

Behavioral Contrast in Short Term Memory: Serial Binary Memory Models or Parallel Continuous Memory Models?

STEPHEN GROSSBERG¹

Department of Mathematics, Boston University, Boston, Massachusetts 02215

This paper develops a model wherein STM primacy as well as recency effects can occur. The STM primacy effects can be used to generate correct immediate recall of short lists that have not been coded in LTM. The properties of the model are interpreted in terms of explicit neural mechanisms. The STM primacy effect is a behavioral contrast effect that is analogous to the behavioral contrast that can occur during discrimination learning. The adaptational mechanism that accounts for these effects is also implicated in data on reaction time, retinal adaptation, ratio scales in choice behavior, and von Restorff-type effects. Its ubiquitous appearance is due to the fact that it solves a universal problem concerning the parallel processing of patterned data by noisy cells with finitely many excitable sites. It is argued that the STM primacy effect is not measured in interference experiments because it is masked by competitive STM interactions. These competitive interactions do not prevent the LTM primacy effect from influencing performance. The paper criticizes recent models of STM that use computer analogies to justify binary codes, serial STM buffers, and serial scanning procedures. Several deficiencies of serial models in dealing with psychological and neural processing are overcome by a model in which continuous STM activities and parallel real-time operations play an important role.

1. INTRODUCTION: SERIAL AND BINARY MEMORY PROCESSES OR PARALLEL AND CONTINUOUS MEMORY PROCESSES?

A great deal of experimental and theoretical work (e.g., Melton & Martin, 1972; Restle, *et al.*, 1975; Tulving & Donaldson, 1972) has been done on problems relating to how learning subjects store data in short-term memory (STM) before it is transcribed into long-term memory (LTM) or otherwise transformed. Many experimental findings have been interpreted, either explicitly or implicitly, in terms of computer-like constructs, such as binary codes (Anderson & Bower, 1974; Atkinson & Shiffrin, 1968), serial buffers (Atkinson & Shiffrin, 1968), and serial scanning procedures (Sternberg, 1966). This paper suggests that the computer analogy has led to several basic difficulties. It also suggests an alternative theory to explain how order information in STM evolves in real time. This theory predicts a new experimental phenomenon, behavioral contrast in time, analogous to the phenomenon of behavioral contrast in space that occurs during discrimination learning (Bloomfield, 1969), and explains both phenomena using collective properties of well-known neural mechanisms. The theory is

¹Supported in part by the National Science Foundation (NSF MCS 77-02958).

illustrated by examples concerning free recall, discrimination learning, reaction time, perceptual adaptation, and von Restorff-type effects.

Discrete and serial memory models have an immediate appeal in situations where behavioral responses are counted as they occur one at a time. However, discrete and serial behavioral properties do not imply that the processes which control them are also discrete and serial. Townsend (1974) has, for example, noted that the reaction times found in the Sternberg paradigm do not imply a serial process by describing statistical parallel processing completion times that are indistinguishable from their serial processing counterparts. It can furthermore be argued that accepting a discrete serial model precludes the study of some basic processes of learning and perception. Even in simple behavioral tasks, both continuous and discrete elements are evident. Many perceptions seem to be continuous; for example, colors or sounds seem to vary continuously in intensity and quality. Yet the language with which we describe them seems to be much more discrete; for example, letters such as *A* or *B* seem, in daily speech and listening, to be indecomposable units of behavior, and all of our language utterances seem to be built up from finitely many such units. To understand the process of seeing a color and describing it by language, we must face the problem of how seemingly continuous representations can interact, or be transformed, into seemingly discrete representations. We must be able to discuss the "degree of continuity" at all levels of this transformation.

The relationship between seemingly continuous and seemingly discrete events is a deep one especially because the same behavior can seem to have either type of representation depending on how familiar it is to us. The process of learning to walk or to talk is illustrative. Before we can walk, attention is paid almost continuously to the complex coordinations that are required. Yet after we know how to walk, much of this coordination is automatic, so that we can simply start to walk, pay attention to other things, and then decide to stop walking. The control of walking eventually approximates a binary on-off switch, except for some steering and object avoidance. Thus the process of learning can alter the control of walking from a relatively continuous representation to a relatively discrete one. A similar process occurs in many learning tasks wherein some form of "abstraction" occurs. Yet it would be wrong to believe that, after such a task is learned, its representation is "really" discrete rather than "seemingly" or "relatively" discrete, since the brain waves that occur during familiar speech or walking fluctuate continuously through time across billions of cells (Donchin & Lindsley, 1969). Moreover the sound spectrogram of familiar speech is an almost continuous flow of sound despite our impression that it is a series of discrete words (Lenneberg, 1967). In fact, an unfamiliar foreign language does sound like an almost continuous flow of sound. The process of learning makes the sounds seem to be discrete by perceptually grouping them into learned units. Thus if one accepts a binary representation of familiar events by fiat, then one must in principle miss vital ingredients of the learning process. In effects, the consensual impression of the event then blinds us to its functional representation.

Similar considerations make it clear that even in tasks that appear serial, such as serial learning, important underlying control processes are parallel processes. For example, the code that controls performance after a serial list is repetitively practiced are not just the individual list items. As Young (1959, 1961, 1963, 1968) noted, if they were, prior

serial practice of a list *A-B-C-D-...* should yield marked positive transfer for later learning of the paired associates *A-B*, *C-D*, ..., but it does not. Horowitz and Izawa (1963) suggested that more than one item can be the functional stimulus for a given response in a serial list, in particular that several items preceding the response serve as its functional stimulus. This viewpoint illustrates the familiar idea that a series of items can be chunked together (Miller, 1956) to form a new code whereby a series of behaviors can be more efficiently performed. Such a chunking process is based on the simultaneous availability of all the individual units, and is thus a parallel process.

2. BOWED SERIAL POSITION CURVE IN FREE RECALL

A basic datum about STM is the bowed serial position curve that is found in free recall experiments (Fig. 1). When a subject repeats a sufficiently long standard list of items in any order after hearing it once, the items near the beginning and end of the list are performed earliest and with the highest probability (Atkinson & Shiffrin, 1971). The advantage of items near the list beginning is called a *primacy* effect, that of items near the list end is called a *recency* effect. A computer analogy to explain these effects can be developed as follows; cf. Atkinson and Shiffrin (1968). Let a list r_1, r_2, \dots, r_n of behaviorally matched items be presented to a subject. It is supposed that each item is either in an STM buffer, or is not in the buffer, at a given time. That is, assume that a binary code exists such that 1 is assigned to r_i if r_i is in the buffer, and 0 is assigned to r_i if r_i is not in the buffer. If $k > 1$ items are in the buffer at time t , one cannot determine the order in which they entered the buffer by looking at their 0's and 1's, since all k items that are in the buffer have the value 1. Thus a binary code carries no order information. If there did not exist any internal trace of the order in which items occurred, there would be no way to encode this information in LTM. Given a binary code, some mechanism other than an item's activity (0 or 1) is needed to code order information. A serial STM buffer is therefore assumed to exist. Suppose that this buffer contains m serially organized slots s_1, s_2, \dots, s_m . The first item r_1 enters s_1 . When r_2 occurs, it enters s_1 and displaces r_1 from s_1 to s_2 . Then r_3 displaces r_2 from s_1 to s_2 , and r_1 from s_2 to s_3 .

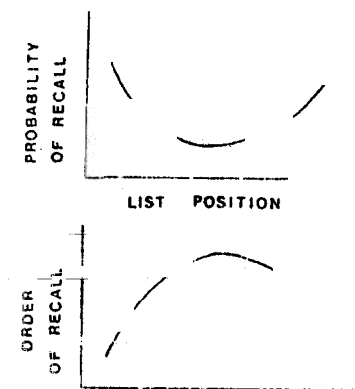


FIG. 1. Probability and order of recall in free recall experiment.

One can then tell which items occurred first by testing their relative positions in the buffer. The above process continues until m items are presented. Item r_{m+1} knocks r_1 out of the buffer from slot s_m , and each successive item eliminates the earliest remaining item from the buffer at slot s_m . In all, at any time there will be a block of successive items in the buffer, each with activity 1. Thus, given a binary code, a serially organized STM buffer is needed to store order information.

The binary buffer concept does not, however, explain the data in Fig. 1. If the buffer worked in a deterministic fashion, then each item could be perfectly performed, and each item would be performed in its correct order. The two bowed curves in Fig. 1 would be replaced by a horizontal line and an increasing straight line, respectively. Consequently, the buffer cannot work in a deterministic fashion. It must work probabilistically, if it exists at all. Introducing probabilities brings continuous variables back into the model, and creates a hybrid mixture of computer and probabilistic ideas. To explain the recency effect, this hybrid model makes two more related assumptions, both of which say that the buffer works badly in a prescribed way. First, one says that the buffer is *leaky* in the sense that an item can fall out of the buffer even before it reaches s_m . Since the probability of falling out increases as a function of how long an item is in the buffer, this makes it most probable that the most recent items are still in the buffer (Fig. 2). A recency effect for the probability of being in the buffer is achieved by averaging across subjects. In each subject, however, items that remain in the buffer all still have activity 1. Probabilistic models of STM usually stop at this point. They fail to ask a crucial question whose answer casts doubt on the binary code assumption. How is the probability distribution of being in the STM buffer translated into the real-time performance of individual items?

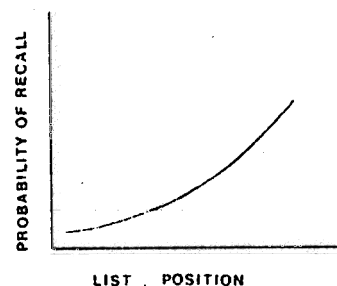


FIG. 2. Recency effect is ascribed to lesser probability of remaining in a serial STM buffer as a function of duration in the buffer.

Given the above framework, it follows that *any* item r_i can be performed from *any* buffer position s_j in order to derive a recency effect from subject performance. Otherwise there would be 0 probability of performance of r_i from any s_j from which r_i could not be elicited.

This latter assumption implies that the motor code for performing any item can be read out from any buffer position. This imposes extraordinary demands on system design, and makes a definite statement about neural dynamics. It says that the motor codes for eliciting an item, although not initially learned at all positions s_j , can be per-

formed at all positions s_j . The code is *shift-invariant*. An even more demanding implication is that all codes that are hierarchically built up from the STM buffer must also be shift-invariant. To predict the recency effect from a binary code model, one is hereby led to conclude that the buffer is so poorly designed that it leaks, but is so exquisitely designed that its entire hierarchy of codes is shift-invariant. Usually this implication is ignored, possibly because it is so disturbing, and possibly because probabilistic modelers often overlook real-time constraints on performing individual items when they construct their models.

Given the considerable machinery that is needed to produce the recency effect using a binary code, we ask whether it is necessary? In particular, the enormous amount of neural data on continuously fluctuating potentials, spiking frequencies, and the like, leads one to question the binary assumption itself. If the binary code is abandoned, then all of the above difficulties evaporate. Items r_i can then have fixed internal representations v_i that are innate or built up by learning; their codes need not move around in a buffer. Thus there need be no shift-invariant code. Moreover, the v_i need not be leaky, and each subject can possess a recency gradient, rather than assuming the recency gradient is a statistical property of a pool of subjects, as in a binary theory.

In a continuous theory, a recency gradient exists if the most recent items have the greatest STM activity in their representations v_i , say because their STM activity has had less opportunity to decay, either spontaneously or due to interference. If greater STM activities translate into faster reaction times of item performance, then a recency effect in performance can be achieved without the need to move items around in a serially organized buffer. In other words, if continuous STM activities exist, then they already carry order information. Below we suggest some STM interactions that can occur at a single level of input processing. Section 7 discusses how LTM feedback from a higher network level can modify these properties. The paper Grossberg (1977a) develops a more complete theory in which several levels of STM and LTM processing are needed to self-organize complex behavioral codes, maps, and plans.

How does a continuous mechanism work? Suppose for definiteness that each item r_i has an internal representation v_i with STM activity x_i . If the most recently presented items have the largest STM traces, and if r_i is the last item to have occurred, then $x_1 < x_2 < x_3 < \dots < x_i$. The storage of these STM activities must be distinguished from their overt rehearsal. How does rehearsal translate differential STM activity into a prescribed order of performance, in particular the order $r_i, r_{i-1}, r_{i-2}, \dots, r_1$, in the case when $x_1 < x_2 < x_3 < \dots < x_i$? In many neural examples, a nonspecific rehearsal, or arousal, wave can accomplish this. Such a mechanism simultaneously amplifies all STM activities so that they can exceed an output threshold, or alternatively lowers the output threshold until it is exceeded by the STM activities (Fig. 3). The largest STM activity x_i exceeds the output threshold first, and thereby elicits the fastest output signal. This output signal controls performance of item r_i . If this signal was not self-terminating, then perseverative performance of item r_i would occur. Under normal circumstances, the output signal generates feedback inhibition that self-inhibits, or resets, its STM activity. Then the state v_{i-1} is most active, so that its output signal can elicit performance of item r_{i-1} . This process of STM arousal and reset continues until all of the items are

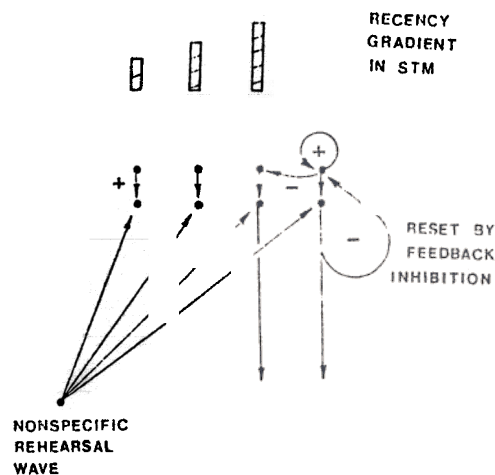


FIG. 3. Readout of STM by a nonspecific rehearsal wave, and reset of STM by feedback inhibition.

performed in the order r_i, r_{i-1}, \dots, r_1 . The role of nonspecific arousal and reset as rehearsal mechanisms are further discussed in Grossberg (1977a). This paper also describes how a list can be grouped into parts which can all be performed in their correct order.

In the above example, the nonspecific rehearsal mechanism is a parallel operation that simultaneously influences all representations v_i , despite the fact that item performance is serial. Serial properties do not imply serial mechanisms.

3. PRIMACY EFFECT

The binary model also implies questionable conclusions to explain the primacy effect of Fig. 1. The binary model assumes that a further operation is activated by items while they are in the STM buffer. This operation is described as coding the item in LTM. Although the binary model does not describe the LTM coding mechanism, clearly more coding can occur the longer an item is in the buffer, other things equal. Consequently the earliest items can be coded in LTM better than more recent items. This LTM process is claimed to produce the primacy effect. Thus early items produce a primacy effect via LTM, whereas late items produce a recency effect via STM.

Several types of evidence are compatible with this view. For example, if STM storage is inhibited by an interfering task, then the primacy effect remains but the recency effect is obliterated (Atkinson & Shiffrin, 1971). The similarity of primacy effects with or without interference is a main source of the belief that STM does not contribute to primacy. This is only indirect evidence, however, and it is argued in Section 7 that competitive interactions acting in parallel across internal representations can effectively mask any STM primacy effects that might exist, leaving the impression that only LTM influences performance. Other experiments are based on the premise that rehearsal should strengthen

LTM, so that a good correlation between the number of rehearsals at different list positions and the size of the LTM contribution to recall at these positions should argue for LTM as a basis for the primacy effect (Rundus, 1971). This argument does not help if STM primacy effects are masked. Moreover, it has been shown that the size of the LTM effect can depend on the type of rehearsal (maintenance vs elaborative), and on whether performance is measured by recognition or recall (Craig & Watkins, 1973; Woodward, Bjork, & Jongeward, 1973). In Grossberg (1977a, Sects. 31 and 47) a coding mechanism is described wherein mere repetition of items, improved recognition, and improved recall are distinguished. Improved recognition can result from new code formation and sustained STM activity of these new codes even when individual item codes are rapidly reset in STM. Improved recall can result from the formation of new motor associations using the new codes as sampling sources. Because the sampling sources must be synthesized before the motor associations can be learned, recognition often improves before recall does.

Using the LTM primacy hypothesis, the binary model can fit some of the interference and rehearsal data, but is also forced into the counterintuitive idea that items near the beginning of the list can only be performed in their correct order by being read out of LTM. This idea overlooks the fact that a telephone number can be perfectly repeated immediately after hearing it, yet it could have been obliterated from memory by a distracting event before performance occurred, so presumably was not stored in LTM. Indeed, amnesic patients with Korsakoff syndrome have no LTM capability, but exhibit essentially normal digit span performance (Baddeley & Warrington, 1970; Milner, 1956). These examples question whether LTM is *necessary* to produce a primacy effect. The data used to support the LTM contribution to the primacy effect do not disprove that STM also contributes to primacy, and sometimes without a large LTM contribution.

4. STM PRIMACY EFFECT

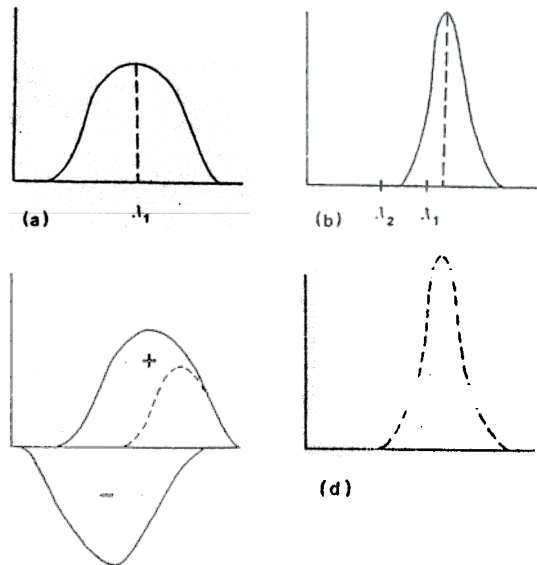
How can performance of a telephone number due to STM but not LTM be achieved? This is easy in the continuous model if the earliest items have the largest STM activities, since these items will be performed first when the STM field is amplified by a nonspecific rehearsal wave. However, if the earliest items have the largest STM strengths in a short list, then how can the most recent items also have large STM strengths in longer lists to produce a bowed STM pattern across list representations, and thus the bowed order of recall in Fig. 1? We will illustrate below how an STM primacy effect can be generated in short lists, but becomes an STM bow in longer lists, such that the STM recency effect becomes progressively stronger as list length increases. The list length at which the bow appears is called the transient memory span (TMS). The TMS can depend on such factors as a subject's attentional and motivational state, but it can be proved that the TMS is no longer than the more familiar immediate memory span (Miller, 1956) under rather general circumstances (Grossberg, 1977a, Sect. 32).

Perhaps the belief that the STM activities of earlier items should always be weaker supported the idea that only LTM can ever generate a primacy effect, despite its unfortunate implications for the immediate recall of short lists. To counter this belief, we

note that an STM primacy effect has already been found in some free recall data. Baddeley and Warrington (1970) study amnesic Korsakoff patients whose STM is intact, but whose LTM is nonfunctional. In free recall tasks, these patients produce a bowed probability of recall curve that is due to STM alone. Hogan and Hogan (1975) theoretically disentangle STM and LTM contributions in their free recall data for normal subjects, and find an STM primacy effect which they mention without mechanistic interpretation. Furthermore, we will suggest that STM primacy is a temporal analog of a phenomenon which is more familiar experimentally, but which until recently was theoretically paradoxical; namely, behavioral contrast in discrimination learning experiments (Bloomfield, 1969).

5. BEHAVIORAL CONTRAST IN SPACE

A typical example of behavioral contrast is this. If a pigeon is rewarded on errorless discrimination trials for pecking on a key illuminated by a light of wavelength λ , then during extinction trials, when the pigeon is allowed to peck in responses to keys illuminated by various wavelengths, a generalization gradient of pecks centered at λ is generated (Fig. 4a). By contrast, if the pigeon is rewarded for pecking a key illuminated at wavelength λ_1 , and punished for pecking at a nearby wavelength $\lambda_2 < \lambda_1$, then during extinction, the pigeon pecks most vigorously at wavelength $\lambda_3 > \lambda_1$ (i.e., a peak shift occurs). Remarkably, the pigeon pecks λ_3 more vigorously than it would have pecked λ_1 if λ_2 had not occurred (Fig. 4b); that is, behavioral contrast occurs. Behavioral contrast is



paradoxical because the punishing λ_2 causes the pigeon to peck at the unrewarded λ_3 more than it would have pecked at the rewarded λ_1 in the absence of λ_2 . The difficulty in explaining behavioral contrast is this: Suppose that reward at λ_1 generates a positive generalization gradient centered at λ_1 , and punishment at λ_2 generates a negative generalization gradient centered at λ_2 . If performance at λ_3 is due to the net gradient, then a peak shift will occur, but pecking at λ_3 should be less vigorous than pecking at λ_2 (Fig. 4c). What then causes behavioral contrast?

Grossberg (1975) suggests that behavioral contrast follows from a property of cell populations that undergo mass action interactions in recurrent on-center off-surround anatomies (Fig. 5). Grossberg (1973) derives networks of this type as a solution to a

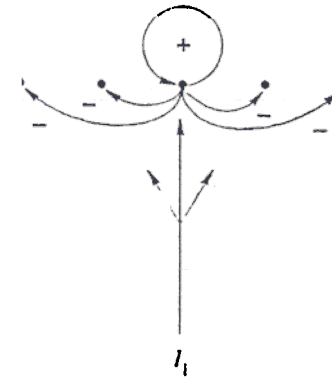


FIG. 5. Recurrent on-center off-surround anatomy can suppress noise, contrast enhance suprathreshold patterns, and store them in STM without saturation.

universal problem concerning how patterned information can be processed by noisy cells with finitely many excitable sites; cf., Levine and Grossberg (1976) or Grossberg (1977b) for a review. In the present context, the populations are sensitive to different hues. The on-center (excitatory feedback) defines an excitatory generalization gradient to nearby hues, and makes possible STM storage after external inputs cease. The off-surround (inhibitory feedback) maintains network sensitivity to relative input sizes even in response to large inputs; otherwise expressed, automatic gain control, driven by the off-surround, prevents cell saturation by adapting network responses to different background activity levels. The mass action laws reduce to the familiar equations (Hodgkin, 1964; Katz, 1966)

$$C \frac{\partial V}{\partial t} = (V_+ - V)g_+ + (V_- - V)g_- + (V_p - V)g_p \quad (1)$$

designates a passive channel. The convention is also accepted that depolarization makes $V(t)$ more positive.

It has been proved that such networks tend to conserve the total potential of all cells in the network. This property is called *normalization*, and is a form of network adaptation due to automatic gain control by its off-surround. In Fig. 4c, the net gradient is narrower than the gradient in Fig. 4a. The total potential, or area under the curve in Fig. 4a, is approximately conserved when the excitatory gradient is replaced by the narrower net gradient. Normalization therefore amplifies the net gradient to produce the higher and steeper normalized net gradient of Fig. 4d. Thus behavioral contrast can be explained as the result of a net gradient normalized by a network that is capable of storing cues in STM without saturation.

The simplest example of this phenomenon occurs in feedforward networks (Fig. 6).

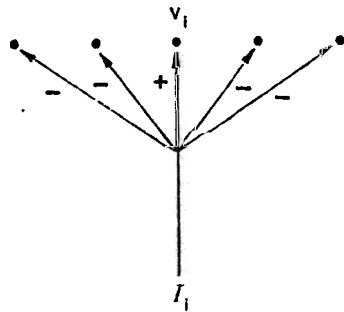


FIG. 6. Nonrecurrent on-center off-surround anatomy.

Let n cells (or cell populations) v_i , $i = 1, 2, \dots, n$, be given, and let $x_i(t)$ be the potential of v_i . In (1), replace $V(t)$ by $x_i(t)$ and choose the constant parameters $C = 1$, $V_+ = B > 0$, $V_- = V_p = 0$, and $g_p = A > 0$ for simplicity. The conductance g_+ is influenced by an on-center input $g_+ = I_i$ and the conductance g_- is influenced by an off-surround input $g_- = \sum_{k \neq i} I_k$. Then at the cell v_i , (1) becomes

$$\frac{d}{dt} x_i = -A x_i + (B - x_i) I_i - x_i \sum_{k \neq i} I_k. \quad (2)$$

If the inputs remain constant for awhile, then the system approaches equilibrium. At equilibrium $(d/dt) x_i = 0$ so that the equilibrium potential of v_i is

$$x_i = B I_i / (A + I_i), \quad (3)$$

where $I = \sum_{k=1}^n I_k$. Letting $\theta_i = I_i I^{-1}$ be the relative input to v_i , (3) can be written as

$$x_i = \theta_i B I / (A + I), \quad (4)$$

which shows that each v_i retains its sensitivity to θ_i even as I is parametrically increased. The dependence of x_i on the ratio θ_i is a form of adaptation to changing background

activity levels. Without such adaptation, each x_i would rapidly saturate at B as I_i increased. Ratio theories have often been suggested by perceptual or learning data. For example, Zeiler (1963) developed an adaptation-level theory in which the subject's perception of a stimulus depends on its ratio with respect to an internal norm, or adaptation level. Luce (1959) developed a theory in which choice behavior depends on the ratio of two reaction tendencies. Color theories are often based on ratios that represent the reflectances of external objects (Cornsweet, 1970). I suggest that ratios appear in such a great variety of situations to deal with the ubiquitous saturation problem. Even in the simplest case of Eq. (4), however, the ratio influence is modified by a term $B I (A + I)^{-1}$ which is of Weber-Fechner form (Cornsweet, 1970, p. 249). In other examples of on-center off-surround networks, only ratios above an adaptation level can cause positive x_i values (Grossberg, 1977c), or there can be complicated hysteresis, normative, decision, and related types of behavior (Grossberg, 1977a, b).

Wherever ratios appear in individual activities x_i , the total activity $x = \sum_{k=1}^n x_k$ obeys a normalization rule. In (4), $x = \sum_{k=1}^n x_k = B I (A + I)^{-1}$ is always less than B . It thus has an upper bound that is independent of the number n of cells and the total input I . This is normalization in a feedforward network. In a recurrent, or feedback, network the normalization property is strengthened. The normalized inequality $x \leq B$ is replaced by normalized equalities.

The law (4) has another important property; namely, x_i , plotted as a function of the logarithm of its on-center $K = \ln I_i$ and its off-surround $L = \sum_{k \neq i} I_k$ obeys

$$x_i(K, L) = B e^K / (A + e^K + L). \quad (5)$$

Thus, if the off-surround input is shifted from $L = L_1$ to $L = L_2$, the whole curve (5) is shifted by an amount $S = \ln[(A + L_1)(A + L_2)^{-1}]$, since

$$x_i(K + S, L_1) = x_i(K, L_2) \quad \text{for all } K \geq 0. \quad (6)$$

A similar shift occurs, for example, in bipolar cell responses in the mudpuppy retina (Werblin, 1971; Grossberg, 1977b, c). The shift relocates where x_i is most sensitive.

The above properties are summarized herein to emphasize three points. First, in the laws (1) and (2), inputs exercise their effects by *multiplying* potentials. Hence these laws are called *shunting* laws. Additive models cannot generate these effects. Second, in (2), all cells v_k , $k \neq i$, inhibit v_i with equal strength. If this is not true, say because inhibitory interactions become weaker as the distance between populations increases, then the normalization effect becomes partial, and the total potential can grow to a finite asymptote as more populations are excited; that is, the adaptation effect is only partial, and saturation starts to set in as the background input becomes large. Third, the normalization rule helps to clarify from a neurophysiological perspective why probability axioms often model behavioral data so well. The normalization rule plays the role of summing all the probabilities to 1, and the shunting laws play the role of multiplying the probabilities of independent events. However, even in Eq. (4), ratios do not appear alone, and the normalization rule can often hold only partially, as we will see below.

6. BEHAVIORAL CONTRAST IN TIME

Bowing in STM will now be explained as a behavioral contrast effect that evolves as items are presented in time, rather than across space. Before developing the ideas in general, consider the simplest example as an illustration. Suppose that total activity is normalized. Set it equal to 1, for definiteness, when some item is active in STM. Also suppose that when a new item occurs, the old item's STM activity is reduced by a multiplicative factor w due to shunting inhibition. When item r_1 occurs, its activity x_1 equals 1. When item r_2 occurs, x_1 is changed to w . By normalization, $x_1 = w$ and $x_2 = 1 - w$. If $w > \frac{1}{2}$, then $x_1 > x_2$; that is, an STM primacy effect occurs. A large value of w means that the reverberating STM activity x_1 of v_1 can substantially inhibit v_2 when v_2 is receiving an input due to presentation of r_2 . When r_3 occurs, the old STM activities are again multiplied by w , so that $x_1 = w^2$ and $x_2 = w(1 - w)$. By normalization, $x_3 = 1 - w$. Note that $x_1 > x_2$ and $x_3 > x_2$. A bow has occurred at v_2 . As new items r_i are presented, $i > 3$, the bow remains at position v_2 , but a pronounced recency effect develops due to normalization. In particular, given any list of length $k > 1$, the last item to enter STM always has STM activity $1 - w$. Below we will show that a bow can arise at any list position if network parameters are properly chosen.

The behavioral contrast mechanism can be derived from three concepts. The first concept is operationally described by saying that new items change the STM activities of old items by a multiplicative factor. This mechanism is the simplest rule for making rigorous the idea that shunting interactions join the network populations together. There exists a deeper justification for using the multiplicative rule. Grossberg (1977a) develops a theory of neural coding, in which it is shown (Sect. 25) that the multiplicative rule leaves invariant the codes of old items as new items occur and activate new codes. This concept is needed to prevent each new item from destabilizing the internal representations of all the old items. It says that new items do not deny the fact that old items occurred, even if they alter their importance, or even totally inhibit them. The rule is therefore called a *Principle of Code Invariance*. The theory hereby establishes a conceptual bridge between statements about STM interactions—via laws describing cellular potentials and signals—and statements about LTM interactions—via laws describing cellular learning. This bridge shows in a precise formal way how each type of law is adapted to the needs of the other.

The multiplicative operations of the Invariance Principle have the following effect. Let item r_i enter the network with STM activity u_i . Let the i th item multiplicatively modify the STM activity of all previous items r_1, r_2, \dots, r_{i-1} by a factor w_i . Suppose r_1 enters with weight $x_1 = u_1$. After r_2 occurs, $x_1 = w_2 u_1$ and $x_2 = u_2$. After r_3 occurs, $x_1 = w_3 w_2 u_1$, $x_2 = w_3 u_2$, and $x_3 = u_3$. And so on. The total STM strength S_i after r_i occurs is thus

$$S_i = \sum_{m=1}^i \prod_{r=m+1}^i w_r u_m. \quad (7)$$

The second hypothesis is the *Normalization Rule*. This says that total STM strength grows in a negatively accelerated way from a minimum of u_1 , when only r_1 occurs, to

some finite maximum M . The case $u_1 = M$ characterizes complete normalization. An analogous experimental phenomenon is that pupil diameter increases in a negatively accelerated way as a function of the number of items presented to a subject (Kahneman & Beatty, 1966), and one might try to use this paradigm to estimate u_i and M in particular cases. Mathematically, the Normalization Rule says that

$$S_i = u_i \theta^{i-1} + M(1 - \theta^{i-1}), \quad (8)$$

where $M \geq u_1 \geq 0$ and $0 \leq \theta \leq 1$. The parameters u_1 , M , and θ can depend on the geometry of the network as well as on attentional and motivational factors that can retune network interactions (Grossberg, 1976a, b). Our goal is to solve for the weights w_k , $k \geq 2$, in terms of the parameters θ , M , and u_i , $i = 1, 2, \dots$. That this can be done is summarized in the next statement.

(I) Suppose that the Invariance Principle and the Normalization Rule both hold; that is, let both (7) and (8) hold. Then the shunting parameters can be explicitly determined. They are

$$w_k = \frac{u_1 \theta^{k-1} + M(1 - \theta^{k-1}) - u_k}{u_1 \theta^{k-2} + M(1 - \theta^{k-2})}, \quad k > 1. \quad (9)$$

Since the STM activity of v_i after item r_i occurs is, by the Invariance Principle,

$$x_i = u_i \prod_{k=i+1}^j w_k, \quad i \leq j, \quad (10)$$

the STM code can be completely solved by specifying the STM weight u_i of the most recent item r_i , $i = 1, 2, \dots$. We assume that u_i estimates how much attention is paid to item r_i as it is presented. Then (9) and (10) show that, once u_1 , M , and θ estimate the code geometry, and (presumably constant) performance variables, it suffices to specify how much attention is paid to each item as it is presented.

The following result also holds if each u_i depends only on r_i and possibly events r_1, r_2, \dots, r_{i-1} that have preceded it.

(II) Suppose that the Invariance Principle holds. If an STM bow occurs at position J in a list of length K , then it also occurs at position J given a list of any length $k \geq K$. This strong property notes that the factors w_k in (10) change the relative sizes of past STM strengths, but not where local maxima or minima occur in the STM pattern across old items. It is this property that allows us to define a TMS for lists composed of matched items presented under fixed performance conditions.

To derive further information about the code, we now impose some natural constraints on the u_i . These constraints do not hold if attentional conditions vary in an arbitrary fashion as new items are presented. They summarize in mathematical terms various stable attentional conditions. Intuition suggests that if we pay equal attention to each item as it is presented, then $u_1 \geq u_2 \geq u_3 \geq \dots$ (equality might be destroyed by negative feedback acting on later items) and that the u_i equilibrate at some positive value u_∞ as longer lists are used. This idea simply says that the STM strength of the last item should

always be positive, if one always pays attention to the most recent item, and can only get weaker as i increases due to greater total inhibition from the larger numbers of items that are already in STM. For example, if all $u_i = u > 0$, then a fixed amount of STM strength is always given to the last item. In this case, if $M > u$, it can be proved that the bow in STM occurs at item r_j , where J is the maximal index j such that

$$(u^{-1}M - 1)(1 - \theta)^{j-2} > 1, \quad (11)$$

Thus a bow can occur at any list position if u , M , and θ are suitably chosen.

Another plausible rule for the u_i sequence is: set $u_1 = u$ and $u_k = u_\infty$, $k \geq 2$, where $u_\infty < u$. In other words, if r_1 occurs alone, then no inhibition occurs since no others items are reverberating in STM to supply it. Once more than one item occurs, the most recent item always has a fixed amount of STM strength due to a balancing out of excitation and inhibition across all items. This occurs in the special case where $u_1 = M$ and all $w_i = w < 1$. Both of these rules are a special case of the rule that

$$u_i = u_1 \phi^{i-1} + u_\infty (1 - \phi^{i-1}), \quad (12)$$

$i \geq 1$, with $u_1 \geq u_\infty > 0$ and $0 \leq \phi \leq 1$. In other words, u_i is negatively accelerated function of i . Choosing $\phi = 1$ yields the rule $u_i = u_1$, $i \geq 1$; choosing $\phi = 0$ yields the rule $u_1 > u_k = u_\infty$, $k \geq 2$. The analysis can be generalized further by letting a twice-differentiable function $u(t)$, $t \geq 1$, interpolate the sequence u_1, u_2, \dots ; that is, let $u(k) = u_k$, $k \geq 1$, and by placing hypotheses on $u(t)$. A natural generalization of (12) is

or that $u(t)$ is a nonincreasing logarithmically convex function. Using these constraints on the u_i , we can prove the following general statements. The existence of a limiting u_∞ is sufficient to prove that a recency effect always occurs in sufficiently long lists:

(III) If the Invariance Principle and Normalization Rule both hold, and u_∞ exists, then in all sufficiently long lists, a recency effect develops. This follows from (9) and (10). By (10), a recency effect develops if the function

$$G(j) = w_j u_{j-1} - u_j \quad (14)$$

becomes negative as j becomes sufficiently large. To prove this using (9) and (10), one shows that $G(j) < 0$ is equivalent to

$$B_j - B_{j-1} < 1 + u_{j-1}^{-1}, \quad (15)$$

where $B_j = u_j^{-1} S_j$. Since $\lim_{j \rightarrow \infty} B_j = u_\infty^{-1} M$ exists, the left-hand side of (15) approaches 0 as $j \rightarrow \infty$ while the right-hand side exceeds 1.

Properties (II) and (III) indicate the tendency for the STM primacy effect, if it exists, to become weaker as list length increases. This is because more and more of the normalized total STM activity S_i ($\leq M$) gets devoted to the recency effect as $i \rightarrow \infty$.

Special hypotheses on $u(t)$ are needed only to show that no more than one bow occurs in the STM patterns (x_1, x_2, \dots, x_L) no matter how long the list length L is chosen. These hypotheses constrain the differential amounts of attention that can be paid to items without creating more than one bow. Without such constraints, a unimodal STM bow is not guaranteed, and this fact is experimentally important, as will be illustrated in Section 8. A sufficient condition that an STM bow is unimodal, if it occurs, is given by the hypotheses (13):

(IV) If the Invariance Principle and Normalization Rule both hold, and $u(t)$ satisfies (13), then all the STM patterns (x_1, x_2, \dots, x_L) are either increasing, decreasing, or possess a unimodal bow. To prove that the bow is unimodal, function $G(j)$ is extended to a function $G(t)$ of a continuous variable $t \geq 1$. It is then verified using (7), (8), and (13) that if $G(T) = 0$ then $(d/dt) G(T) \leq 0$; that is, once the primacy effect becomes a recency effect, it can never flip back to a primacy effect.

To assert that a unimodal bow occurs if (7), (8), and (13) hold, it therefore suffices to guarantee that a primacy effect exists. By (10), this is true if and only if

$$G(2) = w_2 u_1 - u_2 > 0. \quad (16)$$

If (9) and (12) hold, (16) is true if and only if

$$u_1 \theta + M(1 - \theta) > 2[u_1 \phi + u_\infty (1 - \phi)], \quad (17)$$

If (16) holds, the bow occurs at that index $j = J$ where $G(j) = w_j u_{j-1} - u_j$ changes sign from positive to negative.

7. MASKING OF STM PRIMACY BY NORMALIZATION

The above sections note a behavioral rationale, a physiological mechanism, and some data in which an STM primacy effect is implicated. Why then is there so little evidence of STM primacy in interference experiments, wherein the primacy effect is little changed before or after interference with STM, yet the recency effect is almost entirely obliterated? One factor is that the STM primacy effect becomes smaller as the list length increases. I suggest, however, that even when a large STM primacy effect exists, it can be masked due to normalization.

To understand how this can happen, a brief review of how LTM is encoded in the present framework is needed. A psychophysiological theory of LTM encoding in response to a list of items is successively developed in Grossberg (1969), Grossberg and Pepe (1971), Grossberg (1974), and Grossberg (1977a). Many of these results were applied to the study of serial learning and paired associate learning, but they are readily adapted to the free recall paradigm. To fix ideas, suppose that two fields $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ of populations $v_i^{(1)}$ and $v_j^{(2)}$, respectively, are given. Suppose that the populations $\mathcal{F}^{(1)}$ can send signals to

the populations $\mathcal{F}^{(2)}$ over directed pathways, or axons. Let an LTM trace z_{ij} be computed at the end of the pathway from $v_i^{(1)}$ to $v_j^{(2)}$. Assume that z_{ij} obeys the equation

$$(d/dt)z_{ij} = -C_{ij}z_{ij} + S_{ij}(x_i^{(1)})x_j^{(2)}, \quad (15)$$

where C_{ij} is the LTM decay rate, $S_{ij}(x_i^{(1)})$ is the sampling signal from $v_i^{(1)}$ to $v_j^{(2)}$, $x_i^{(1)}$ is the STM trace of $v_i^{(1)}$, and $x_j^{(2)}$ is the STM trace of $v_j^{(2)}$ (Fig. 7). If $x_i^{(1)}$ is sufficiently large to make $S_{ij}(x_i^{(1)}) > 0$, then z_{ij} can sample the STM trace $x_j^{(2)}$. Thus LTM in the model depends on a nonlinear mechanism that time-averages (via the term $-C_{ij}z_{ij}$) the products of sampling signals and STM traces (via the term $S_{ij}(x_i^{(1)})x_j^{(2)}$). Each $v_i^{(1)}$ controls all the LTM traces $z_{i1}, z_{i2}, z_{i3}, \dots$ via its sampling signals $S_{i1}(x_i^{(1)}), S_{i2}(x_i^{(1)}), S_{i3}(x_i^{(1)}), \dots$. We therefore say that $v_i^{(1)}$ controls the LTM pattern $z_i = (z_{i1}, z_{i2}, z_{i3}, \dots)$. All of the LTM patterns z_i can be different. These patterns are generated, via the products $S_{ij}(x_i^{(1)})x_j^{(2)}$, by the distributions of STM activity that evolve across $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ through time. The experimental paradigms of serial learning, paired associate learning, and free recall can all produce different STM patterns, and hence different LTM patterns.

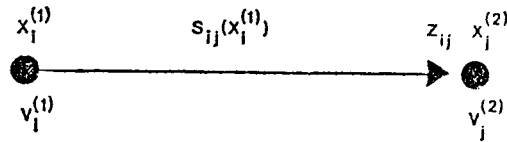


FIG. 7. The signal $S_{ij}(x_i^{(1)})$ from $v_i^{(1)}$ is gated by z_{ij} on its way to $v_j^{(2)}$.

How is the LTM pattern z_i read out to influence performance at some population $v_j^{(2)}$? This can only happen if $v_i^{(1)}$ is activated enough to elicit a sampling signal $S_{ij}(x_i^{(1)})$. This activation can, for example, be due to a probe stimulus that excites the STM trace $x_i^{(1)}$ of $v_i^{(1)}$, or to a lingering STM activity $x_i^{(1)}$ that is due to prior stimuli. Different experimental paradigms can generate different sequences of probes and hence different performance characteristics.

Readout from the LTM trace z_{ij} occurs when the sampling signal $S_{ij}(x_i^{(1)})$ from $v_i^{(1)}$ is gated by z_{ij} on its way to $v_j^{(2)}$. The net signal to $v_j^{(2)}$ is $S_{ij}(x_i^{(1)})z_{ij}$. This gating mechanism is also a nonlinear process. Every active population in $\mathcal{F}^{(1)}$ can produce such a gated signal to $v_j^{(2)}$. The total effect of $\mathcal{F}^{(1)}$ on $v_j^{(2)}$ is given by the sum of these gated signals; namely, by

$$T_j = \sum_i S_{ij}(x_i^{(1)})z_{ij}. \quad (19)$$

Thus LTM can affect $\mathcal{F}^{(2)}$ only if it is activated by STM-driven signals from $\mathcal{F}^{(1)}$.

The signal T_j in (19) influences $v_j^{(2)}$ by activating its STM trace $x_j^{(2)}$. Such a signal is received by every $v_j^{(2)}$, and hereby generates a pattern (T_1, T_2, T_3, \dots) of inputs across $\mathcal{F}^{(2)}$. Competitive interactions within $\mathcal{F}^{(2)}$ transform this input pattern to produce the final STM pattern $(x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, \dots)$ across $\mathcal{F}^{(2)}$. Normalization of STM within $\mathcal{F}^{(2)}$ is a particular consequence of these competitive interactions. After the populations in $\mathcal{F}^{(2)}$ compete for STM activity, the effects of this competition feed back into each LTM

pattern due to the STM term $x_j^{(2)}$ in (18). The STM competition hereby tends to produce ratio scales in LTM as well as in STM; cf., Eq. (4). This (approximate) LTM ratio scale helps to explain competitive retrieval rules in free recall experiments (Rundus, 1973). Note also that the sampling signal $S_{ij}(x_i^{(1)})$ both controls performance, via (19), and strengthens learning, via (18). This helps to physiologically explain how test trials can act as training trials (Lachman & Laughery, 1968; Tulving, 1967), since making $S_{ij}(x_i^{(1)})$ large enough to elicit performance also makes it large enough to strengthen the LTM trace z_{ij} . In more general physiological models, unbiased simultaneous sampling by many cues of the same event is impossible unless the performance signal is large only if the learning signal is also large. This constraint is called the *local flow* condition (Grossberg, 1972; 1974, Sect. VI).

What types of LTM patterns can evolve? Suppose for simplicity that each $v_i^{(1)}$ sends the same sampling signal $S_{ij}(x_i^{(1)})$ to all cells in $\mathcal{F}^{(2)}$. Also suppose that the STM patterns across $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ exhibit either an STM recency gradient or an STM bow due to presentation of the list $r_1, r_2, r_3, \dots, r_L$. Then the pattern z_1 learns an LTM primacy gradient; namely, $z_{11} > z_{12} > z_{13} > \dots$ (The proof is in Grossberg (1977a, Sect. 32)). If a probe stimulus excites $v_1^{(1)}$ on performance trials, then the signals from $v_1^{(1)}$ to $\mathcal{F}^{(2)}$ satisfy $S_{11}(x_1^{(1)})z_{11} > S_{12}(x_1^{(1)})z_{12} > S_{13}(x_1^{(1)})z_{13} > \dots$. These signals elicit an STM primacy effect across $\mathcal{F}^{(2)}$. Thus, even if $v_1^{(1)}$ samples an STM recency gradient across $\mathcal{F}^{(2)}$ on learning trials, it can perform an STM primacy gradient across $\mathcal{F}^{(2)}$ on performance trials. This is due to the nonlinear nature of Eqs. (18) and (19). The other populations $v_i^{(1)}$, $i \neq 1$, usually do not learn an LTM primacy gradient. For example, if $v_i^{(1)}$ and $v_1^{(2)}$ are simultaneously excited, $i = 1, 2, \dots, L$, and both experience STM recency gradients, then the maximum LTM trace in pattern z_i is z_{ii} , and the other traces z_{ij} decrease as a function of $|i - j|$ to produce a generalization gradient that is centered at $v_i^{(2)}$. Moreover, if the STM recency gradients are inhibited by interfering activities right after the last list item is presented, then $z_{11} > z_{22} > z_{33} > \dots$. Most LTM storage is therefore concentrated at the populations that are excited by the beginning of the list, especially in the LTM primacy gradient controlled by $v_1^{(2)}$. (This is not true in serial learning, where a bow in the LTM pattern of correct associations can occur.)

Given the above summary, we can now see how an STM primacy gradient can be masked by normalization. To fix ideas, suppose that there exist three fields $\mathcal{F}^{(1)}$, $\mathcal{F}^{(2)}$, and $\mathcal{F}^{(3)}$ of populations. Let $\mathcal{F}^{(1)}$ consist of acoustically coded populations, $\mathcal{F}^{(2)}$ consist of semantically coded populations, and $\mathcal{F}^{(3)}$ consist of motor control populations (Baddeley & Warrington, 1970; Bartlett & Tulving, 1974; Craik, 1970; Craik & Lockhart, 1972; Jacoby & Bartz, 1972; Maskorince & Brown, 1974). Suppose that both $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ can send signals to $\mathcal{F}^{(3)}$, from which performance is controlled by a nonspecific source of motor arousal. Also let LTM traces z_{ij} occur in the pathways from $\mathcal{F}^{(2)}$ to $\mathcal{F}^{(3)}$, with z_1 coding a primacy gradient and the other z_i , $i \neq 1$, coding the type of gradients summarized above. Suppose moreover that prior presentation of a list r_1, r_2, \dots, r_L establishes either an STM recency gradient or an STM bow across $\mathcal{F}^{(1)}$, and either an STM primacy gradient or an STM bow across $\mathcal{F}^{(2)}$ (Fig. 8a). By (19), the STM primacy gradient across $\mathcal{F}^{(2)}$ magnifies the LTM primacy gradient that it reads out of LTM into STM at $\mathcal{F}^{(3)}$. This magnified STM primacy gradient at $\mathcal{F}^{(3)}$ has to compete, however, with the STM

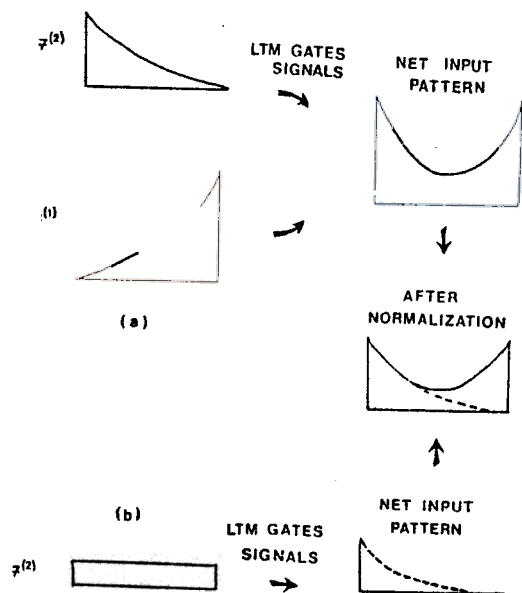


FIG. 8. Normalization can mask an STM primacy gradient by competition with an STM recency gradient.

recency gradient that $\mathcal{F}^{(1)}$ reads into $\mathcal{F}^{(3)}$. After normalization takes place across $\mathcal{F}^{(3)}$, the magnified STM primacy gradient is reduced due to competition by the STM recency gradient.

By contrast, consider what happens if an interfering event inhibits all STM in $\mathcal{F}^{(1)}$, $\mathcal{F}^{(2)}$, and $\mathcal{F}^{(3)}$. Correct performance is then impossible unless $\mathcal{F}^{(2)}$ is activated, since only the pathways from $\mathcal{F}^{(2)}$ to $\mathcal{F}^{(3)}$ contain LTM traces. Suppose that $\mathcal{F}^{(2)}$ is activated either randomly, or uniformly using a nonspecific sampling signal. In either case, the LTM primacy gradient will be read into an STM primacy gradient across $\mathcal{F}^{(3)}$ (Fig. 8b). This LTM primacy gradient is not reinforced by an STM primacy gradient across $\mathcal{F}^{(2)}$, so the net effect on $\mathcal{F}^{(3)}$ is smaller than before. However, the STM primacy gradient across $\mathcal{F}^{(3)}$ does not have to compete with an STM recency gradient. Hence normalization will amplify this STM primacy gradient, just as it amplifies the net gradient in Fig. 4. In all, normalization can mask an STM primacy gradient by differentially suppressing its effect when an STM recency gradient exists.

8. OTHER BINARY CODE DIFFICULTIES

The serial binary code assumption leads to several other beliefs that might be stated too strongly. For example, if STM storage is thought of as a binary event, then one can readily conclude that the number of rehearsals is the crucial parameter that determines whether an item remains or is reinstated in the STM buffer. For example, Bower writes "if the person is told during study of an item that its later recall will be worth a lot of

money, he will concentrate harder (rehearse more, maintain that item in STM for a longer time) and remember it better" (Hilgard & Bower, 1975, p. 580). However "concentrating harder," "rehearsing more," and "maintaining that item in STM for a longer time" can all be achieved by distinct mechanisms. Concentrating harder can, for example, generate an unusually large arousal level that supplements the item's usual input to STM (Grossberg, 1975). The two input sources acting together can create an unusually large STM strength that, other things equal (which they usually are not), increases the probability of saying the item, and of saying it out of order; in particular, saying it at an earlier recall position, than would otherwise occur. These effects can be generated without rehearsing this item any more than any other item.

A boost in STM strength of one item can depress the STM strengths of related items that are simultaneously in STM; cf., Ellis *et al.*, 1971. This von Restorff-type of depression can sometimes be due to the Normalization Rule, rather than to less rehearsal.

The fact that reaction time increases with the number of items in STM does not imply that recognition memory is realized by a serial scanning process (Sternberg, 1966). In a normalized STM field, each item in the field—except possibly the last—has a smaller STM trace if a longer list perturbs the field. If reaction time increases as STM activity decreases, then reaction time will depend on how many items are stored in STM, even though the rehearsal operation is a parallel operation that simultaneously influences all populations in the STM field.

9. CONCLUDING REMARKS

During the last decade, experimental and theoretical studies of STM and LTM have been remarkably vigorous and productive. The use of theoretical analogies from other disciplines, such as the computer analogy, indicates a healthy desire to conceptually organize the vast array of experimental findings. However, the binary and serial nature of computer concepts leads to conceptual difficulties in the many situations where continuous and parallel brain processes are operative. Computer modelers often claim that details like whether a code is binary or continuous are unimportant, because the same global strategies, or wiring diagrams, will occur despite differences in the individual components. Similarly, Townsend's (1974) result showing the equivalence of serial and parallel reaction time models has sometimes been interpreted as saying that it does not matter which type of model is used when memory processes are studied. The above examples are a few of the growing number that show binary serial models and continuous parallel models to be fundamentally different in design and properties.

REFERENCES

- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington, D. C.: Hemisphere, 1974.
- ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Advances in the psychology of learning and motivation, research and theory*, Vol. II. New York: Academic Press, 1968.

- ATKINSON, R. C., & SHIFFRIN, R. M. The control of short term memory. *Scientific American*, August, 1971, 82.
- 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.
- BARTLETT, J. C., & TULVING, E. Effect of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 297-309.
- BLOOMFIELD, T. M. Behavioral contrast and the peak shift. In R. M. Gilbert & N. S. Sutherland (Eds.), *Animal discrimination learning*. New York: Academic Press, 1969. Pp. 215-241.
- CORNWELL, T. N. *Visual perception*. New York: Academic, 1970.
- CRAIK, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 143-148.
- CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.
- CRAIK, F. I. M., & WATKINS, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 599-607.
- DONCHIN, E., & LINDSLEY, D. B. *Average evoked potentials: Methods, results, and evaluations*. Washington, D. C.: NASA, 1969.
- ELLIS, N. R., DETTERMAN, D. K., RUNCIE, D., MCCARVER, R. B., & CRAIG, E. M. Amnesic effects in short-term memory. *Journal of Experimental Psychology*, 1971, 89, 357-361.
- GROSSBERG, S. On the serial learning of lists. *Mathematical Biosciences*, 1969, 4, 201-253.
- GROSSBERG, S. Pattern learning by functional-differential neural networks with arbitrary path weights. In K. Schmitt (Ed.), *Delay and functional differential equations and their applications*. New York: Academic Press, 1972. Pp. 121-160.
- GROSSBERG, S. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 1973, 52, 213-257.
- GROSSBERG, S. Classical and instrumental learning by neural networks. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology*. New York: Academic Press, 1974. Pp. 51-141.
- GROSSBERG, S. A neural model of attention, reinforcement, and discrimination learning. In C. Pfeiffer (Ed.), *International Review of Neurobiology*, 1975, 18, 263-327.
- GROSSBERG, S. Adaptive pattern classification and universal recoding. I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 1976a, 23, 121-134.
- GROSSBERG, S. Adaptive pattern classification and universal recoding. II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 1976b, 23, 187-202.
- GROSSBERG, S. 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*. New York: Academic Press, 1977a.
- GROSSBERG, S. Pattern formation by the global limits of a nonlinear competitive interaction in n dimensions. *Journal of Mathematical Biology*, 1977b, 4, 237-256.
- GROSSBERG, S. A theory of visual coding, memory and development. In E. Leeuwenberg & H. Buffart (Eds.), *Formal theories of visual perception*. New York: Wiley, 1977c.
- GROSSBERG, S. Communication, memory, and development. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology*. New York: Academic Press, 1977d.
- 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, 263-327.
- GROSSBERG, S., & PEPE, J. Spiking threshold and over-arousal effects in serial learning. *Journal of Statistical Physics*, 1971, 3, 95-125.
- HILGARD, E. R., & BOWER, G. H. *Theories of Learning*, 4th ed. Englewood Cliffs, N. J.: Prentice-Hall, 1975.
- HODGKIN, A. L. *The conduction of the nervous impulse*. Springfield, Ill.: C. C. Thomas, 1964.
- HOGAN, R. M., & HOGAN, M. M. Structural and transient components of memory. *Memory and Cognition*, 1975, 3, 210-215.
- HOROWITZ, L. W., & IZAWA, C. Comparison of serial and paired-associate learning. *Journal of Experimental Psychology*, 1963, 65, 352-361.
- JACOBY, L. L., & BARTZ, W. H. Rehearsal and transfer to long term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 561-565.
- KAHNEMAN, D., & BEATTY, J. Pupil diameter and load on memory. *Science*, 1966, 154, 1583-1585.
- KATZ, B. *Nerve, muscle, and synapse*. New York: McGraw-Hill, 1966.
- LACHMAN, R., & LAUGHERY, K. R. Is a test trial a training trial in free recall learning? *Journal of Experimental Psychology*, 1968, 76, 40-50.
- LENNEBERG, E. *Biological foundations of language*. New York: Wiley, 1967.
- LEVINE, D. S., & GROSSBERG, S. Visual illusions in neural networks: Line neutralization, tilt after effect, and angle expansion. *Journal of Theoretical Biology*, 1976, 61, 477-504.
- LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.
- MASKORINEC, A. S., & BROWN, S. C. Positive and negative recency effects in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1974, 16, 328-334.
- MELTON, A. W., & MARTIN, E. (Eds.). *Coding processes in human memory*. Washington, D. C.: Winston, 1972.
- MILLER, G. A. The magic number seven, plus or minus two. *Psychological Review*, 1956, 63, 81-97.
- MILNER, B. Amnesia following operation on the temporal lobes. In C. W. M. Whitty & O. L. Zangwill (Eds.), *Amnesia*. London: Butterworths, 1956.
- RESTLE, F., SHIFFRIN, R. M., COSTELLAN, N. J., LINDMAN, H. R., & PISONI, D. B. (Eds.), *Cognitive theory*, Vol. 1. Hillsdale, N. J.: Erlbaum, 1975.
- RUNDUS, D. J. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, 89, 63-77.
- RUNDUS, D. Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 43-50.
- STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, 153, 652-657.
- TOWNSEND, J. T. Issues and models concerning the processing of a finite number of inputs. In Kantowitz, B. H. (Ed.), *Human information processing: Tutorials in performance and cognition*. Potomac, Md.: Erlbaum, 1974. P. 133.
- TULVING, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 175-184.
- TULVING, E., & DONALDSON, W. (Eds.). *Organization of memory*. New York: Academic Press, 1972.
- WERBLIN, F. S. Adaptation in a vertebrate retina: Intracellular recording in Necturus. *Journal of Neurophysiology*, 1971, 34, 228-241.
- WOODWARD, A. E., BJORK, R. A., & JONGEWARD, R. H., JR. Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 608-617.
- YOUNG, R. K. A comparison of two methods of learning serial associations. *American Journal of Psychology*, 1959, 72, 554-559.
- YOUNG, R. K. The stimulus in serial learning. *American Journal of Psychology*, 1961, 74, 517-528.
- YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *Journal of Experimental Psychology*, 1963, 63, 307-313.
- YOUNG, R. K. Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal Learning and General Behavior Theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. 122-148.
- ZEILER, M. D. The ratio theory of intermediate size discrimination. *Psychological Review*, 1963, 70, 516-533.

RECEIVED: July 15, 1977