler et al. (1982) analyzed the hit rates and false alarm rates for the L and H new and old words that were used during the recognition test. They showed that the  $d'$  for the L-L comparison was larger than that for the H-H comparison, whereas the  $d'$  for the H-L comparison was larger than that for the L-H comparison. Also the  $d'$  for the H-L comparison was larger than that for the H-H comparison.

The model of Mandler (1980) introduces refinements and modifications of the Underwood and Freund (1970) model but also new difficulties. Mandler replaces the notion of situational frequency and the slower-than-linear increase of discriminability with frequency by introducing his concept of familiarity. Mandler discusses how an event's familiarity and a retrieval process can work together to determine recognition. He lets " $F$  = the probability that an event will be called *old* on the basis of its familiarity value;  $R$  = the probability that an event will be called *old* as a result of retrieval processes;  $Rg$  = the probability that an event will be called *old*" (1980, p. 257). These probabilities obey the equation

$$Rg = F + R - FR. \quad (5)$$

Both Mandler (1980) and Mandler et al. (1982) argued that retrieval processes are not rate limiting in determining the reversal that occurs during word recognition. Consequently, they devel-

oped the properties of the familiarity concept to explain the word frequency effect.

Familiarity of an event is determined by the integration, perceptual distinctiveness, and *internal structure* of that event . . . and by the amount of attention expended on the event or item *itself* [italics ours]. Retrievability . . . is determined by interevent relationships and the elaboration of the target event in the context of other events or items. (Mandler et al., 1982, p. 33)

Thus intraitem changes in familiarity bear the total burden of explaining the recognition reversal in the Mandler (1980) model, unlike the role of IARs in the Underwood and Freund (1970) model.

The actual implementation of the familiarity concept to explain the word frequency effect faces several difficulties. Despite these difficulties, the intuitions that led to the Mandler (1980) theory are instructive. Hence we describe both the model's intuitive basis and its formal difficulties before suggesting how our theory overcomes these difficulties.

Mandler (1980) assumed that every word has a "base familiarity" before it is presented in a recognition experiment. Typically the base familiarity  $F_0$  of a high-frequency word is larger than the base familiarity value  $f_0$  of a low-frequency word. In a theory that explains judgments of recent occurrence using a familiarity concept, a base familiarity value *must* be defined or one could not even compare the familiarities of old-word targets and new-word distractors. Mandler (1980) and Mandler et al. (1982) acknowledged this need with their discussions of the Underwood and Freund (1970) experiment and the Glanzer and Bowles (1976) experiment. The latter experiment revealed the basic fact that

false alarms demonstrate the dominance of high-frequency words; that is, hit rates are higher for low-frequency words, but false alarm rates are higher for high-frequency words. In other words, in the absence of retrievability the recognition judgement (for distractors) depends on the familiarity of the item. (Mandler, 1980, p. 267)

The crucial step in a theory based on familiarity is to explain how base familiarity is altered when a word is presented, or how a new word becomes an old word. Because

$$F_0 > f_0, \quad (6)$$

one needs an operation that can reverse the effect of word frequency on the base familiarity value. Formally, the question becomes, How do increments  $\Delta F$  and  $\Delta f$  in familiarity alter the base familiarity values  $F_0$  and  $f_0$  to generate new familiarity values  $F_1$  and  $f_1$  such that

$$F_1 < f_1? \quad (7)$$

Mandler (1980) makes the plausible assumption that "the increment in familiarity (integration) for all words is a constant function of the amount of time that the item is presented" (p. 268). In other words,

$$d = \Delta F = \Delta f. \quad (8)$$

Given Assumption 8, no obvious additive model can convert Inequality 6 into Inequality 7. Mandler therefore chooses a ratio model. He suggests that "the operative  $F$  value for a word

be  $d/(d + F)$ , where  $F$  is the preexperimental base familiarity value of the word" (1980, p. 268). In other words,

$$F_1 = \frac{d}{d + F_0}, \quad (9)$$

and

$$f_1 = \frac{d}{d + f_0} \quad (10)$$

Mandler goes on to apply this definition by using Equation 5 to determine when

$$f_1 + r - rf_1 > F_1 + R - RF_1 \quad (11)$$

(1980, p. 268) under the condition that the retrievability ( $r$ ) of low-frequency words and that of high-frequency words ( $R$ ) are equal. He notes that Inequality 6 is the basis for deriving Inequalities 7 and 11.

Despite the ingenuity of this familiarity concept, it is not entirely satisfactory. For example, a comparison of Inequalities 6 and 11 shows that a single study trial, no matter how short, must reverse the Inequality 6 governing base familiarities, which are determined by a large number of prior word exposures under natural conditions. This paradoxical conclusion follows because for any positive  $\Delta f = \Delta F$ , no matter how small,  $F_1 < f_1$  if  $F_0 > f_0$ . Another way to state this problem is as follows. One might expect the operative familiarity value to approach the base familiarity value as the increment in familiarity approaches zero: that is,

$$f_1 \rightarrow f_0 \text{ as } \Delta f \rightarrow 0, \quad (12)$$

and

$$F_1 \rightarrow F_0 \text{ as } \Delta F \rightarrow 0. \quad (13)$$

Instead, the Definitions 9 and 10 imply that

$$f_1 \rightarrow 0 \text{ as } \Delta f \rightarrow 0 \quad (14)$$

and

$$F_1 \rightarrow 0 \text{ as } \Delta F \rightarrow 0. \quad (15)$$

One might wish to salvage the situation by replying that if an item has not been studied in the experiment, it has a zero familiarity value. However one cannot then explain how new distractors generate false alarms on the basis of word frequency in the Underwood and Freund (1970) and Glanzer and Bowles (1976) experiments.

Moreover, the Mandler et al. (1982) analysis of  $d'$  scores in their definition task experiment is inconsistent with the Underwood and Freund (1970) data in an important respect. Mandler et al. (1982) assume that familiarity is an inraitem variable and that forced-choice responses are based on selection of the more familiar word in each pair. Under such assumptions, a  $d'$  analysis requires that the lines between points (L-H, H-H) and points (L-L, H-L) be parallel. This is not true in the Underwood and Freund (1970) data (Figure 7). Thus, although the Mandler (1980) model escapes its worst difficulties when it does not combine old words with new words, it fails to be able to make such a comparison both formally and in important data. These difficulties suggest that the Mandler (1980) familiarity concept is insufficient to capture major properties of subjects' ability to judge previous occurrences.

One can escape the limit problem expressed in Properties 12 and 13 by redefining  $F_1$  and  $f_1$  as follows. Let

$$F_1 = F_0 + \frac{\lambda d}{d + F_0}, \quad (16)$$

and

$$f_1 = f_0 + \frac{\lambda d}{d + f_0}, \quad (17)$$

where  $\lambda$  is a positive constant. Then the desired limit Properties 12 and 13 hold. Moreover, if  $\lambda$  is sufficiently large and  $d$  is not too small, then  $F_1 < f_1$  even if  $F_0 > f_0$ . Using this new model, one can explain the Underwood and Freund (1970) data as follows. To explain the greater errors for H-H than for L-L, we need to show that

$$F_1 - F_0 < f_1 - f_0. \quad (18)$$

By Equations 16 and 17, this reduces to the inequality

$$F_0 > f_0 \quad (6)$$

for the base familiarities. To explain the greater errors for L-H than for H-L, we need to show that

$$F_1 - f_0 > f_1 - F_0. \quad (19)$$

This inequality reduces to

$$2 > \frac{\lambda d}{(d + f_0)(d + F_0)}, \quad (20)$$

which is attainable for a range of  $d$  values, given any  $\lambda$ ,  $f_0$ , and  $F_0$ .

Although the new definitions of familiarity (Definitions 16 and 17) escape some problems of the Mandler (1980) model, they face challenges of a more subtle nature. Mandler himself has emphasized the cumulative nature of familiarity: "Each additional presentation and processing of an event adds some specified degree of integration to the target. . . . Repeated recognition tests (presentations) not only prevent loss of familiarity but actually increment it" (Mandler, 1980, pps. 267, 269). If presentations alter an item's familiarity by incrementing the integration of an item's internal representation, then where do incremental terms such as  $d/(d + F)$  leave off and new base familiarities begin? In other words, no matter how large  $d$  gets in a formula such as

$$F_1 = F_0 + \frac{\lambda d}{d + F_0}, \quad (16)$$

whether because of a single sustained presentation or many brief presentations, the formula does not show how an old "base" familiarity  $F_0$  generates a new "base" familiarity  $F_1$ . It makes no physical sense to arbitrarily write

$$F_2 = F_1 + \frac{\lambda d}{d + F_1} \quad (17)$$

for the next round of recognition experiments. A formula such as Equation 16 fails to explain how the cumulative effects of many presentations determine the influence of word frequency on the base familiarities  $F_0$  and  $f_0$ . It is also difficult to understand how a subject could separately store, as part of an item's "integration," an increment  $d$  and a base value  $F_0$  for 24 hr before comparing them on later recognition trials. Because of the fundamental nature of these issues in explaining the word frequency

effect, a different theory must be sought that captures the insights of the Underwood and Freund (1970) and Mandler (1980) models but escapes their pitfalls.

### 15. Role of Intralist Restructuring and Contextual Associations

Our explanation of the word frequency effect contains elements in common with both the Underwood and Freund (1970) model and the Mandler (1980) model in that we suggest both interitem and intraitem organizational changes subserve this effect. As in our interpretation of the IARs of the Underwood and Freund (1970) model, we note that contextual associations between  $V^*$  and  $\{A_4, A_5\}$ , as well as between  $A_4$  and  $A_5$ , can form as a result of studying a list of old words. This conclusion does not require any new assumptions within our theory. Such LTM associations always form when the relevant item and list representations are simultaneously activated in STM. Such associations can also be quickly restructured by competing LTM associations or masked by competitive STM interactions unless their triggering environmental events can utilize or form distinctive list representations, buffered by their own top-down templates, between which to build these new associative bonds (Grossberg, 1978a).

Thus a unique visual experience, via contextually mediated bonds between  $V^*$  and  $\{A_4, A_5\}$ , can have an enduring effect on word recognition. In a similar way, embedding an item in a unique verbal list can generate strong contextual effects due to the formation of new list representations within  $\{A_4, A_5\}$ . Such contextual effects are not usually important in a lexical decision task because such a task does not define lists of old items and of new items. Contextual effects are important, however, in experiments studying the word frequency effect, serial verbal learning, and the like.

To understand how contextual associations can be differentially formed with old L words and old H words, we first need to understand how L words and H words differentially activate their internal representations in STM. We show how such differential STM activations can differentially alter the intralist LTM organization or "integration," to use Mandler's phrase, of the corresponding list representations. These differential STM activations can also differentially form new interlist LTM associations with contextual representations in  $V^*$  and, under appropriate experimental conditions, with  $A_5$ . Such interlist LTM reactions are closer to the Underwood and Freund concepts of IARs. Both of these types of LTM changes cooperate to alter the total reaction to old words on test trials.

Even in an experiment in which old words are not divided into two classes, such as L and H, contextual associations can still contribute an increment in integration over and above the increment caused within the internal representations of the old words. This contextual increment helps to reduce the overall error rate in recognizing old words as distinct from new words.

We suggest that subjects use relatively simple STM indices to make these recognition judgments. In experiments in which presentation of a memory item can activate an informative contextual impression, say, via an  $A_4 \rightarrow A_5 \rightarrow V^*$  pathway or a  $V^* \rightarrow \{A_4, A_5\} \rightarrow V^*$  pathway, then differential STM activations of the contextual representations themselves may be used as cues. In experiments in which the experimental context is the same

on study trials and test trials, context can differentially act primarily via  $V^* \rightarrow \{A_4, A_5\}$  associations. Then the subject is reduced to using STM indices such as the differential sizes of STM bursts or equilibration values to judge old from new. These are the same types of STM indices that subjects use to make their judgments in a lexical decision task (see Section 11). Our theory hereby unifies the explanation of lexical decision and word frequency data by showing how different types of experiments can differentially probe the same perceptual and cognitive mechanisms.

### 16. Explaining Recognition and Recall Differences

To start our explanation of the main recognition and recall differences described in Sections 12–14, we note the obvious fact that both L and H words can be recognized as words by experimental subjects. Both types of words possess unitized internal representations at the list level (see Figure 5). Their differences, or lack of differences, in recognition and recall properties thus cannot be attributed to unitization or lack thereof per se. However, several quantitative differences in the learned encoding and interlist interactions between H and L words are relevant to our discussion.

The first difference can be understood by considering a property of the bottom-up feature tuning process described in Section 7. There we concluded that an LTM Vector  $z_j(t)$  equals a time-average

$$\sum_k a_j^{(k)}(t) S_j^{(k)}$$

of all the bottom-up signal patterns  $S_j^{(k)}$  that its node  $v_j$  can ever sample. If a particular pattern  $S_j^{(M)}$  appears with a high relative frequency, then its weight  $a_j^{(M)}(t)$  in Summation 21 becomes relatively large. Once  $z_j(t)$  approximately equals  $S_j^{(M)}$  further presentations of  $S_j^{(M)}$  cause relatively small changes in  $z_j(t)$ . The converse is also true. If a particular  $S_j^{(M)}$  has not been occurring frequently, then its reoccurrence can begin to cause a significant change in  $z_j(t)$  toward  $S_j^{(M)}$ . Thus presentation of infrequent patterns can begin to significantly retune LTM, other things being equal.

This conclusion depends on the fact that the pattern of signals  $T_j = S_j^{(M)} \cdot z_j$  that is caused by an infrequent pattern  $S_j^{(M)}$  is sharpened by contrast enhancement before being stored as a pattern of STM activities  $x_j$ . Thus relatively infrequent patterns  $S_j^{(M)}$  can cause relatively large  $x_j$ 's in those encoding populations  $v_j$  whose STM activities survive the process of contrast enhancement. These surviving STM activities can then drive learning within the corresponding LTM vectors  $z_j$  via the learning equation (Equation 1).

The contrast enhancement property helps to explain the insignificant difference between H and L word recognition in lexical decision tasks using a backward mask. The differences that may exist between the tuning of the LTM vectors to H and L words tend to be offset by contrast enhancement in STM. Nonwords, by contrast, cannot easily activate any list representations.

This conclusion no longer follows when the bottom-up activation process can read out top-down templates, as in lexical decision tasks that do not use backward masks. Then the same template-matching property that implies the list-item error trade-

off (see Section 11) progressively amplifies the small differences between H and L word LTM tuning that exist in the bottom-up and top-down pathways to generate large STM differences in the list representations of H and L words. Thus, whereas initial contrast enhancement within a list level tends to reduce word frequency effects, asymptotic template match-mismatch differences between list and item levels tend to amplify word frequency effects. A significant difference in the speed of H and L word recognition is also generated by the better top-down matching, and hence larger and faster STM activation, that occurs in response to H words than to L words.

This difference in the size and degree of sharpening of the H word list representations also helps to generate better H word recall than L word recall. In recall, as opposed to recognition, the task is to generate old words, whether H words or L words. On a study trial, H words generate greater STM activations than do L words. This difference is reflected on later recognition trials in the lower error rates of H-L than of L-L. During a study trial, the greater activation of H words facilitates the formation of interitem chunks (with  $A_5$ ) and contextual associations (with  $V^*$ ). On a later recall trials, those chunks and contextual associations that survive intervening competitive recoding can lead to better retrieval of old H words than of old L words. The better tuning of the sensory-motor associative map  $A_4 \rightarrow A_3 \rightarrow M_3$  (see Section 10) to H words than to L words also contributes to this effect.

We can use these properties to begin our explanation of why old L words paired with new L words are recognized better than old H words paired with new H words yet old H words paired with new L words are recognized better than old L words paired with new H words. Before turning to the role of contextual associations, we consider the interaction of two properties: (1) the greater LTM retuning of list representations caused by L word presentation than H word presentation and (2) the greater STM activations, both of a word's list representation and its interlist associates, caused by an H word than an L word.

Property 1 may be compared with the Mandler (1980) concept of a change in familiarity. As in Mandler's model, L words cause a larger tuning change than do H words. Unlike Mandler's model, a very brief training trial need not reverse the base "familiarity" values. Property 2 may be compared with the IAR concept of Underwood and Freund (1970). As in their model, we assert that learned interlist interactions play a role. Unlike their model, we do not associate values of 1 with all the IARs of a study item. Instead we focus on the effect of study conditions on the total STM activation generated by a new or old L word or H word.

Using these properties, we draw the following conclusions. The old-H-new-H comparison causes a relatively high number of errors, in part because study of an H word causes relatively little LTM tuning of its chunk (Property 1). Hence on a later recognition trial, both the old H words and the new H words cause large and similar amounts of total STM activation (Property 2). The old-L-new-L case causes a relatively low number of errors because the study of an L word causes relatively rapid LTM tuning of its chunk (Property 1). Hence on a later recognition trial, an old L word causes a relatively large STM activation, whereas a new L word causes a relatively small STM activation (Property 2). The better recognition of an old H word paired with a new L word compared with an old H word paired with

a new H word is also easily understood from this perspective. To understand the reversal effect, note that an old H word generates significantly more STM activation than does a new L word (Property 2), whereas the additional STM activation caused by an old L word (Property 1) is offset by the large STM activation caused by a new H word (Property 2).

Property 3 in our explanation is the formation of contextual associations to old L word and H word list representations. Contextual associations can form to all old words and their associates but not to new words that are not associates of old words. Contextual associations can hereby lower the overall error rate in a recognition task by augmenting the STM activations of all old words.

Some finer learned interactions can occur among the list representations of study items and with contextual list representations when the study period is organized in a way that approximates serial verbal learning or paired associative learning conditions (Grossberg, 1969d, 1982a, 1982c; Grossberg & Pepe, 1971), such as when lists of L words are studied together (Underwood & Freund, 1970). Although such learned interactions have not been needed to explain the main comparisons within the data described above, they do provide a clearer understanding of how H items become associatively linked with many list representations, and they may eventually help to explain why certain H-H, L-L, H-L, and L-H comparisons are not invariant across experimental conditions. Finally, other mechanisms described in this article may help to expand the data base explained by our theory. For example, the competitive interaction of all viable sublist representations in the masking field (see Section 10) could explain why new distractors that are compounds of previously presented words produce increased false alarm rates (Ghatala, Levin, Bell, Truman, & Lodico, 1978).

## 17. Concluding Remarks

The present article describes a macrotheory and a microtheory capable of modifying and unifying several models of language-related behavior and of characterizing the relations between different types of language-related data that are often treated separately from one another. Development of circular reactions, analysis by synthesis, motor theory of speech perception, serial and paired associate verbal learning, free recall, categorical perception, selective adaptation, auditory contrast, word superiority effects, word frequency effects on recognition and recall, and lexical decision tasks can now all be analyzed using a single processing theory (see Figure 5). The core of this theory consists of a few basic design principles such as the temporal chunking problem, the LTM invariance principle, and the factorization of pattern and energy (see Sections 9 and 10). These principles are realized by real-time networks built up from a few basic mechanisms such as bottom-up adaptive filters, top-down learned templates, and cooperative-competitive interactions of one sort or another.

No alternative theory has yet been shown to have a comparable explanatory and predictive range. In particular, as the debate continues concerning the relative virtues of matrix models and convolution models (Anderson, Silverstein, Ritz, & Jones, 1977; Eich, 1985; Murdock, 1983, 1985; Pike, 1984), it should be realized that the embedding field theory, which is assimilated within

the adaptive resonance theory, long ago provided a detailed analysis of the classical bowed and skewed serial position curve and of the error distributions found in serial verbal learning (Grossberg, 1969d, 1982c; Grossberg & Pepe, 1971). These fundamental data have not yet been explained by either the matrix model or the convolution model. We trace this explanatory gap to the absence within these models of the very sorts of design principles and nonlinear mechanisms that we have used to explain data about word recognition and recall. We suggest that any future theory that may supplant the present one must also include such design principles and nonlinear mechanisms.

Superimposed upon these design principles and mechanisms are a number of new functional ideas that can be used in a model-independent way to think about difficult data. For example, the idea of adaptive resonance provides a new vantage point for understanding how learned codes are stabilized against chaotic recoding by the "blooming buzzing confusion" of irrelevant experience, and for thinking about processing stages that interact via rapidly cycling feedback interactions. The concept of resonant equilibration provides a helpful way to think about verification and attentional focussing without being led into a serial processing metaphor that seems to have no plausible physical realization. The list-item error trade-off provides a new perspective for analyzing certain deviations from speed-accuracy trade-off, especially in situations wherein matching due to top-down feedback can compensate for initial error tendencies. The concepts of top-down subliminal control and bottom-up supraliminal control rationalize the distinction between attentional gain control and attentional priming and indicate how to supplant the intuitive concepts of "automatic activation" and "conscious control" by a mechanistic understanding. Such known principles, mechanisms, and functional ideas enable a large data base to be integrated concerning how humans learn and perform simple language skills and provide a foundation for future studies of the dynamical transformations whereby higher language skills are self-organized.

### References

- Anderson, J. A., Silverstein, J. W., Ritz, S. R., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413-451.
- Antos, S. J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 527-545.
- Basar, E., Flohr, H., Haken, H., & Mandell, A. J. (Eds.). (1983). *Synergetics of the brain*. New York: Springer-Verlag.
- Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 556-566.
- Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 389-401.
- Becker, C. A., Schvaneveldt, R. W., & Gomez, L. (1973, November). *Semantic, graphemic, and phonetic factors in word recognition*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, MO.
- Carpenter, G. A., & Grossberg, S. (1985). Neural dynamics of category learning and recognition: Attention, memory consolidation, and amnesia. In J. Davis, R. Newburgh, & E. Wegman (Eds.), *Brain structure, learning, and memory*. Washington, DC: American Association for the Advancement of Science Symposium Series.
- Carpenter, G. A., & Grossberg, S. (1986). Neural dynamics of category learning and recognition: Structural invariants, reinforcement, and evoked potentials. In M. L. Commons, S. M. Kosslyn, & R. J. Herrnstein (Eds.), *Pattern recognition in animals, people, and machines*. Hillsdale, NJ: Erlbaum.
- Cermak, L. S., & Craik, F. I. M. (Eds.). (1979). *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.
- Cohen, M. A., & Grossberg, S. (1984a). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, *36*, 428-456.
- Cohen, M. A., & Grossberg, S. (1984b). Some global properties of binocular resonances: Disparity matching, filling-in, and figure-ground synthesis. In P. Dodwell & T. Caelli (Eds.), *Figural synthesis*. Hillsdale, NJ: Erlbaum.
- Cohen, M. A., & Grossberg, S. (in press). Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology*.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). New York: Academic Press.
- Cooper, W. E. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.
- Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review*, *92*, 1-38.
- Ghatala, E. S., Levin, J. R., Bell, J. A., Truman, D. L., & Lodico, M. G. (1978). The effect of semantic and nonsemantic factors on the integration of verbal units in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 647-655.
- Glanzer, M., & Bowles, N. (1976). Analysis of the word frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 21-31.
- Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, *61*, 23-29.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Kreiger.
- Grossberg, S. (1969a). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, *1*, 319-350.
- Grossberg, S. (1969b). On learning, information, lateral inhibition, and transmitters. *Mathematical Biosciences*, *4*, 255-310.
- Grossberg, S. (1969c). On learning of spatiotemporal patterns by networks with ordered sensory and motor components: I. Excitatory components of the cerebellum. *Studies in Applied Mathematics*, *48*, 105-132.
- Grossberg, S. (1969d). On the serial learning of lists. *Mathematical Biosciences*, *4*, 201-253.
- Grossberg, S. (1970a). Neural pattern discrimination. *Journal of Theoretical Biology*, *27*, 291-337.
- Grossberg, S. (1970b). Some networks that can learn, remember, and reproduce any number of complicated space-time patterns: II. *Studies in Applied Mathematics*, *49*, 135-166.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *52*, 217-257.
- Grossberg, S. (1975). A neural model of attention, reinforcement, and

- discrimination learning. *International Review of Neurobiology*, 18, 263–327.
- Grossberg, S. (1976a). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121–134.
- Grossberg, S. (1976b). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187–202.
- Grossberg, S. (1978a). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.
- Grossberg, S. (1978b). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 17, 199–219.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1–51.
- Grossberg, S. (1982a). Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns. In S. I. Amari & M. Arbib (Eds.), *Competition and cooperation in neural networks* (pp. 295–341). New York: Springer-Verlag.
- Grossberg, S. (1982b). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, 89, 529–572.
- Grossberg, S. (1982c). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*. Boston: Reidel Press.
- Grossberg, S. (1983). The quantized geometry of visual space: The coherent computation of depth, form, and lightness. *The Behavioral and Brain Sciences*, 6, 625–692.
- Grossberg, S. (1984). Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory*, 7, 263–283.
- Grossberg, S. (1985). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines* (Vol. 1). New York: Academic Press.
- Grossberg, S. (1986a). *Competitive learning: From interactive activation to adaptive resonance*. Manuscript submitted for publication.
- Grossberg, S. (1986b). *The adaptive brain: I. Learning, reinforcement, motivation, and rhythm*. Amsterdam: North-Holland.
- Grossberg, S. (1986c). *The adaptive brain: II. Vision, speech, language, and motor control*. Amsterdam: North-Holland.
- Grossberg, S., & Kuperstein, M. (1985). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements*. Amsterdam: North-Holland.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92, 173–211.
- Grossberg, S., & Mingolla, E. (in press). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*.
- Grossberg, S., & Pepe, J. (1971). Spiking threshold and overarousal effects in serial learning. *Journal of Statistical Physics*, 3, 95–125.
- Grossberg, S., & Stone, G. (1985). *Neural dynamics of attention switching and temporal order information in short-term memory*. Manuscript submitted for publication.
- Halle, M., & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IRE Transactions and Information Theory*, IT-8, 155–159.
- Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. (1971). Recognition time for information stored in long-term memory. *Perception and Psychophysics*, 10, 8–14.
- Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review*, 74, 496–504.
- Landauer, T., & Freedman, J. (1968). Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, 7, 291–295.
- Lapinski, R. H., & Tweedy, J. R. (1976, August). *Associate-like nonwords in a lexical-decision task: Paradoxical semantic context effects*. Paper presented at the annual meeting of the Society for Mathematical Psychology, New York University.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.
- Lewis, J. (1970). Semantic processing of unattended messages using dichoptic listening. *Journal of Experimental Psychology*, 85, 225–228.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Lieberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Tuerber (Eds.), *Handbook of sensory physiology* (Vol. 8). Heidelberg: Springer-Verlag.
- MacKay, D. G. (1973). Aspects of the theory of comprehension, memory, and attention. *Quarterly Journal of Experimental Psychology*, 25, 22–40.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271.
- Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory and Cognition*, 10, 33–42.
- Mandler, G., Pearlstone, Z., & Koopmans, H. J. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 8, 410–423.
- Manelis, J. (1977). Frequency and meaningfulness in tachistoscopic word perception. *American Journal of Psychology*, 99, 269–280.
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548–558.
- Marcel, A. (1980). Conscious and preconscious recognition of polysemous words: Locating the selective effects of prior verbal context. In R. Nickerson (Ed.), *Attention and performance* (Vol. 8). Hillsdale, NJ: Erlbaum.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.
- McDonald, J. E., & Schvaneveldt, R. (1978, April). *Strategy in a lexical-decision task*. Paper presented at a meeting of the Rocky Mountain Psychological Association, Denver, CO.
- Meyer, D. E., Schvaneveldt, R., & Ruddy, M. G. (1974). Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5). New York: Academic Press.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.
- Murdock, B. B., Jr. (1983). A distributed memory model for serial-order information. *Psychological Review*, 90, 316–338.
- Murdock, B. B., Jr. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review*, 92, 130–132.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, 4, 648–654.
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: The roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, 106, 226–254.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Paap, K. R., & Newsome, S. L. (1980). A perceptual confusion account of the WSE in the target search paradigm. *Perception and Psychophysics*, 27, 444–456.

- Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, *89*, 573-594.
- Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, NJ: Erlbaum.
- Pew, R. W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica*, *30*, 16-26.
- Piaget, J. (1963). *The origins of intelligence in children*. New York: Norton.
- Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, *91*, 281-294.
- Pitts, W., & McCullough, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, *9*, 127-147.
- Posner, M. I., & Snyder, C. R. R. (1975a). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, NJ: Erlbaum.
- Posner, M. I., & Snyder, C. R. R. (1975b). Facilitation and inhibition in the processing of signals. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance* (Vol. 5). New York: Academic Press.
- Ratliff, F. (1965). *Mach bands: Quantitative studies on neural networks in the retina*. New York: Holden-Day.
- Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, *69*, 1154-1163.
- Rubenstein, H., Garfield, L., & Millikan, J. (1970). Homographic centers in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*, 487-494.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60-94.
- Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, *9*, 75-112.
- Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General*, *114*, 50-77.
- Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1982). Length effects in word perception: We is better than I but worse than you or them. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 91-105.
- Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1983). Reply to Matthei: We really is worse than you or them, and so are ma and pa. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 321-322.
- Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1-17.
- Schuberth, R. E. (1978, November). *Context effects in a lexical-decision task*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, Texas.
- Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*, 211-212.
- Schvaneveldt, R. W., & McDonald, J. E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 673-687.
- Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 243-256.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1-29.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Dorder's method. In W. G. Koster (Ed.), *Attention and performance* (Vol. 2). Amsterdam: North-Holland.
- Stevens, K. N. (1972). Segments, features, and analysis by synthesis. In J. V. Cavanaugh & I. G. Mattingly (Eds.), *Language by eye and by ear*. Cambridge, MA: MIT Press.
- Stevens, K. N., & Halle, M. (1964). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), *Proceedings of the AFCRL symposium on models for the perception of speech and visual form*. Cambridge, MA: MIT Press.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*, 234-249.
- Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.
- Underwood, B. J. (1966). *Experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.
- Underwood, B. J., & Freund, J. S. (1970). Word frequency and short term recognition memory. *American Journal of Psychology*, *83*, 343-351.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*, 59-85.
- Young, R. K. (1968). Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general theory*. Englewood Cliffs, NJ: Prentice-Hall.

Received January 30, 1985

Revision received June 11, 1985 ■

# Neural Dynamics of Word Recognition and Recall: Attentional Priming, Learning, and Resonance

Stephen Grossberg and Gregory Stone