COGNITION AND BRAIN THEORY, 1984, 7(3&4), 263-283

Unitization, Automaticity, Temporal Order, and Word Recognition

Stephen Grossberg
Boston University

ABSTRACT

Samuel, van Santen, and Johnston (1982, 1983) reported a word length effect in a word superiority paradigm. A word length effect was predicted in Grossberg (1978a). This article describes the main concepts about the unitization process that led to this prediction. The article also discusses recent data and models of word and letter perception, controlled and automatic information processing, temporal order information in short term memory and in long term memory, spreading activation, and limited capacity due to inhibitory interactions in terms of the unitization process. It is shown that several popular models have been based upon an inadequate definition of the functional units of cognitive processing, and of the principles subserving the unitization process. The dichotomy between automatic processing and limited capacity processing is, for example, based on a fundamental misunderstanding of the unitization process. These problems have caused internal paradoxes and predictive limitations of the models, which have prevented them from being unified into a single processing theory. A "self-organization critique" is applied to some recent models to illustrate their internal difficulties. It is also shown how principles of self-organization can be used to generate a theory wherein these data domains and their empirical models can begin to be unified.

THE WORD LENGTH EFFECT

The recent experiments of Samuel, van Santen, and Johnston (1982, 1983) discovered a word length effect in word superiority studies. That is, a letter is better recognized as it is embedded in longer words of lengths from 1 to 4. A word length effect was predicted in Grossberg (1978a, p. 329; reprinted in 1982a, p. 595). This prediction arose from an analysis of how unitization of new internal representations takes place in real-time. The same design princi-

Requests for reprints should be sent to Stephen Grossberg, Center for Adaptive Systems, 111 Cummington Street, Boston University, Boston, Mass. 02215.

ple is needed to unitize new internal representations in response to sound streams, visual letter arrays, or sequences of motor commands. Thus although the Samuel *et al.* experiments seem to study a narrowly defined information processing issue, my theory suggests that this type of experiment probes a general principle governing the learning of serial order in behavior, and thus should be generally known.

UNITIZATION AND PSYCHOLOGICAL PROGRESS

Samuel, van Santen, and Johnston also wrote that "lexical...theories...have difficulty explaining the length effect in a principled manner" (Samuel, van Santen, and Johnston, 1982, p. 104) and that "lexical theories had not previously included mechanisms that were explicitly length dependent" (Samuel, van Santen, and Johnston, 1983, p. 322). These assertions are true of lexical theories that are concerned entirely with information processing issues, such as letter and word recognition. By contrast, the lexical theory that led to the word length prediction was derived from an analysis of how behaving individuals adapt in real-time to environments whose properties can unpredictably change. Such an analysis leads to design principles and mechanisms that cannot easily be inferred from processing data. Other lexical processing theories did not predict the word length effect because they overlooked fundamental constraints upon the design of behavioral mechanisms.

These design constraints concern the evolutionary process—variously called chunking, unitization, automation, or coding—whereby behavioral fragments are grouped into new control units that become the fragments of still higher behavioral units in a continuing process of hierarchical organization and command synthesis. It is perhaps surprising that lexical theories have been so unconcerned with unitization, since pseudowords can acquire many of the recognition properties of words after just five or six presentations (Salasoo, Shiffrin, and Feustel, 1984).

In the remainder of this article, I will outline the main concepts needed to understand the word length prediction. I will also note some of the internal problems that beset several types of popular information processing models because they do not deal with the unitization issue. These models have arisen independently from one another and contain no principles whereby they can be unified. I will indicate how an analysis of unitization leads to a different theory that is free from these internal problems and also unifies the main insights of the disparate models.

THE TEMPORAL CHUNKING PROBLEM

The critical design problem that leads to the word length prediction is called the *temporal chunking problem*. Suppose that an unfamiliar list of familiar items is sequentially presented; e.g., a novel word composed of familiar letters. In terms of frequency and familiarity, the most familiar units in the list are the items themselves. In order 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 prevents item familiarity from forcing the list to always be processed as a sequence of individual items, rather than eventually as a list as a whole? How does a not-yet-established word representation overcome the salience of well-established letter representations? How does unitization of unfamiliar lists of familiar items ever get off the ground?

Another version of the temporal chunking problem becomes evident by noticing that every sublist of a list is a perfectly good list in its own right. Letters and words are special sublists that have achieved a privileged status due to experience. In order to understand how this privileged status emerges, we need to analyse the processing substrate upon which all possible sublists struggle to be represented even before learning occurs. The design of this processing framework must also enable learning to unfold through time in a stable and self-consistent way. In particular, what design constraints prevent the presentation of new list items from destabilizing the encoding of all past item sublists? What design constraints enable the totality of represented sublists to define a more global and predictive representation of the environment than any individual list chunk could?

The subtlety of this unitization process is reflected even by the trivial fact that novel words composed of familiar letters can be learned. This fact shows that not all sublists have equal prewired weights in the competitive struggle to be represented. Such prewired weights include the number of coding sites in a sublist representation and the strength of the competitive signals that are emitted from each sublist's representation. Somehow a word as a whole can use such prewired processing biases to overcome, or to mask, the learned potency of its constituent items. This is the primary reason in my theory for the existence of a word length effect in word superiority studies.

This conclusion seems, however, to be self-contradictory upon further reflection. If prewired word biases can *inhibit* learned letter biases, then how is perception of letters *facilitated* by a word context, which is the main result of word superiority studies? This paradox can also be resolved through an analysis of the unitization process.

ALL LETTERS ARE SUBLISTS

Some insight into this paradox can be gleaned by further considering what it means to say that every sublist of a list is also a list. In order for sublists of a list to struggle for representational status, sets of individual items of the list need first to be simultaneously represented in STM at some level of processing. For definiteness, call this level $\mathcal{F}_i$, where the index i does not equal 1 because, in the full theory, this level of processing is not the first one. The theory shows how item representations that are simultaneously active in STM across $\mathcal{F}_i$ can be grouped, or chunked, into representations of sublists at the next level of processing $\mathcal{F}_{i+1}$. The sublist representations can then compete with each other for STM activation within $\mathcal{F}_{i+1}$. Once the two levels $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$ are clearly distinguished, it becomes obvious that individual list items, being sublists, can be represented at $\mathcal{F}_{i+1}$ as well as at $\mathcal{F}_i$. In the special case of letters and words, this means that letters are represented at the item level, as well as at the list level. Prewired word biases can inhibit learned letter biases at the level $\mathcal{F}_{i+1}$, but not at the level $\mathcal{F}_i$. That is why I call level $\mathcal{F}_{i+1}$ a *masking field.*

To clearly understand how the item representations at $\mathcal{F}_i$ differ from the sublist representations at $\mathcal{F}_{i+1}$, one must study the theory's processes in some detail. Even without such a study, one can conclude that "all letters are sublists." Indeed, all events capable of being represented at $\mathcal{F}_{i+1}$ exist on an equal dynamical footing. In the full theory, the implications of this conclusion clarify how changes in the context of a verbal item can significantly alter the processing of that item, and why the problem of identifying the functional units of language has proved to be so perplexing (Darwin, 1976; Studdert-Kennedy, 1980; Young, 1968). In $\mathcal{F}_{i+1}$, no simple verbal description of the functional unit, such as phoneme or syllable, has a privileged status. Only the STM patterns that survive a context-sensitive interaction between associative and competitive rules have a concrete existence.

The dictum that "all letters are sublists" helps to explain the data of Wheeler (1970) that were a starting point for the Samuel *et al.* (1982, 1983) experiments. One might intuitively believe that, since a word context can improve the recognition of its constituent letters, letters such as I and A that are also words would be better recognized than other letters. Wheeler (1970) showed that this is not the case. In my theory, this is due to the property that *all* familiar letters have a unitized sublist representation at $\mathcal{F}_{i+1}$, not only letters that are also used as words. The Wheeler (1970) data thus demonstrate how perilous it is to directly translate the distinctions of lay language into the definition of an underlying psychological process. The lay concept of a word is a misleading guidepost for understanding the process whereby all familiar sublists can achieve a unitized status. In fact, letters such as I and A may be reported slightly worse than other letters. The same masking mechanism also helps to explain why word superiority effects do not occur in Chastain's paradigm (Chastain, 1982; Grossberg, 1984, Section 44), although the experimental manipulations that engage the masking mechanism differ in the two paradigms. The masking mechanism thus explains how opposite effects can be generated within closely related performance paradigms as an expression of the unitization process.

EXPECTANCY LEARNING AND PRIMING

To avoid possible misunderstanding, I should promptly say what the dictum "all letters are sublists" does not imply. It is well-known that a human subject can be differentially primed to preferentially respond to letters rather than words, or to numbers rather than letters, and so on. Such a capability involves the activation of learned top-down templates, or expectancies, that selectively sensitize some internal representations more than others. Top-down excitatory feedback, or priming, from $\mathcal{F}_{i+1}$ to $\mathcal{F}_i$ is also used to explain how the word length bias in $\mathcal{F}_{i+1}$ can differentially excite item representations in $\mathcal{F}_i$ to generate the word length effect. This is because the prewired biases of a masking field enable the sublist representations of longer sublists to generate larger top-down excitatory signals, other things being equal. The phrase "all letters are sublists" is thus a conclusion about the local processing laws that letters and words share, not about the global contextual effects that can flexibly modulate the STM and LTM processes that these laws define.

The existence in my theory of learned top-down templates, or expectancies, does not arise from a desire to fit data about word superiority, object superiority, attentional priming, phonemic restoration, and the like. The need for such templates, and the laws that govern their properties, were derived from an analysis of how the unitization process stabilizes itself against adventitious recoding by behaviorally irrelevant environmental events. This analysis led to many unexpected conclusions that have begun to unify a large data base. In Grossberg (1980), for example, these templates were used to analyse how STM is reset by unexpected events in a way that preserves the stability of unitized representations. This analysis led to the prediction (p. 25) that a hippocampal generator of the P300 evoked potential exists. A hippocampal P300 generator has been experimentally reported by Halgren *et al.* (1980). The validity of this prediction can be further tested by performing discrimination learning experiments that should be able to dissociate possible cortical and hippocampal generators of the P300 (Grossberg, 1982b, Section 48).

THE McCLELLAND AND RUMELHART MODEL

Before continuing my theoretical discussion, I should note that the conclusions which have already been drawn have major implications for popular models of letter and word recognition, such as the McClelland and Rumelhart model (McCelland and Rumelhart, 1981; Rumelhart and McClelland, 1982), for which I now sketch a self-organization critique.

By a "self-organization critique" I mean an internal analysis of a model from the viewpoint of whether its information processing mechanisms could, in principle, develop or be learned. A model which cannot, in principle, self-organize must be using certain mechanisms that are physically incorrect. Both the nodal units and the internodal interactions that McClelland and Rumelhart postulate are seriously challenged by a self-organization critique.

For example, McClelland and Rumelhart identify a stage of letter nodes that precedes a stage of word nodes. They use these stages to discuss the processing of letters in 4-letter words. The hypothesis of separate stages for letter and word processing implies that letters are not also represented on the level of words of length four.

In order to be of general applicability, these concepts should certainly be generalizable to words of length less than four, notably to 1-letter words such as A and I. A consistent extension of the McClelland and Rumelhart stages would require that those letters which are also words, such as A and I, are represented on both the letter level and the word level, whereas those letters which are not words, such as E and F, are represented only on the letter level. How this distinction can be learned without using a homunculus is unclear.

This problem of processing units is symptomatic of a more general difficulty. The letter and word levels contain only nodes that represent letters and words. What did these nodes represent before their respective letters and words were learned? Where will the nodes come from to represent the letters and words that the model individual has not yet learned? Are these nodes to be created *de novo*? Are they created *de novo* within the five or six trials that enable a pseudoword to acquire many of the recognition characteristics of a word (Salasso, Shiffrin, and Feustel, 1984)?

These concerns clarify the need to define, once and for all, a processing substrate that can represent the learned units of a subject's internal lexicon before, during, or after they are learned. Such a substrate cannot be defined in terms of letters and words without forcing the untenable conclusion that all letters and words from all possible languages past, present, and future, and *only* these units, have prelabelled nodes awaiting their use in every human brain. The assumption of separate letter and word levels also requires special assumptions to deal with various data, such as the data of Wheeler (1970) and Samuel, van Santen, and Johnston (1982, 1983) concerning word superiority effects. If separate letter and word levels exist, then letters such as A and I which are also words should, as words, be able to prime their letter representations. By contrast, letters such as D and E which are not words should receive no significant priming from the word level. One might therefore expect easier recognition of A and I than of D and E. This is not the case.

The assumption of separate letter and word levels could escape this contradiction by assuming that *all* letters can be recognized so much more quickly than words of length at least two that no priming whatsoever can be received from the word level before letter recognition is complete. This assumption would, however, appear to be incompatible with the word length data of

Samuel, van Santen, and Johnston (1982, 1983). These authors showed that recognition improves if a letter is embedded in words of greater length. Thus a letter that is presented alone for a fixed time before a mask appears is recognized less well than a letter presented for the same amount of time in a word of length 2, 3, or 4. These data cast doubt on any explanation based on speed of processing alone.

A related problem arises due to the manner in which McClelland and Rumelhart have interconnected their letter level and their word level. "Each letter node is assumed to activate all of those word nodes consistent with it and inhibit all other word nodes. Each active word node competes with all other word nodes..." (Rumelhart and McClelland, 1982, p. 61). Knowledge of which letters and words are consistent can only be achieved by learning a particular language. However, when learning mechanisms are superimposed upon these hypotheses, it can be shown that either the learning process whereby the letter-to-word connections are formed cannot get started, so that no word representations are ever learned, or that after learning gets started, a forced oscillation between learning and forgetting is triggered. Thus the model is unstable in a learning mode. This instability problem is one reason why all learned inter-level interactions within the lexical theory of Grossberg (1978a) were chosen to be excitatory.

The instability of learning in the McClelland and Rumelhart (1981) model can be understood by considering combinations of two possible cases: (a) Before learning occurs, strong inhibitory interactions exist from the letter level to the word level. Excitatory connections are learned until net excitatory connections exist from letters to compatible words. (b) Before learning occurs, strong excitatory interactions exist from the letter level to the word level. Inhibitory connections are learned until net inhibitory connections exist from letters to incompatible words. In case (a), the excitatory connections can be learned only if the word nodes to be conditioned can first be activated. They can be activated only by their letter nodes. Since all the strong connections from letter nodes to word nodes are initially inhibitory, the word nodes cannot receive a net excitatory signal, hence conditioning can never get started. In case (b), the inhibitory connections can be learned only if the word nodes to be conditioned can first be activated, since strong excitatory connections exist initially. Suppose, therefore, that conjoint activation of a letter node and a word node strengthens the inhibitory connection from a letter node to a word node. As the inhibitory connection becomes increasingly strong, activating the letter node progressively inhibits its target word node. As the connection strength tracks the size of this progressively decreasing word node activation, it too becomes smaller. As the connection strength becomes smaller, the word node activation can begin to recover. Then the connection strength can also grow larger once more. A cycle of forced learning and forgetting is hereby perpetuated.

In response to these observations, one might say: why not make all the learned inter-level connections excitatory, and let pre-wired intra-level connections be both excitatory and inhibitory. The conditionable inter-level connections can adjust themselves to the pre-wired intra-level connections to achieve the designed consistency and inconsistency relationships as a function of experience. This is, in fact, what the Grossberg (1978a) theory postulates.

Another conceptual difficulty of the McClelland and Rumelhart (1981) model is that it does not contain any principles suggesting how parameter choices that vary with list length, prior learning, or serial order can influence the coding of *individual* lists. Instead, the model assigns the same parameters to all word nodes, and derives all processing differences between words, pseudo-words, and non-words from differences in the *number* of activated words in the network hierarchy. Such an approach also leads to unstable learning, in addition to providing no ready explanations of data such as the word length effect of Samuel, van Santen, and Johnston (1982, 1983). To see why learning in such a network can become unstable, note that a word node corresponding to a word of length 4 can learn a subword of length 2 as quickly as a node corresponding to the subword itself, even in verbal contexts where the entire word is not presented. This property can cause unselective activation and coding of long word nodes by all of its subwords. The noise level in such unselective codes rapidly becomes unmanageable as the complexity of the word set that is to be encoded increases. All of these conceptual problems are overcome in a masking field (Grossberg, 1978a, Sections 36–43; 1984, Sections 37–44).

THE SCHNEIDER AND SHIFFRIN MODEL

The seminal articles of Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) have organized a large and complex data base in terms of the dichotomy between automatic and controlled processing. Concepts of unitization capable of explaining these data, as well as the word length effect, are fundamentally different from those espoused by Schneider and Shiffrin (Grossberg, 1978a). Experimental support for these concepts have accumulated at an accelerating rate during the last few years (Francolini and Egeth, 1980; Hoffman, Nelson, and Houck, 1983; Kahneman and Chajczyk, 1983; Kahneman and Treisman, 1983; Schneider and Fisk, 1984). Although the experimental models that have arisen from these data are also closer to the unitization theory, they have not yet incorporated some of this theory's most important insights.

Schneider and Shiffrin posited two complementary types of information processing to explain a larger data base. *Automatic* processing is said to be a simultaneous parallel, relatively independent detection process. *Controlled* processing is said to be a serial terminating search process. The authors showed that the two types of processing can be experimentally probed using different experimental manipulations. Automatic processing occurs when the subject has practiced at giving a consistent detection response to memory set items that are never distractors, as in detecting digits among letter distractors. This is called a *consistent mapping* (CM) condition. Controlled processing occurs when memory set items and distractors are mixed from trial to trial, as in detecting digits among digit distractors. This is called a *varied mapping* (VM) condition. CM performance is usually better than VM performance. During CM performance, there is little effect of varying the number of distractors in a frame or of memory set size. By contrast, VM performance is monotonically related to each of these variables. Also during CM performance, false alarms (detections when on target is present) increase significantly at fast frame speeds, but this does not occur during VM performance.

The distinction between controlled and automatic processing may be viewed as a contribution to the unitization literature. Roughly speaking, controlled processing is used before an item or task is unitized, whereas automatic processing is used after unitization has occurred. The use of distinct VM and CM paradigms to experimentally probe these different situations provided a static view of unitization by looking at "before" and "after" unitization conditions, but not at the process of unitization itself.

When one considers Schneider and Shiffrin's conception of controlled vs. automatic processing during the unitization process, it is seen to be fraught with difficulties. Consider, for example, the learning of any new list of familiar items, as in the temporal chunking problem of Section 2. According to Schneider and Shiffrin, each familiar item is assumed to be processed by a parallel process, while each unfamiliar inter-item contingency is processed by a serial process. Thus their theory claims that the brain rapidly alternates between parallel and serial processing in this situation. Moreover, as the whole list becomes unitized, their theory suggests that this hybrid of serial and parallel processing somehow switches to exclusively parallel processing.

A similar conceptual difficulty occurs when one considers visual information processing. When a subject views a picture whose left half contains a familiar face and whose right half contains a collection of unfamiliar features, the Schneider and Shiffrin theory would claim that the visual process somehow splits itself into a parallel half and a serial half. As unitization occurs, the visual process then somehow reintegrates itself into a parallel process as the unfamiliar features are unitized.

The conceptually paradoxical nature of these conclusions is matched by unexplained data. Why is it that the "time for automatic search is at least as long as that for an easy controlled search" (Schneider and Shiffrin, 1976)?

Do not such data violate the intuitive understanding of the concept "automatic"?

I claim that these problems arise from associating a serial *process* to the serial *properties* of controlled search, and a parallel *process* to the parallel *properties* of automatic search. By contrast, the unitization theory in Grossberg (1978a) suggests that both types of properties are generated by parallel mechanisms. As unitization proceeds, the distribution of learned bottom-up codes and top-down templates changes in an experimentally dependent fashion. The parallel mechanisms of the unitization theory do not change, but the learning that they control can make the difference between controlled and automatic performance properties.

Below I quote from Grossberg (1978a, Section 61) as a point of departure for further discussion of how concepts about controlled and automatic processing can be modified and thereby integrated into this theory of unitization. The most critical points in the quote occur at its beginning and end. The middle section alludes to mechanisms that have recently been incorporated into some empirical models. I have included bracketed terms to help the reader make the bridge between concepts of the unitization theory and concepts that are being used to explain more recent information processing experiments.

"Below it is argued that both types of processing utilize common parallel operations, and that their apparent differences are due to shifts in the relative balance of these operations that are caused by experimental conditions. In particular, serial *properties* do not necessarily imply serial *operations*... Consider CM [consistent mapping] search. Repeated use of the same memory set gradually generates a higher-order auditory code [category] that can sample the visual codes for all the items over successive trials. When the higher-order code is activated, the visual codes of *all* memory set items can be subliminally activated. Matching with any one of these codes generates a resonant burst [recognition event]. The process therefore seems to be more parallel than VM [varied mapping] search. I claim, however, that this is primarily because the higher-order code must be established before the visual codes of all memory set items can be sampled by a single internal representation...the...*codes* [filters] and *templates* [sets] that are activated in VM and CM conditions are different, but the two conditions otherwise share common mechanisms... Attention enters the search process in several ways. The simplest attentional reaction is amplification of network response to expected items [priming by gain control]...The 'time for automatic search is at least as long as that for a very easy controlled search'. This is paradoxical if CM search is a more efficient processing scheme. Is partial normalization [limited capacity] of the visual template one reason for this? If more cues are subliminally active [subthreshold] during CM than during VM search, then each cue will have less subliminal activity. The reaction time for supraliminal [superthreshold] signals to be generated during a match will then be greater during CM than during VM...Also of interest are the data concerning performance accuracy when a memory set item occurs 0, 1, 2, or 3 frames away from an identical, or different, memory set item...Matching one item does not require reset to match a different item. However, if two identical items occur simultaneously, then the first match can interfere with the registration of the second match...By explaining the Schneider and Shiffrin data in a unified way, we avoid several serious problems of their theory. They claim, and I agree, that automatic processing is used to rapidly code familiar behavioral units so that controlled processing can then build these units into new unitized elements. I disagree that the 'automatic attention response' in the CM condition is a mechanism that is qualitatively different from mechanisms operating in the VM condition. If the two types of conditions use serial vs. parallel *operations,* as Shiffrin and Schneider claim, then how does the brain tirelessly alternate between serial and parallel mechanisms as it practices any new list of unitized elements? How do the serial and parallel processes compete when a visual scene contains both unitized and unfamiliar but relevant objects? How does the switchover from serial to parallel processing take place as an item is unitized? These problems evaporate in the present theoretical framework."

Just as the McClelland and Rumelhart model assigns a letter level and a word level to verbal items that are so labelled by lay language, the Schneider and Shiffrin model assigns a serial process to ostensibly serial behavioral properties and a parallel process to ostensibly parallel behavioral properties. Consideration of how we unitize a novel list of familiar items reveals the paradoxical nature of these conclusions in both the McClelland and Rumelhart model and the Schneider and Shiffrin model, and provides a way to unify the two types of models.

Recent data have led several authors to reconsider the validity of the dichotomy between automatic and controlled processing. Some authors have attempted to save the binary nature of this distinction in a weakened form by introducing epicyclic concepts like *strongly automatic* and *partly automatic* (Kahneman and Chajczyk, 1983). Such epicycles often precede the final breakdown of a conceptual framework.

PARALLEL PROCESSING AND UNLIMITED CAPACITY

Schneider and Shiffrin's dichotomy between controlled processing as a serial process and automatic processing as a parallel process has led to other assumptions that are challenged by a self-organization critique. For example, Hoffman, Nelson, and Houck (1983, p. 380) write: "Stage 1 is characterized as a large interconnected set of nodes or logogens...which are automatically 'activated' by presentation of their corresponding sensory inputs... Processing in this stage is assumed to be parallel and unlimited in capacity." The assumption that "parallel processing" and "unlimited capacity processing" co-exist has been broadly accepted in the literature, and is one reason why it seems natural to identify controlled processing with a serial (that is, non-parallel) mechanism. A self-organization critique seriously challenges whether automatic activation and limited capacity processing form a creditable processing dichotomy, in even an approximate sense. The fundamental inadequacy of this dichotomy becomes clear using a microscopic analysis of the unitization process. Such an analysis shows that the activation of a *single* unitized representation *seems* to be automatic *because* of the action of a limited capacity competitive process. The process that is usually identified as the *antithesis* of automatic activation is *responsible for* the consensual impression of automatic activation. I claim that on the level of microscopic processing, the dichotomy between automatic activation and limited capacity processing is invalid.

To see why this is so, let us again consider the STM level $\mathcal{F}_{i+1}$ that regulates the LTM chunking of sublists. My analysis of this process suggests that sublists of a list competitively struggle for representational status within $\mathcal{F}_{i+1}$. When a familiar word is processed, the sublist representation corresponding to the word rapidly wins the competition *because* the word is familiar. The associative LTM changes that subserve word familiarity have altered the balance of competitive processing in favor of the word representation, but they have not elminated the existence of the competitive process that *could have* chosen a different winning representation in response to different learning conditions. The apparent automaticity of the word interpretation derives from the network's ability to rapidly suppress these alternative sublist parsings using a limited capacity competitive masking process.

This conclusion does not undermine the claim (Grossberg, 1978a, Section 61), that apparent capacity changes can be due to learning of new chunks (or filters) and expectancies (or sets), as well as to several types of attentional mechanisms. These processes enable the network to reorganize its reactions to the same input patterns. A different learned top-down set can, for example, match an input pattern that a previous set mismatched. A different learned bottom-up filter can, for example, match a top-down set that a previous filter mismatched. Attentional gain control can, for example, focus network sensitivity at a subfield where an approximate match occurs, or at a different subfield where a serious mismatch occurs.

THE FUNCTIONAL UNIT OF COGNITIVE PROCESSING: NOT SPREADING ACTIVATION

The above example illustrates my claim that the traditional discussion of unlimited capacity suffers from an inadequate choice of the functional unit of cognitive processing. I suggest that the functional unit is not activation of a single node, or a "spreading activation" among individual nodes. The functional unit is a spatial pattern of activity that is coherently processed across a field of nodes. Once one accepts that the functional unit of processing is a spatially distributed activity pattern, rather than individual nodal activations, then "a large interconnected set of nodes" may simply transform one spatial pattern into another spatial pattern. The popular processing metaphor that directly relates the number of nodes or pathways to processing capacity then collapses.

This processing metaphor has often been used to explain how unitized representations can be automatically activated without capacity limitations. One imagines an appealing picture in which content addressable nodes are automatically activated by signals along labelled pathways. If many nodes exist, then they can process their labelled signals with less interference, other things being equal. Given this metaphor, the antithesis of automatic activation seems to be a limited capacity process in which many nodes compete for a limited activation resource. An analysis of unitization undermines the internal logic behind this assumption. Along the way, it also variates the assumption that the computer is a viable model of human information processing.

CAPACITY VS. MATCHING

If the metaphor that many people use to discuss limited capacity is questionable, then the notion of capacity itself needs reinvestigation. Various experiments have demonstrated, for example, that recognition accuracy and reaction time do not depend on processing load *per se,* but rather on factors like the goodness of match between priming and test cues (Fisher and Craik, 1980; Myers and Lorch, 1980; Schvaneveldt and McDonald, 1981). An increased reaction time is thus not due just to competition for a limited activation resource among many mutually inhibitory nodes. Mutual inhibition can subserve a match (which can speed up reaction time) or a mismatch (which can slow down reaction time) over the same set of activated nodes.

An analysis of the unitization process leads to mechanisms which also have these properties (Grossberg, 1976b, 1980). These mechanisms describe competitive and cooperative internodal interactions that occur at every level of network processing. Such interactions enable each level to sensitively process its patterned functional units without major contamination by internal noise or saturation effects. The masking geometry of the sublist level $\mathcal{F}_{i+1}$ is, in fact, a special case of these interactions. One can view the masking geometry as a competitive-cooperative interaction scheme that developmentally equilibrates to input patterns which vary in spatial scale and processing load (Grossberg, 1984, Sections 42–43).

Satisfying the general need for sensitive registration of patterned functional units at each network level *automatically* leads to properties that are compatible with the aforementioned reaction time data. This is true because an approximate match between a pair of bottom-up and top-down input patterns at a level can enhance its activation, thereby reducing its reaction time. By contrast, a mismatch between a pair of bottom-up and top-down input patterns at a level can suppress its activation, thereby increasing its reaction time. In both the match and the mismatch situations, the same number of nodes can receive inputs, the total input size can be the same, and thus the same network capacity is utilized.

This relationship between matching, activity amplification, and reaction time plays a fundamental role in the theory's explanation of how the stability-plasticity dilemma is solved, and about how a mismatch can trigger a search of associative memory (Grossberg, 1980). The relationship also shows that certain types of matching are more appropriate as cognitive mechanisms than others. In particular, it argues against the use of Euclidean matching algorithms (Grossberg, 1983, Section 22).

These remarks illustrate how a seemingly elementary problem about real-time processing, such as the noise-saturation problem, if carefully posed and quantitatively solved, can have unsuspected implications that ramify into and thereby help to unify a large and difficult experimental literature.

ADAPTIVE FILTER: THE PROCESSING BRIDGE BETWEEN SUBLIST MASKING AND TEMPORAL ORDER INFORMATION OVER ITEM REPRESENTATIONS

A still broader unification of data and models emerges when one considers how signals are relayed from level $\mathcal{F}_i$ to the next level $\mathcal{F}_{i+1}$, and conversely. When a signal from a node in $\mathcal{F}_i$ is carried along a pathway to $\mathcal{F}_{i+1}$, the signal is multiplied, or *gated,* by the pathway's LTM trace. The LTM gated signal then reaches the target node. Each target node sums up all of its LTM gated signals. In this way, a pattern of output signals from $\mathcal{F}_i$ generates a pattern of input signals to $\mathcal{F}_{i+1}$. This transformation is said to define an *adaptive filter.*

The input pattern to $\mathcal{F}_{i+1}$ is itself quickly transformed further by the competitive-cooperative interactions within $\mathcal{F}_{i+1}$. In the simplest example of this process, these interactions choose the node which received the largest input. The choice transformation executes a particularly severe type of contrast enhancement. In a masking geometry such as $\mathcal{F}_{i+1}$, the contrast enhancing transformation is considerably more subtle than a simple choice. The transformed pattern, not the input pattern itself, is then stored in STM. Only nodes which are active in STM across $\mathcal{F}_{i+1}$ can elicit new learning at their contiguous LTM traces.

This type of interaction between associative LTM mechanisms and competitive-cooperative STM mechanisms has many desirable properties. It generalizes the Baysian tendency to minimize risk in a noisy environment. It spontaneously tends to form learned categories. Its categories are stable under several types of perturbations. Its STM patterns are context-sensitive. Its learning at each LTM trace is sensitive at each time to the entire STM pattern that is active at that time, as well as to all prior learning that ever occurred at *all* the LTM traces. The learning capabilities of the choice model are mathematically characterized in Grossberg (1976a). The properties of the masking field model are described in Grossberg (1978a, 1984).

In the special case where the levels are the item level $\mathcal{F}_i$ and the sublist level $\mathcal{F}_{i+1}$, the activity pattern across $\mathcal{F}_i$ encodes temporal order information (TOI) in STM across the item representations of $\mathcal{F}_i$, and the LTM traces in the pathways between $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$ encode temporal order information (TOI) in LTM. The similarity between the patterns of LTM TOI in certain pathways and the pattern of STM TOI that is stored at any moment across the item representations of $\mathcal{F}_i$ helps to determine which unitized sublist representations will be activated across $\mathcal{F}_{i+1}$ by the bottom-up filter.

Unless a model explicitly defines how TOI in LTM is encoded, the model cannot determine how to compute TOI in STM so that a reasonable comparison process between STM and LTM can take place. The reverse conclusion is also true. If one does understand how TOI in LTM is computed, then one can use this information to *derive* laws for the temporal unfolding of TOI in STM. This was done in Grossberg (1978a, 1978b).

These STM laws have many implications for data and models about STM. For example, these STM laws suggest an alternative to serial buffer models such as the Atkinson and Shiffrin (1968, 1971) model of free recall by showing how to encode TOI in STM without using a serial buffer, and by explaining data that are at variance with the classical buffer model, such as data of Lee and Estes (1977), Ratcliff (1981), Reeves and Sperling (1984), and Sperling and Reeves (1980). The model also provides a principled derivation of mechanisms similar to those in the empirically derived Reeves and Sperling (1984) Generalized Attention Gating Model (GAGM), and raises processing issues that have not yet been addressed by experiments. The fact of greatest importance is that this approach shows how temporal order information of items in STM, temporal order information of sublist chunks and templates in LTM (filters and sets), and competitive masking of sublist chunks in STM are designed together as parts of the unified processing module that regulates unitization. The next sections indicate how this unified processing module is designed, and discusses some related experimental issues.

THE LTM INVARIANCE PRINCIPLE: TEMPORAL ORDER INFORMATION WITHOUT A SERIAL BUFFER

I now summarize how the adaptive filter is used to constrain the law of STM TOI. To do this, I again consider how we learn a novel list of familiar items. Suppose that list items $r_1, r_2, \ldots, r_j$ have already been presented. Suppose that these items have generated a spatial pattern of STM activation across the item representations of $\mathcal{F}_i$. This STM pattern represents "past" order information. I assume that a new list item r_{j+1} can alter the total pattern of STM across $\mathcal{F}_i$, but that this new STM pattern does not cause LTM recoding of that part of the pattern which represents past order information. For example, learning a novel word does not force unlearning of its constituent letters. New events are permitted to weaken the influence of LTM codes representing past order information on STM decision-making within $\mathcal{F}_{i+1}$, but not to deny the fact that the past events occurred. This hypothesis prevents the LTM record of past order information from being destroyed by every future event that happens to occur.

To translate this intuitive discussion into a precise computation, let us again recognize that every sublist of the list $r_1, r_2, \ldots, r_j$ is a perfectly good list in its own right. Every such sublist can, in principle, be encoded by LTM patterns in the adaptive filter from $\mathcal{F}_i$ to $\mathcal{F}_{i+1}$. To prevent a future event r_{j+1} from destroying these past list encodings, I assume that the following principle holds (Grossberg, 1978a, 1978b):

LTM INVARIANCE PRINCIPLE: The spatial patterns of STM TOI across $\mathcal{F}_i$ are generated by a sequentially presented list in such a way as to leave the LTM codes of past events invariant.

The LTM Invariance Principle is instantiated by choosing STM activities across $\mathcal{F}_i$ so that the *relative* activities of *all possible filterings* of a past event sequence $r_1, r_2, \ldots, r_j$ are left invariant by a future event r_{j+1}. It turns out that this property is also generated by a suitably designed competitive-cooperative interaction across $\mathcal{F}_i$, in keeping with general requirements that pattern processing across $\mathcal{F}_i$ be free from massive noise or saturation. Some of the most important properties of the STM TOI patterns that can arise in $\mathcal{F}_i$ are the following ones.

Primacy gradients, recency gradients, and bowed gradients in STM can occur. Primacy gradients can be generated by sufficiently short lists. Direct read-out of TOI from STM can then be accomplished by the combination of a reaction time rule that reads-out the largest activities first, and a self-inhibitory reset rule that prevents read-out of a single item from perseverating for all time. Several recent empirical models have used variants of these rules (Reeves and Sperling, 1984; Rumelhart and Norman, 1982). The STM gradients, and thus the TOI, that develop through time are sensitive to the amount of attention that an item receives when it enters STM and to the subsequent transformation of these STM activities by lateral inhibition. The GAGM model of Reeves and Sperling (1984) also makes this point.

The existence of a primacy gradient in STM raises an issue that has not yet been addressed by experimentalists. Interference experiments (Rundus, 1971) suggest that a primacy gradient in STM does not exist in the free recall paradigm. Such data have been used to support models of free recall in which the only primacy gradient in free recall is due to LTM (Atkinson and Shiffrin, 1968, 1971). The possibility of recalling a short list correctly out of STM suggests, by contrast, that a primacy gradient in STM can sometimes exist during free recall. Free recall data of Korsakoff amnesics (Baddeley and Warrington, 1970) and of normals (Hogan and Hogan, 1975) also support this conclusion. In Grossberg (1978b, Section 7), I showed how this apparent contradiction can be theoretically explained. My explanation suggests that a limited capacity competitive process prevents a primacy gradient in STM from being measured in an interference experiment, even in cases where it exists. Moreover, this limited capacity process is a parallel process, not a serial process. This explanation has not yet been experimentally tested. It illustrates that, even though the words "limited capacity process" and "parallel process" are freely used in the experimental literature, their implications are not widely understood.

Another important issue is raised by this STM TOI model. The model is capable of generating STM TOI without the use of a serial buffer. The TOI evolves through time as it does across item representations due to the network's competitive rules. The GAGM model of Reeves and Sperling (1984) also works without a serial buffer using mechanisms similar to those introduced in my theory. Classical serial buffers, by contrast, such as those of Atkinson and Shiffrin (1968, 1971) and Raaijmakers and Shiffrin (1981), do not fare well when they are analysed from the viewpoint of the unitization process (Grossberg, 1978b). The interplay of factors relating to attention, competition, serial buffers, and primacy gradients in STM require much more experimental study.

A related set of remarks can be made about ideas concerning TOI in LTM, notably the LTM TOI that evolves within the top-down conditionable pathways from $\mathcal{F}_{i+1}$ to $\mathcal{F}_i$ during serial verbal learning and paired associate learning. The bowed and skewed serial position effect and related verbal learning data were analysed using such a buffer-free interaction between STM and LTM in Grossberg (1969) and Grossberg and Pepe (1970, 1971). These LTM TOI rules turned out to have the right properties to build up a theory of how goal-oriented cognitive plans are self-organized (Grossberg, 1978a). A number of predictions concerning how the bowed serial position curve should change with state variables like arousal were made in 1970–71, but still have not been experimentally tested, despite their importance for understanding verbal learning, cognitive planning, and the transition to abnormal overaroused-attentive states such as those found in schizophrenia (Maher, 1977). A nontechnical review of these serial learning concepts is found in Grossberg (1982c). Murdock (1979) has been using related ideas about cross-correlation to analyse verbal learning data, but his model's computations have not yet enabled him to explain the bowed and skewed serial position curve.

SPATIAL FREQUENCY ANALYSIS OF TEMPORAL ORDER INFORMATION

The discussion of STM TOI at $\mathcal{F}_i$ and of sublist masking at $\mathcal{F}_{i+1}$ shows that both levels $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$ are designed as competitive networks, even though they accomplish different functional tasks. The fact that both $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$ possess a "limited capacity" provides little insight into how they work, or how they work so differently; notably how $\mathcal{F}_{i+1}$, but not $\mathcal{F}_i$, is capable of computing a "magic number seven" (Miller, 1957). One of the important tasks of cognitive science is, I believe, to classify specialized competitive networks according to the functional transformations that these networks can compute. A great deal is now known about these transformations (Grossberg, 1982a).

To end this discussion, I will now indicate how an analysis of unitization leads to the conclusion that the competitive masking process in $\mathcal{F}_{i+1}$ does a type of spatial frequency analysis of the LTM-filtered STM TOI that it receives from $\mathcal{F}_i$. This observation shows that mechanisms which are more familiar in visual, or more generally spatial, processing are also important in language or, more generally temporal, processing. The interactions between experimentalists in these two areas should thus be stronger than they are at present.

The last section indicated how the LTM Invariance Principle can be used to generate STM TOI across item representations in $\mathcal{F}_i$. A spatial pattern of STM activity over a set of item representations encodes this information. As more items are presented, new spatial patterns are registered that include larger regions of the item field, up to some maximal list length. Thus the *temporal* processing of items is converted into a succession of expanding *spatial* patterns.

Given this insight, the temporal chunking problem can be rephrased as follows. How do sublist chunks in $\mathcal{F}_{i+1}$ that encode broader regions of the item field mask sublist chunks that encode narrower regions of the item field? When I asked this question about language processing in 1974, I already knew the answer due to work on visual masking that my colleague Dan Levine and I were just finishing (Grossberg and Levine, 1975; Levine and Grossberg, 1976). We had shown how to define competitive networks that are composed of masking subfields. Each masking subfield was characterized by a different choice of numerical parameters. At the risk of oversimplifying the analysis, we found that subfields whose cell populations have broader spatial frequencies and more coding sites can mask STM activation of subfields with narrower spatial frequencies and fewer coding sites. The temporal chunking problem then suggested how to put together results about STM TOI and competitive masking by suggesting the following design principle, whose relevance to the word length effect of Samuel *et al.* (1982, 1983) should now be obvious.

SEQUENCE MASKING PRINCIPLE: Broader regions of the item field $\mathcal{F}_i$ are filtered in such a way that they selectively excite nodes in $\mathcal{F}_{i+1}$ with larger masking parameters.

The sequence masking principle is capable of organizing a series of simple design rules for the integrated construction of the network module consisting of $\mathcal{F}_i$, $\mathcal{F}_{i+1}$, and their mutual interactions. Many predictions about cognitive processing, neural development, neuroanatomy, and neurophysiology are consequences of this construction. See Grossberg (1984) for a recent description of these and related properties.

CONCLUSION

This article has avoided most of the technical considerations that are needed to precisely characterize the dynamics of unitization. Instead it has focused on a few of the intuitive ideas that motivate a larger theory. These ideas illustrate how models can be strengthened and unified by analysing their internal structure from the viewpoint of the unitization process, and indicate that this process of unification is already well underway.

ACKNOWLEDGMENTS

Supported in part by the Office of Naval Research (ONR-N00014-83-K0337).

REFERENCES

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, DC: V. H. Winston and Sons, 1973.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *Advances in the psychology of learning and motivation research and theory* (Vol. 2). New York: Academic Press, 1968.

Atkinson, R. C., & Shiffrin, R. M. The control of short term memory. *Scientific American,* August, 1971.

Baddeley, A. D., & Warrington, E. K. Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1970, *9,* 176–189.

Chastain, G. Scanning, holistic encoding, and the word-superiority effect. *Memory and Cognition,* 1982, *10,* 232–236.

Cole, R. A., & Rudnicky, A. I. What's new in speech perception? The research and ideas of William Chandler Bagley, 1874–1946. *Psychological Review,* 1983, *90,* 94–101.

Darwin, C. J. The perception of speech. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception, Vol. VII: Language and speech.* New York: Academic Press, 1976.

Fisher, R. P., & Craik, F. I. M. The effects of elaboration on recognition memory. *Memory and Cognition,* 1980, *8,* 400–404.

Francolini, C. M., & Egeth, H. On the non-automaticity of "automatic" activation: Evidence of selective seeing. *Perception and Psychophysics,* 1980, *27,* 331–342.

Grossberg, S. On the serial learning of lists. *Mathematical Biosciences,* 1969, *4,* 201-253.
Grossberg, S. Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics,* 1976, *23,* 121-134. (a)
Grossberg, S. Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics,* 1976, *23,* 187-202. (b)
Grossberg, S. A theory of human memory: Self-organization of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), *Progress in theoretical biology,* Vol. 5, pp. 223-374. New York: Academic Press, 1978. (a)
Grossberg, S. Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology,* 1978, *17,* 199-219. (b)
Grossberg, S. How does a brain build a cognitive code? *Psychological Review,* 1980, *87,* 1-51.
Grossberg, S. *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston: Reidel Press, 1982. (a)
Grossberg, S. Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review,* 1982, *89,* 529-572. (b)
Grossberg, S. Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns. In S. I. Amari and M. Arbib (Eds.), *Competition and cooperation in neural networks.* New York: Springer-Verlag, 1982. (c).
Grossberg, S. The quantized geometry of visual space: The coherent computation of depth, form, and lightness. *The Behavioral and Brain Sciences,* 1983, *6,* 625-692.
Grossberg, S. The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab and H. C. Nusbaum (Eds.), *Perception of speech and visual form: Theoretical issues, models, and research.* New York: Academic Press, 1984.
Grossberg, S., & Levine, D. S. Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology,* 1975, *53,* 341-380.
Grossberg, S., & Pepe, J. Schizophrenia: Possible dependence of associational span, bowing, and primacy vs. recency on spiking threshold. *Behavioral Science,* 1970, *15,* 359-362.
Grossberg, S., & Pepe, J. Spiking threshold and overarousal effects in serial learning. *Journal of Statistical Physics,* 1971, *3,* 95-125.
Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., & Crandall, P. H. Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science,* 1980, *210,* 803-805.
Hoffman, J. E., Nelson, B., & Houck, M. R. The role of attentional resources in automatic detection. *Cognitive Psychology,* 1983, *51,* 379-410.
Hogan, R. M., & Hogan, M. M. Structural and transient components of memory. *Memory and Cognition,* 1975, *3,* 210-215.
Kahneman, D., & Chajczyk, D. Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception and Performance,* 1983, *9,* 497-509.
Kahneman, D., & Treisman, A. Changing views of attention and automaticity. In R. Parasuramen, R. Davies, & J. Beatty (Eds.), *Varieties of attention.* New York: Academic Press, 1983.
Lee, C., & Estes, W. K. Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior,* 1977, *16,* 395-418.
Levine, D. S., & Grossberg, S. Visual illusion in neural networks: Line neutralization, tilt aftereffect, and angle expansion. *Journal of Theoretical Biology,* 1976, *61,* 477-504.
Macchi, G., & Rinvik, E. Thalamo-telencephalic circuits: A neuroanatomical survey. In A. Rémond (Ed.), *Handbook of electroencephalography and clinical neurophysiology* (Vol. 2, part A). Amsterdam: Elsevier/North-Holland, 1976.
Maher, B. A. *Contributions to the psychopathology of schizophrenia.* New York: Academic Press, 1977.

McClelland, J. L., & Rumelhart, D. E. An interactive activation model of context effects in letter perception, Part 1: An account of basic findings. *Psychological Review,* 1981, *88,* 375–407.

Miller, G. A. The magic number seven plus or minus two. *Psychological Review,* 1956, *63,* 81.

Murdock, B. B. Convolution and correlation in perception and memory. In L. G. Nilsson (Ed.), *Perspectives in memory research.* Hillsdale, NJ: Erlbaum Press, 1979.

Myers, J. L., & Lorch, R. F. Jr. Interference and facilitation effects of primes upon verification processes. *Memory and Cognition,* 1980, *8,* 405–414.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Posner, M. I. *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum Press, 1978.

Raaijmakers, J. G. W., & Shiffrin, R. M. Search of associative memory. *Psychological Review,* 1981, *88,* 93–134.

Ratcliff, R. A theory of order relations in pérceptual matching. *Psychological Review,* 1981, *88,* 552–572.

Reeves, A., & Sperling, G. Attentional theory of order information in short-term visual memory. Preprint, 1984.

Rumelhart, D. E., & McClelland, J. L. 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,* 1982, *89,* 60–94.

Rumelhart, D. E., & Norman, D. A. Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science,* 1982, 6, 1–36.

Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology,* 1971, *89,* 63–77.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. Building permanent memory codes: Codification and repetition effects in word identification. Preprint, 1984.

Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. Length effects in word perception: We is better than I but worse than you or them. *Journal of Experimental Psychology: Human Perception and Performance,* 1982, *8,* 91–105.

Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. 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,* 1983, *9,* 321–322.

Schneider, W., & Fisk, A. D. Automatic category search and its transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 1984, *10,* 1–15.

Schneider, W., & Shiffrin, R. M. Automatic and controlled information processing in vision. In D. LaBarge and S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension.* Hillsdale, NJ: Erlbaum, 1976.

Schneider, W., & Shiffrin, R. M. Controlled and automatic information processing, I: Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Schvaneveldt, R. W., & McDonald, J. E. Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance,* 1981, *3,* 673–687.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing, II: Perceptual learning, automatic attending and a general theory. *Psychological Review,* 1977, *84,* 127–190.

Sperling, G., & Reeves, A. Measuring the reaction time of a shift of visual attention. In R. Nickerson (Ed.), *Attention and performance VIII.* Hillsdale, NJ: Erlbam Press, 1980.

Studdert-Kennedy, M. Speech perception. *Language and Speech,* 1980, *23,* 45–65.

Tsumoto, T., Creutzfeldt, O. D., & Legéndy, C. R. Functional organization of the corticofugal system from visual cortex to lateral geniculate body of the cat. *Experimental Brain Research,* 1978, *25,* 291–306.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology,* 1970, *1,* 59–85.

Young, R. K. Serial learning. In T. R. Dixon and D. L Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, NJ: Prentice-Hall, 1968.