On the Serial Learning of Lists

STEPHEN GROSSBERG
*Department of Mathematics, Massachusetts Institute of Technology, Cambridge
Massachusetts*

Communicated by Richard Bellman

ABSTRACT

A mathematical model with both a psychological and neurophysiological interpretation is introduced to qualitatively explain data about serial learning of lists. Phenomena such as bowing, anchoring, chunking, backward learning, all-or-none versus gradualist learning, anticipatory versus perseverative errors, accumulation of inhibition, and their dependence on intratrial interval, intertrial interval, list length, list position, and reaction time have mathematical analogs in this model. The proper definitions of a list's beginning, middle, and end for the purposes of learning theory are seen to depend crucially on temporal as well as geometrical factors.

1. INTRODUCTION

A recent series of papers [1–11] has introduced a new theory of learning in a rigorous mathematical setting. This theory investigates a collection of mathematical machines $\mathcal{M}$ whose laws are derived from simple psychological postulates [5]. How some of these machines learn, remember, and recall in various experimental situations has been studied in previous papers. In this paper we study how one of these machines learns lists of symbols of arbitrary length presented in a serial fashion at an arbitrary speed. It is well known that varying list length, intratrial interval, and intertrial interval in a serial learning experiment can dramatically change the way in which a list is learned by a human subject. We will show that qualitatively similar changes often occur in our machine. Among the phenomena for which we will find a mathematical analog in the machine are bowing [12], chaining [13], and chunking [14]. The classical Hovland curves showing different relative benefits of small and large

intertrial intervals when the presentation rate of the list is fast or slow qualitatively hold in this machine. The serial response oscillation curve also qualitatively holds. It is shown that anticipatory errors are more frequent than perseverative errors for an item in the middle of a long list, and that very remote errors are roughly equally probable in both the forward and backward directions.

We see qualitatively in the machine that increasing list length rapidly increases learning difficulty, and that the degree of skewness of the mean error curves decreases as list length increases. We also see that one-trial learning often occurs on short lists.

An empirically familiar order of learning at different list positions is found, in particular the anchoring effect, which seems to occur around the first list position. Familiar effects of interpolating a short well-learned list in a long list arise, particularly a clustering effect in the learning of new list items, which occurs in both the forward and backward directions around the short list. Generalizations to bimodal and multimodal bowing effects are made, and applications to the relative benefits of whole and part learning are discussed. Various other serial effects also follow from the same line of reasoning.

The mathematical variables of our machines have been given a neurophysiological and anatomical interpretation [5–7]. Thus all of the above-mentioned psychological phenomena can readily be interpreted as properties of ensembles of nerve cells. We find, for example (as Hull hypothesized thirty years ago [12]), that a kind of "accumulation of inhibition" is responsible for bowing in the middle of sufficiently long lists. We have provided a rigorous mathematical mechanism that gives rise to this inhibition (the details of which are, to be sure, very different from those envisaged by Hull) and have shown that this inhibition is lateral inhibition of the neural variety [15], since our equations reduce in a special steady-state case to the empirically derived Hartline–Ratliff equation for lateral inhibition [3, 6]. A derivation of this lateral inhibitory mechanism based on learning postulates has been given elsewhere [6].

Substantially more important than specific predictions is our analysis of the concepts of psychological space and time that underly them. These concepts apply to a much broader class of phenomena than serial learning data. They are particularly evident in serial learning data only because these data constitute one of the simplest and most "homogeneous" examples wherein behavioral units are integrated into a reproducible spatiotemporal pattern.

2. THE MACHINE

The laws of our machine are expressed in the mathematical language of nonlinear difference-differential equations. Our discussion here aims at the simplest possible qualitative derivation of our conclusions in order to clearly show how the various phenomena we have mentioned arise from a single mathematical mechanism. We will therefore avoid a technical exposition and will instead study approximations to our machine that reveal in a simple way the mathematical reasons behind the effects that we seek to explain. These approximations are, moreover, compatible with all the rigorous results described in earlier papers.

Before approximating our machine, we must define it precisely. This is done in the following way. Let us for convenience denote the symbols from which our lists will be constructed by $r_1, r_2, \ldots, r_n$. For example, if our symbols are the letters A, B, ..., Z of the alphabet, then $n = 26$ and we can identify r_1 with A, r_2 with B, and so on until we reach r_{26} and Z. We will call the collection of all the symbols r_i, $i = 1, 2, \ldots, n$, the *alphabet $\mathcal{A}$*. We let $\mathcal{L}(\mathcal{A})$ be the collection of all lists constructed from the alphabet $\mathcal{A}$, subject to the constraint that no letter in $\mathcal{A}$ appears more than once in any list. For example, when $r_1 = A, r_2 = B, \ldots, r_{26} = Z$, the list ABCDE is in $\mathcal{L}(\mathcal{A})$ but the list ABACA is not. We wish to construct the smallest machine that can, in principle, learn any list in $\mathcal{L}(\mathcal{A})$. We denote this machine by $\mathcal{M}(\mathcal{A})$.

The first step in defining $\mathcal{M}(\mathcal{A})$ is to assign to each symbol r_i in $\mathcal{A}$ a *point* (or *vertex*) v_i in $\mathcal{M}(\mathcal{A})$. We then connect every point v_i to every *distinct* point $v_j, j \neq i$, by a directed *arrow* (or *edge*) e_{ij} in $\mathcal{M}(\mathcal{A})$ that leads from v_i to v_j (see [5]).

To every point v_i we assign a function $x_i(t)$ that describes a process taking place in $\mathcal{M}(\mathcal{A})$ at v_i. The function $x_i(t)$ is called the *point strength of v_i* or, alternatively, the *stimulus trace of v_i*. We call the arrowhead of e_{ij} the *node N_{ij} of e_{ij}*, and suppose that a process, denoted by $y_{ij}(t)$, is taking place in $\mathcal{M}(\mathcal{A})$ at N_{ij}. Process $y_{ij}(t)$ is called the *nodal strength of N_{ij}*, or alternatively the *associational strength of N_{ij}*. These assignments are pictured in Fig. 1. The point strengths and nodal strengths are defined
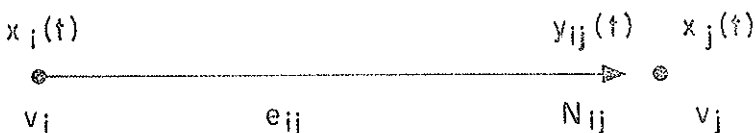

FIG. 1

14

by the following equations, which we designate system (*).

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{m=1}^{n} x_m(t - \tau) y_{mi}(t) + I_i(t), \tag{1}$$

$$y_{jk}(t) = z_{jk}(t) \left[\sum_{m=1}^{n} z_{jm}(t) \right]^{-1}, \tag{2}$$

$$\dot{z}_{jk}(t) = -u z_{jk}(t) + \beta x_j(t - \tau) x_k(t), \qquad j \neq k, \tag{3}$$

and

$$z_{jj}(t) \equiv 0, \tag{4}$$

$i, j, k = 1, 2, \ldots, n$. The functions $z_{jk}(t)$ can be entirely eliminated from this definition by observing that (2)–(4) are equivalent to

$$y_{jk}(t) = \frac{z_{jk}(0) + \beta \int_0^t x_j(v - \tau) x_k(v) e^{uv}\, dv}{\sum_{m \neq j} \left[z_{jm}(0) + \beta \int_0^t x_j(v - \tau) x_m(v) e^{uv}\, dv \right]}, \qquad j \neq k,$$

and $y_{jj}(t) = 0$, in which $z_{jk}(t)$ appears only through the initial data $z_{jk}(0)$. In reference [5] we derived the system (*), among others, from some simple psychological postulates, and showed how to interpret it as a flow over the vertices v_i and edges e_{jk} with cross-correlations occurring at the nodes N_{jk}. Such a system is called an *embedding field*. The inputs $I_i(t)$ in (*) describe an experiment performed on $\mathcal{M}(\mathcal{A})$ by an experimenter $\mathcal{E}$. Each choice of these inputs prescribes a different experiment.

3. THE SERIAL LEARNING PARADIGM

We wish to study how $\mathcal{M}(\mathcal{A})$ learns any list from $\mathcal{L}(\mathcal{A})$ that is presented to $\mathcal{M}(\mathcal{A})$ in a serial fashion. Suppose, for example, that our list is $\mathcal{L} = r_1 r_2 \cdots r_L$. Denote the event of presenting the symbol r_i to $\mathcal{M}(\mathcal{A})$ at time t by $\pi_i(t)$. The serial learning of $\mathcal{L}$ by $\mathcal{M}(\mathcal{A})$ can now be concisely described as the following collection of events.

$$\pi_1(0), \qquad \pi_2(w), \qquad \pi_3(2w), \qquad \ldots, \pi_L((L-1)w);$$

$$\pi_1((L-1)w+W), \qquad \pi_2(Lw+W), \qquad \pi_3((L+1)w+W), \ldots, \pi_L(2(L-1)w+W);$$

$$\pi_1(2(L-1)w+2W), \quad \pi_2((2L-1)w+2W), \quad \pi_3(2Lw+2W), \quad \ldots, \pi_L(3(L-1)w+2W);$$

.
.
.

and so on. That is, $\mathcal{L}$ is presented once to $\mathcal{M}(\mathcal{A})$ at a rate w^{-1}. A pause of W time units then occurs before the list is presented once again at rate

w^{-1}. This process is repeated until a suitable criterion of perfect learning is satisfied. The entire collection of events is denoted by $\mathcal{E}(w, W; L)$. The first row of this collection, namely,

$$\pi_1(0), \pi_2(w), \pi_3(2w), \ldots, \pi_L((L-1)w),$$

is called the first *trial* of $\mathcal{E}(w, W; L)$ and is denoted by $\mathcal{E}_1(w, W; L)$. Similarly, the ith row is called the ith *trial* of $\mathcal{E}(w, W; L)$ and is denoted by $\mathcal{E}_i(w, W; L)$. w is therefore called the *intratrial interval* (or *presentation interval*) of $\mathcal{E}(w, W; L)$, and W is called the *intertrial interval* (or *rest period*). These are, of course, just the usual empirical definitions given in a concise mathematical shorthand.
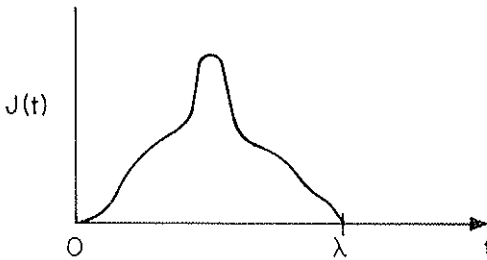

FIG. 2

An event $\pi_i(t_0)$ has the following effect on $\mathcal{M}(\mathcal{A})$. Suppose that a function $J(t)$ with a graph of the general form given in Fig. 2 is defined once and for all. That is, $J(t)$ is a nonnegative and continuous function that is positive in an interval having $t = 0$ as its left-hand end point. We call any nonnegative and continuous function $f(t)$ that is positive in an interval (λ, μ), $\lambda < \mu$, an *input pulse*, and λ is called the *onset time* of $f(t)$. Thus $J(t)$ is an input pulse with onset time zero. The event $\pi_i(t_0)$ is realized in $\mathcal{M}(\mathcal{A})$ as the input pulse $J(t - t_0)$ *delivered to* v_i. Input pulse $J(t - t_0)$ has onset time $t = t_0$ corresponding to the fact that r_i is presented to $\mathcal{M}(\mathcal{A})$ at time $t = t_0$. In a serial experiment $\mathcal{E}(w, W; L)$ consisting of N trials, the total input to v_i is simply the sum of the individual input pulses created by the events

$$\pi_i((i-1)w + (L-1)mw + mW), \qquad m = 0, 1, \ldots, N-1,$$

corresponding to the N presentations of r_i to $\mathcal{M}(\mathcal{A})$. That is, for $i = 1, 2, \ldots, L$,

$$I_i(t) = \sum_{m=0}^{N-1} J(t - (i-1)w - (L-1)mw - mW) \qquad (5)$$

in (1), whereas for $i = k + 1, \ldots, N$,

$$I_i(t) \equiv 0. \tag{6}$$

Any choice of inputs such as (5) and (6), for any fixed number $N \geqslant 1$ of trials and any fixed list length L, $L = 2, 3, \ldots, n$, defines a serial learning task $\mathcal{E}(w, W; L)$ for $\mathcal{M}(\mathcal{A})$.

4. REVIEW OF SOME EXPERIMENTAL DATA

We will study the effects on $\mathcal{M}(\mathcal{A})$ of varying the time intervals w and W while keeping the list length L fixed, and of varying L while keeping w and W fixed. Some of the qualitative effects that we will find are depicted in the classical experimental curves of Fig. 3, with the associated Table 1.
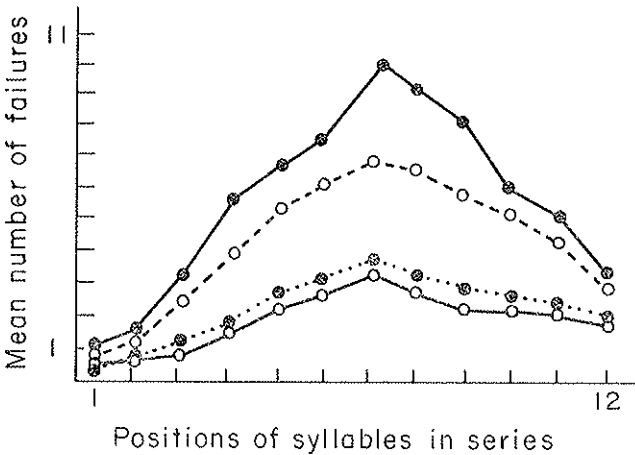

FIG. 3. From C. I. Hovland ([12], page 506).

TABLE 1

●———●	$W = 6$ sec, $w = 2$ sec
○- - - -○	$W = 2$ min 6 sec, $w = 2$ sec
●- - - -●	$W = 6$ sec, $w = 4$ sec
○———○	$W = 2$ min 6 sec, $w = 4$ sec

These curves say, speaking roughly, that the middle of a long list is harder to learn than its beginning and end, and that the beginning is easier to learn than the end. Moreover, increasing the rest interval W between

trials improves learning in the list's middle more when the presentation rate w^{-1} is fast than when it is slow. If the presentation rate is sufficiently slow, then increasing the rest period by even a large factor will have a negligible effect on the learning rate.

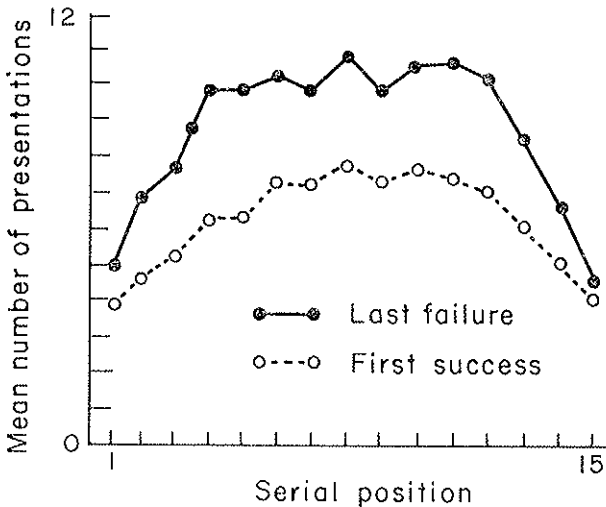A related experimental phenomenon is described by Fig. 4, which

FIG. 4. From C. L. Hull *et al.* ([12], page 503).

shows a "response oscillation" curve; this curve says, speaking roughly, both that it takes longer to begin to show learning on an item in the middle of a long list, and that incorrect guessing persists longest in the middle of the list after some evidence of learning appears. It is also known from such studies that: incorrect guesses at the middle of a list can come from a much broader set of response items than errors at either end of the list; the frequency of a given error is a monotonically decreasing function of increases in the remoteness of the error from the stimulus item; anticipatory errors are more probable than perseverative errors of a given remoteness; and very remote errors of equal remoteness are approximately equally probable in both the anticipatory and perseverative categories. See Osgood ([12], Chapter 12) for further experimental details.

The very existence of these data immediately broaches some rather deep general questions concerning psychological space and time. In a "long" list, for example, we typically find the bowing effect of Fig. 3 if the

presentation rate w^{-1} is sufficiently fast. In a "short" list, such as AB, or even ABC, manifestly no such pronounced bowing effect occurs. How long must a long list be before bowing occurs? Bowing is a phenomenon associated with a list's middle. Yet ABC has a "middle," namely B, and no significant bowing occurs while a standard adult learns ABC. Does this mean that, in some sense, the list ABC really has no middle? Moreover, the amount of bowing depends on the presentation rate w^{-1} or on temporal factors. Does the definition of a list's middle, which seems to be a purely geometrical problem, really depend on time as well? By asking questions of this kind, we are led to realize that there exist at least two answers to such seemingly innocent geometrical questions as the following. (1) How long is a long list? (2) Where is a list's beginning, middle, and end? (3) Where is a list's boundary and interior?

Given a list $r_{i_1} r_{i_2} \cdots r_{i_m}$, the first—and trivial—answers to these questions are: (1) length $= m$; (2) beginning $= \{r_{i_1}\}$, end $= \{r_{i_m}\}$, middle $= \{r_{i_{m/2}}\}$ when m is even and $\{r_{i_{(m+1)/2}}\}$ when m is odd; and (3) boundary $= \{r_{i_1}, r_{i_m}\}$, interior $= \{r_{i_2}, r_{i_3}, \ldots, r_{i_{m-1}}\}$. Significant answers to these questions must, by contrast, try to characterize those features of the learning process that will guarantee that a long list takes many trials to learn, that the middle of a list elicits a bowing effect, and that the boundary of a list is learned at a different rate from its interior. Answers of the latter kind are in the spirit of Figs. 3 and 4, and we now propose some answers in this spirit.

5. VARYING INTRATRIAL AND INTERTRIAL INTERVALS IN THE BARE FIELD

For simplicity, we always consider a machine that is initially at rest and in a state of maximal ignorance. "At rest" means that the initial data $x_i(v) = 0$, for $v \in [-\tau, 0]$ and $i = 1, 2, \ldots, n$. That is, the experimental inputs of $\mathcal{E}(w, W; L)$ are the first to perturb $\mathcal{M}(\mathcal{A})$ in a very long time. "Maximal ignorance" means that $z_{jk}(0) = \delta > 0$, $j \neq k$, and $z_{jj}(0) = 0$. That is, all associational strengths $y_{jk}(0)$ with $j \neq k$ are initially the same. We now investigate how such a machine learns for various choices of w, W, and L. A previous paper [5] studied a special case of this problem, namely, the learning of long lists (say, $L = n > 2$) whose intertrial and intratrial intervals are the same (i.e., $w = W$). That paper showed that such lists are learned very poorly indeed if w is either too small or too large. By this we mean the following. We must always evaluate the size of w relative to at least three factors: (i) the "reaction time" τ of $\mathcal{M}(\mathcal{A})$;

(ii) the duration of any event, namely, sup $\{t: J(t) > 0\}$; and (iii) the rate with which the effects of an event wear off, namely, the relaxation time of the $x_i(t)$. These three factors determine time scales by which events are measured within $\mathcal{M}(\mathcal{A})$. When we say that w is "small," we will always mean "small relative to these factors." That is, w is small if events are presented so quickly to $\mathcal{M}(\mathcal{A})$ that a second event arrives long before the effects *within* $\mathcal{M}(\mathcal{A})$ of the first wear off.

Suppose, for example, that $w = 0$. Then $I_1(t) \equiv I_2(t) \equiv \cdots \equiv I_n(t)$, and since $x_i(0) = x_j(0)$ and $z_{ij}(0) = z_{km}(0)$ whenever $i \neq j$ and $k \neq m$, we immediately conclude, by symmetry, that all associational strengths $y_{jk}(t)$ with $j \neq k$ remain equal for all $t > 0$. Thus no learning occurs in $\mathcal{M}(\mathcal{A})$.

Suppose w is very large; for example, $w \gg \tau$. Then a second event occurs in $\mathcal{M}(\mathcal{A})$ long after the effect of a prior event wears off. In this situation, an input to any given v_j, $j = 1, 2, \ldots, n - 1$, creates equal signals in each edge e_{jk}, $k \neq j$, which reach the node N_{jk} τ time units later. The effects of the input gradually wear off and the signals also decay to zero. The next input to $\mathcal{M}(\mathcal{A})$ reaches v_{j+1} w time units later, which is long after the signals have become very small. As a result, all the cross-correlations $z_{jk}(t)$ with $k \neq j$ of (3) always remain comparable, and so all the associational strengths $y_{jk}(t)$ of (2) always remain approximately the same. Again very little learning occurs. The earlier article [5] went on to show that learning is better when $w \cong \tau$, since then a signal created by an input to v_j reaches $N_{j,j+1}$ as the input I_{j+1} becomes large at v_{j+1}.

We will now consider more general choices of w, W, and L. Our previous remarks illustrate that a variation of w and W can have profound effects on how $\mathcal{M}(\mathcal{A})$ learns. In order to study this variation in its simplest setting, we now consider the way in which each point v_i separately responds to serial inputs. To accomplish this, we temporarily remove all the interactions between points in $\mathcal{M}(\mathcal{A})$. That is, we remove all the edges e_{mi} from $\mathcal{M}(\mathcal{A})$ and all the interaction terms

$$\beta \sum_{m=1}^{n} x_m(t - \tau) y_{mi}(t)$$

from (1). Then (1) becomes the following system of linear differential equations.

$$\dot{x}_i(t) = -\alpha x_i(t) + I_i(t), \qquad i = 1, 2, \ldots, n, \tag{7}$$

where $x_i(0) = 0$ and $I_i(t)$ satisfies (5) and (6). The linear equations (7) are collectively called the *bare field of* $\mathcal{M}(\mathcal{A})$.

We will always, for simplicity, constrain the input pulse $J(t)$ from which the inputs $I_i(t)$ are constructed by the following conditions.

I. $J(t)$ is positive only in $(0, \lambda)$ where $\lambda < \tau$.

That is, the duration of a "simple" event such as $\pi_i(t_0)$ is less than the "reaction time" of $\mathcal{M}(\mathcal{A})$.

II. $J(t)$ increases monotonically to a finite maximum and then decreases monotonically to zero.

That is, since $J(t)$ represents the occurrence of a simple event, it rises to a single maximum and then decays.

We will also impose the following condition on w and W.

III. $w \leq W$.

That is, the rest period W between trials is no shorter than the interval w between successive item presentations.

Each of the constraints I, II, and III can eventually be removed and leads to obvious modifications in the details of the following analysis. Consider the first trial $\mathcal{E}_1(w, W; L)$ of a serial learning experiment performed on the bare field (7). Since the events

$$\pi_1(0), \pi_2(w), \pi_3(2w), \ldots, \text{ and } \pi_L((L - 1)w)$$

occur in succession, the input

$$I_i(t) = J(t - (i - 1)w), \qquad i = 1, 2, \ldots, L,$$

occurs at v_i with onset time $t = (i - 1)w$. Since also $x_i(0) = 0$, (7) implies that

$$x_i(t) = \exp(-\alpha t) \int_0^t \exp(\alpha v) J(v - (i - 1)w) \, dv,$$

which after a change of variables becomes

$$x_i(t) = \begin{cases} 0, & 0 \leq t \leq (i - 1)w \\ \exp\{-\alpha[t - (i - 1)w]\} \displaystyle\int_0^{t-(i-1)w} \exp(\alpha v) J(v) \, dv, \\ & (i - 1)w < t \leq (L - 1)w + W. \end{cases} \qquad (8)$$

In particular,

$$x_i(t) = x_1(t - (i-1)w) \qquad \text{for } t \in [0, (L-1)w + W] \qquad (9)$$

and all $i = 1, 2, \ldots, L$.

The input pulse $J(v)$ has duration λ, and thus for all $t \in [(i-1)w + \lambda,$ $(L-1)w + W]$, (8) implies

$$x_i(t) = \Lambda \exp\{-\alpha[t - (i-1)w]\} \qquad (10)$$

where

$$\Lambda = \int_0^\lambda \exp(\alpha v) J(v) \, dv.$$

That is, $x_i(t)$ decays at the exponential rate α λ time units after $\pi_i((i-1)w)$ occurs. In short, first x_1 begins to grow at time $t = 0$, then x_2 begins to grow at time $t = w$, and in general $x_i(t)$ begins to grow at time $t = (i-1)w$, for all $i = 1, 2, \ldots, L$. No more than λ time units after its growth begins, each $x_i(t)$ begins to decay at an exponential rate.

We can say more concerning the pattern of growth and decay of each $x_i(t)$ because $J(t)$ rises to a maximum and then decays. Indeed, we now show that $x_i(t)$ also grows *monotonically* to its maximum and then decreases *monotonically* to zero. By (9), it suffices to prove this for $i = 1$. By (7) and (8), we have on trial 1 that

$$\dot{x}_1(t) = -\alpha \exp(-\alpha t) \int_0^t \exp(\alpha v) J(v) \, dv + J(t)$$

$$= -\exp(-\alpha t) \int_0^t J(v) \frac{d}{dv} \exp(\alpha v) \, dv + J(t).$$

Integrating by parts, we find

$$\dot{x}_1(t) = -\exp(-\alpha t)\left[J(t) \exp(\alpha t) - J(0) - \int_0^t \exp(\alpha v) \dot{J}(v) \, dv \right] + J(t)$$

or

$$\dot{x}_1(t) = \exp(-\alpha t) \int_0^t \exp(\alpha v) \dot{J}(v) \, dv. \qquad (11)$$

We now invoke hypothesis II concerning $J(v)$, namely, that $\dot{J}(v) \geqslant 0$ for small v and $\dot{J}(v) \leqslant 0$ for all large v. Thus by (11), $\dot{x}_1(t) \geqslant 0$ until after $\dot{J}(v)$ changes sign. Thereafter $\dot{x}_1(t)$ changes sign at most once, since $\dot{J}(v)$ changes sign at most once; $\dot{x}_1(t)$ changes sign precisely once since $x_1(t)$ decreases to zero for all $t > \lambda$. We summarize these simple facts in the following proposition.

Proposition 1. On trial $\mathcal{E}_1(w, W; L)$, $x_i(t) = x_1(t - (i - 1)w)$, $i = 1, 2, \ldots, L$, *where*

$$\text{(a)} \ x_1(t) = \begin{cases} 0, & t \leqslant 0, \\ \exp(-\alpha t) \int_0^t \exp(\alpha v) J(v) \, dv, & 0 < t \leqslant (L - 1)w + W; \end{cases}$$

(b) $x_1(t)$ *rises monotonically to its maximum in* $(0, \lambda)$; *and*

(c) $x_1(t)$ *decays monotonically toward zero thereafter, and at the exponential rate α in* $[\lambda, (L - 1)w + W]$.

Moreover, $x_i(t) \equiv 0$, $i = L + 1, L + 2, \ldots, n$.

One of the most important variables determining learning effects in $\mathcal{M}(\mathcal{A})$ is the number of $x_i(t)$ that are large at any time t. The reason for this, as we will soon prove, is that the associational strength $y_{jk}(t)$ for any $j \neq k$ measures the size of $x_j(t - \tau)x_k(t)$ relative to the other $x_j(t - \tau)x_m(t)$ values, $m \neq k$, in $\mathcal{M}(\mathcal{A})$ at time t. If only $x_j(t - \tau)x_k(t)$ is large, then $y_{jk}(t)$ will grow rapidly and learning of the association from r_j to r_k is enhanced. If, on the other hand, many $x_j(t - \tau)x_m(t)$ are large, then $y_{jk}(t)$ might well decrease, and learning of the association from r_j to r_k is jeopardized.

Since the number of $x_m(t)$ that are large at any time t is relevant to learning in $\mathcal{M}(\mathcal{A})$ at that time, we now introduce some simple devices whereby this number can be discussed in a precise way as a function of time and of the parameters w, W, and L.

6. ACCUMULATION SETS

For fixed $\varepsilon > 0$, let $A_\varepsilon(w, W, L; t)$ denote the collection of indices i such that $x_i(t) \geqslant \varepsilon$. That is, $A_\varepsilon(w, W, L; t)$ tells us which x_i are at least as large as ε at time t. For simplicity, we write $A_\varepsilon(w, W, L; t)$ as $A_\varepsilon(t)$, and also let $|A_\varepsilon(t)|$ be the number of indices in $A_\varepsilon(t)$. $A_\varepsilon(t)$ is called the

ε-accumulation set at time t, since it contains the indices of all points v_i that have accumulated at least an amount ε of point strength at time t. We always suppose in the following that ε is fixed in such a way that $0 < \varepsilon < \max\{x_1(t): t \geqslant 0\}$ to avoid trivialities. The following basic facts concerning $A_\varepsilon(t)$ on trial $\mathcal{E}_1(w, W; L)$ are easy consequences of Proposition 1.

The set $|A_\varepsilon(t)|$ remains zero until the first time $t = t_\varepsilon$ at which $x_1(t) = \varepsilon$. Then $|A_\varepsilon(t_\varepsilon)| = 1$. The index 1 remains in $A_\varepsilon(t)$ until the time $t = T_\varepsilon$ at which $x_1(t) = \varepsilon$ for the last time, since by Proposition 1 we can also assert that $x_1(t) \geqslant \varepsilon$ for all t in $[t_\varepsilon, T_\varepsilon]$. Since $x_i(t) = x_1(t - (i - 1)w)$ for all $i = 1, 2, \ldots, L$, the index 2 enters $A_\varepsilon(t)$ at time $t = t_\varepsilon + w$, the index 3 enters $A_\varepsilon(t)$ at time $t = t_\varepsilon + 2w$, and in general the index i enters $A_\varepsilon(t)$ at time $t = t_\varepsilon + (i - 1)w$, $i = 1, 2, \ldots, L$. Each of these indices remains in $A_\varepsilon(t)$ for $T_\varepsilon - t_\varepsilon$ time units, and none of the indices $i = L + 1, \ldots, n$ ever enters $A_\varepsilon(t)$.

The overall behavior of $A_\varepsilon(t)$ as t varies within $[0, Lw]$ depends on two factors, for fixed w, W, and L. These are the amount of time $S_\varepsilon = T_\varepsilon - t_\varepsilon$ that a single index remains in $A_\varepsilon(t)$, and the number of new indices that are added to $A_\varepsilon(t)$ during this time. To describe the interplay of these quantities in a precise way, we introduce the following notation.

For any $\xi \geqslant 0$, let $[\xi]$ be the greatest integer that does not exceed ξ. Now let

$$G_\varepsilon(w) = \left[\frac{S_\varepsilon}{w}\right];$$

$G_\varepsilon(w)$ measures the number of new indices that can be added to $A_\varepsilon(t)$ before an old index drops out. Since S_ε is independent of w, $G_\varepsilon(w)$ is a *monotone decreasing* function of w. We will find that the existence or nonexistence of a bowing effect in $\mathcal{M}(\mathcal{A})$ during the learning of a given list $\mathcal{L} = r_1 r_2 \cdots r_L$ can be qualitatively decided by examining the absolute and relative sizes of $G_\varepsilon(w)$, L, and W. To do this, we must distinguish two cases.

Case 1. $G_\varepsilon(w) < L - 1$.

In this case, $A_\varepsilon(t)$ accumulates the indices $1, 2, \ldots, G_\varepsilon(w) + 1$ at a linear rate at the times $t = t_\varepsilon, t_\varepsilon + w, \ldots, t_\varepsilon + wG_\varepsilon(w)$. In particular, $|A_\varepsilon(t)|$ jumps by 1 every w time units from its initial value 0 until it reaches $G_\varepsilon(w) + 1$.

After time $t = T_\varepsilon$, the "old" index 1 drops out of $A_\varepsilon(t)$, but at time $t = w(G_\varepsilon(w) + 1)$ the "new" index $G_\varepsilon(w) + 2$ enters $A_\varepsilon(t)$. Thereafter, one old index leaves $A_\varepsilon(t)$ and one new index enters $A_\varepsilon(t)$ every w time units
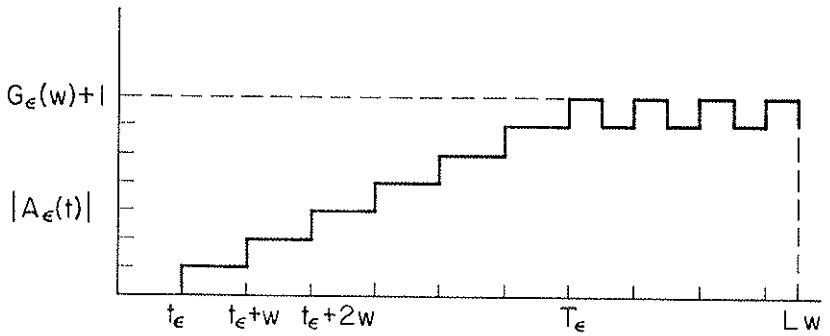
FIG. 5

until all indices $i = 1, 2, \ldots, L$ have entered $A_\varepsilon(t)$. In particular, after $|A_\varepsilon(t)|$ climbs at a linear rate to its maximum value $G_\varepsilon(w) + 1$, it thereafter oscillates with period w between $G_\varepsilon(w) + 1$ and $G_\varepsilon(w)$ until $t = Lw$, as in Fig. 5.

Case 2. $G_\varepsilon(w) \geqslant L - 1$.

This case can be treated just as Case 1 was with the following all-important difference. The list indices $i = 1, 2, \ldots, L$ all enter $A_\varepsilon(t)$ as $|A_\varepsilon(t)|$ climbs to the value L. Thus, $|A_\varepsilon(t)|$ climbs at a linear rate to the maximum value L, and there is no steady-state oscillatory behavior with period w, as in Fig. 6.

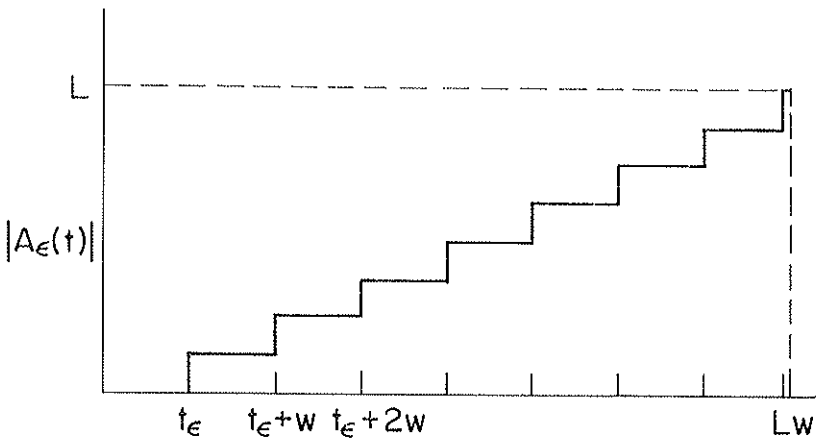
FIG. 6

Cases 1 and 2 exhaust all the possibilities for t in $[0, Lw]$, so that in all cases, $A_\varepsilon(t)$ is a *connected* set of indices of the form

$$A_\varepsilon(t) = \{k_\varepsilon(t), k_\varepsilon(t) + 1, \ldots, k_\varepsilon(t) + r_\varepsilon(t)\}$$

where $k_\varepsilon(t)$ and $r_\varepsilon(t)$ depend on w and L. We summarize these facts in the following proposition.

Proposition 2. On trial $\mathcal{E}_1(w, W; L)$, for t in $[0, Lw]$, $A_\varepsilon(t)$ is a connected set of indices such that

(a) *indices are added to $A_\varepsilon(t)$ in chronological order at times $t = t_\varepsilon$, $t_\varepsilon + w, \ldots, t_\varepsilon + (L-1)w$, where $t_\varepsilon = \min\{t: x_1(t) = \varepsilon\}$, and*

(b) *each index remains in $A_\varepsilon(t)$ for $T_\varepsilon - t_\varepsilon$ time units, where $T_\varepsilon = \max\{t: x_1(t) = \varepsilon\}$. In particular, letting $G_\varepsilon(w) = [(T_\varepsilon - t_\varepsilon)/w]$, if $G_\varepsilon(w) < L - 1$, then $|A_\varepsilon(t)|$ increases in unit steps every w time units until $|A_\varepsilon(t)| = G_\varepsilon(w) + 1$. Thereafter $|A_\varepsilon(t)|$ oscillates between $G_\varepsilon(w) + 1$ and $G_\varepsilon(w)$ with period w. If $G_\varepsilon(w) \geqslant L - 1$, then $|A_\varepsilon(t)|$ increases in unit steps every w time units until $|A_\varepsilon(t)| = L$.*

We now use Proposition 2 to study how changes in w and L produce changes in the associational strengths $y_{jk}(t)$ through time. The main fact needed for this study is that $y_{jk}(t)$ compares $x_j(t - \tau)x_k(t)$ with all $x_j(t - \tau)x_m(t)$, $m \neq j$. When $x_j(t - \tau)x_k(t)$ is much larger than all $x_\varepsilon(t - \tau)x_m(t)$, $m \neq j, k$, then $y_{jk}(t)$ grows quickly and the association from r_j to r_k is quickly learned. When $x_j(t - \tau)x_k(t)$ is comparable in size or smaller than several $x_j(t - \tau)x_m(t)$, then $y_{jk}(t)$ grows slowly, if at all, and learning from r_j to r_k is negligible. We now translate these facts into corresponding facts about accumulation sets $A_\varepsilon(t)$, and then classify the choices of w and L that produce accumulation sets that have a prescribed effect on learning in $\mathcal{M}(\mathcal{A})$.

7. MASSED VERSUS DISTRIBUTED PRACTICE

The association $y_{jk}(t)$ from r_j to r_k grows quickly at times t for which

$$x_j(t - \tau)x_k(t) \gg x_j(t - \tau)x_m(t), \qquad m \neq j, k.$$

This condition can be achieved if $x_j(t - \tau)$ and $x_k(t)$ are large, whereas all $x_m(t)$ are small, $m \neq j, k$. This means that j is in $A_\varepsilon(t - \tau)$, k is in $A_\varepsilon(t)$, and all $m \neq j, k$ are not in $A_\varepsilon(t)$, for some sufficiently large ε that we fix once and for all. In particular, $|A_\varepsilon(t)|$ is a small number, since only j and k can be in $A_\varepsilon(t)$.

How can we guarantee that $|A_\varepsilon(t)|$ be a small number? By Proposition

2, the maximum of $|A_\varepsilon(t)|$ in $[0, Lw]$ is $G_\varepsilon(w) + 1$. We need therefore merely require that $G_\varepsilon(w)$ be small. But $G_\varepsilon(w) = [S_\varepsilon/w]$, which is monotone decreasing in w. Therefore, $|A_\varepsilon(t)|$ will remain small for all t in $[0, Lw]$ if w is taken sufficiently large. One way of speeding up learning in $\mathcal{M}(\mathcal{A})$ is thus to slow down the rate with which list symbols are presented; that is, to "distribute practice."

Conversely, if the presentation rate of the list is very fast, $G_\varepsilon(w)$ will be large, and there will exist times t in $[0, Lw]$ at which $|A_\varepsilon(t)|$ is large. Indices that are in $A_\varepsilon(t)$ at these times will not rapidly be incorporated into new associations. Thus "massed practice" can slow the learning rate.

Distributing practice will not always facilitate learning. By Section 5, choosing $w \gg \tau$ yields very bad learning even though such a choice of w certainly distributes practice. The good or bad effects on learning of increasing w correspond to two different factors, distinguishability and correlations.

Distinguishability. Increasing w decreases $G_\varepsilon(w)$ and thereby keeps $|A_\varepsilon(t)|$ small in $[0, Lw]$. Thus only a few $x_i(t)$ are large at any time, and these can easily be distinguished from the many small $x_i(t)$ by the associational strengths.

Correlations. By Section 4, choosing $w \gg \tau$ means that all products $x_j(t - \tau)x_k(t)$ with $k \neq j$ are always either approximately equal or small, and thus all $y_{jk}(t)$ remain approximately constant.

Distributing practice helps learning only if good distinguishability *and* good correlations prevail; that is, if $G_\varepsilon(w)$ is small *and* $w \cong \tau$. Since $G_\varepsilon(w) = [S_\varepsilon/w]$, $G_\varepsilon(w)$ is small and $w \cong \tau$ only if $S_\varepsilon \cong w \cong \tau$.

These conditions impose constraints on the duration λ of an input pulse and on the rate α with which point strengths respond to inputs. We derive these constraints in the following way. By definition, $S_\varepsilon = T_\varepsilon - t_\varepsilon$. Since $x_i(t)$ grows monotonically to a maximum and then decays monotonically toward zero, S_ε estimates the amount of time needed for $x_1(t)$ to respond to an input pulse and then to decay toward zero. Clearly, $x_1(t)$ needs a little longer to respond fully to an input pulse than the duration λ of the pulse. Since $S_\varepsilon \cong \tau$, we must choose $\lambda < \tau$ if we desire good learning to occur for *some* choice of w. This is why we imposed this condition on λ in Section 5. Point strength $x_1(t)$ decays no faster than the exponential $e^{-\alpha t}$. Since $e^{-\alpha t}$ requires $T = \alpha \log 2$ time units to decay to half its initial value, we must also choose $\alpha \cong \tau$ to create conditions optimal for rapid learning in $\mathcal{M}(\mathcal{A})$.

The foregoing considerations show how easily we can be misled by the dictum "distributing practice helps learning," since as $w \to \infty$, the relative improvement in distinguishability diminishes rapidly while the relative disintegration of good correlations increases rapidly.

8. CONTIGUITY VERSUS CONNECTEDNESS

We can now easily see that *contiguous* symbols, such as r_{j-1}, r_j, and r_{j+1}, are most likely to enter into associations with one another. This is because $A_\varepsilon(t)$ is always a *connected* set. For example, in order that $y_{jk}(t)$ grow rapidly, $x_j(t - \tau)$ and $x_k(t)$ must be large, and $|A_\varepsilon(t)|$ must be small. By Section 7, this is best guaranteed when $S_\varepsilon \cong \tau$, so that no index i remains in $A_\varepsilon(t)$ much longer than τ time units. By Proposition 2, we also know that indices are added to $A_\varepsilon(t)$ in chronological order. Since fast learning requires that j be in $A_\varepsilon(t - \tau)$ and k be in $A_\varepsilon(t)$, we conclude that $y_{jk}(t)$ will grow fastest if $k \cong j + 1$; that is, if r_j and r_k are *contiguous*.

When w is small (and $G_\varepsilon(w)$ is large), $A_\varepsilon(t)$ is still a connected set, even though there exist times t when it contains many indices. Once again we can assert that contiguous associations are the strongest ones, but we must interpret this assertion in the weak sense that associations form best at any time t among the indices in the connected set $A_\varepsilon(t)$. In particular, associations such as $y_{j,j+2}(t)$, $y_{j,j+3}(t)$, and $y_{j,j-1}(t)$ might well be of substantial size, thereby reducing the size of $y_{j,j+1}(t)$. Backward learning effects, such as an increase in $y_{j,j-1}(t)$, are discussed in [5].

9. THE BEGINNING AND THE MIDDLE OF A LIST

We are now in a position to define the beginning and middle of a list in a way that takes into account some of the temporal and learning factors mentioned in Section 4. For example, let w and L be chosen in such a way that $L \gg 1$ and $G_\varepsilon(w) < L - 1$. Then by Proposition 2, $|A_\varepsilon(t)|$ grows in unit steps every w time units from 0 to $G_\varepsilon(w) + 1$, and thereafter oscillates between $G_\varepsilon(w)$ and $G_\varepsilon(w) + 1$ until $t = Lw$. We must distinguish two cases.

Case 3. $G_\varepsilon(w) \gg 0$.

In this case, $A_\varepsilon(t)$ goes through two phases: (1) a *transient* phase at times t corresponding to the monotonic increase of $|A_\varepsilon(t)|$ from 0 to $G_\varepsilon(w) + 1 \gg 0$; and (2) a *steady-state* phase at times t corresponding to

the periodic oscillation of $|A_\varepsilon(t)|$ between $G_\varepsilon(w)$ and $G_\varepsilon(w) + 1$. See Fig. 5.

Case 4. $G_\varepsilon(w) \cong 0$.

In this case, $A_\varepsilon(t)$ goes through essentially only one phase, since $|A_\varepsilon(t)|$ oscillates between $G_\varepsilon(w) \cong 0$ and $G_\varepsilon(w) + 1 \cong 1$ at *all* times t in $[0, Lw]$. See Fig. 7.
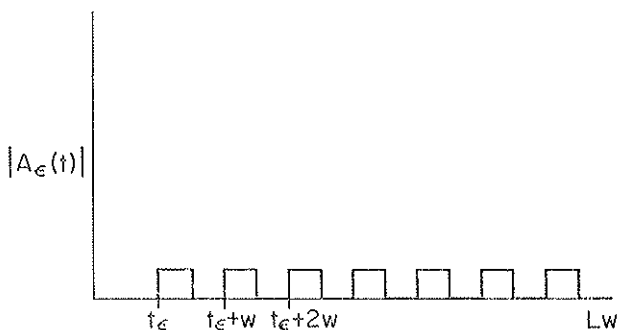

FIG. 7

We can now define the (*dynamical*) *beginning* of the list $r_1 r_2 \cdots r_L$ for fixed w and L at times t in $[0, Lw]$ as being the set of symbols r_i whose indices i are in the *same phase* of $A_\varepsilon(t)$'s development *as the index 1* is. The (*dynamical*) *middle* of the list is the set of symbols r_i corresponding to the *second* phase of $A_\varepsilon(t)$'s development, whenever this phase exists. We denote the set of symbols in the dynamical beginning by $B_\varepsilon \equiv B_\varepsilon(w, L)$, and those in the dynamical middle by $M_\varepsilon \equiv M_\varepsilon(w, L)$. When $G_\varepsilon(w)$ is large, several symbols will be in both B_ε and M_ε. This ambiguity is in the nature of the problem.

The foregoing definitions of B_ε and M_ε have some unusual, but nonetheless useful, consequences. For example, (a) the numerical length of a list's dynamical beginning is a function of w, and (b) there exists lists that have no dynamical middle and whose symbols all belong to the list's dynamical beginning.

Various experimental bowing effects can be conveniently explained in terms of these definitions; for example, (c) symbols in the list's dynamical middle are harder to learn than symbols in the dynamical beginning. This is because every symbol r_i in M_ε has a large $x_i(t)$ value when $|A_\varepsilon(t)|$ equals one of the *large* numbers $G_\varepsilon(w)$ or $G_\varepsilon(w) + 1$. Thus $x_i(t)$ cannot be easily distinguished and associations are not easily formed with r_i.

By contrast, even if $G_\varepsilon(w)$ is large, $x_1(t)$ is large when $|A_\varepsilon(t)|$ is small, $x_2(t)$ is large when $|A_\varepsilon(t)|$ is only slightly larger, and so on. Learning in B_ε is therefore certainly faster than learning in M_ε if $G_\varepsilon(w)$ is large.

If $G_\varepsilon(w)$ is small, say $\cong 0$, learning is fast throughout B_ε, which now includes all symbols r_i in the list, since $|A_\varepsilon(t)|$ remains small for all t in $[0, Lw]$.

The assertion that learning is faster in B_ε than in M_ε confirms the first half of the experimental bowed learning curve of Fig. 3. The assertion that Fig. 6 is transformed into Fig. 5 and finally into Fig. 7 as w increases agrees with the experimental fact that slowing the presentation rate flattens the first part of the bowed curve.

The definitions B_ε and M_ε merely illustrate the interplay of temporal and geometrical factors in determining ease of learning. The size of $|A_\varepsilon(t)|$ when each $x_i(t)$ is large is the crucial fact to determine.

10. HOW LONG IS A LONG LIST?

Just as a list's dynamical beginning varies with w, a list's *dynamical length* does also. Intuition bids us to say that a list is "dynamically long" only if it takes a long time to learn. Since an increase in w causes a decrease in $G_\varepsilon(w)$ and a consequent increase in learning speed, increasing w also decreases the dynamical length of the list.

Is a list of short numerical length (say $L \cong 2$) dynamically short? The answer is "yes" whenever $1 \cong G_\varepsilon(w) \cong L - 1$, since then good correlations and good distinguishability occur together. Consider, by contrast, the case $1 \ll G_\varepsilon(w) \leqslant L - 1$. Then the list has a beginning and a middle, and is consequently long. Letting w increase until $1 \cong G_\varepsilon(w) \ll L - 1$ eliminates the middle and thereby shortens the list.

Decreasing w until $G_\varepsilon(w) \geqslant L - 1$ has a qualitatively different effect. By Proposition 2, $A_\varepsilon(t)$ then has only a transient phase in which $|A_\varepsilon(t)|$ increases from 0 to L for t in $[0, Lw]$. Again B_ε exhausts all list symbols, but now point strengths $x_i(t)$ with $i \cong L$ are large only when $|A_\varepsilon(t)| \cong L$. Thus if $L \gg 1$, the symbols r_i with $i \cong L$ do not easily enter new associations for t in $[0, Lw]$. Once $G_\varepsilon(w) \geqslant L - 1$ for some $w = w_0$, decreasing w further cannot increase learning difficulty by changing the phases of $A_\varepsilon(t)$. In this sense, the numerical length L of a list places an upper bound on the difficulty of learning at any presentation rate. Nonetheless, decreasing w beyond $w = w_0$ makes it increasingly likely that all correlations will be approximately equal, by symmetry.

11. WHERE IS THE END OF A LIST?

We have thus far considered the behavior of $A_\varepsilon(t)$ only for t in $[0, Lw]$, and have based our definitions of B_ε and M_ε on this behavior. *No dynamical end exists in the list before time $t = Lw$, even though all list items have been presented to $\mathcal{M}(\mathcal{A})$ before time $t = Lw$!* The dynamical end of a list is created only *after* the list has been presented to $\mathcal{M}(\mathcal{A})$, and is due to the interactions of stimulus traces $x_i(t)$ and associations $y_{jk}(t)$ within $\mathcal{M}(\mathcal{A})$ before the list is presented for the second time. To see this, let us now consider $A_\varepsilon(t)$ throughout trial $\mathcal{E}_1(w, W; L)$, where as usual $w \leqslant W$; that is, throughout the time interval $(0, (L - 1)w + W)$. It suffices to consider the interval $(Lw, (L - 1)w + W)$. This is readily done, since all $x_i(t) \equiv 0$, $i = L + 1, \ldots, n$, by (7), and thus no new indices enter $A_\varepsilon(t)$ for t in $(Lw, (L - 1)w + W)$. Old indices continue to drop out of $A_\varepsilon(t)$, however, and consequently $|A_\varepsilon(t)|$ decreases in unit steps every w time units. We can again distinguish two cases.

Case 5. $G_\varepsilon(w) \gg 0$.

The steady-state phase of $A_\varepsilon(t)$ is followed by a second transient phase during the times t at which $|A_\varepsilon(t)|$ decreases in unit steps at the rate w.

Case 6. $G_\varepsilon(w) \cong 0$.

Since $|A_\varepsilon(t)|$ is always small in $[0, Lw]$, any decrease in $|A_\varepsilon(t)|$ due to an uncompensated dropping out of indices is negligible, and so once again $A_\varepsilon(t)$ has essentially only one phase.

We now define the *(dynamical) end* of a list as the set of symbols, if any, whose indices appear in $A_\varepsilon(t)$ during its *second* transient phase. We denote the symbols in the dynamical end by $E_\varepsilon \equiv E_\varepsilon(w, W; L)$. We immediately conclude that (a) learning is faster in the dynamical end of a list than in its dynamical middle. The reasoning is the same that showed the advantage of the beginning over the middle. We have hereby shown that the middle is harder to learn than the beginning and end simply by counting the number of large stimulus traces $x_i(t)$ that the associations $y_{jk}(t)$ must distinguish and correlate with $x_j(t - \tau)$ at any time t.

The distinction between a list's dynamical beginning, middle, and end is highly ambiguous when $G_\varepsilon(w) \cong 0$. This is because $A_\varepsilon(t)$ goes through essentially only one phase, and the most we can say heuristically is that *all* list symbols are either in the beginning, or the middle, or the end, and that learning is satisfactory if also $w \cong \tau$. (This statement will be modified

to take into account the numerical length L of the list when the interactions are replaced.) This ambiguity therefore implies that (b) the bowed curve flattens both at its beginning and its end as the intratrial interval increases, as Fig. 3 shows.

12. THE DEPENDENCE OF A LIST'S END ON THE INTERTRIAL INTERVAL

The intertrial interval W has a profound effect on $E_\varepsilon(w, W; L)$ because $|A_\varepsilon(t)|$ has less opportunity to decrease when W is small. For example, suppose that $W = w$. Then trial $\mathcal{E}_2(w, w; L)$ begins right after trial $\mathcal{E}_1(w, w; L)$ ends. Vertex v_1 receives its second input pulse $J(t - Lw)$ at time $t = Lw$, and thus

$$x_1(t) = \exp[-\alpha(t - Lw)]\left\{ x_1(Lw) + \int_{Lw}^{t} \exp[\alpha(v - Lw)]J(v - Lw)\, dv \right\}$$

for t in $[Lw, 2Lw]$. A change of variables shows that

$$x_1(t) = \exp[-\alpha(t - Lw)]x_1(Lw) + \exp[-\alpha(t - Lw)]\int_{0}^{t-Lw} \exp(\alpha v)J(v)\, dv.$$

$$(11)$$

In particular

$$x_1(t) \geqslant \exp[-\alpha(t - Lw)]\int_{0}^{t-Lw} \exp(\alpha v)J(v)\, dv$$

$$= x_1(t - Lw).$$

The index 1 therefore enters $A_\varepsilon(t)$ on trial $\mathcal{E}_2(w, w; L)$ not longer than w time units after L enters $A_\varepsilon(t)$ on trial $\mathcal{E}_1(w, w; L)$. Since each $x_i(t)$ with $i = 1, 2, \ldots, L$ satisfies

$$x_i(t) = x_1(t - (i - 1)w),$$

the indices $1, 2, 3, \ldots, L$ enter $A_\varepsilon(t)$ on trial $\mathcal{E}_2(w, w; L)$ in chronological order at rate w. We conclude that $|A_\varepsilon(t)|$ never has a chance to decay after the inputs of trial $\mathcal{E}_1(w, w; L)$ cease. We can readily extend this argument to show that $|A_\varepsilon(t)|$ never has a chance to decay on the first $N - 1$ trials of any serial experiment $\mathcal{E}(w, w; L)$ of N trials. The phases that occur again can be classified into two cases.

Case 7. $G_\varepsilon(w) \gg 0$.

The set $A_\varepsilon(t)$ goes through a transient phase on trial $\mathcal{E}_1(w, w; L)$ during which $|A_\varepsilon(t)|$ increases from 0 to $G_\varepsilon(w) + 1$. This phase is followed by a

steady-state phase of oscillations in $|A_\varepsilon(t)|$ as old indices leave $A_\varepsilon(t)$ and new indices enter $A_\varepsilon(t)$ in chronological order, modulo L. This phase lasts from trial $\mathcal{E}_1(w, w; L)$ through trial $\mathcal{E}_{N-1}(w, w; L)$. On trial $\mathcal{E}_N(w, w; L)$, a second transient phase is entered during which $|A_\varepsilon(t)|$ decreases monotonically to zero.

Case 8. $G_\varepsilon(w) \cong 0$.

$A_\varepsilon(t)$ goes through a single steady-state phase (see Fig. 8).
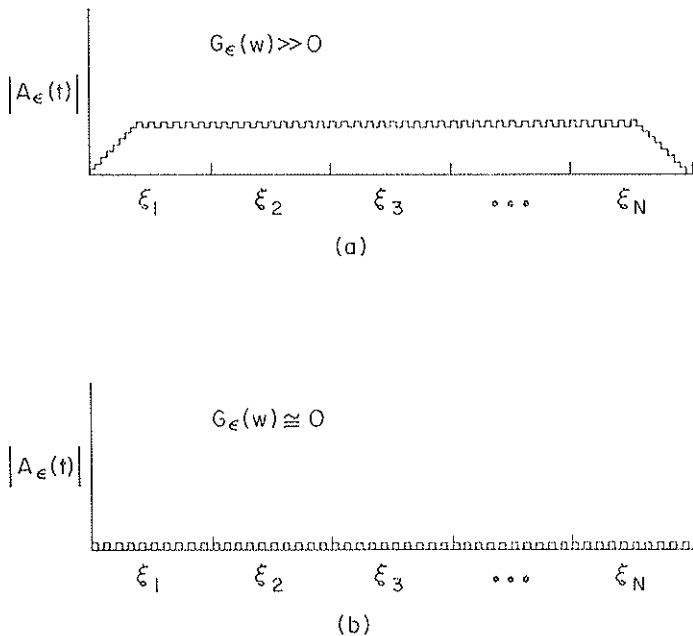

(a)

(b)

FIG. 8

Case 7 shows that a symbol that is in one dynamical part of a list on a given trial can be in a *different* part on another trial. For example, r_1 is in B_ε on trial $\mathcal{E}_1(w, w; L)$. By contrast, when r_1 enters $A_\varepsilon(t)$ on trial $\mathcal{E}_2(w, w; L)$, $|A_\varepsilon(t)| = G_\varepsilon(w) \gg 0$ and $A_\varepsilon(t)$ is in its steady-state phase. Thus r_1 is in the list's dynamical *middle* on trial $\mathcal{E}_2(w, w; L)$! In fact *all* list symbols are in the dynamical middle on trials $\mathcal{E}_i(w, w; L)$, where $i = 2, 3, \ldots, L - 1$. Because r_1 and r_2 are in B_ε on trial $\mathcal{E}_1(w, w; L)$, associations such as $y_{12}(t)$ grow quickly on this trial. But since r_1 and r_2 are in M_ε during the next few trials, the initial advantage to $y_{12}(t)$ is gradually washed away during these trials.

In Case 7, $A_\varepsilon(t)$ has the same phases no matter how many trials exist in $\mathcal{E}(w, w; L)$. Thus N presentations of $r_1 r_2 \cdots r_L$ in such an experiment are properly thought of as *one* presentation of the cyclic list

$$\underbrace{r_1 r_2 \cdots r_L r_1 \cdots r_L r_1 \cdots r_L r_1 \cdots r_L r_1 \cdots r_L.}_{N \text{ times}}$$

We now consider the effect of increasing W step by step when the list has a middle; that is, when $0 \ll G_\varepsilon(w) < L - 1$. If $W = 2w$, then $|A_\varepsilon(t)|$ decreases by 1 after its steady-state phase on trial $\mathcal{E}_1(w, 2w; L)$.
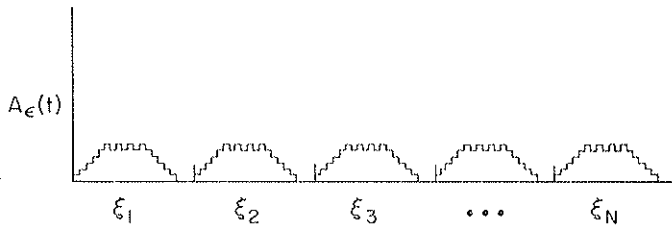

FIG. 9

Trial $\mathcal{E}_2(w, 2w; L)$ then begins and $|A_\varepsilon(t)|$ quickly rises once again to its steady-state phase. The advantage to $y_{12}(t)$ of trial $\mathcal{E}_1(w, 2w; L)$ is thus not entirely destroyed on trial $\mathcal{E}_2(w, 2w; L)$, but the advantage to $y_{23}(t)$ on trial $\mathcal{E}_2(w, 2w; L)$ is slight.

Let us now increase W step by step. Then for a fixed value of W, $|A_\varepsilon(t)|$ decreases step by step at a rate w for t in $(Lw, (L-1)w + W)$ to a minimum value of

$$\max\left\{ 0, G_\varepsilon(w) + 1 - \left[\frac{W}{w}\right] \right\}.$$

If W is chosen so large that

$$\left[\frac{W}{w}\right] = G_\varepsilon(w) + 1,$$

then $|A_\varepsilon(t)|$ decreases to 0 before trial $\mathcal{E}_2(w, W; L)$ begins. $A_\varepsilon(t)$ will therefore have essentially the same phases on trials $\mathcal{E}_2(w, W; L)$ through $\mathcal{E}_N(w, W; L)$ as it had on trial $\mathcal{E}_1(w, W; L)$. In particular, symbols that are in B_ε, M_ε, or E_ε on one trial will be in the same dynamical part of the list on all trials, and the effects on associations that characterize a given list part will have a cumulative effect on them as more and more trials occur (see Fig. 9).

Using these remarks, we can easily account for the major bowing effects of Fig. 3 in $\mathcal{M}(\mathcal{A})$. For example,

(a) if $L \cong 2$ and $w \cong \tau$, bowing does not occur, since $|A_\varepsilon(t)|$ is always small;

(b) bowing occurs only when $0 \ll L \leqslant G_\varepsilon(w)$ or $0 \ll G_\varepsilon(w) < L$, since only then does $|A_\varepsilon(t)|$ ever achieve large values;

(c) the bowed curve is flattened, but *raised*, when $0 \ll L \leqslant G_\varepsilon(w)$ or $0 \ll G_\varepsilon(w) < L$ if $W \cong w$, since then all list symbols are usually in M_ε;

(d) if for fixed $W > w$ bowing does occur, then increasing W lowers the bowed curve near its numerical middle by increasing the numerical length of B_ε and E_ε;

(e) increasing W by a fixed amount has less of a lowering effect if w is large than if w is small, because $G_\varepsilon(w)$ is monotone decreasing in w;

(f) for fixed w, increasing W beyond a W_0 such that

$$\left[\frac{W_0}{w} \right] = G_\varepsilon(w) + 1$$

has little lowering effect on the bowed curve since $|A_\varepsilon(t)|$ decays to zero at the end of each trial for *all* such W; and

(g) if bowing occurs but $1 \ll L - 1 < G_\varepsilon(w)$, then increasing the list's numerical length L while keeping w and W fixed can decrease the skewness of the bowed curve by increasing the numerical length of M_ε.

13. RESPONSE OSCILLATION AND REMOTENESS

We now have enough information at our disposal to qualitatively understand how some features of the response oscillation curve of Fig. 4 arise in $\mathcal{M}(\mathcal{A})$. Suppose, for example, that $1 \ll G_\varepsilon(w) < L - 1$ and that W is sufficiently large for some bowing to occur.

At times t when $|A_\varepsilon(t)|$ is small, the formation of new associations will be restricted to a small number of indices. Thus learning will begin to show its effects faster in B_ε and E_ε than in M_ε and competing responses are restricted to a relatively small set of list symbols. By contrast, for r_i in M_ε, i is in $A_\varepsilon(t)$ when $|A_\varepsilon(t)|$ is large. Competing responses to a symbol in M_ε are therefore broadly distributed throughout the list. Learning therefore takes relatively long to show its effects in M_ε, and a long time is needed to eliminate the large collection of competing responses after learning begins. These are the main effects of Fig. 4.

These remarks are all special cases of the heuristic observation that associations between symbols become harder to form as the "dynamical remoteness" of the symbols increases. A convenient measure of the *dynamical remoteness* $R_{ij}^{(\varepsilon)}$ from r_i to r_j on an experiment consisting of N trials is

$$R_{ij}^{(\varepsilon)} = \frac{N[(L-1)w + W]}{\sum_{k=1}^{N} U_{ij}^{(\varepsilon)}(k)}$$

where $U_{ij}^{(\varepsilon)}(k)$ is the total amount of time during which i and j are *both* in $A_\varepsilon(t)$ on trial $\delta_k(w, W; L)$. We can now say that (a) the very remote errors of a symbol in the middle of a long list are approximately equally probable in both the anticipatory and the preseverative directions. For example, if $0 \ll G_\varepsilon(w) \ll L$ and

$$\left[\frac{W}{w}\right] \cong G_\varepsilon(w) + 1,$$

then for any r_i in M_ε

$$R_{i1}^{(\varepsilon)} = R_{iL}^{(\varepsilon)} = \infty,$$

since neither $x_1(t)$ and $x_i(t)$, nor $x_L(t)$ and $x_i(t)$, are ever large simultaneously. Nonetheless, given the same choices of w, W, and L, (b) nonremote errors of a symbol in the middle of a long list are stronger in the anticipatory than in the perseverative direction, as was shown in the discussion of backward learning in [5].

These conclusions do not hold for symbols in B_ε and E_ε, however. For example, given the same w, W, and L,

$$R_{12}^{(\varepsilon)} \ll \infty, \qquad \text{whereas} \qquad R_{1L}^{(\varepsilon)} = \infty.$$

Thus (c) anticipatory errors are more probable than perseverative errors at the beginning of a long list if significant bowing occurs.

By contrast,

$$R_{L,L-1}^{(\varepsilon)} \ll \infty, \qquad \text{whereas} \qquad R_{L1}^{(\varepsilon)} = \infty,$$

so that (d) perseverative errors are more probable at the end of a long list in which bowing occurs than they are at the beginning of the list. The anticipatory bias of B_ε and the perseverative bias of E_ε are mirrored by the fact that $|A_\varepsilon(t)|$ *increases* in B_ε and *decreases* in E_ε. That is, the behavior of $|A_\varepsilon(t)|$ in E_ε is (roughly) the "time-reversed" behavior of $|A_\varepsilon(t)|$ in B_ε.

Of course, when (say) $W = w$ and $L - 1 > G_\varepsilon(w) \gg 1$, the conclusions above are no longer valid, since then all symbols are in M_ε at practically all times during $\delta(w, W; L)$.

14. ASSOCIATIONAL STRENGTHS IN THE BARE FIELD

We are now ready to introduce the associational strengths into the bare field. That is, we place $y_{jk}(t)$ at the arrowhead N_{jk} and let it compute the $x_j(t - \tau)$ and $x_m(t)$ pulses, $m \neq j$, without allowing these pulses to interact along the edges. By this we mean that Eqs. (7), (2), (3), and (4) hold for $i, j, k = 1, 2, \ldots, n$.

We also let $u = 0$. The reason for this is readily seen as follows. Let all $x_i(t) = 0$. Then (3) becomes

$$\dot{z}_{jk}(t) = -u z_{jk}(t),$$

so that if $u > 0$, $z_{jk}(t) = z_{jk}(0)e^{-ut}$. Thus by (2), $y_{jk}(t) = y_{jk}(0)$ for all $t \geqslant 0$. In other words, if all stimulus traces are zero, then the memory of $\mathcal{M}(\mathcal{A})$ is perfect. On the other hand, each $z_{jk}(t)$ decays exponentially at a rate u, which creates the paradoxical situation that the *observable* associations $y_{jk}(t)$ stay fixed whereas the unobservable correlations $z_{jk}(t)$ decay. The choice $u = 0$ eliminates this difficulty by keeping both $z_{jk}(t)$ and $y_{jk}(t)$ constant when all stimulus traces $x_i(t)$ are zero. The use of $y_{jk}(t)$ instead of $z_{jk}(t)$ as an association is helpful to guarantee that $\mathcal{M}(\mathcal{A})$ be able to learn lists *perfectly* given enough practice, as shown in [5]. The $y_{jk}(t)$'s accomplish this by adding a mutual inhibition to the associations leading from v_j, since *increasing* $y_{jk}(t)$ *decreases* $y_{jm}(t)$, $m \neq k$. In [6], this inhibition between associations is replaced by lateral inhibition coupled to the $z_{jk}(t)$'s, which is the finite-rate analog of mutual inhibition between the $y_{jk}(t)$'s. When this is done, the two functions $y_{jk}(t)$ and $z_{jk}(t)$ are merged into one, and the constancy of either function therefore automatically implies that of the other. The choice $u = 0$ comes as close to this merging of functions as the present level of approximation permits.

Our previous remarks concerning accumulation sets have used the hypothesis that $y_{jk}(t)$ compares $x_j(t - \tau)x_k(t)$ with all $x_j(t - \tau)x_m(t)$, $m \neq j$. The next lemma makes this assertion precise, and will permit a somewhat more exact analysis of bowing than that provided by accumulation sets.

LEMMA 1. *For any distinct values of j and k,*

$$y_{jk}(t) = \frac{y_{jk}(0) + K_j \int_0^t x_j(v - \tau)x_k(v)\, dv}{1 + K_j \int_0^t x_j(v - \tau)x^{(j)}(v)\, dv} \tag{13}$$

where $x^{(j)} = \sum_{m \neq j} x_m$ and $K_j = \beta [\sum_{m \neq j} z_{jm}(0)]^{-1}$.

Proof. Let $u = 0$ and integrate (3). Then

$$z_{jk}(t) = z_{jk}(0) + \beta \int_0^t x_j(v - \tau)x_k(v)\, dv. \tag{14}$$

Letting $z^{(j)}(t) = \sum_{m \neq j} z_{jm}(t)$ and $x^{(j)}(t) = \sum_{m \neq j} x_m(t)$, we find that (14) implies

$$z^{(j)}(t) = z^{(j)}(0) + \beta \int_0^t x_j(v - \tau)x^{(j)}(v)\, dv. \tag{15}$$

By (2) and (4), $y_{jk}(t) = z_{jk}(t)[z^{(j)}(t)]^{-1}$, which by (14) and (15) becomes

$$y_{jk}(t) = \frac{z_{jk}(0) + \beta \int_0^t x_j(v - \tau)x_k(v)\, dv}{z^{(j)}(0) + \beta \int_0^t x_j(v - \tau)x^{(j)}(v)\, dv}. \tag{16}$$

Division of numerator and denominator in (16) by $z^{(j)}(0)$ yields (13).

Equation (13) shows that $y_{jk}(t)$ compares $x_j(v - \tau)x_k(v)$ with each $x_j(v - \tau)x_m(v)$, $m \neq j$, for *all* $v \leqslant t$. Our analysis using accumulation sets considered explicitly only the value $v = t$. This analysis is nonetheless qualitatively correct because it claims that $y_{jk}(t)$ will grow substantially at *any* time for which

$$x_j(t - \tau)x_k(t) \gg x_j(t - \tau)x_m(t), \qquad m \neq j, k,$$

as (13) shows. Equation (13) also claims that the changes during *all* such values of t (namely, all $v \leqslant t$) are cumulative.

The system (7), (2), (3), (4) can be solved to yield explicit formulas for the associations $y_{jk}(t)$ in terms of the input pulses $I_i(v)$, $v \leqslant t$. Before deriving these formulas, we show heuristically how such phenomena as bowing occur in the associations $y_{jk}(t)$, using an argument that naturally extends the idea of accumulation sets. We therefore consider a list $\mathcal{L} = r_1 r_2 \cdots r_L$ presented to $\mathcal{M}(\mathcal{A})$ with an intratrial interval $w = \tau$ that satisfies $1 \ll G_\varepsilon(w) < L - 1$. We also let all $y_{jk}(0) = 1/(n - 1)$, $j \neq k$, and $K_i = K$ to start $\mathcal{M}(\mathcal{A})$ in a state of maximal ignorance.

15. TWO FACTORS YIELDING DECREASE IN ASSOCIATIONAL STRENGTH WITH INCREASED REMOTENESS

Consider the associations $y_{1i}(t)$, $i = 2, 3, \ldots, L$ on trial $\mathcal{E}_1(\tau, W; L)$ for t in $[0, L\tau)$.

A. *The Time Interval* $[0, \tau)$

Symbol r_1 is presented to $\mathcal{M}(\mathcal{A})$ at time $t = 0$ and creates a positive input signal $I_1(t) = J(t)$ at v_1 within the time interval $(0, \lambda)$. Point strength $x_1(t)$ thereupon grows and equal signals $\beta x_1(t)$ are sent to all nodes N_{1j}, $j \neq 1$. These signals reach the nodes N_{1j}, $j \neq 1$, at time $t = \tau$. By (7), only $x_1(t)$ is positive for t in $[0, \tau)$. Thus by (13), all $y_{1i}(t) = 1/(n - 1)$ for t in $[0, \tau)$, and no associational changes occur.

B. *The Time Interval* $[\tau, 2\tau)$

Symbol r_2 is presented to $\mathcal{M}(\mathcal{A})$ at time $t = \tau$ and creates a positive input signal $I_2(t) = J(t - \tau)$ at v_2 within the time interval $(\tau, \tau + \lambda)$. Point strength $x_2(t)$ thereupon grows. By (7), $x_1(t)$ is positive for all $t > 0$ and $x_2(t)$ is positive for all $t > \tau$. Since the signal $I_1(t)$ to v_1 is large for t in $(0, \lambda)$, $x_1(t)$ achieves its largest values in $(0, \lambda)$ and thereafter decays at the exponential rate α. Similarly, $x_2(t)$ achieves its largest values in $(\tau, \tau + \lambda)$ and thereafter decays exponentially. In particular $x_1(t - \tau)$ and $x_2(t)$ achieve their largest values during the common time interval $(\tau, \tau + \lambda)$, and thus $x_1(t - \tau)x_2(t)$ is large during this interval. By contrast, all $x_j(t)$ with $j \neq 1, 2$ remain equal to zero for $t < 2\tau$, as (7) shows. Thus for $t < 2\tau$, (13) yields

$$y_{12}(t) = \frac{(n - 1)^{-1} + K \int_{\tau}^{t} x_1(v - \tau)x_2(v)\, dv}{1 + K \int_{\tau}^{t} x_1(v - \tau)x_2(v)\, dv}, \tag{17}$$

and $y_{12}(t)$ grows substantially for $\tau < t < 2\tau$. All other $y_{1i}(t)$ decrease by equal amounts for $\tau < t < 2\tau$ since $x_i(t) = 0$ during these times.

C. *The Time Interval* $[2\tau, 3\tau)$

Point strength $x_3(t)$ becomes positive for $t > 2\tau$ because r_3 is presented to $\mathcal{M}(\mathcal{A})$ at time $t = 2\tau$. $x_3(t)$ achieves its largest values for t in $(2\tau, 2\tau + \lambda)$, by the usual argument. Only $x_1(t)$, $x_2(t)$, and $x_3(t)$ are positive for $t < 3\tau$, and thus by (13),

$$y_{13}(t) = \frac{(n - 1)^{-1} + K \int_{2\tau}^{t} x_1(v - \tau)x_3(v)\, dv}{1 + K \left[\int_{\tau}^{t} x_1(v - \tau)x_2(v)\, dv + \int_{2\tau}^{t} x_1(v - \tau)x_3(v)\, dv \right]} \tag{18}$$

for $2\tau < t < 3\tau$. We now show that $y_{13}(t)$ grows less for $2\tau < t < 3\tau$ than $y_{12}(t)$ did for $\tau < t < 2\tau$. This is due to two factors.

1. *Competition by* $y_{12}(t)$. Just as the growth of $y_{12}(t)$ in (17) is accomplished by increasing

$$\int_{\tau}^{t} x_1(v - \tau)x_2(v)\, dv, \qquad (19)$$

the growth of $y_{13}(t)$ in (18) is accomplished by increasing

$$\int_{2\tau}^{t} x_1(v - \tau)x_3(v)\, dv. \qquad (20)$$

The term (19) in (18), however, diminishes $y_{13}(t)$ growth, and no such competitive term diminishes $y_{12}(t)$ in (17) for $\tau < t < 2\tau$.

2. *Poor correlation of stimulus traces.* The process $x_1(v - \tau)$ attains its largest values for v in $(\tau, \tau + \lambda)$ and decays exponentially thereafter, whereas $x_3(v)$ attains its largest values in $(2\tau, 2\tau + \lambda)$, where $2\tau > \tau + \lambda$. Thus the overlap between large $x_1(v - \tau)$ and $x_3(v)$ values in (18) is not as good as the overlap between large $x_1(v - \tau)$ and $x_2(v)$ values in (17). Since also $x_3(v) = x_2(v - \tau)$, the boost by $x_1(v - \tau)x_3(v)$ to $y_{13}(t)$ growth for t in $[2\tau, 3\tau)$ is not as large as the boost by $x_1(v - \tau)x_2(v)$ to $y_{12}(t)$ growth for t in $[\tau, 2\tau)$.

We conclude that the "higher-order" association $y_{13}(t)$ that forms for t in $[2\tau, 3\tau)$ is not as strong as the "correct" association $y_{12}(t)$.

D. *The Time Interval* $[3\tau, 4\tau)$

For $3\tau < t < 4\tau$, the association $y_{14}(t)$ also receives a boost, since by (13),

$$y_{14}(t) = \frac{(n - 1)^{-1} + K\int_{0}^{t} x_1(v - \tau)x_4(v)\, dv}{1 + K\int_{0}^{t} x_1(v - \tau)(x_2(v) + x_3(v) + x_4(v))\, dv} \qquad (21)$$

for t in $[3\tau, 4\tau)$, and $x_4(v)$ becomes positive after r_4 is presented at time $t = 3\tau$. Nonetheless, $y_{14}(t)$ grows even less than $y_{13}(t - \tau)$ did both because (i) the overlap between the large positive values of $x_1(v - \tau)$ and $x_4(v)$ is worse still than that between $x_1(v - \tau)$ and $x_3(v)$, and (ii) $x_1(v - \tau)x_4(v)$ must compete with the *two* products $x_1(v - \tau)x_2(v)$ and $x_1(v - \tau)x_3(v)$ in the denominator of (21) during its period of maximal growth.

By continuing this argument in successive intervals of the form $[(m - 1)\tau, m\tau)$, where $m \leqslant L$, we find that the strength of associations $y_{1m}(t)$ decreases as m increases; that is, decreases with an increase in the *remoteness* of the list positions r_1 and r_m, both because of an increasingly bad overlap of the stimulus traces $x_1(v - \tau)$ and $x_m(v)$ as m increases, and because the correlation $x_1(v - \tau)x_m(v)$ must compete with all correlations $x_1(v - \tau)x_i(v)$, $i = 2, 3, \ldots, m - 1$, during its period of maximal growth. This increased competition between correlations is interpreted as increased lateral inhibition between the v_i, $i = 1, 2, \ldots, m$, in [6].

16. THE LISTS'S DYNAMICAL MIDDLE: INHIBITION BY BOTH PAST AND FUTURE FIELDS

Consider associations $y_{Ri}(t)$ for all $i \neq R$ where R is fixed once and for all to satisfy $G_\varepsilon(\tau) + 1 \leqslant R < L - 1$ and t varies in $[0, L\tau)$; that is, v_R is in the list's dynamical middle within $[0, L\tau)$.

Element r_R is first presented to $\mathcal{M}(\mathcal{A})$ at time $t = (R - 1)\tau$. Since $I_R(t)$ equals $J(t - (R - 1)\tau)$, $I_R(t)$ is positive for t in $((R - 1)\tau, (R - 1)\tau + \lambda)$. Thus by (7), $x_R(t)$ becomes positive after $t = (L - 1)\tau$ and attains its largest values before $t = (L - 1)\tau + \lambda$, after which $x_R(t)$ decays at an exponential rate α. By (13), the various associations $y_{Ri}(t)$ cannot change until $t > R\tau$. Since $x_{R+1}(t)$ becomes positive after $t = R\tau$, and all $x_i(t)$, $i = 1, 2, \ldots, R - 1$, become positive before $t = R\tau$, all the associations $y_{Ri}(t)$, $i = 1, 2, \ldots, R - 1$, $R + 1$, are susceptible to change for $t > R\tau$. Indeed *only* these associations, for fixed R, can change for t in $(R\tau, (R + 1)\tau)$. By (13), the "correct" association $y_{R,R+1}(t)$ satisfies the equation

$$y_{R,R+1}(t) = \frac{(n - 1)^{-1} + K \int_{R\tau}^{t} x_R(v - \tau)x_{R+1}(v)\,dv}{1 + K \int_{R\tau}^{t} x_R(v - \tau)[\sum_{i=1}^{R-1} x_i(v) + x_{R+1}(v)]\,dv} \tag{22}$$

for t in $(R\tau, (R + 1)\tau)$, whereas the "backward" associations $y_{Ri}(t)$, $i = 1, 2, \ldots, R - 1$, satisfy

$$y_{Ri}(t) = \frac{(n - 1)^{-1} + K \int_{R\tau}^{t} x_R(v - \tau)x_i(v)\,dv}{1 + K \int_{R\tau}^{t} x_R(v - \tau)[\sum_{i=1}^{R-1} x_i(v) + x_{R+1}(v)]\,dv} \tag{23}$$

for t in $(R\tau, (R + 1)\tau)$. The interval $(R\tau, (R + 1)\tau)$ is obviously the interval of maximal overlap of the stimulus traces $x_R(v - \tau)$ and $x_{R+1}(v)$ and is therefore the interval of maximal growth of $y_{R,R+1}(t)$. By (20) and (21), however, the *backward* associations $y_{Ri}(t)$, $i = 1, 2, \ldots, R - 1$, seriously compete with $y_{R,R+1}(t)$ growth in this interval. Thus, although the overlap between $x_R(v - \tau)$ and $x_{R+1}(v)$ is just as good for v in $(R\tau, (R + 1)\tau)$ as was the overlap between $x_1(v - \tau)$ and $x_2(v)$ for v in $(\tau, 2\tau)$, the growth of $y_{R,R+1}(t)$ for t in $(R\tau, (R + 1)\tau)$ is less than that for $y_{12}(t)$ in $(\tau, 2\tau)$ due to the added competition of backward associations; that is, due to increased lateral inhibition by previously excited stimulus traces (or membrane potentials).

Clearly, the backward associations $y_{Ri}(t)$, $1 \leqslant i \leqslant R - 1$, that compete most vigorously with $y_{R,R+1}(t)$ are those for which $|i - R|$ is small (i.e., "contiguous" associations), since the overlap between $x_R(v - \tau)$ and $x_i(v)$ gets worse as $|i - R|$ increases.

The associations $y_{1,1+i}(t)$, $i \geqslant 1$, at the beginning of the list thus grow faster than the associations $y_{R,R+i}(t)$, $i \geqslant 1$, near the middle of the list simply because there are more competing backward associations when the middle of the list is presented. This fact can be conveniently summarized using the following terminology.

For fixed $\varepsilon > 0$, let the set of indices

$$P_\varepsilon(R) = \bigcup \{A_\varepsilon(t) \cap \{1, 2, \ldots, R - 1\} : R \in A_\varepsilon(t),$$
$$0 \leqslant t \leqslant (L - 1)w + W\},$$

be the *past field* of r_R (or alternatively of $y_R(t)$) on trial $\mathcal{E}_1(w, W; L)$, and let

$$F_\varepsilon(R) = \bigcup \{A_\varepsilon(t) \cap \{R + 1, R + 2, \ldots, L\} : R \in A_\varepsilon(t),$$
$$0 \leqslant t \leqslant (L - 1)w + W\},$$

be the *future field* of r_R (or alternatively of $y_R(t)$) on trial $\mathcal{E}_1(w, W; L)$. The set $P_\varepsilon(R)$ tells how many "past" $x_i(t)$, $1 \leqslant i \leqslant R - 1$, are large when $x_R(t)$ is large on trial $\mathcal{E}_1(w, W; L)$, whereas $F_\varepsilon(R)$ tells how many "future" $x_i(t)$, $R + 1 \leqslant i \leqslant L$, are large when $x_R(t)$ is large on trial $\mathcal{E}_1(w, W; L)$. Correct associations $y_{R,R+1}(t)$ at the beginning of the list are larger than correct associations in the middle of the list simply because only correct associations at the middle must compete with associations leading to the past field, whereas associations in the beginning and middle of the list must compete with the future field.

17. DECREASE OF INHIBITION BY THE FUTURE FIELD FOR ASSOCIATIONS
    AT THE LIST'S DYNAMICAL END

Consider the associations $y_{L-1,i}(t)$, $i = 1, 2, \ldots, L - 2, L$, for t in $[0, L\tau)$. By (13), these associations remain constant until $x_{L-1}(v - \tau)$ becomes positive at $v = (L - 1)\tau$. For t in $((L - 1)\tau, L\tau)$, (13) implies

$$y_{L-1,L}(t) = \frac{(n - 1)^{-1} + K \int_{(L-1)\tau}^{t} x_{L-1}(v - \tau) x_L(v) \, dv}{1 + K \int_{(L-1)\tau}^{t} x_{L-1}(v - \tau) [\sum_{i=1}^{L-2} x_i(v) + x_L(v)] \, dv},$$

which shows that the inhibition due to backward associations $y_{ij}(t)$, $j = 1, 2, \ldots, i - 1$, during the interval $(i\tau, (i + 1)\tau)$ of $y_{i,i+1}(t)$'s greatest growth is *maximal* if $i = L - 1$. In other words, until t exceeds $L\tau$, the associations $y_{L-1,i}(t)$ are in the dynamical *middle* of the list. It is not until after *all* list items are presented to $\mathcal{M}(\mathcal{A})$ and an extra w time units go by that $\mathcal{M}(\mathcal{A})$ can possibly know that r_L is the last item in $\mathcal{L}$. Viewed by a psychological experimenter, the fact that *later* information affects $\mathcal{M}(\mathcal{A})$'s processing of a *past* event (past on the psychologist's time scale!) means that a dynamical "time reversal" has occurred.

The inhibition that the past field of r_{L-1} exerts on $y_{L-1,L}(t)$ for t in $((L - 1)\tau, L\tau)$ will be comparable to the inhibition exerted by the past field of r_i on $y_{i,i+1}(t)$ for t in $(i\tau, (i + 1)\tau)$ for all i such that $G_\varepsilon(w) < i < L - 1$, since the maximal size of $|A_\varepsilon(t)|$ is $G_\varepsilon(w) + 1$, and after this size is reached $A_\varepsilon(t)$ adds and subtracts competing vertices one at a time in chronological order.

Now consider $y_{L-1,L}(t)$ for t in $(L\tau, (L - 1)\tau + W)$. *No new indices* enter the future field of r_{L-1} during these times. Thus, although associations in the list's middle and end can receive comparable amounts of inhibition from their respective past fields, the future field of a position at the list's end exerts less competition than a position in the list's middle. Of course, the reader must qualify these remarks by always choosing W sufficiently large, or else new indices will enter the future field of r_{L-1} from the list's numerical beginning on trial $\mathcal{E}_2(w, W; L)$ before the indices remaining from trial $\mathcal{E}_1(w, W; L)$ can decay out of sight.

In summary, the heuristical explanation of bowing on trial $\mathcal{E}_1(w, W; L)$ becomes: (a) associations in the dynamical beginning must compete primarily with their future field; (b) associations in the dynamical end must compete primarily with their past field; whereas (c) associations in

the dynamical middle must compete both with their past and future fields.

The three major heuristic ideas behind these conclusions are (1) presentation of list items creates fluctuations in stimulus traces that persist in time; (2) associations correlate these stimulus traces; and (3) associations mutually inhibit one another, by a process that is identified in [6] with lateral inhibition.

The details of our analysis require only an estimate of how good the overlap between correlated stimulus traces is and, relatively speaking, how much inhibition is created by stimulus traces that are simultaneously large. With these ideas in mind, the reader can now easily work out relative sizes of the remaining associations $y_{jk}(t)$ on trial $\mathcal{E}_1(w, W; L)$.

Before proceeding to trial $\mathcal{E}_2(w, W; L)$, we present the formulas for all $y_{jk}(t)$ on trial $\mathcal{E}_1(w, W; L)$. Again we choose $w = \tau$, and for simplicity we let $W = rw$ for some $r \geqslant 1$. It then suffices to compute $y_{jk}(m\tau)$ for all m such that $0 \leqslant m \leqslant L + r - 1$ and all $j \neq k$, $j, k = 1, 2, \ldots, L$.

THEOREM 1.   *Let* $u = 0$, $\lambda < \tau$,

$$\Lambda = \int_0^\lambda e^{\alpha\xi} J(\xi)\, d\xi, \tag{24}$$

$$H = \int_0^\lambda e^{-2\alpha v} \int_0^v e^{\alpha\xi} J(\xi)\, d\xi\, dv, \tag{25}$$

and

$$G = \int_0^\lambda e^{-2\alpha v} \left(\int_0^v e^{\alpha\xi} J(\xi)\, d\xi \right)^2 dv. \tag{26}$$

Then on trial $\mathcal{E}_1(\tau, r\tau; L)$,

$$y_{jk}(m\tau) = \frac{(n-1)^{-1} + K f_{jk}(m\tau)}{1 + K(g_j(m\tau) + f_{j,j+1}(m\tau) + h_j(m\tau))}, \tag{27}$$

where for all $j \neq k$

$$f_{jk}(m\tau) = 0 \qquad\qquad\qquad \text{if } j \geqslant m \text{ or } k \geqslant m + 1,$$

$$= \Lambda \exp[-(j - k + 1)\alpha\tau]$$

$$\times \left(H + \frac{\Lambda}{2\alpha} \{\exp(-2\alpha\lambda) - \exp[-(m - j)2\alpha\tau]\} \right)$$

$$\text{if } k < j < m,$$

$$= G + \frac{\Lambda^2}{2\alpha} \{\exp(-2\alpha\lambda) - \exp[-(m-j)2\alpha\tau]\}$$

$$\text{if } k - 1 = j < m,$$

$$= \Lambda \exp[-(k-j-1)\alpha\tau]$$

$$\times \left(H + \frac{\Lambda}{2\alpha} \{\exp(-2\alpha\lambda) - \exp[-(m-k+1)2\alpha\tau]\} \right)$$

$$\text{if } j < k - 1 < m;$$

$$g_j(m\tau) = 0 \qquad \text{if } j \geqslant m,$$

$$= \left[\frac{\exp(-\alpha\tau) - \exp(-j\alpha\tau)}{\exp(\alpha\tau) - 1} \right]$$

$$\times \left(\Lambda H + \frac{\Lambda^2}{2\alpha} \{\exp(-2\alpha\lambda) - \exp[-(m-j)2\alpha\tau]\} \right)$$

$$\text{if } m > j;$$

and

$$h_j(m\tau) = 0 \qquad \text{if } j + 1 \geqslant m,$$

$$= \left\{ \frac{\exp(-\alpha\tau) - \exp[(j-m+1)\alpha\tau]}{1 - \exp(-\alpha\tau)} \right\} \left[\Lambda H + \frac{\Lambda^2}{2\alpha} \exp(-2\alpha\lambda) \right]$$

$$- \frac{\Lambda^2}{2\alpha} \left\{ \frac{\exp[(j-m-1)\alpha\tau] - \exp[(2j-2m+1)\alpha\tau]}{\exp(\alpha\tau) - 1} \right\}$$

$$\text{if } j < m \leqslant L + 1,$$

$$= \left\{ \frac{\exp(-\alpha\tau) - \exp[(j-L)\alpha\tau]}{1 - \exp(-\alpha\tau)} \right\} \left[\Lambda H + \frac{\Lambda^2}{2\alpha} \exp(-2\alpha\lambda) \right]$$

$$- \frac{\Lambda^2}{2\alpha} \exp(-m2\alpha\tau) \left\{ \frac{\exp[(j+L)\alpha\tau] - \exp[(2j+1)\alpha\tau]}{\exp(\alpha\tau) - 1} \right\}$$

$$\text{if } j < m > L.$$

Proof. By (13),

$$y_{jk}(m\tau) = \frac{(n-1)^{-1} + K \int_0^{m\tau} x_j(v-\tau)x_k(v) \, dv}{1 + K \sum_{i \neq j} \int_0^{m\tau} x_j(v-\tau)x_i(v) \, dv} .$$

It therefore suffices to compute the integrals

$$f_{jk}(t) \equiv \int_0^t x_j(v-\tau)x_k(v) \, dv \tag{28}$$

at values $t = m\tau$ for all $j \neq k$. This is done using simple changes of variable along with the identities

$$x_j(t) = x_1(t - (j - 1)\tau), \qquad j = 1, 2, \ldots, L, \tag{29}$$

and

$$x_1(t) = \begin{cases} 0, & t \leqslant 0, \\ e^{-\alpha t} \int_0^t e^{\alpha v} J(v) \, dv, & 0 \leqslant t \leqslant (L + r - 1)\tau, \end{cases} \tag{30}$$

of Proposition 1. By (29), (28) can be written

$$f_{jk}(t) = \int_0^t x_1(v - j\tau) x_1(v - (k - 1)\tau) \, dv. \tag{31}$$

Suppose, for example, that $k < j$. Then by (30), $f_{jk}(t) = 0$ for $t < j\tau$, whereas for $t > j\tau$,

$$f_{jk}(t) = \int_{j\tau}^t x_1(v - j\tau) x_1(v - (k - 1)\tau) \, dv$$

$$= \int_0^{t-j\tau} x_1(v) x_1(v + (j - k + 1)\tau) \, dv,$$

which by (30) yields

$$f_{jk}(t) = \exp[-(j - k + 1)\alpha\tau] \int_0^{t-j\tau} \exp(-2\alpha v)$$

$$\times \int_0^v \exp(\alpha\xi) J(\xi) \, d\xi \int_0^{v+(j-k+1)\tau} \exp(\alpha\eta) J(\eta) \, d\eta \, dv. \tag{32}$$

In (32), $J(\eta) = 0$ for $\eta > \lambda$, where $\lambda < \tau$, and $v + (j - k + 1)\tau > \tau$ for all $v \geqslant 0$. Thus by (24)

$$f_{jk}(t) = \Lambda \exp[-(j - k + 1)\alpha\tau] \int_0^{t-j\tau} \exp(-2\alpha v) \int_0^v \exp(\alpha\xi) J(\xi) \, d\xi \, dv.$$

Now let $t = m\tau$, where $m > \tau$. Since $J(\xi)$ is positive only for ξ in $(0, \lambda)$,

$$\int_0^v e^{\alpha\xi} J(\xi) \, d\xi = \Lambda \qquad \text{for all } v \geqslant \lambda.$$

Thus

$$f_{jk}(m\tau) = \Lambda \exp[-(j - k + 1)\alpha\tau]$$
$$\times \left[\int_0^\lambda + \int_\lambda^{(m-j)\tau}\right] \exp(-2\alpha v)\int_0^v \exp(\alpha\xi)J(\xi)\,d\xi\,dv,$$

which by (25) is the same as

$$f_{jk}(m\tau) = \Lambda \exp[-(j - k + 1)\alpha\tau]$$
$$\times \left(H + \frac{\Lambda}{2\alpha}\{\exp(-2\alpha\lambda) - \exp[-(m - j)\alpha\tau]\}\right).$$

The case $k = j + 1$ is treated similarly. Indeed,

$$f_{j,j+1}(t) = \int_0^t x_j(v - \tau)x_{j+1}(v)\,dv$$

$$= \int_0^t x_{j+1}^2(v)\,dv$$

$$= \int_{j\tau}^t x_1^2(v - j\tau)\,dv$$

$$= \int_0^{t-j\tau} x_1^2(v)\,dv$$

$$= \int_0^{t-j\tau} e^{-2\alpha v}\left(\int_0^v e^{\alpha\xi}J(\xi)\,d\xi\right)^2 dv.$$

Let $t = m\tau$ with $m > j$. Then by (24),

$$f_{j,j+1}(m\tau) = \left[\int_0^\lambda + \int_\lambda^{(m-j)\tau}\right] \exp(-2\alpha v)\left(\int_0^v \exp(\alpha\xi)J(\xi)\,d\xi\right)^2 dv$$

$$= G + \frac{\Lambda^2}{2\alpha}\{\exp(-2\alpha\lambda) - \exp[-(m - j)\alpha\tau]\}.$$

The case $k > j - 1$ differs from the preceding cases only in that the integral

$$f_{jk}(t) = \int_{(k-1)\tau}^t x_1(v - j\tau)x_1(v - (k - 1)\tau)\,dv$$

is cut off at $(k - 1)\tau$ rather than at $j\tau$.

The quantities $g_j(m\tau)$ and $h_j(m\tau)$ are derived from the formulas for $f_{jk}(m\tau)$ by the definitions

$$g_j(m\tau) = \sum_{k=1}^{j-1} f_{jk}(m\tau) \quad \text{and} \quad h_j(m\tau) = \sum_{k=j+2}^{L} f_{jk}(m\tau).$$

These summations can readily be performed by the reader.

We now use Theorem 1 to check some of our previous heuristic conclusions in more quantitative detail. For example, consider the "correct" associations $y_{j,j+1}(t)$ right after their interval of greatest growth; that is, let $t = (j+1)\tau$. Our heuristic arguments suggest that $y_{j,j+1}((j+1)\tau)$ is a monotone decreasing function of $j = 1, 2, \ldots, L - 1$. Corollary 1 confirms this expectation. We will set $K = 1$ for simplicity in this and later corollaries.

COROLLARY 1. $y_{j,j+1}((j+1)\tau)$ is a positively accelerated, monotone decreasing function of $j = 1, 2, \ldots, L - 1$.

Proof. By Theorem 1, $h_j((j+1)\tau) = 0$. Thus by (27),

$$y_{j,j+1}((j+1)\tau) = \frac{(n-1)^{-1} + f_{j,j+1}((j+1)\tau)}{1 + g_j((j+1)\tau) + f_{j,j+1}((j+1)\tau)},$$

which is the same as

$$y_{j,j+1}((j+1)\tau) = \frac{(n-1)^{-1} + A}{1 + A + B(e^{-\alpha\tau} - e^{-j\alpha\tau})} \tag{33}$$

where

$$A = G + \frac{\Lambda^2}{2\alpha}(e^{-2\alpha\lambda} - e^{-2\alpha\tau})$$

and

$$B = (e^{\alpha\tau} - 1)^{-1}\left[\Lambda H + \frac{\Lambda^2}{2\alpha}(e^{-2\alpha\lambda} - e^{-2\alpha\tau})\right].$$

Function (33) is readily seen to be monotone decreasing in j. The positive acceleration of $y_{j,j+1}((j+1)\tau)$ is proved by defining

$$\lambda(t) = \frac{(n+1)^{-1} + A}{1 + A + B(e^{-\alpha\tau} - e^{-t\alpha\tau})}.$$

The quantity $\lambda(t)$ interpolates $y_{j,j+1}((j+1)\tau)$ in the sense that $\lambda(j) = y_{j,j+1}((j+1)\tau)$, and it is easy to see that $\ddot{\lambda}(t) > 0$.

Our heuristic arguments also lead us to expect that the "correct" association $y_{L-1,L}(m\tau)$ is facilitated for $m \geqslant L + 1$, since then no further list items are presented to create an inhibitory future field. Corollary 2 proves this.

COROLLARY 2. $y_{L-1,L}(m\tau)$ *is monotone increasing in m for* $L + 1 \leqslant$ $m \leqslant L + r + 1$.

Proof. By Theorem 1, $h_{L-1}(m\tau) = 0$, $m \geqslant L + 1$. By (27),

$$y_{L-1,L}(m\tau) = \frac{(n-1)^{-1} + f_{L-1,L}(m\tau)}{1 + g_{L-1}(m\tau) + f_{L-1,L}(m\tau)},$$

which yields

$$y_{L-1,L}(m\tau) = \frac{(n-1)^{-1} + A - B\exp(-m2\alpha\tau)}{1 + A - B\exp(-m2\alpha\tau) + C - D\exp(-m2\alpha\tau)} \quad (34)$$

where

$$A = G + \frac{\Lambda^2}{2\alpha}\exp(-2\alpha\lambda), \quad (35)$$

$$B = \frac{\Lambda^2}{2\alpha}\exp[(L-1)2\alpha\tau], \quad (36)$$

$$C = \left\{\frac{\exp(-\alpha\tau) - \exp[-(L-1)\alpha\tau]}{\exp(\alpha\tau) - 1}\right\}\left[\Lambda H + \frac{\Lambda^2\exp(-2\alpha\lambda)}{2\alpha}\right], \quad (37)$$

and

$$D = \left\{\frac{\exp(-\alpha\tau) - \exp[-(L-1)\alpha\tau]}{\exp(\alpha\tau) - 1}\right\}\frac{\Lambda^2}{2\alpha}\exp[(L-1)2\alpha\tau]. \quad (38)$$

To show that $y_{L-1,L}(m\tau)$ is monotone increasing in m, we define $\omega(t)$ by replacing m in (34) by t; that is, let

$$\omega(t) = y_{L-1,L}(t\tau),$$

and show that $\dot\omega(t) > 0$ for $L + 1 \leqslant t \leqslant L + r + 1$.

Since the denominator of $\dot\omega(t)$ is positive, the sign of $\dot\omega(t)$ is the same as the sign of the numerator of $\dot\omega(t)$ where

$$\text{sign}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0, \end{cases}$$

for every real x. We therefore find that

$$\begin{aligned}
\text{sign}(\dot\omega(t)) &= \text{sign}\{(1 + A - Be^{-2\alpha\tau t} + C - De^{-2\alpha\tau t})(2\alpha\tau Be^{-2\alpha\tau t}) \\
&\quad - [(n-1)^{-1} + A - Be^{-2\alpha\tau t}](2\alpha\tau Be^{-2\alpha\tau t} + 2\alpha\tau De^{-2\alpha\tau t})\} \\
&= \text{sign}\{B(1 + A - Be^{-2\alpha\tau t} + C - De^{-2\alpha\tau t}) \\
&\quad - (B + D)[(n-1)^{-1} + A - Be^{-2\alpha\tau t}]\} \\
&= \text{sign}\{B[1 - (n-1)^{-1}] + BC - [A + (n-1)^{-1}{}_{,}D\}.
\end{aligned}$$

By (35)–(38),

$$BC - [A + (n - 1)^{-1}]D$$

$$= \frac{\Lambda^2}{2\alpha} \exp[(L - 1)2\alpha\tau]$$

$$\times \left\{ \frac{\exp(-\alpha\tau) - \exp[-(L - 1)\alpha\tau]}{\exp(\alpha\tau) - 1} \right\} [\Lambda H - G - (n - 1)^{-1}]. \quad (39)$$

By the nonnegativity of $J(\xi)$ and the definitions (24)–(26), it readily follows that

$$\Lambda H \geqslant G.$$

Hence (39) yields

$$BC - [A + (n - 1)^{-1}]D \geqslant -(n - 1)^{-1} \frac{\Lambda^2}{2\alpha} \exp[(L - 1)2\alpha\tau]$$

$$\times \left\{ \frac{\exp(-\alpha\tau) - \exp[-(L - 1)\alpha\tau]}{\exp(\alpha\tau) - 1} \right\}. \quad (40)$$

By (40) along with (36),

$$B[1 - (n - 1)^{-1}] + BC - [A + (n - 1)^{-1}]D$$

$$\geqslant \frac{\Lambda^2}{2\alpha} \exp[(L - 1)2\alpha\tau]$$

$$\times \left(1 - (n - 1)^{-1} \left\{ 1 + \frac{\exp(-\alpha\tau) - \exp[-(L - 1)\alpha\tau]}{\exp(\alpha\tau) - 1} \right\} \right). \quad (41)$$

As a function of $\alpha\tau \geqslant 0$,

$$\frac{\exp(-\alpha\tau) - \exp[-(L - 1)\alpha\tau]}{\exp(\alpha\tau) - 1}$$

has a maximum of $L - 2$. Inequality (41) therefore implies

$$B[1 - (n - 1)^{-1}] + BC - [A + (n - 1)^{-1}]D$$

$$\geqslant \frac{n - L}{n - 1} \frac{\Lambda^2}{2\alpha} \exp[(L - 1)2\alpha\tau]$$

$$> 0,$$

and thus $\mathrm{sign}(\dot{\omega}(t)) > 0$.

We will now prove, as an illustration, one of the formulas showing that bowing occurs. Corollary 1 shows that $y_{i,i+1}(t)$ is monotone decreasing in i if $t = (i + 1)\tau$. Corollary 2 shows that a kind of facilitation appears in $y_{L-1,L}(t)$ if $t = m\tau$ and m is allowed to increase beyond the value $m = L$. The next corollary shows that this facilitation effect propagates backward into the list and produces bowing. We can only

expect this effect to occur if m is taken sufficiently large. We let m approach infinity to simplify our formulas. For m at finite values, the bow occurs nearer to v_L than to v_1.

COROLLARY 3 (*Asymptotic Bowing*).   *The function*

$$B(i) \equiv \lim_{m \to \infty} y_{i,i+1}(m\tau)$$

first decreases monotonically to a minimum and then increases monotonically as i increases from 1 to $L - 1$. The minimum occurs for $i = \frac{1}{2}(L - 1)$ if L is odd, and for $i = \frac{1}{2}(L - 1)$ or $i = \frac{1}{2}L$ if L is even.

Proof.   By Theorem 1, for $t > L$,

$$y_{t,t+1}(m\tau) = \frac{\mathcal{N}_i(t, m)}{\mathfrak{D}_i(t, m)} \tag{42}$$

with

$$\mathcal{N}_i(t, m) = (n - 1)^{-1} + A - Be^{2\alpha t},$$

and

$$\mathfrak{D}_i(t, m) = 1 + A - Be^{2\alpha t} + (C - De^{-\alpha t})(E - Fe^{2\alpha t})$$
$$+ (I - Je^{\alpha t}) - Ke^{\alpha t} + Le^{2\alpha t}$$

where

$$A = G + \frac{\Lambda^2}{2\alpha} e^{-2\alpha\lambda}, \tag{43}$$

$$B = \frac{\Lambda^2}{2\alpha} e^{-m2\alpha t}, \tag{44}$$

$$C = \frac{e^{-\alpha\tau}}{e^{\alpha\tau} - 1}, \tag{45}$$

$$D = (e^{\alpha\tau} - 1)^{-1}, \tag{46}$$

$$E = \Lambda H + \frac{\Lambda^2}{2\alpha} e^{-2\alpha\lambda}, \tag{47}$$

$$F = \frac{\Lambda^2}{2\alpha} e^{-m2\alpha\tau}, \tag{48}$$

$$I = \left(\Lambda H + \frac{\Lambda^2 e^{-2\alpha\lambda}}{2\alpha}\right) \frac{e^{-\alpha\tau}}{1 - e^{-\alpha\tau}}, \tag{49}$$

$$J = \left(\Lambda H + \frac{\Lambda^2 e^{-2\alpha\lambda}}{2\alpha}\right) \frac{e^{-L\alpha\tau}}{1 - e^{-\alpha\tau}}, \tag{50}$$

$$K = \frac{\Lambda}{2\alpha} e^{-m2\alpha\tau} \frac{e^{L\alpha\tau}}{e^{\alpha\tau} - 1}, \tag{51}$$

and

$$L = \frac{\Lambda}{2\alpha} \, e^{-m2\alpha\tau} \frac{e^{\alpha\tau}}{e^{\alpha\tau} - 1} \, . \tag{52}$$

We wish to show that the function

$$\mathcal{B}(t) = \lim_{m \to \infty} \text{sign}\left(\frac{d}{dt} \, y_{t,t+1}(m\tau)\right)$$

is negative when $t = 1$, positive when $t = L - 1$, and zero exactly once for t in $(1, L - 1)$. Since the denominator of $\dot{y}_{t,t+1}(m\tau)$ is positive, it suffices to consider the numerator of $\dot{y}_{t,t+1}(m\tau)$ to find the sign of $\dot{y}_{t,t+1}(m\tau)$. Proceeding as in Corollary 2, we find that

$$\text{sign}(\dot{y}_{t,t+1}(m\tau))$$
$$= \text{sign}(Be^{2\alpha\tau t}\{-2[1 - (n-1)^{-1}] + 3DEe^{-\alpha\tau t} - 2CE - 2I + (J + K)e^{\alpha\tau t}\}$$
$$+ [(n-1)^{-1} + A][2Be^{2\alpha\tau t} - DEe^{-\alpha\tau t} - DFe^{\alpha\tau t}$$
$$+ 2FCe^{2\alpha\tau t} + (J + K)e^{\alpha\tau t} - 2Le^{2\alpha\tau t}]). \tag{53}$$

We now let m approach infinity. By (44), (48), (51), and (52), we find that $B \to 0$, $F \to 0$, $K \to 0$, and $L \to 0$ as $m \to \infty$. Since also $(n + 1)^{-1} + A > 0$ by (43), (53) yields

$$\mathcal{B}(t) = \text{sign}(-DEe^{-\alpha\tau t} + Je^{\alpha\tau t}).$$

Since $J \neq 0 \neq DE$, $\mathcal{B}(t)$ has at most one zero. Since $J > 0$, $\mathcal{B}(t) > 0$ for all sufficiently large t. It remains only to show that $\mathcal{B}(1) < 0$ and $\mathcal{B}(L - 1) > 0$.

By (46), (47), and (50),

$$\mathcal{B}(t) = \text{sign}\left(\frac{e^{\alpha\tau t}e^{-L\alpha\tau}}{1 - e^{-\alpha\tau}} - \frac{e^{-\alpha\tau t}}{e^{\alpha\tau} - 1}\right)$$
$$= \text{sign}(e^{\alpha\tau t}e^{-L\alpha\tau} - e^{-\alpha\tau}e^{-\alpha\tau t}).$$

Thus

$$\mathcal{B}(1) = \text{sign}\{\exp[-(L - 1)\alpha\tau] - \exp(-2\alpha\tau)\},$$

which for any $L \geqslant 3$ is negative (and for $L < 3$ manifestly no bowing occurs), whereas

$$\mathcal{B}(L - 1) = \text{sign}\{\exp(-\alpha\tau) - \exp(-\alpha\tau)\exp[-(L - 1)\alpha\tau]\}$$
$$= \text{sign}\{1 - \exp[-(L - 1)\alpha\tau]\},$$

which is positive for all $L \geqslant 2$. Note also that $\mathcal{B}(t)$ changes sign for $t = \frac{1}{2}(L - 1)$. The proof that bowing occurs for sufficiently large m (and thus for sufficiently large W) is therefore complete.

The next corollary shows that the strength of associations $y_{j,j+k}(m\tau)$ and $y_{j,j-q}(m\tau)$ always decreases as the numerical remoteness $k > 1$ and $q \geqslant 1$ increases. The inequality $y_{j,j+1}(m\tau) > y_{j,j+2}(m\tau)$ between first- and second-order associations holds, however, only if the product $\alpha\tau$ of decay rate and reaction time or the ratio τ/λ of reaction time to input duration is sufficiently large.

COROLLARY 4.   *Association $y_{j,j+k}(m\tau)$ is an associational strength $y_{j,j+k}$ $(m\tau)$ is a monotone decreasing function of k, $1 < k \leqslant L - j$, $j = 1, 2, \ldots, L - 1$. $y_{j,j+1}(m\tau) > y_{j,j+2}(m\tau)$ if (say) $\tau > 2\lambda$ or $G > e^{-\alpha\tau}\Lambda H$. $y_{j,j-k}(m\tau)$ is an associational strength $y_{j,j-k}(m\tau)$ is a monotone decreasing function of k, $1 \leqslant k \leqslant j - 1$, $j = 2, 3, \ldots, L$.*

Proof. The denominator of $y_{j,j+k}(m\tau)$ in (25) is independent of $k \geqslant 1$. Hence it suffices to prove our claim for the numerator $(n - 1)^{-1} + f_{j,j+k}(m\tau)$ of $y_{j,j+k}(m\tau)$, and in particular for $f_{j,j+k}(m\tau)$ alone. By Theorem 1,

$$f_{j,j+k}(m\tau) = 0 \qquad \text{if } j + k \geqslant m + 1,$$

$$= G + \frac{\Lambda^2}{2\alpha}\{\exp(-2\alpha\lambda) - \exp[-(m - j)2\alpha\tau]\}$$

$$\text{if } k = 1 \text{ and } j < m,$$

$$= \left\{\Lambda H + \frac{\Lambda^2}{2\alpha}\exp(-2\alpha\lambda)\right\}\exp[-(k - 1)\alpha\tau]$$

$$- \frac{\Lambda^2}{2\alpha}\exp[-(m - j)2\alpha\tau]\exp[(k - 1)\alpha\tau]$$

$$\text{if } k > 1 \text{ and } j + k < m + 1.$$

To show that $f_{j,j+1}(m\tau) > f_{j,j+2}(m\tau)$, it suffices to prove that

$$G + \frac{\Lambda^2}{2\alpha}\{\exp(-2\alpha\lambda) - \exp[-(m - j)2\alpha\tau]\}$$

$$> \left[\Lambda H + \frac{\Lambda^2}{2\alpha}\exp(-2\alpha\lambda)\right]\exp(-\alpha\tau) - \frac{\Lambda^2}{2\alpha}\exp[-(m - j)2\alpha\tau]\exp(\alpha\tau).$$

A rearrangement of terms shows that this is true if

$$\frac{\Lambda^2}{2\alpha} \exp(-2\alpha\lambda)[1 - \exp(-\alpha\tau)] > \Lambda H \exp(-\alpha\tau) - G, \qquad (54)$$

which certainly holds if $G > \Lambda H \exp(-\alpha\tau)$. Inequality (54) can also be satisfied if $\tau > 2\lambda$, because the inequality

$$\Lambda H e^{-\alpha\tau} < \frac{\Lambda^2}{2\alpha} e^{-\alpha\tau}(1 - e^{-2\alpha\lambda})$$

shows that (54) holds if

$$G > \frac{\Lambda^2}{2\alpha}(e^{-\alpha\tau} - e^{-2\alpha\lambda}),$$

which is certainly true if $\tau > 2\lambda$, since $G > 0$.

It remains only to show that $f_{j,j+k}(m\tau)$ decreases as a function of $k \geqslant 2$ wherever it does not vanish. For $k \geqslant 2$ and $k < m - j + 1$,

$$f_{j,j+k}(m\tau) = \left[\Lambda H + \frac{\Lambda^2}{2\alpha}\exp(-2\alpha\lambda)\right]\exp[-(k-1)\alpha\tau]$$

$$- \frac{\Lambda^2}{2\alpha}\exp[-(m-j)2\alpha\tau]\exp[(k-1)\alpha\tau],$$

from which our assertion is obvious.

A similar proof goes through for $f_{j,j-k}(m\tau)$, $k \geqslant 1$.

The forward association $y_{j,j+1}(m\tau)$ will dominate the backward association $y_{j,j-1}(m\tau)$ only if similar constraints are placed on α, or τ, or λ.

COROLLARY 5. *The inequality $f_{j,j+1}(m\tau) > f_{j,j-1}(m\tau)$ holds if and only if*

$$\frac{\Lambda^2}{2\alpha}\{\exp(-2\alpha\lambda) - \exp[-(m-j)2\alpha\tau]\}[1 - \exp(-2\alpha\tau)]$$

$$> \Lambda H \exp(-2\alpha\tau) - G,$$

and thus if

$$G > \Lambda H \exp(-2\alpha\tau).$$

The corollary is proved by comparing the numerators of $f_{j,j+1}(m\tau)$ and $f_{j,j-1}(m\tau)$, and rearranging terms. The inequality $G > \Lambda H \exp(-\alpha\tau)$ of Corollary 4 thus implies $f_{j,j+1}(m\tau) > f_{j,j-1}(m\tau)$.

18. FASTER LEARNING IN THE DYNAMICAL BEGINNING THAN IN THE DYNAMICAL END

Very often, as in Fig. 3, the beginning of a long list is learned more rapidly than the end of the list, even though both beginning and end have an advantage over the middle. The main reason for this in $\mathcal{M}(\mathcal{A})$ is simply the following.

Consider $y_{L-1,L}(t)$ versus $y_{12}(t)$, for example; $y_{L-1,L}(t)$ first grows for t in $((L-1)\tau, L\tau)$, but must compete with a maximally large past field during these times. $y_{12}(t)$ first grows for t in $(\tau, 2\tau)$ and suffers no competition during this time interval. Association $y_{L-1,L}(t)$ has no future field with which to compete for $L\tau \leqslant t \leqslant (L-1)\tau + W$, but it continues to compete with its *entire* exponentially decaying past field as $x_{L-1}(t - \tau)$ also decays exponentially. By contrast, $y_{12}(t)$ for $2\tau \leqslant t \leqslant (L-1)\tau + W$ competes with associations in a future field that gains new entries only one at a time every τ time units, so that $x_1(t - \tau)$ is very small before many competing associations can be activated by successive inputs. These two factors help to guarantee faster learning in the beginning than the end of the list.

If this argument is extended to all intermediate associations $y_{i,i+1}(t)$, by "continuity" with respect to i, we readily find that the bowed curve often decreases over a smaller number of list items than the number of items over which it increases, as in Fig. 3. The length of the intermediate flat portion of the curve is determined, as was pointed out in Sections 6 and 12, by the relative sizes of $G_\varepsilon(w)$ and L. If $1 \ll G_\varepsilon(w) \leqslant L$, then many indices are in $A_\varepsilon(t)$ during its steady-state phase, and the bowed curve is flattened over these indices. If, by contrast $1 \ll G_\varepsilon(w) \cong L$, then $|A_\varepsilon(t)|$ rises to a large maximum $G_\varepsilon(w) \cong L$ and then falls to zero (supposing W is sufficiently large). The flat portion of the bowed curve is consequently very narrow.

19. GROWING CHAINS OF ASSOCIATIONS AROUND AN ANCHOR SYMBOL

We now consider learning on the later trials $\mathcal{E}_i(w, W; L)$, $i > 1$. We saw in Section 12 that the learning produced on trial $\mathcal{E}_1(w, W; L)$ can be erased on later trials if w and W are small and L is large. We henceforth suppose that W is chosen sufficiently large that all $x_j(t)$ can decay to small values after trial $\mathcal{E}_i(w, W; L)$ ends and before trial $\mathcal{E}_{i+1}(w, W; L)$ begins, $i = 1, 2, \ldots$.

Given this assumption, trial $\mathcal{E}_2(w, W; L)$ differs from trial $\mathcal{E}_1(w, W; L)$ only in that its initial data $y_{jk}((L - 1)w + W)$ are no longer uniformly distributed, as the values $y_{jk}(0) = (n - 1)^{-1}$ at the beginning of trial $\mathcal{E}_1(w, W; L)$ were. Instead, some learning has already occurred, especially at the list's beginning and end. By repeating the arguments leading to Theorem 1, we easily find that on trial $\mathcal{E}_2(w, W; L)$, learning is again faster at the beginning and the end of $\mathcal{L}$, and the effects of learning on the first two trials are *cumulative*. This fact suffices to show that the list seems to be learned in growing *chains* of associations, as the next paragraph shows.
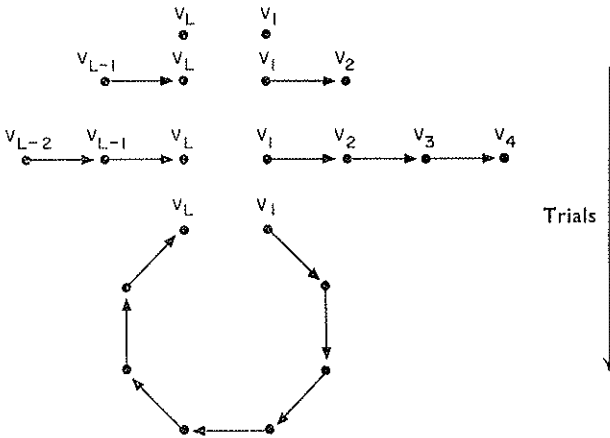

FIG. 10

Suppose, for example, that a given association $y_{i,i+1}(t)$ produces a behavioral guess of r_{i+1}, given r_i, only if $y_{i,i+1}(t)$ exceeds some given threshold $1 - \varepsilon$, $0 < \varepsilon < 1$. (See, e.g., [6] and [8], where response thresholds and some of their physiological causes are discussed.) Clearly $y_{12}(t)$ reaches threshold before $y_{23}(t)$ does, and $y_{23}(t)$ reaches threshold before $y_{34}(t)$, and so on. Similarly, $y_{L-1,L}(t)$ reaches threshold before $y_{L-2,L-1}(t)$ does, and $y_{L-2,L-1}(t)$ reaches threshold before $y_{L-3,L-2}(t)$ does, and so on. In other words, $\mathcal{M}(\mathcal{A})$ seems to learn the list in both the forward and backward directions around the "anchor stimulus" r_1, as has been experimentally reported [13]. "Chains" of associations seem to sprout from the anchor r_1 and grow in length in both the forward and backward directions as more trials occur, until the two chains meet (see Fig. 10).

20. FORWARD AND BACKWARD FUNNELING REDUCES INHIBITION ON
    LATER TRIALS

We now discuss heuristically the flow of events in the *nonlinear* system
(1)–(4) that leads to the process of "chaining" around an anchor stimulus.
Trial $\mathcal{E}_1(w, W; L)$ causes $y_{12}(t)$ to exceed its initial value $(n-1)^{-1}$ and
all other values $y_{1i}(t)$ at the beginning of trial $\mathcal{E}_2(w, W; L)$. Let r_1 occur on
trial $\mathcal{E}_2(w, W; L)$ at time $t = (L-1)w + W$. The input signal from r_1 to
r_i for times t in $(Lw + W, (L+1)w + W)$ is

$$\beta x_1(t - \tau) y_{1i}(t),$$

as (1) shows. Since $y_{12}(t) > y_{12}(0) = \cdots = y_{1i}(0) > y_{1i}(t)$ during these
times, the input received by v_2 from v_1 on trial $\mathcal{E}_2(w, W; L)$ exceeds the
input that it received $(L-1)w + W$ time units ago on trial $\mathcal{E}_1(w, W; L)$.
Growth of the association $y_{12}(t)$ is hereby enhanced on trial $\mathcal{E}_2(w, W; L)$.
Another way of saying this is that the association $y_{12}(t)$, due to past
training, has *funneled* a disproportionately large fraction of the total signal
$\beta x_1(t - \tau)$ from v_1 to the "correct" response point v_2. A similar argument
shows that $y_{23}(t)$ funnels a large fraction of $\beta x_2(t - \tau)$ from v_2 to v_3, but
not quite as large a fraction as $y_{12}(t)$ funneled from v_1 to v_2. As more and
more trials occur, the stimulus trace $x_i(t)$ of each list item r_i is funneled
with ever greater effectiveness to v_{i+1}, $i = 1, 2, \ldots, L-1$. *Funneling
hereby reduces competition between the correct associations $y_{i,i+1}(t)$ and
the incorrect associations $y_{ik}(t)$, $k \neq i + 1$,* by keeping the stimulus traces
$x_k(t)$ small unless they are perturbed by experimental inputs. In physio-
logical terms, *funneling diminishes lateral inhibition by the incorrect
vertices v_k, $k \neq i + 1$, of the correct vertex v_{i+1}.* (See the sections on
spatiotemporal masking and reaction time in [6] for a more detailed
physiological discussion.)

A similar argument shows that funneling also occurs in the backward
direction; that is, into the past field. Namely, since $y_{L-1,L}(t)$ is relatively
large when trial $\mathcal{E}_2(w, W; L)$ begins, the stimulus trace $x_{L-1}(t)$ is funneled
preferentially to v_L. The future field of v_{L-1} is thereby restricted primarily
to v_L; that is, to the "correct" vertex, instead of to all vertices. Once
$y_{L-2,L-1}(t)$ becomes large, it can funnel the stimulus trace $x_{L-2}(t)$ primarily
to v_{L-1}, and thereby decrease competition due to incorrect vertices by
eliminating these vertices from the future field of v_{L-2}. The argument can
be extended successively to $y_{L-3,L-2}(t)$, $y_{L-4,L-3}(t)$, and so on. The
heuristic point is always the same: the growing correct associations

funnel their stimulus traces to correct response points, and thereby eliminate competition via lateral inhibition by stimulus traces that have not been recently excited by experimental inputs.

The existence of competing vertices *before* the list is learned is, of course, necessary, since otherwise $\mathcal{M}(\mathcal{A})$ would not be capable of learning many different possible lists [5]. Once a given list $\mathcal{L}$ is presented to $\mathcal{M}(\mathcal{A})$, however, funneling eliminates those alternatives that are incorrect, given $\mathcal{L}$, and lets learning proceed with an ever lessening degree of competitive inhibition.

21. TRAVELING WAVES OF EXCITATION IN REFLEX ARCS

After many trials have occurred, and all $y_{i,i+1}(t) \cong 1$, $i = 1, 2, \ldots,$ $L - 1$, the funneling process carries stimulus traces along a chain from v_1 to v_2, from v_2 to v_3, and so on until v_{L-1} and v_L are reached; that is, a *traveling wave* of excitation is set up that inhibits out incorrect alternatives as it flows from point to successor point. In [6] it is shown that "feedback inputs" are needed to keep the traveling wave from dying out on later links in a long chain. These inputs are created in the medium surrounding $\mathcal{M}(\mathcal{A})$ much as we "hear ourselves talk."

The chain of associations $y_{i,i+1}(t) \cong 1$, $i = 1, 2, \ldots, L - 1$, can be thought of as a "reflex arc," since exciting v_i causes only v_{i+1} to be excited, and all other "choices" v_k, $k \neq i + 1$, have been inhibited away. In other words, the associations form a "Pavlovian circuit."

A later paper will show that these traveling waves of excitation, bolstered by lateral inhibition of incorrect vertices, are closely related to brain waves.

22. CLUSTERING AROUND A KNOWN SUBLIST OF A LONG LIST

The funneling argument can be used to explain why learning occurs, or is "clustered," in a forward and backward direction around a short well-learned list placed in the middle of a long poorly learned list, or why several bows can be created in a very long list if it is learned with a long rest pause near its numerical middle.

Clustering can be explained as in Fig. 11. Let the short well-learned list $\mathcal{L}' \equiv r_i r_{i+1} \cdots r_{i+k}$ be interpolated in the dynamical middle of $\mathcal{L} = r_1 r_2 \cdots r_L$, where we have supposed that conditions suitable to the occurrence of bowing are satisfied by w, W, and L; for example, $1 \ll G_\varepsilon(w) + 1 < i \ll L$ and $\tau = w \ll W$. Consider the presentation of
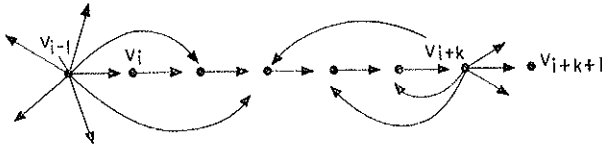
$r_1, \ldots, r_i$ to $\mathcal{M}(\mathcal{A})$ on trial $\mathcal{E}_1(w, W; L)$. Right after r_i is presented, the past field of r_i is very large, since $|A_\varepsilon(t)| \cong G_\varepsilon(w) \gg 1$ at this time. Since $\mathcal{L}'$ is well known to $\mathcal{M}(\mathcal{A})$, $y_{i,i+1}(t) \cong y_{i+1,i+2}(t) \cong \cdots \cong y_{i+k-1,i+k}(t) \cong 1$. Two effects now occur.

1. The stimulus traces $x_j(t)$, $j = 1, 2, \ldots, i - 1$, decay exponentially to zero in the time interval $((i - 1)\tau, (i + k)\tau)$ during which $\mathcal{L}'$ is presented to $\mathcal{M}(\mathcal{A})$.

2. The stimulus traces $x_j(t)$, $j = i, i + 1, \ldots, i + k - 1$, corresponding to symbols in $\mathcal{L}'$ are all funneled to their *correct* successor v_{j+1}.

By (1) and (2), if the time interval of length $(k + 1)\tau$ during which $\mathcal{L}'$ is presented to $\mathcal{M}(\mathcal{A})$ is sufficiently long and the decay rate α of the stimulus traces is sufficiently large, then $|A_\varepsilon((i + k)\tau)|$ will be far smaller than $|A_\varepsilon((i - 1)\tau)|$ was. Thus the association $y_{i+k,i+k+1}(t)$ will suffer much less interference from its past field than it would have had $\mathcal{L}'$ been originally unlearned by $\mathcal{M}(\mathcal{A})$.

This argument also shows why $y_{i-1,i}(t)$, then $y_{i-2,i-1}(t)$, and so on, are learned quickly by $\mathcal{M}(\mathcal{A})$. Let r_{i-1} and then r_i be presented to $\mathcal{M}(\mathcal{A})$. The trace $x_i(t)$ of r_i is funneled only to r_{i+1}, instead of to all r_k, $k \neq i$. The future field of r_i is thereby decreased in size, and the number of competing associations is reduced. Once $y_{i-1,i}(t)$ becomes large, it can be thought of as an extension of the known list $\mathcal{L}'$, and our argument above can then be applied to $x_{i-1}(t)$ instead of to $x_i(t)$.

Interpolating a short well-learned list $\mathcal{L}'$ in $\mathcal{L}$ has much the same effect on the learning of $\mathcal{L}$ as would interpolating a long rest period between the initial segment $\mathcal{L}_I = r_1 r_2 \cdots r_{i-1}$ and the terminal segment $\mathcal{L}_T = r_{i+k+1} \cdots r_L$ of $\mathcal{L}$, as we now show.

23. MULTIMODAL BOWING (WHOLE VERSUS PART LEARNING)

Given the list $\mathcal{L} = r_1 r_2 \cdots r_L$, choose w, W, L, and i as above. Present r_{i+1} after r_i with a time lag of $W \gg w$ instead of w. Then $|A_\varepsilon(t)|$ sinks from

$G_\varepsilon(w)$ or $G_\varepsilon(w) + 1$ to zero in the time interval $((i - 1)\tau, (i - 1)\tau + W)$. Element r_i is therefore in the dynamical end of the sublist $\mathcal{L}_I = r_1 r_2 \cdots r_i$ whereas r_{i+1} is in the dynamical beginning of the sublist $\mathcal{L}_{I'} = r_{i+1} r_{i+2} \cdots r_L$. Two bows therefore occur, one for $\mathcal{L}_I$ and one for $\mathcal{L}_{I'}$. The two sublists $\mathcal{L}_I$ plus $\mathcal{L}_{I'}$ are clearly learned more rapidly than the single serial list $\mathcal{L}$. That is, *part* learning is better than *whole* learning, as experiments have shown [13].

24. MASSING VERSUS DISTRIBUTING INTRATRIAL INTERVAL

The argument above suggests that a suitable mixing of small or large values of w on different trials might well speed up learning.

Before learning occurs, a small (or massed) w can slow the speed of learning by creating a large $G_\varepsilon(w)$ value at certain list positions. Thus distributing practice (or keeping w large) might well speed up learning on the first few trials.

Once some learning has occurred, funneling sets in. Thus the associations help to reduce the amount of scatter of stimulus traces caused by a small w. Massing practice on later trials might well therefore have less harmful effects than on early trials.

Similarly, a presentation on *each* trial of items at the beginning and end of the list with a small w and those near the middle of the list with a large w might well help to flatten the bowed curve of errors.

25. FEWER ALTERNATIVES AND MORE COMPETITION

If for fixed i, all $y_{ij}(t)$, $j \neq i$, are equal, then all r_j, $j \neq i$, are equally good responses to an isolated presentation of r_i, but it is highly unlikely that any of these responses will be produced by $\mathcal{M}(\mathcal{A})$, since the condition of equal $y_{ij}(t)$ means that lateral inhibition of *outputs* from the v_j is maximal [6]. If, say, only $y_{i,i+1}(t)$ and $y_{i,i+2}(t)$ are large, then only r_{i+1} and r_{i+2} are likely responses to r_i, but *each* of these responses can occur at different times if $y_{i,i+1}(t)$ and $y_{i,i+2}(t)$ are sufficiently large, since each can satisfy a response criterion. This fact helps to explain the seeming paradox that a large number of equivalent response alternatives might interfere less with *responding* than a small number of privileged response alternatives. Because of this fact, distributing practice can create rapid learning on the first few trials by keeping $G_\varepsilon(w)$ everywhere small, only to be followed by

several trials of strong competition between a small number of large contiguous associations.

26. PAIRED ASSOCIATES VERSUS SERIAL LEARNING (GRADUALIST VERSUS ALL-OR-NONE)

These remarks apply equally well to the case of paired-associate learning, since we need only test the size of $|A_\varepsilon(t)|$ produced by the paired-associate paradigm through time. Let the following paired associate paradigm be given, for example.

$$\pi_1(0),$$
$$\pi_1(w),\ \pi_2(w + W),$$
$$\pi_3(2w + W),$$
$$\pi_3(3w + W),\ \pi_4(3w + 2W),$$
$$\cdot$$
$$\cdot$$
$$\cdot$$

and so on. That is, r_1 is presented at time $t = 0$; $\mathcal{M}(\mathcal{A})$ is given w time units in which to guess r_2. Then r_1 and r_2 are presented with a time

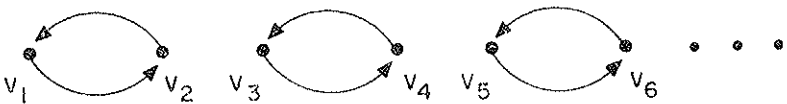

FIG. 12

separation of W. Item r_3 is then presented, after which the correct pair r_3 and r_4 is given, and so on. Clearly the paradigm distributes practice in the sense that only r_1 and r_2 occur in the time interval $(0, 2w + W)$, only r_3 and r_4 occur in the time interval $(2w + W, 4w + 2W)$, and so on. The set $|A_\varepsilon(t)|$ is therefore small for all $t \geqslant 0$, and the pairs (r_1, r_2), (r_3, r_4), $\cdots$, are learned by $\mathcal{M}(\mathcal{A})$ much as a sequence of lists of two letters each might be. As shown in [5], strong forward and backward learning occurs within each pair (r_1, r_2), (r_3, r_4), ..., and a sequence of associational "dipoles" is created, as in Fig. 12. Suppose that the pairs are changed after several trials of the experiment to form a new paired-associate paradigm (r_2, r_3), (r_4, r_5), ... Although r_2 is merely a "response" symbol to r_1 in the first

paradigm, it has by this time formed a strong backward association $y_{21}(t)$ due to the distribution of practice between pairs. This backward association will thereupon interfere dramatically with formation of the forward association $y_{23}(t)$ in the new paradigm, where r_2 is a "stimulus" symbol for r_3.

The situation in a serial list is quite different. For example, given an item r_i in the dynamical middle of a long list, the associations $y_{ik}(t)$, $k \neq i$, remain uniformly distributed over several trials until funneling reaches r_i. We can therefore interchange two points r_i and r_j in the dynamical middle *before* funneling reaches these points without altering associational strengths, since then $y_{ik}(t) \cong y_{jk}$, $j \neq k \neq i$, and $y_{ij}(t) \cong y_{ji}(t)$; that is, *without* creating interference due to prior learning. A later note will study such effects systematically to show that all-or-none *and* gradualist learning effects can occur at different list positions on successive trials of the *same* experiment! We will thereby demonstrate that the explanation of all-or-none learning does not require a mathematical model in which discrete jumps in associational strength occur.

Serial learning also differs from paired associate learning in that entire *sequences* of symbols in a serial list can act like simple symbols (or "stimuli") after a sufficient amount of practice occurs, as is discussed in some detail in [6]. This fact does not require the introduction of any special "serial learning strategy," but depends merely on an analysis of the overlap of stimulus traces produced by the serial paradigm and the rates of growth of associational strengths produced thereby. As the sequences of letters begin to act like "simple" symbols, it seems that the original symbols are *chunked* together. A process of chunking has been experimentally reported [14]. This "chunking" process is apparent even as individual points are tied together by a single chain of associations as learning goes on.

27. UPPER BOUND ON LENGTH OF LEARNABLE LISTS

When all interactions are turned off and w is chosen to satisfy $G_\varepsilon(w) \cong 0$, $A_\varepsilon(t)$ goes through only one phase in which $|A_\varepsilon(t)| \cong 0$ or 1. Thus distinguishability is good at *all* list positions. If also $w \cong \tau$, then correlations between contiguous list items are good at all list positions. This conclusion is independent of the numerical length L of the list. Given this conclusion, why is it not possible to learn lists of *any* numerical length L, just so long as $G_\varepsilon(w) \cong 0$ and $w \cong \tau$?

We readily find a reason by turning on the interactions. Then point strength flows from each vertex v_i to *every* vertex v_j with $j \neq i$ after v_i receives an input. This creates a positive level of "background noise" at incorrect vertices $v_j, j \neq i + 1$. As more and more vertices are perturbed on a given trial, the level of background noise at each vertex gradually accumulates. Competition between associations (or lateral inhibition) works to annihilate the background noise, but it might not be completely successful without annihilating the correct succession of stimulus traces as well, since by choosing $w \cong \tau$ we guarantee that background noise from a given vertex v_i arrives at the vertices v_j just as the input to the correct vertex v_{i+1} occurs.

Thus, if L is taken too large, the following danger to learning the list occurs. When vertices v_i with $i \cong L$ recieve their inputs, the level of background noise in the past field of v_i can be so high that the input signal I_i is lost in it; that is, the *signal to noise ratio* becomes unfavorable to learning. The background noise level when I_i occurs can, of course, be lowered by increasing w. Doing this might, however, make good *correlations* impossible to achieve, since w must be increased as L is increased.

A finite upper bound on the numerical length L of lists that can be learned exists, therefore, since given a presentation rate for which good distinguishability *and* correlations exist at *some* vertices, accumulating background noise decreases the distinguishability of vertices v_i with $i \cong L$ as L increases.

REFERENCES

1 S. Grossberg, Nonlinear difference-differential equations in prediction and learning theory, *Proc. Natl. Acad. Sci. USA* **58**(1967), 1329–1334.
2 S. Grossberg, Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity, *Proc. Natl. Acad. Sci. USA* **59**(1968), 368–372.
3 S. Grossberg, Some physiological and biochemical consequences of psychological postulates, *Proc. Natl. Acad. Sci. USA* **60**(1968), 758–765.
4 S. Grossberg, Global ratio limit theorems for some nonlinear functional-differential equations: I, II; *Bull. Amer. Math. Soc.* **74**(1968), 95–105.
5 S. Grossberg, Embedding fields: A theory of learning with physiological implications, *J. Math. Psychol.* **6**(1969).

6 S. Grossberg, On learning, information, lateral inhibition, and transmitters, *Math. Biosci.* **4** (1969).

7 S. Grossberg, On the production and release of chemical transmitters and related topics in cellular control, *J. Theoret. Biol.* (in press).

8 S. Grossberg, A prediction theory for some nonlinear functional-differential equations, I: Learning of lists, *J. Math. Anal. Appl.* **21**(1968), 643–694.

9 S. Grossberg, A prediction theory for some nonlinear functional-differential equations, II: Learning of Patterns, *J. Math. Anal. Appl.* **22**(1968), 490–522.

10 S. Grossberg, On the global limits and oscillations of a system of nonlinear differential equations describing a flow on a probabilistic network, *J. Diff. Eqs.* (in press).

11 S. Grossberg, On the variational systems of some nonlinear difference-differential equations, *J. Diff. Eqs.* (in press).

12 C. E. Osgood, *Method and theory in experimental psychology*, Oxford Univ. Press, London and New York, 1953.

13 A. R. Jensen, *J. Psychol.* **53**(1962), 127.

14 G. A. Miller, *Psychol. Rev.* **63**(1956), 81.

15 F. Ratliff, *Mach bands: Quantitative studies on neural networks in the retina*, Holden-Day, San Francisco, 1965.