Reprinted from JOURNAL OF STATISTICAL PHYSICS Vol. 3, No. 2, Summer 1971
Printed in Belgium

Spiking Threshold and Overarousal Effects in Serial Learning

Stephen Grossberg[1] and James Pepe[1,2]

Received June 22, 1970

Possible dependencies of serial learning data on physiological parameters such as spiking thresholds, arousal level, and decay rate of potentials are considered in a rigorous learning model. Influence of these parameters on the inverted U in learning, skewing of the bowed curve, primacy vs. recency, associational span, distribution of remote associations, and growth of associations is studied. A smooth variation of parameters leads from phenomena characteristic of normal subjects to abnormal phenomena, which can be interpreted in terms of increased response interference and consequent poor paying attention in the presence of overarousal. The study involves a type of biological many-body problem including dynamical time-reversals due to macroscopically nonlocal interactions.

KEY WORDS: Learning; stimulus sampling; underarousal and overarousal; paying attention; primacy vs. recency; skewing of mean error curve; spiking thresholds; neural networks; remote associations; response interference; whole vs. part learning; many-body problems.

1. INTRODUCTION

This paper studies some possible dependencies of serial learning data on several underlying physiological parameters. We suggest that a smooth variation of these parameters can lead from serial data characteristic of normal learning subjects to data analogous to certain forms of abnormal behavior. These facts have been previously announced.[1]

Our studies analyze a rigorously defined learning network $\mathscr{M}$ having a suggestive psychological, neurophysiological, anatomical, and biochemical interpretation[2–4]. $\mathscr{M}$ was derived[2] from an analysis of how a machine, or learning subject, could

Supported in part by the A. P. Sloan Foundation (71609), the NSF (GP-13778), and the ONR (N00014-67-A-0204-00-0051).

[1] Massachusetts Institute of Technology, Cambridge, Massachusetts.

[2] Supported in part by the ONR 4102 (02).

eventually predict an event B in response to an event A after practicing the sequence AB sufficiently often. Given such an $\mathcal{M}$, qualitatively new phenomena occur when a long list, such as the alphabet $ABC \cdots XYZ$, is presented. These include anchoring (learning associations nearest to the "anchor stimulus" A earliest), bowing (middle of the list harder to learn than either end), and chunking (simple list items aggregate into composite items).[5] Thus, suitable analysis of "two-body" predictions (AB) automatically yields phenomena familiar in "many-body" predictions ($ABC \cdots XYZ$).

The microscopic mechanism that produces these macroscopic many-body effects is still speculative, although it has been mathematically proven capable of discriminating, learning, and performing complicated tasks and yields effects analogous to related experimental data.[2–10] Its interpretation describes control of presynaptic transmitter production in terms of the cross-correlation of presynaptic spiking frequency and postsynaptic potential.[8] Alternative chemical interpretations are possible, but they are strongly constrained by the functional form of the learning equations.

We will investigate the dependence of serial learning phenomena on the spiking threshold Γ of $\mathcal{M}$. Analogous statements can be made concerning serial learning dependence on the level of persistent physiological excitation ("arousal") and on the strength of lateral inhibitory mechanisms.[1] The case $\Gamma = 0$ has previously been studied.[5] Here, we choose $\Gamma \geqslant 0$. In the case $\Gamma = 0$, a phenomenon unlike normal learning occurs; namely, asymptotic recency dominates primacy. That is, after a sufficiently long time, associations near the end of the list are stronger than associations near the list's beginning. The reverse situation commonly occurs in the data of normal subjects.[5] This abnormal phenomenon is due to the buildup of response interference, especially in response to items at the list's beginning, as a result of the persistent presentation of events in the serial paradigm. When $\Gamma = 0$, each list position can form associations with all other list positions, albeit perhaps of different strengths. Thus, as ever more serial events are presented, correct associations near the list's beginning are eventually competitively inhibited by remote incorrect associations. Raising the threshold prevents remote forward associations from forming at the list's beginning. At the list's end, however, remote backward associations can form given any Γ for which some learning is possible. A previous note[1] suggests that lowering signal thresholds can hereby cause difficulties in paying attention and can yield punning behavior, because none but the most recent events can then overcome the buildup of interfering responses and influence future behavior.

In this paper, we show that as Γ increases, asymptotic primacy eventually dominates recency, as the data of normal subjects suggests.[5] We also discuss the dependence of bowing (in particular, the degree of skewness of the bowed curve) and associational span (duration in which associations can form with a given event) on spiking thresholds. The phenomenon of bowing is of particular interest from the statistical mechanical point of view, since it describes a kind of dynamical time-reversal on the time scale of the macroscopic experimentalist due to interactions which are nonlocal on the space of macroscopically observable data.[5] The "wobble" between asymptotic primacy and recency, on the other hand, describes macroscopic asymmetries in the structure of forward and backward associations.

The paper ends with some parametric computer studies, which in fact first pointed out that asymptotic recency dominates primacy if $\Gamma = 0$.

2. THE LEARNING NETWORK

A brief description of $\mathscr{M}$ will now be given for completeness. Let $\mathfrak{A} = \{r_1, r_2, ..., r_n\}$ denote the collection of n events, or list items, that are represented in $\mathscr{M}$. To each r_i is assigned a state (or vertex, or cell body cluster) v_i in $\mathscr{M}$. A directed edge (or interaction pathway, or axon cluster) e_{ij} leads from each v_i to every distinct v_j, $j \neq i$. The arrowhead (or synaptic knob cluster) of e_{ij}, which touches v_j, is denoted by N_{ij}.

Processes occur in all v_i and N_{ij}. The vertex function (or stimulus trace, or average membrane potential) $x_i(t)$ fluctuates in v_i, and the interaction function (or associational strength, or memory trace, or excitatory transmitter level) $z_{ij}(t)$ fluctuates in N_{ij}. For present purposes, it suffices to define these functions by the system

$$\dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{m \neq i} [x_m(t - \tau) - \Gamma]^+ z_{mi}(t) + I_i(t) \tag{1}$$

and

$$\dot{z}_{jk}(t) = -\gamma z_{jk}(t) + \delta[x_j(t - \tau) - \Gamma]^+ x_k(t), \qquad j \neq k \tag{2}$$

where $[\xi]^+ = \max(\xi, 0)$ for any real number ξ, and $i, j, k = 1, 2, ..., n$. The system (1)–(2) is an example of an *embedding field*.[6]

The list $\mathscr{L} = r_1 r_2 \cdots r_L$ will be presented once to $\mathscr{M}$ in a serial manner with τ time units (the intratrial interval) between presentation of each r_i and r_{i+1}, $i = 1, 2, ..., L - 1$. Effects of different intratrial intervals and of successive list presentations have been previously discussed.[2,5]

Our results discuss the functions

$$y_{jk}(t) = z_{jk}(t) \left[\sum_{m \neq j} z_{jm}(t)\right]^{-1} \tag{3}$$

defined for $k \neq j$ and all $j = 1, 2, ..., n$. The function $y_{jk}(t)$ measures the strength of the association $r_j \to r_k$ relative to the strength of all competing associations $r_j \to r_m$, $m \neq k \neq j$, through time. Thus, $y_{jk}(t)$ measures the distinguishability of the association $r_j \to r_k$ during recall trials. Strong competing associations $r_j \to r_m$ can annihilate behavioral effects of $r_j \to r_k$ via laterial inhibition if $y_{jk}(t)$ is too small. In some previous papers, the functions $y_{jk}(t)$ are built into the dynamics of $\mathscr{M}$,[2,5] but this creates unnecessary disadvantages.[8]

Direct computation of (3) in the nonlinear system (1)–(2) is not presently possible. Such a computation can be carried out in a more linear version of system (1)–(2) that studies the primary effect of serial inputs on the functions $y_{jk}(t)$ while ignoring higher-order interaction effects. This system is called the *bare field* of $\mathscr{M}$,[5] and is defined by (2), (3), and

$$\dot{x}_i(t) = -\alpha x_i(t) + I_i(t) \tag{4}$$

where the serial nature of the inputs implies

$$I_i(t) = \begin{cases} I_{i+1}(t + \tau), & i = 1, 2, ..., L - 1 \\ 0, & i = L + 1, L + 2, ..., n \end{cases} \tag{5}$$

To simplify notation, we write $I_1(t) \equiv J(t)$. Natural auxiliary conditions will be assumed for convenience, as done previously,[5] and can be removed with obvious modifications

(a) $\mathscr{M}$ is initially at rest and in a state of maximal ignorance; i.e., all $x_i(t) = 0$ for $t \in [-\tau, 0]$, and all $z_{jk}(0) = \epsilon > 0, j \neq k$;

(b) $J(t)$ is positive only in an interval $(0, \lambda)$, with $\lambda < \tau$, and is zero elsewhere;

(c) $J(t)$ is continuous and has a single maximum;

(d) $\int_0^\lambda e^{-\alpha(\lambda - v)} J(v)\, dv > \Gamma$;

(e) $\gamma = 0$: the decay rate of associations is small compared to the time scale of the transients to be studied.

3. COMPUTATION OF RELATIVE ASSOCIATIONAL STRENGTHS

It will be convenient to use the following notation in computing $y_{jk}(t)$:

$$A = \int_0^\lambda e^{\alpha v} J(v)\, dv \tag{6}$$

$$B(t, p) = \int_0^t e^{-2\alpha(v+p)} \int_0^{v+p} e^{\alpha w} J(w)\, dw\, dv \tag{7}$$

$$C(t, p) = \int_0^t e^{-\alpha(v+p)} \int_0^{v+p} e^{\alpha w} J(w)\, dw\, dv \tag{8}$$

$$D(t, p) = \int_0^t e^{-2\alpha(v+p)} \left[\int_0^{v+p} e^{\alpha w} J(w)\, dw\right]^2 dv \tag{9}$$

$$E(x) = e^{-\alpha\tau x} \tag{10}$$

$$E(x, y) = (1/\alpha)(e^{-\alpha x} - e^{-\alpha y}) \tag{11}$$

$$T_1 = \inf\{t : x_1(t) > \Gamma\} \tag{12}$$

$$T_2 = \sup\{t : x_2(t) > \Gamma\} \tag{13}$$

$$F_{jk}(t) = \min\{t - j\tau - T_1, T_2 - T_1\} \tag{14}$$

$$G_{jk}(t) = \min\{t - (k-1)\tau, (j - k + 1)\tau + T_2\} \tag{15}$$

$$f_{jk}(t) = \int_0^t [x_j(v - \tau) - \Gamma]^+ x_k(v)\, dv \tag{16}$$

$$g_j(t) = \sum_{k=1}^{j-1} f_{jk}(t) \tag{17}$$

and

$$h_j(t) = \sum_{k=j+2}^{\sigma_j} f_{jk}(t) \tag{18}$$

with

$$\sigma_j = \min(L, j + [\![T_2/\tau]\!] + 1, [t/\tau] + 1) \tag{19}$$

where $[\xi]$ denotes the greatest integer less than or equal to ξ. Henceforth the j on σ_j will be omitted without any danger of confusion. In terms of this notation, the following formula for $y_{jk}(t)$ can be stated.

Theorem 1. Under conditions (a)–(e),

$$y_{jk}(t) = \frac{[1/(n-1)] + \rho f_{jk}(t)}{1 + \rho(f_{j,j+1}(t) + g_j(t) + h_j(t))} \tag{20}$$

where $\rho = \delta[\sum_{m \neq j} z_{jm}(0)]^{-1}$. The equations for $f_{jk}(t)$ are

$$f_{jk}(t) = AE(j-k+1)[B(t-j\tau-T_1, T_1) + \Gamma E(t-j\tau, T_1)] \tag{21}$$

if $t - j\tau \leqslant \lambda$ and $j > k$,

$$f_{jk}(t) = AE(j-k+1)[B(\lambda - T_1, T_1) + \Gamma E(t-j\tau, T_1) + \tfrac{1}{2}AE(2\lambda, 2(t-j\tau))] \tag{22}$$

if $\lambda \leqslant t - j\tau \leqslant T_2$ and $j > k$,

$$f_{jk}(t) = AE(j-k+1)[B(\lambda - T_1, T_1) + \Gamma E(T_2, T_1) + \tfrac{1}{2}AE(2\lambda, 2T_2)] \tag{23}$$

if $t - j\tau \geqslant T_2$ and $j > k$,

$$f_{jk}(t) = 0 \tag{24}$$

if $j\tau + T_2 \leqslant (k-1)\tau$ and $j + 1 < k$,

$$f_{jk}(t) = AE(k-j-1)\, B(t-(k-1)\tau, 0) - \Gamma C(t-(k-1)\tau, 0) \tag{25}$$

if $0 \leqslant t-(k-1)\tau \leqslant j\tau + T_2 - (k-1)\tau \leqslant \lambda$ and $j+1<k$ or if $t-(k-1)\tau \leqslant \lambda \leqslant j\tau + T_2 - (k-1)\tau$ and $j+1<k$,

$$f_{jk}(t) = AE(k-j-1)\, B(j\tau + T_2 - (k-1)\tau, 0) - \Gamma C(j\tau + T_2 - (k-1)\tau, 0) \tag{26}$$

if $0 \leqslant j\tau + T_2 - (k-1)\tau \leqslant t-(k-1)\tau \leqslant \lambda$ and $j+1<k$, or if $0 \leqslant j\tau + T_2 - (k-1)\tau \leqslant \lambda \leqslant t-(k-1)\tau$ and $j+1<k$,

$$\begin{aligned} f_{jk}(t) = {} & AE(k-j-1)\{B(\lambda, 0) + \tfrac{1}{2}AE(2\lambda, 2[t-(k-1)\tau])\} \\ & + \Gamma\{AE(t-[k-1]\tau, \lambda) - C(\lambda, 0)\} \end{aligned} \tag{27}$$

if $\lambda \leqslant t-(k-1)\tau \leqslant j\tau + T_2 - (k-1)\tau$ and $j+1<k$,

$$\begin{aligned} f_{jk}(t) = {} & AE(k-j-1)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + \Gamma[(\Gamma/2\alpha)\, E(j-k+1) - C(\lambda, 0) - (A/\alpha)\, e^{-\alpha\lambda}] \end{aligned} \tag{28}$$

if $\lambda \leqslant j\tau + T_2 - (k-1)\tau \leqslant t-(k-1)\tau$ and $j+1<k$,

$$f_{jk}(t) = D(t-j\tau-T_1, T_1) - \Gamma C(t-j\tau-T_1, T_1) \tag{29}$$

if $t - j\tau \leqslant \lambda$ and $j = k-1$,

$$f_{jk}(t) = D(\lambda - T_1, T_1) + \tfrac{1}{2}A^2E(2\lambda, 2[t - j\tau]) + \Gamma[AE(t - j\tau, \lambda) - C(\lambda - T_1, T_1)] \tag{30}$$

if $\lambda \leqslant t - j\tau \leqslant T_2$ and $j = k - 1$,

$$f_{jk}(t) = D(\lambda - T_1, T_1) + \tfrac{1}{2}A^2E(2\lambda, 2T_2) + \Gamma[AE(T_2, \lambda) - C(\lambda - T_1, T_1)] \tag{31}$$

if $t - j\tau \geqslant T_2$ and $j = k - 1$. The equations for $g_j(t)$ are

$$g_j(t) = AE^{-1}(-\tau, 0)\, E(\tau, j\tau)[B(t - j\tau - T_1, T_1) + \Gamma E(t - j\tau, T_1)] \tag{32}$$

if $t - j\tau \leqslant \lambda$,

$$g_j(t) = AE^{-1}(-\tau, 0)\, E(\tau, j\tau)\{B(\lambda - T_1, T_1) + \tfrac{1}{2}AE(2\lambda, 2[t - j\tau]) + \Gamma E(t - j\tau, T_1)\} \tag{33}$$

if $\lambda \leqslant t - j\tau \leqslant T_2$,

$$g_j(t) = AE^{-1}(-\tau, 0)\, E(\tau, j\tau)[B(\lambda - T_1, T_1) + \tfrac{1}{2}AE(2\lambda, 2T_2) + \Gamma E(T_2, T_1)] \tag{34}$$

if $t - j\tau \geqslant T_2$. The equations for $h_j(t)$ are

$$h_j(t) = 0 \tag{35}$$

if $\sigma < j + 2$,

$$\begin{aligned} h_j(t) = {} & AE^{-1}(0, \tau)\, E(\tau, [\sigma - j - 1]\, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + (A/\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma - 1]\, \tau, -[j + 1]\, \tau)[\Gamma e^{-\alpha t} - \tfrac{1}{2}AE(-j)\, e^{-2\alpha t}] \\ & - \Gamma[C(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}](\sigma - j - 2) \\ & + AE(\sigma - j - 1)\, B(t - [\sigma - 1]\, \tau, 0) - \Gamma C(t - [\sigma - 1]\, \tau, 0) \end{aligned} \tag{36}$$

if $0 \leqslant t - (\sigma - 1)\,\tau \leqslant j\tau + T_2 - (\sigma - 1)\,\tau \leqslant \lambda$ or if $t - (\sigma - 1)\,\tau \leqslant \lambda \leqslant j\tau + T_2 - (\sigma - 1)\,\tau$,

$$\begin{aligned} h_j(t) = {} & AE^{-1}(0, \tau)\, E(\tau, [\sigma - j]\, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + (A/\alpha)\, E^{-1}(-\tau, 0)\, E(-\sigma\tau, -[j + 1]\, \tau)[\Gamma e^{-\alpha t} - \tfrac{1}{2}AE(-j)\, e^{-2\alpha t}] \\ & - \Gamma[C(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}](\sigma - j - 1) \end{aligned} \tag{37}$$

if $\lambda \leqslant t - (\sigma - 1)\,\tau \leqslant j\tau + T_2 - (\sigma - 1)\,\tau$,

$$\begin{aligned} h_j(t) = {} & AE^{-1}(0, \tau)\, E(\tau, [\sigma - j - 1]\, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + (\Gamma^2/2\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma - j - 1]\, \tau, -\tau) - \Gamma[C(\lambda, 0) \\ & + (A/\alpha)\, e^{-\alpha\lambda}](\sigma - j - 2) \\ & + AE(\sigma - j - 1)\, B(j\tau + T_2 - [\sigma - 1]\, \tau, 0) - \Gamma C(j\tau + T_2 - [\sigma - 1]\, \tau, 0) \end{aligned} \tag{38}$$

if $0 \leqslant j\tau + T_2 - (\sigma - 1)\,\tau \leqslant \lambda \leqslant t - (\sigma - 1)\,\tau$ or $0 \leqslant j\tau + T_2 - (\sigma - 1)\,\tau \leqslant t - (\sigma - 1)\,\tau \leqslant \lambda$,

$$\begin{aligned} h_j(t) = {} & AE(0, \tau)\, E(\tau, [\sigma - j]\, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + (\Gamma^2/2\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma - j]\, \tau, -\tau) \\ & - \Gamma[C(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}](\sigma - j - 1) \end{aligned} \tag{39}$$

if $\lambda \leqslant j\tau + T_2 - (\sigma - 1)\,\tau \leqslant t - (\sigma - 1)\,\tau$.

Proof. Integration of (2) under conditions (a) and (e) followed by substitution of (2) into (3) yields the equations

$$y_{jk}(t) = \frac{[1/(n-1)] + \rho f_{jk}(t)}{1 + \rho \sum_{m \neq j} f_{jm}(t)}, \qquad j \neq k \tag{40}$$

It therefore suffices to evaluate all functions $f_{jk}(t)$, $j \neq k$. By the serial nature of the inputs, $I_i(t) = I_1(t - [i - 1]\,\tau)$ for all $i = 1,..., L$. Thus, by (4) and condition (a),

$$x_i(t) = x_1(t - [i - 1]\,\tau), \qquad i = 1,..., L \tag{41}$$

where

$$x_1(t) = \begin{cases} 0, & \text{if } \ t < 0 \\ \int_0^t e^{-\alpha(t-v)}\, J(v)\, dv, & \text{if } \ 0 \leqslant t \leqslant \lambda \\ Ae^{-\alpha\tau}, & \text{if } \ \lambda \leqslant t \end{cases} \tag{42}$$

In particular, by (16) and (1), we find

$$f_{jk}(t) = \int_0^t [x_1(v - j\tau) - \Gamma]^+\, x_1(v - [k - 1]\tau)\, dv \tag{43}$$

Evaluation of (43) must be done in several cases, since the limits of integration over which the integrand is positive depend on the relative sizes of j, k, and t, and of x_1 relative to Γ. The following computations illustrate this fact.

Let $j > k$. Since $x_1(v - j\tau) \leqslant \Gamma$ if $v \leqslant j\tau + T_1$, (43) shows that $f_{jk}(t) = 0$ if $t \leqslant j\tau + T_1$, and thus that

$$f_{jk}(t) = \int_{j\tau+T_1}^t [x_1(v - j\tau) - \Gamma]^+\, x_1(v - [k - 1]\tau)\, dv \tag{44}$$

for $t \geqslant j\tau + T_1$. The expression $[x_1(v - j\tau) - \Gamma]^+$ is positive only in the interval $(j\tau + T_1, j\tau + T_2)$, by (12), (13), (42), and conditions (b) and (c). A change of variables in (44) therefore shows that

$$f_{jk}(t) = \int_0^{F_{jk}(t)} [x_1(v + T_1) - \Gamma]\, x_1(v + T_1 + [j - k + 1]\tau)\, dv \tag{45}$$

where $F_{jk}(t)$ is defined as in (14). Note that no superscript "+" occurs in (45).

Condition (b) along with the hypothesis $j > k$ implies that

$$v + T_1 + (j - k + 1)\,\tau > \lambda$$

for all $v \geqslant 0$. Therefore, (42) and (45) imply

$$f_{jk}(t) = AE(j - k + 1) \int_0^{F_{jk}(t)} e^{-\alpha(v+T_1)} \left[e^{-\alpha(v+T_1)} \int_0^{v+T_1} e^{\alpha w} J(w)\, dw - \Gamma\right] dv \tag{46}$$

Suppose that $t - j\tau \leqslant \lambda$. It remains only to apply (42) to (46) given special choices of t. By condition (d), we also have $\lambda \leqslant T_2$, and so $F_{jk}(t) = t - j\tau - T_1$. Equation (21) follows readily. Suppose that $\lambda \leqslant t - j\tau \leqslant T_2$. Again $F_{jk}(t) = t - j\tau - T_1$, but the last two cases of (42) both arise if $T_2 > \lambda$. Breaking up the integral from 0 to $F_{jk}(t)$ at $t = \lambda - T_1$ isolates these two cases. Then term-by-term evaluation yields (22). Finally, suppose $T_2 < t - j\tau$. Now, $F_{jk}(t) = T_2 - T_1$ and substitution of T_2 for $t - j\tau$ in (22) yields (23).

The sum $g_j(t)$ in (17) can also readily be evaluated, since all $f_{jk}(t)$ with $k = 1, 2,..., j - 1$ simultaneously obey the same equations (21), then (22), then (23) as t increases. A simple application of the identity $\sum_{i=0}^{k} x^i = (1 - x^{k+1})(1 - x)^{-1}$ for suitable x therefore completes the computation.

Suppose that $j < k - 1$. Again the functions $f_{jk}(t)$ are computed in several cases. More cases occur, however, than when $j > k$, because now the list items r_k occur after r_j does rather than before. This asymmetry in the treatment of forward vs. backward associations is clearly seen in (43). For $j > k$, $x_k(t) > 0$ whenever $x_j(t - \tau) > \Gamma$, so that if any past r_k influences $y_{jk}(t)$, then all do. By contrast, the product $[x_j(t - \tau) - \Gamma]^+ x_k(t)$ can be identically zero for some $k > j + 1$ but not others. We now consider the cases that arise if $j < k - 1$.

The product $[x_1(t - j\tau) - \Gamma]^+ x_1(t - [k - 1]\,\tau)$ is identically zero if $j\tau + T_2 \leqslant (k - 1)\,\tau$. Thus, $f_{jk}(t) \equiv 0$ by (43), yielding (24). Again by (43) and the hypothesis $j < k - 1$, $f_{jk}(t)$ can only be positive if $t \geqslant (k - 1)\,\tau$, and in this case (43) can be written as

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

Since $v + (k - j - 1)\,\tau \geqslant \lambda$ for all $v \geqslant 0$, condition (b) implies

$$f_{jk}(t) = \int_0^{t-(k-1)\tau} [AE(k - j - 1)\, e^{-\alpha v} - \Gamma]^+ x_1(v)\, dv \tag{47}$$

Two pairs of cases now occur depending on whether $[AE(k - j - 1)\, e^{-\alpha v} - \Gamma]^+$ is positive or equal to zero, and whether $x_1(v)$ obeys the second or third case of (42); that is, if $v \leqslant \lambda$ or $v \geqslant \lambda$. To guarantee that $[AE(k - j - 1)\, e^{-\alpha v} - \Gamma]^+$ is positive, suppose that $t \leqslant \min\{\lambda + (k - 1)\,\tau, j\tau + T_2\}$. Then the superscript "+" in (47) can be removed, and (47) becomes

$$f_{jk}(t) = \int_0^{t-(k-1)\tau} [AE(k - j - 1)\, e^{-\alpha v} - \Gamma]\, e^{-\alpha v} \int_0^{v} e^{\alpha w} J(w)\, dw\, dv$$

which can be evaluated term-by-term to yield (25). If, however, either

$$j\tau + T_2 < t \leqslant \lambda + (k-1)\tau, \qquad \text{or} \quad j\tau + T_2 \leqslant \lambda + (k-1)\tau < t,$$

then $[AE(k-j-1)e^{-\alpha v} - \Gamma]^+$ is positive only when $v \leqslant \lambda$. Hence, $f_{jk}(t)$ satisfies (25) with $t = j\tau + T_2$, yielding (26).

Now we consider the cases in which $[AE(k-j-1)e^{-\alpha v} - \Gamma]^+$ is positive even if $v > \lambda$. Suppose $(k-1)\tau + \lambda < t \leqslant j\tau + T_2$. Then the integral from 0 to $t - (k-1)\tau$ in (47) is broken up at $v = \lambda$ to distinguish the second and third cases of (42) in evaluating $x_1(v)$. Term-by-term integration now yields (27). Supposing that $\lambda + (k-1)\tau < j\tau + T_2 \leqslant t$ now yields (28) by letting $t = j\tau + T_2$ in (27). This completes the case $j < k-1$. The remaining case $j = k-1$ can be carried out in a similar way.

It remains only to compute the sum $\bar{h}_j(t) = \sum_{m=j+2}^{L} f_{jm}(t)$ in the denominator of (40). This sum equals $h_j(t)$ as given by (18), because $[x_1(t-j\tau) - \Gamma]^+ = 0$ for all $t > j + \tau^{-1}T_2$. As with evaluating $f_{jk}(t)$ for $j < k-1$, more cases can occur than when evaluating $g_j(t)$.

Equation (35) results if $j\tau + T_2 \leqslant (k-1)\tau$ because (18) and (24) show $h_j(t) = 0$. However, if $t \leqslant \min\{\lambda + (\sigma-1)\tau, j\tau + T_2\}$, $h_j(t)$ is found by summing $f_{jk}(t)$, given by (27), for $k = j+2, j+3, ..., \sigma - 1$, and $f_{j\sigma}(t)$, given by (25). The resulting sum is (36). For $\lambda + (\sigma-1)\tau \leqslant t \leqslant j\tau + T_2$, $f_{jk}(t)$ is given by (27). Equation (37) results from summing these $f_{jk}(t)$ for $k = j+2, j+3, ..., \sigma$. When

$$j\tau + T_2 \leqslant \min\{\lambda + (\sigma-1)\tau, t\}$$

$f_{jk}(t)$ is given by (28) for $k = j+2, j+3, ..., \sigma - 1$, and $f_{j\sigma}$ is given by (26). The resulting value of $h_j(t)$ is Eq. (38). Finally, in the case of $\lambda + (\sigma-1)\tau \leqslant j\tau + T_2 \leqslant t$, $h_j(t)$ is computed using (28) instead of (27) for $f_{jk}(t)$, yielding (39). This completes the proof of Theorem 1.

4. ASSOCIATIONAL SPAN

The associational span is defined heuristically as the maximum duration during which associations can be formed between a given r_i and other events r_k. Alternatively, it can be defined as the number of r_k with which r_i can form an association. Using this concept, we will make precise the important fact that v_i can sample all v_k with $k \leqslant i-1$, but not necessarily any v_k with $k \geqslant i+1$ other than v_{i+1}. That is to say, when associations are being formed with r_i, different information is available in the network concerning the past than the future.

We will use the following definitions of associational span, and of the related concept of associational interval, for simplicity.

Definition. The interval $(T_1 + i\tau, T_2 + i\tau)$ is the *associational interval* of r_i, and $\mathcal{A} = T_2 - T_1$ is the *associational span* of all r_i.

These definitions are justified as follows. By (12), (13), and (42), $[x_1(t) - \Gamma]^+$ is positive only for $t \in (T_1, T_2)$. Since

$$z_{ij}(t) = \epsilon + v \int_0^t [x_1(v - i\tau) - \Gamma]^+ x_1(v - [j-1]\tau)\, dv$$

$z_{ij}(t)$ can only change for $t \in (T_1 + i\tau, T_2 + i\tau)$; namely, in the associational interval of r_i, whose length $\mathcal{A}$ is the associational span.

If $J(t)$ is a rectangular input pulse of intensity J and duration λ, then

$$\mathcal{A} = \lambda + (1/\alpha) \log\{[(J/\alpha\Gamma) - 1](1 - e^{-\alpha\lambda})\} \tag{48}$$

which is monotone decreasing in Γ. As Γ decreases, more forward associations $r_i \to r_k$, $k > i + 1$, can form, thereby reducing the relative strength of $r_i \to r_{i+1}$. This does not mean, however, that increasing Γ always improves learning of $r_i \to r_{i+1}$. If Γ is too large, then even though no forward associations can compete with $r_i \to r_{i+1}$, nonetheless $[x_1(t) - \Gamma]^+$ is usually zero or small in value, so that little learning of $r_i \to r_{i+1}$ occurs. Thus, there exists an optimal region of threshold choice that reduces response interference without unduly diminishing the rate of learning. Alternatively expressed, this optimal region maximizes distinguishability of the correct association while providing enough energy to drive the learning process. An analogous region of optimal performance has been suggested by recent experiments.[11]

Notice that decreasing J in (48) has the same qualitative effect as increasing Γ. Thus, all of our statements concerning threshold regulation given fixed levels of physiological excitation can be transformed into corresponding statements concerning variations in the level of excitation as it compares with the system's fixed threshold parameters. The interplay between J and inhibitory interaction strength is, by contrast, more subtle.[9]

5. BOWING

As mentioned in the introduction, bowing means that the middle of the list is harder to learn than either end. This is the net result of two effects. First, as list position i increases, there always exist more backward associations $r_i \to r_k$, $k < i$, that compete with $r_i \to r_{i+1}$, thereby increasing learning difficulty. Second, there exist fewer forward associations $r_i \to r_{i+1}$, thereby decreasing learning difficulty. However, by varying the associational span, we can guarantee that no forward association ever competes with $r_i \to r_{i+1}$ for any i. For example, as in Section 4, choose Γ so large that $[x_i(t) - \Gamma]^+ = 0$ whenever $x_k(t) > 0$ and $k > i + 1$. Then, the associations $r_i \to r_k$ never form, and consequently the major effect on the association $r_i \to r_{i+1}$ as i increases is simply to increase response interference due to increasing numbers of backward response alternatives. Apart from such degenerate cases, however, bowing always occurs in the bare field, as we will shortly prove.

We will first consider "asymptotic" bowing. Namely, letting

$$\mathcal{B}(i, \Gamma) \equiv \lim_{t\to\infty} y_{i,i+1}(t), \qquad i = 1, 2, \ldots, L-1,$$

we will prove that for any fixed $\Gamma \geqslant 0$, $\mathscr{B}(i, \Gamma)$ either first decreases and then increases as i increases from 1 to L, or the degenerate case occurs in which $\mathscr{B}(i, \Gamma)$ is monotone decreasing. By definition, for fixed Γ, the bow occurs at the list position $M(\Gamma)$ for which $\mathscr{B}(i, \Gamma)$ is a minimum. If there exists more than one such position, we let $M(\Gamma)$ be the largest one, since in the presence of nonlinear interactions, background noise can only increase as more events are presented.

In the bare field, $M(\Gamma)$ is a monotone-increasing function of Γ. Furthermore, $M(0) = \frac{1}{2}(L - 1)$ if L is odd and $M(0) = \frac{1}{2}(L - 2)$ or $\frac{1}{2}L$ if L is even.[5] In the degenerate case above, $M(\Gamma) = L$ for sufficiently large Γ. Thus, maximal difficulty in learning can occur at any list position greater than the list's numerical middle. Since "normal" learning requires a positive Γ, the bow will occur nearer to the end than to the beginning of the list, and the bowed curve will therefore be skewed. This also occurs *in vivo*.[5]

At times $t < \infty$, let $\mathscr{B}(i, \Gamma, t) = y_{i,i+1}(t)$, and suppose that $\min_i \mathscr{B}(i, \Gamma, t)$ occurs at list position $M(t, \Gamma)$ for every fixed t and Γ. Then, for fixed Γ, $M(t, \Gamma)$ ultimately decreases from $M(t, \Gamma) = L$ to $M(t, \Gamma) = M(\Gamma)$ as t increases beyond the time at which r_L is presented to infinity. This happens because the *non*occurrence of the events $r_{L+1}, r_{L+2}, \ldots, r_n$ gradually decreases the relative amount of response interference to $r_{L-1} \to r_L$ growth, since the future associations $r_{L-1} \to r_k$, $k > L$, never form as t increases. Thus, skewing can depend both on Γ and on the intertrial interval.[5] Of course, if Γ is very large, the intertrial interval effect will be negligible.

If we consider only the influence of past associations, then the distribution of associational strengths is much simpler to understand. For example, the function $y_{i,i+1}(t + [i - 1]\,\tau)$, $0 \leqslant t \leqslant 2\tau$, is a negatively accelerated, monotone-decreasing function of i. This function measures the increase with list position of competing remote past associations after each correct association has had an equal time to develop.

The following notation will be convenient for our study of bowing. Let

$$\begin{aligned}\mathscr{C} = \tau E^{-1}(-\tau, 0)[AB(\lambda - T_1, T_1) + A\Gamma E(T_2, T_1) \\ + \tfrac{1}{2}A^2 E(2\lambda, 2T_1) - (\Gamma^2/2\alpha)\, E(-\sigma)]\end{aligned} \tag{49}$$

$$\mathscr{D} = A\tau E(\sigma)\, E^{-1}(0, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \tag{50}$$

and

$$\mathscr{E} = \Gamma[C(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}] \tag{51}$$

Theorem 2 (*Asymptotic Bowing*). Under conditions (a)–(e), either asymptotic bowing or the degenerate case occurs, with

$$M(\Gamma) = \frac{1}{\alpha\tau} \log \left[\frac{\mathscr{E} + (\mathscr{E}^2 + 4\mathscr{C}\mathscr{D})^{1/2}}{2\mathscr{D}}\right] \tag{52}$$

and $M(\Gamma)$ monotone increasing in Γ.

Corollary 1 (*No Future Field*). If $T_2 < 2\tau$, the degenerate case occurs.

Proof of Asymptotic Bowing. We will study $\text{sign}(\partial\mathscr{B}(j, \Gamma)/\partial j)$ as a function of $j = 1, 2, ..., L - 1$, where

$$\text{sign}(w) = \begin{cases} 1 & \text{if} \quad w > 0 \\ 0 & \text{if} \quad w = 0 \\ -1 & \text{if} \quad w < 0 \end{cases}$$

and where we let j vary continuously to compute the derivative. Note in (20), for $t > T_2 + (L - 1)\tau$, that all $f_{jk}(t)$ are constant, and moreover all $f_{j,j+1}(t)$ are equal, $j = 1, 2, ..., L - 1$, say to the constant N. Thus, by (20),

$$\text{sign}[\partial\mathscr{B}(j, \Gamma)/\partial j] = -\text{sign}[\partial\mathscr{G}(j, \Gamma)/\partial j] \tag{53}$$

where $\mathscr{G}(j, \Gamma)$ is the denominator of (20) for $t > T_2 + (L - 1)\tau$; namely,

$$\mathscr{G}(j, \Gamma) = 1 + \rho[N + g_j(\infty) + h_j(\infty)] \tag{54}$$

Thus, setting $\rho = 1$ for convenience,

$$\partial\mathscr{G}(j, \Gamma)/\partial j = \partial(g_j(\infty) + h_j(\infty))/\partial j$$

and direct computation using (34) and (39) shows that

$$\partial\mathscr{G}(j, \Gamma)/\partial j = \mathscr{C}E(j) - \mathscr{D}E(-j) + \mathscr{E} \tag{55}$$

Setting $\partial\mathscr{G}(j, \Gamma)/\partial j = 0$ yields the quadratic equation

$$\mathscr{D}z^2 - \mathscr{E}z - \mathscr{C} = 0$$

in $z = E(-j)$, which has at most two real roots. Thus, there exist at most two values of j, j_1 and $j_2 (j_1 \leqslant j_2)$, at which a bow can occur. Four possible bowing cases can therefore be distinguished:

$$\text{(I)} \quad \partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 \geqslant 0 \quad \text{and} \quad \partial^2\mathscr{G}(j_2, \Gamma)/\partial j^2 > 0;$$

$$\text{(II)} \quad \partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 \leqslant 0 \quad \text{and} \quad \partial^2\mathscr{G}(j_2, \Gamma)/\partial j^2 > 0;$$

$$\text{(III)} \quad \partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 > 0 \quad \text{and} \quad \partial^2\mathscr{G}(j_2, \Gamma)/\partial j^2 \leqslant 0;$$

and

$$\text{(IV)} \quad \partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 \leqslant 0 \quad \text{and} \quad \partial^2\mathscr{G}(j_2, \Gamma)/\partial j^2 \leqslant 0.$$

Cases I and II are readily eliminated using (55): since $\mathscr{D} > 0$, $\partial\mathscr{G}(j, \Gamma)/\partial j$ is negative for all sufficiently large j. Consider case III. Since

$$\partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 = -\tfrac{1}{2}\alpha\tau\mathscr{D}^{-1}E(j)[\mathscr{E}^2 + 4\mathscr{C}\mathscr{D} - \mathscr{E}(\mathscr{E}^2 + 4\mathscr{C}\mathscr{D})^{1/2}]$$

and $\partial^2\mathscr{G}(j_1, \Gamma)/\partial j^2 > 0$, it follows that $\mathscr{C} < 0$. If $\mathscr{C} < 0$, then by (55),

$$\lim_{j\to-\infty} \text{sign}\ \partial\mathscr{G}(j, \Gamma)/\partial j = -1 \tag{56}$$

Also, by (55),

$$\lim_{j\to\infty} \text{sign } \partial\mathscr{G}(j, \Gamma)/\partial j = -1 \tag{57}$$

From (56) and (57) along with the defining conditions of case III, it follows that $\partial\mathscr{G}(j, \Gamma)/\partial j$ has a single and double root, which is impossible by (55).

Only case IV remains. To prove bowing, it will suffice to show that $j_1 < 1 < j_2$. In the degenerate case, the inequalitites $j_1 < 1 \leqslant j_2$ will hold. To prove $j_1 < 1$, it will be convenient to consider the graph of $\partial\mathscr{G}(j, \Gamma)/\partial j$ in Fig. 1. Suppose $[T_2/\tau] \leqslant L - 1$. Then, (38) and (39) show that $h_j(\infty)$ is constant at $j = 1$, and (34) shows that $g_j(\infty)$ is monotone increasing in j. Hence, $\mathscr{G}(j, \Gamma)$, as given by (54), is monotone increasing at $j = 1$. Thus, $j_1 < 1$ by Fig. 1. Suppose $[T_2/\tau] \geqslant L$. We will show that

$$\mathscr{E} < 2\mathscr{D}e^{\alpha\tau} \tag{58}$$

By (55), we also know that

$$e^{\alpha\tau j_1} = (\mathscr{E}/2\mathscr{D}) - (1/2\mathscr{D})(\mathscr{E}^2 + 4\mathscr{C}\mathscr{D})^{1/2}$$

which proves $j_1 < 1$.

The inequality (58) will now be established. By (13), (50), and (51), this inequality is equivalent to

$$\mathscr{C}(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda} < 2\tau E[L - 1 - (T_2/\tau)]\, E^{-1}(0, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}]$$

Since $T_2 \geqslant L\tau$, $\tau > \lambda$, and $\tau E^{-1}(0, \tau) > 1$, it suffices to show that

$$C(\lambda, 0) < 2E(-1)\, B(\lambda, 0)$$

which is obvious. Thus, in all cases, $j_1 < 1$.

It remains to show that $j_2 \geqslant 1$, and that $j_2 > 1$ for sufficiently long lists. Our argument will depend on the study of the roots of $\partial\mathscr{G}(j, \Gamma)/\partial j = 0$ as a function of Γ. Hence, we write j_2 as $M(\Gamma)$, defined by (52), to emphasize this dependence.

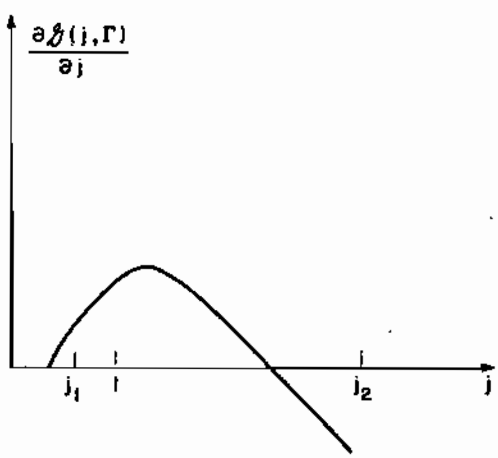

Fig. 1

A previous paper[5] shows that the only root of $\partial\mathscr{G}(j,0)/\partial j = 0$ is $j = \frac{1}{2}(L-1)$. Since we consider only the nontrivial case $L \geqslant 3$, $M(0) \geqslant 1$. If $L \geqslant 5$, $M(0) > 1$. We will now prove that (i) $dM/d\Gamma \geqslant 0$, and moreover that (ii) $dM(0)/d\Gamma > 0$. Thus, $M(\Gamma) \geqslant 1$ for all $\Gamma \geqslant 0$ and $L \geqslant 3$, and $M(\Gamma) > 1$ for all $\Gamma \geqslant 0$ and $L \geqslant 5$. To prove (i), we will directly verify (ii) and show in addition that (iii) $d^2M/d\Gamma^2 \geqslant 0$ whenever $dM/d\Gamma = 0$.

We prove (ii) as follows. By (52),

$$\frac{dM}{d\Gamma} = \frac{1}{\alpha\tau}\frac{(d\mathscr{E}/d\Gamma) + (d\mathscr{C}/d\Gamma)E(j)}{(\mathscr{E}^2 + 4\mathscr{C}\mathscr{D})^{1/2}}$$

where

$$d\mathscr{C}/d\Gamma = \tau E^{-1}(-\tau, 0)[AE(T_2, T_1) - (\Gamma/\alpha)\, E(-\sigma)]$$

and

$$d\mathscr{E}/d\Gamma = \mathscr{C}(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}$$

Thus, it suffices to show that $\Gamma = 0$ implies $d\mathscr{E}/d\Gamma + (d\mathscr{C}/d\Gamma)E(j) > 0$, which is the same as

$$[\alpha C(\lambda, 0)/A] + e^{-\alpha\lambda} - \tau E[j + (T_1/\tau)]\, E^{-1}(-\tau, 0) > 0.$$

This is surely true, however, since even

$$e^{-\alpha\lambda} > e^{-\alpha\tau} > \tau E(j)\, E^{-1}(-\tau, 0).$$

The proof of (iii) uses (i) and two additional facts: (iv) $\operatorname{sign}(d^2M/d\Gamma^2)$ is constant at all points where $dM/d\Gamma = 0$, and (v) for sufficiently large Γ, $\partial\mathscr{G}(j,\Gamma)/\partial j > 0$, and thus $M(\Gamma) = L$.

By (i) and (iv), if $d^2M/d\Gamma^2 \geqslant 0$ at any point where $dM/d\Gamma = 0$, then we are done. By (iv), if $d^2M(\Gamma_0)/d\Gamma^2 < 0$, then $d^2M(\Gamma)/d\Gamma^2 < 0$ for all $\Gamma \geqslant \Gamma_0$. In particular, (v) cannot occur, thereby yielding a contradiction.

We prove (iv) by noting that if $dM/d\Gamma = 0$, then

$$\frac{d^2M}{d\Gamma^2} = \frac{1}{\alpha\tau}\frac{(d^2\mathscr{C}/d\Gamma^2)E(j)}{(\mathscr{E}^2 + 4\mathscr{C}\mathscr{D})^{1/2}}$$

where

$$\frac{d^2\mathscr{C}}{d\Gamma^2} = \frac{\tau E^{-1}(-\tau, 0)}{\alpha}\left[1 + \alpha A e^{-\alpha T_1}\frac{\partial T_1}{\partial \Gamma} - E(-\sigma)\right]$$

We prove (v) as follows. By (13),

$$T_2(\Gamma) = (1/\alpha)\log\left\{\left[\int_0^\lambda e^{\alpha v}J(v)\,dv\right]\Big/\Gamma\right|$$

Hence, $T_2(\Gamma)$ is monotone decreasing in Γ and $\lim_{\Gamma\to\infty} T_2(\Gamma) = 0$. Thus, there exists a Γ_0 such that $T_2(\Gamma_0) = 2\tau$, and in particular $T_2(\Gamma) \leqslant 2\tau$ for $\Gamma \geqslant \Gamma_0$. By (35), however, $T_2 \leqslant 2\tau$ implies $h_j(\infty) = 0$. By (31), $f_{j,j+1}(\infty)$ is independent of

$j = 1, 2, \ldots, L-1$. To prove that $\partial\mathscr{G}(j, \Gamma)/\partial j > 0$, it thus suffices to show that $\partial g_j(\infty)/\partial j > 0$. This follows readily from (34), and completes the proof.

Proposition 1. The function

$$y(i, t) = y_{i,i+1}(t + [i-1]\,\tau)$$

for fixed $t \in [0, 2\tau]$, is positively accelerated and monotone decreasing in i.

Proof. By (35), $h_i(t) \equiv 0$. Thus, by (20),

$$y(i, t) = \frac{[1/(n-1)] + f_{i,i+1}(t + [i-1]\tau)}{1 + f_{i,i+1}(t + [i-1]\tau) + g_i(t + [i-1]\tau)} \tag{59}$$

Henceforth, we write $f = f_{i,i+1}(t + [i-1]\,\tau)$ and $g = g_i(t + [i-1]\,\tau)$ for simplicity.

By (30), f is independent of i. By (33),

$$g = \delta E(\tau, i\tau) \tag{60}$$

where

$$\delta = AE^{-1}(-\tau, 0)[B(\lambda - T_1, T_1) + \tfrac{1}{2}AE(2\lambda, 2[t-\tau]) + \Gamma E(t-\tau, T_1)]$$

Thus, by (59) and (60), letting i vary continuously, we find that

$$\frac{dy}{di} = -\frac{\delta\tau\{[1/(n-1)] + f\}E(i)}{(1 + f + g)^2} < 0$$

and

$$\frac{d^2y}{di^2} = \frac{2\delta^2\tau^2E(2i)\{[1/(n-1)] + f\}}{(1 + f + g)^3} + \frac{\alpha\delta\tau^2E(i)\{[1/(n-1)] + f\}}{(1 + f + g)^2} > 0$$

6. INVERTED *U* IN PERFORMANCE, PAYING ATTENTION, AND PRIMACY VS. RECENCY

This section discusses three interrelated themes. The "inverted *U* in performance" refers to empirical curves of the type shown in Fig. 2. Figure 2 points out a feature

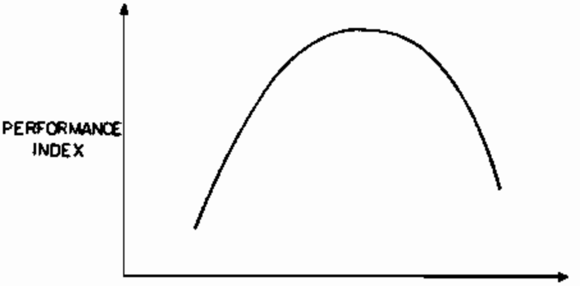

Fig. 2

characteristic of many experimental situations in which a fixed performance index is plotted against a prescribed arousal parameter, and it is found that performance suffers due both to underarousal and overarousal.[11–13] We will study the closely related "inverted *U* in learning," for which both underarousal and overarousal diminish the rate of learning, and thereby influence the accuracy of performance. Underarousal will slow learning by providing too little energy to drive the learning process. Overarousal is characterized by an ample supply of energy, but too much response interference from incorrect response alternatives. See Section 4 for related comments.

At least two mechanisms exist in our system for creating overarousal: first, a large amplification of all inputs to the list-representing states by a nonspecific arousal source; second, a pathological reduction in the thresholds of list-representing states. Experimental data[11,12] has suggested that the former mechanism can operate *in vivo*, with the reticular formation as a possible nonspecific arousal source. The second mechanism could in principle by brought into play by pathological changes in ion binding in the list-representing cells, none of which is necessarily a nonspecific arousal source.

The second theme concerns itself with one possible mechanism for increasing the difficulty of paying attention. Experimental evidence[12] has suggested that suitable forms of overarousal can interfere with paying attention. In the present case, overarousal allows the most recent events to competitively inhibit the trace of past events. Thus, only the most recent events can influence future behavior. If such a mechanism operated *in vivo*, by the time a long sentence could be fully presented, the trace of the sentence's beginning would be inhibited by the presentation of the end. Thus, it would be very difficult to comprehend the meaning of the entire sentence. Indeed, a predominance of low-order associations to the last few words of the sentence would be available. This fact is compatible with the existence of punning behavior in suitable mentally ill patients who are presumed to be in a continual state of overarousal.[11,14]

These facts suggest the third theme. Namely, we are led to study the influence of arousal level on the relative strength of the traces of past and recent events that are presented consecutively in time. In particular, in the serial learning situation, we vary the level of nonspecific arousal, or equivalently, the common threshold of all list-representing states, and study the relative strength of associations at the beginning and the end of the list. In normal subjects, the associations at the list's beginning are stronger than the associations at the list's end (that is, primacy dominates recency). Our previous remarks suggest that as the threshold becomes too low, ultimately the end of the list will have stronger associations than it's beginning (that is, recency dominates primacy). This is indeed the case in the present model. Whether varying arousal in this way actually causes a reversal of the normal (primacy/recency) ratio *in vivo* is unknown to us and therefore stands as a conjecture.

We therefore study the functions $P_{jk}(t, \Gamma) = y_{jk}(t + [j-1]\tau, \Gamma)$ and the functions

$$Q_{jk}(t, \Gamma) = y_{j,j+1}(t + [j-1]\tau, \Gamma)/y_{k,k+1}(t + [k-1]\tau, \Gamma) \tag{61}$$

$j, k = 1, 2, \ldots, L-1$, which compare the jth associational strength with the kth associational strength at corresponding times after the associations are activated by inputs. We will be particularly interested in $Q_{1,L-1}(t, \Gamma)$ to study primacy–recency, and denote this function by $Q(t, \Gamma)$ for brevity. The following facts will be established.

Proposition 2 (*Relative Initial Associational Growth*). For all $\Gamma \geqslant 0$ which are sufficiently small that some learning occurs, $\partial Q_{j,j+1}(0, \Gamma)/\partial \Gamma > 0, j = 1, 2, \ldots, L-1$. That is, the initial growth rate of each correct association decreases as a function of list position. This is due to the increase in response interference due to prior items in the list as j increases.

Theorem 3 (*Time-Evolution of Associations*). For all $\Gamma \geqslant 0$, $\partial P_{j,j+1}(t, \Gamma)/\partial \Gamma$ changes sign from positive to nonpositive at most once as $t \to \infty, j = 1, 2, \ldots, L-1$. In particular, for all $\Gamma \geqslant 0$, $\partial P_{L-1,L}(t, \Gamma)/\partial t \geqslant 0$. That is, response interference due to successively presented future items can build up at a given list position until initial associational growth becomes asymptotic decay. On the other hand, no future items occur on a given trial at the end of the list, and at this list position, associational strength cannot decrease as $t \to \infty$.

Theorem 4 (*Asymptotic Primacy–Recency vs. Under–Overarousal*).

$$Q(\infty, 0) < 1. \; \partial Q(\infty, \Gamma)/\partial \Gamma > 0$$

as Γ increases from 0 to a unique value $\Gamma = \Gamma_1$ such that $Q(\infty, \Gamma_1) = 1$. As Γ increases above Γ_1, $\partial Q(\infty, \Gamma)/\partial \Gamma > 0$ until a unique value Γ_2 is reached. For $\Gamma \geqslant \Gamma_2$, $\partial Q(\infty, \Gamma)/\partial \Gamma \leqslant 0$, but $Q(\infty, \Gamma) \geqslant 1$ for all $\Gamma \geqslant \Gamma_1$. Clearly, $Q(\infty, \Gamma) = 1$ for sufficiently large Γ, since no learning can occur if Γ is too large.

That is, if the threshold is too small, recency dominates primacy due to overarousal. As threshold increases, primacy eventually dominates recency, but if the threshold is too large, the list-representing states cannot emit sampling signals at all, so that no learning occurs, and (degenerately) primacy equals recency. See Section 4 for related comments.

Proposition 4 (*Asymptotic Associational Strength As a Function of Remoteness*). At any fixed time after full presentation of the list, both forward and backward associational strengths at any given list position decrease as a negatively accelerated function of remoteness.

Remark (*Maximal vs. Asymptotic Primacy vs. Recency*). It sometimes occurs that

$$\max_t P_{12}(t, \Gamma) > \max_t P_{L-1,L}(t, \Gamma)$$

despite that fact that

$$P_{12}(\infty, \Gamma) < P_{L-1,L}(\infty, \Gamma)$$

That is, there is actually a reversal in primacy dominating recency as a result of the increase in response interference as $t \to \infty$. On the other hand, it can occur that

$$\max_t P_{12}(t, \Gamma) < \max_t P_{L-1,L}(t, \Gamma)$$

and

$$P_{12}(\infty, \tau) < P_{L-1,L}(\infty, \Gamma)$$

Computer studies show that, for fixed Γ, one can proced from the former to the latter case simply by increasing the local decay rate α at all list-representing states.

It should therefore be clearly realized that primacy dominates recency for sufficiently large Γ because, for *all* $\alpha > 0$, the *initial* growth rate of associations at the beginning dominates the growth rate at the end, and that increasing Γ prevents sampling of too many remote associations.

7. PROOFS OF THRESHOLD-DEPENDENT FACTS

Proposition 2. Under conditions (a)–(e), $\partial Q_{j,j+1}(0, \Gamma)/\partial t > 0$ for all $j = 1, 2, \ldots, L-1$.

Proof. By condition (a), it suffices to show that

$$\partial P_{j,j+1}(0, \Gamma)/\partial t > \partial P_{j+1,j+2}(0, \Gamma)/\partial t.$$

Note that by condition (a), $P_{j,j+1}(0, \Gamma) = P_{j+1,j+2}(0, \Gamma)$, whereas by Proposition 1,

$$P_{j,j+1}(t, \Gamma) > P_{j+1,j+2}(t, \Gamma) \text{ for sufficiently small } t > 0.$$

Theorem 4. Under conditions (a)–(e), $\partial P_{j,j+1}(t, \Gamma)/\partial t$ changes sign from positive to nonpositive at most once as $t \to \infty$, $j = 1, 2, \ldots, L-1$. In particular,

$$\partial P_{L-1,L}(t, \Gamma)/\partial t \geqslant 0 \qquad \text{for all} \quad t \geqslant 0.$$

Proof. First we show that $\partial P_{j,j+1}(t, \Gamma)/\partial t$ changes sign from positive to nonpositive at most once. Clearly, $\partial P_{j,j+1}(0, \Gamma)/\partial t > 0$. It remains only to show that $\partial P_{j,j+1}(T, \Gamma)/\partial t \leqslant 0$ for any $T \geqslant 0$ implies $\partial P_{j,j+1}(t, \Gamma)/\partial t \leqslant 0$ for all $t \geqslant T$. First we compute $\partial P_{j,j+1}(t, \Gamma)/\partial t$. By (20),

$$P_{j,j+1}(t, \Gamma) = \frac{[1/(n-1)] + f_{j,j+1}(t-\tau)}{1 + f_{j,j+1}(t-\tau) + g_j(t-\tau) + h_j(t-\tau)}$$

where we let $\rho = 1$ without loss of generality. Henceforth, all subscripts j on the functions f, g, and h will be omitted for convenience. Clearly,

$$\operatorname{sign} \frac{\partial P_{j,j+1}(t+\tau, \Gamma)}{\partial t} = \operatorname{sign}\left[\left(\frac{n-2}{n-1} + g + h\right)\dot{f} - \left(\frac{1}{n-1} + f\right)(\dot{g} + \dot{h})\right] \qquad (62)$$

Computation of $\dot{f}$ and $\dot{g}$ readily follows from (30) and (33), respectively. We find

$$\dot{f} = Ae^{-\alpha t}[Ae^{-\alpha t} - \Gamma] \quad (> 0) \tag{63}$$

and

$$\dot{g} = E^{-1}(-\tau, 0)\, E(\tau, j\tau)\, \dot{f} \tag{64}$$

Computation of $\dot{h}$ requires consideration of Eqs. (35)–(37).

Consider (35). Then $\dot{h} = 0$. We will prove that this implies

$$\begin{aligned} \partial P_{j,j+1}(t + \tau, \Gamma)/\partial t &> 0 \qquad \text{if} \quad T_1 - \tau < t < T_2 - \tau \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{65}$$

By (62)–(64),

$$\text{sign}\ \partial P_{j,j+1}(t + \tau, \Gamma)/\partial t = \text{sign}\ \dot{f}$$

from which our claim is obvious. Note in particular that $j = L - 1$ implies that $h \equiv 0$. Hence, (65) holds whenever $j = L - 1$.

Consider (36). Then, h can be broken up into the sum

$$h = h^{(1)} + h^{(2)} \tag{66}$$

where

$$\begin{aligned} h^{(1)} = {} & AE^{-1}(0, \tau)\, E(\tau, [\sigma - j - 1]\, \tau)[B(\lambda, 0) + (A/2\alpha)\, e^{-2\alpha\lambda}] \\ & + (A/\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma - j - 1]\, \tau, -\tau)[\Gamma e^{-\alpha t} - \tfrac{1}{2}Ae^{-2\alpha t}] \\ & - \Gamma[C(\lambda, 0) + (A/\alpha)\, e^{-\alpha\lambda}](\sigma - j - 2) \end{aligned} \tag{67}$$

and

$$h^{(2)} = AE(\sigma - j - 1)\, B(t - [\sigma - j - 1]\, \tau, 0) - \Gamma C(t - [\sigma - j - 1]\, \tau, 0) \tag{68}$$

Differentiating (66), we find

$$\dot{h} = \dot{h}^{(1)} + \dot{h}^{(2)}$$

where, by (63),

$$\dot{h}^{(1)} = E^{-1}(-\tau, 0)\, E(-[\sigma - j - 1]\, \tau, -\tau)\, \dot{f} \tag{69}$$

and

$$\dot{h}^{(2)} = A^{-1}e^{\alpha t}\dot{f}x_1(t - [\sigma - j - 1]\, \tau) \tag{70}$$

Substituting (69) and (70) into (62) using the notation

$$\mu = E^{-1}(-\tau, 0)\{E(\tau, j\tau) + E(-[\sigma - j - 1]\, \tau, -\tau)\}$$

we find

$$\text{sign}\ \partial P_{j,j+1}(t + \tau, \Gamma)/\partial t = \text{sign}\ S \tag{71}$$

where

$$S = \frac{n-2}{n-1} + g + h - \left(\frac{1}{n-1} + f\right)\{\mu + A^{-1}e^{\alpha t}x_1(t - [\sigma - j - 1]\tau)\} \tag{72}$$

The case of Eq. (36) will be completed by showing that S is monotone decreasing. Write S as the sum

$$S = S^{(1)} + S^{(2)}$$

where

$$S^{(1)} = \frac{n-2}{n-1} + g + h^{(1)} - \mu\left(\frac{1}{n-1} + f\right) \tag{73}$$

and

$$S^{(2)} = h^{(2)} - A^{-1}e^{\alpha t}\left(\frac{1}{n-1} + f\right) x_1(t - [\sigma - j - 1]\,\tau) \tag{74}$$

It can be shown that $S^{(1)}$ is constant. It remains only to show that $S^{(2)}$ is monotone decreasing.

Note that

$$\operatorname{sign} \dot{S}^{(2)} = \operatorname{sign}[\dot{h}^{(2)} f - \dot{f} h^{(2)}] \tag{75}$$

where

$$\dot{f} = -\alpha A e^{-\alpha t}[2Ae^{-\alpha t} - \Gamma] \tag{76}$$

and

$$\dot{h}^{(2)} = -\alpha A e^{-\alpha t} x_1(t - [\sigma - j - 1]\,\tau) + [Ae^{-\alpha t} - \Gamma]\,\dot{x}_1(t - [\sigma - j - 1]\,\tau) \tag{77}$$

By (63), (70) and (75)-(77), the condition $\dot{S}^{(2)} \leqslant 0$ is equivalent to the trivial inequality $\dot{x}_1 \geqslant -\alpha x_1$.

Consider Eq. (37). In this case, the proof proceeds just as in the previous case, but is simplified by the absence of the terms $h^{(2)}$ and $S^{(2)}$. Theorem 3 is therefore proved.

Theorem 4 will be stated in terms of three increasingly prescribed classes of inputs. Class $J^{(1)}$ satisfies conditions (b)–(d). Class $J^{(2)}$ is the subclass of $J^{(1)}$ satisfying the inequality

$$dI(T_1)/dt \leqslant 0 \tag{78}$$

i.e., inputs that are already decreasing when v_1 begins to sample. $J^{(3)}$ is the subclass of $J^{(2)}$ consisting of rectangular inputs.

Theorem 4. (I) For any input in $J^{(1)}$, $Q(\infty, 0) < 1$. (II) For any input in $J^{(2)}$, there exists a unique Γ_1 such that $Q(\infty, \Gamma_1) = 1$. (III) For any input in $J^{(3)}$, $\partial Q(\infty, \Gamma)/\partial \Gamma$ changes sign once from positive to nonpositive.

Remark. Computer runs in Section 8 indicate that II and III both hold for larger classes of input.

Proof. Point I follows by iterating (60). Point II will be proved by showing that the function $U(\Gamma) = h(\Gamma) - g(\Gamma)$ has one root under conditions (a)–(e). [$U(\Gamma) = 0$ iff $Q(\infty, \Gamma) = 1$.] The graph of U will be shown to have the form given in Fig. 3, where Γ_2 will be seen to be so large as to violate condition (d).

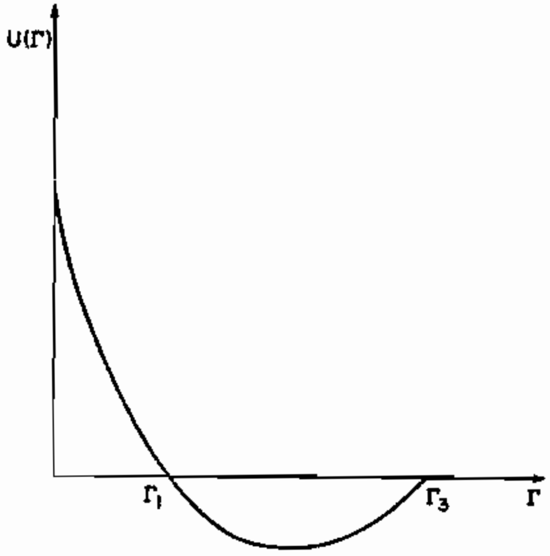

Fig 3

Figure 3 will be established in five steps. First, it is clear that $U(\Gamma)$ is a continuous function of Γ. Second, $U(\Gamma) = 0$ for all sufficiently large Γ, since each $f_{jk}(\infty, \Gamma) = 0$ for all sufficiently large Γ. Third, $U(\Gamma) < 0$ for all Γ in a sufficiently small left-neighborhood of Γ_3. This is because $h(\Gamma) = 0$ in such a neighborhood (degenerate bowing occurs), and thus $U(\Gamma) = -g(\Gamma) < 0$ in this neighborhood. Fourth, we will show that $dU(0)/d\Gamma < 0$, and fifth we will prove that $d^2U(\Gamma)/d\Gamma^2$ is monotone decreasing in Γ.

To prove that $dU(0)/d\Gamma < 0$, note by (34), (38), and (39) that

$$dg(\Gamma)/d\Gamma = AE^{-1}(-\tau, 0)\, E(\tau, [L-1]\,\tau)\, E(T_2, T_1) \tag{79}$$

and

$$\begin{aligned} dh(\Gamma)/d\Gamma &= (\Gamma/\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma-2]\,\tau, -\tau) - [C(\lambda, 0) + (Ae^{-\alpha\lambda}/\alpha)](\sigma - 3) \\ &\quad - C(T_2 + [\sigma - 2]\,\tau, 0), \qquad \text{if} \quad 0 \leqslant T_2 - (\sigma - 2)\,\tau \leqslant \lambda \\ &= (\Gamma/\alpha)\, E^{-1}(-\tau, 0)\, E(-[\sigma-1]\,\tau, -\tau) - [C(\lambda, 0) + (Ae^{-\alpha\lambda}/\alpha)](\sigma - 2), \\ &\qquad\qquad \text{if} \quad \lambda \leqslant T_2 - (\sigma - 2)\,\tau \end{aligned} \tag{80}$$

Thus, $dU(0)/d\Gamma < 0$ is equivalent to

$$(A/\alpha)\, E^{-1}(-\tau, 0)\, E(\tau, [L-1]\,\tau) < (\sigma - 2)[C(\lambda, 0) + (Ae^{-\alpha\lambda}/\alpha)]$$

Since $C(\lambda, 0)$ and A/α are positive, and $\tau > \lambda$, this inequality is implied by $(L-2) > E^{-1}(-\tau, 0)\, E(0, [L-2]\,\tau)$, which is obvious.

To prove $d^2U(\Gamma)/d\Gamma^2$ is monotone decreasing, we will directly compute that

$$d^3h(\Gamma)/d\Gamma^3 \leqslant d^3g(\Gamma)/d\Gamma^3 \tag{81}$$

By (79) and (80),

$$d^2g(\Gamma)/d\Gamma^2 = -AE^{-1}(-\tau, 0)\, E(\tau, [L-1]\,\tau)\, e^{-\alpha T_1}(\partial T_1/\partial\Gamma)^2\, dI(T_1)/dt \qquad (82)$$

and

$$\begin{aligned} d^2h(\Gamma)/d\Gamma^2 &= -(1/\alpha^2\Gamma^2)\, I(T_2 - [\sigma - 2]\,\tau), && \text{if } 0 \leqslant T_2 - (\sigma - 2)\,\tau \leqslant \lambda \\ &= 0, && \text{if } \lambda \leqslant T_2 - (\sigma - 2)\,\tau \end{aligned} \qquad (83)$$

Thus, the inequality (81) is implied by (78).

The proof of III is given in eight steps. In this proof, we will introduce two comparison functions $V(\Gamma)$ and $W(\Gamma)$ to help us determine the properties of $\partial Q(\infty, \Gamma)/\partial\Gamma$ and $\partial^2 Q(\infty, \Gamma)/\partial\Gamma^2$ for different values of Γ. Steps 1–3 are remarks about $V(\Gamma)$ and $W(\Gamma)$. Using these remarks, steps 4–6 prove that $Q(\infty, \Gamma)$ has no critical points for $\Gamma < \Gamma_2$. Finally, steps 7–8 prove that $Q(\infty, \Gamma)$ has one critical point for $\Gamma \geqslant \Gamma_2$. The steps are as follows:

(1) sign $\partial Q(\infty, \Gamma)/\partial\Gamma = \text{sign}[Q(\infty, \Gamma) - V(\Gamma)]$. Thus, if $\partial Q(\infty, \Gamma)/\partial\Gamma > 0$ between two critical points of $Q(\infty, \Gamma)$, then $V(\Gamma)$ has a critical point at which $V(\Gamma)$ is smaller than $Q(\infty, \Gamma)$ between the two critical points of $Q(\infty, \Gamma)$. See Fig. 4. Similarly, if $\partial Q(\infty, \Gamma)/\partial\Gamma < 0$ between two critical points of $Q(\infty, \Gamma)$, then $V(\Gamma)$ has a critical point at which $V(\Gamma)$ is greater than $Q(\infty, \Gamma)$ between the two critical points of $Q(\infty, \Gamma)$.

(2) $dW(\Gamma)/d\Gamma < 0$ for $\Gamma < \Gamma_2$, and $W(\Gamma)$ has at most one critical point for $\Gamma > \Gamma_2$.

(3) sign $dV(\Gamma)/d\Gamma = \text{sign}[V(\Gamma) - W(\Gamma)]$.

(4) There exists a $\Gamma_2 > \Gamma_1$ such that $\partial Q(\infty, \Gamma)/\partial\Gamma > 0$ for $\Gamma \in (\Gamma_1, \Gamma_2)$ and $\partial Q(\infty, \Gamma_2)/\partial\Gamma = 0$.

(5) If $\partial Q(\infty, \Gamma)/\partial\Gamma = 0$ at some $\Gamma < \Gamma_2$, then at this point, $Q(\infty, \Gamma) < 1$, $\partial^2 Q(\infty, \Gamma)/\partial\Gamma^2 > 0$, and $W(\Gamma) > Q(\Gamma)$.

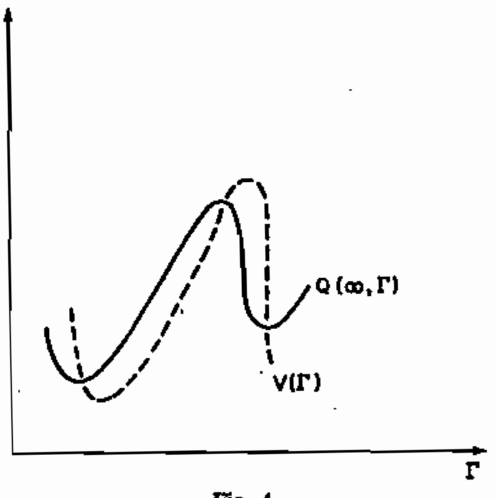

Fig. 4

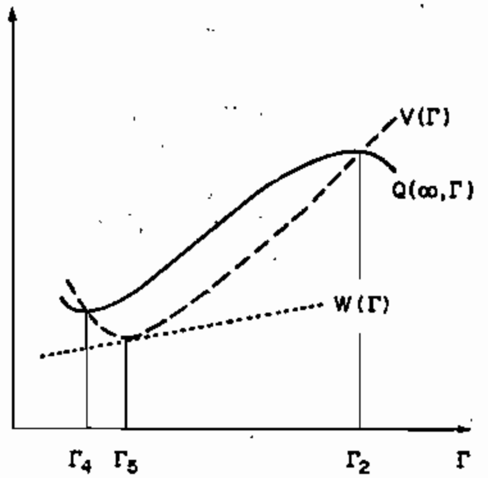

Fig. 5

(6) If at some $\Gamma < \Gamma_2$, say Γ_4, $Q(\infty, \Gamma_4) < 1$, $\partial Q(\infty, \Gamma_4)/\partial\Gamma = 0$, and $\partial^2 Q(\infty, \Gamma_4)/\partial\Gamma^2 > 0$, then steps 1 and 4 show $V(\Gamma)$ has a critical point, say Γ_5, such that $V(\Gamma_5) < Q(\infty, \Gamma_4)$. By step 3, $V(\Gamma_5) = W(\Gamma_5)$. Then, by step 2, $W(\Gamma_4) < Q(\Gamma_4)$. This last fact contradicts step 5 and shows that $Q(\infty, \Gamma)$ can have no critical point for $\Gamma < \Gamma_2$. See Fig. 5.

(7) Since $Q(\infty, \Gamma)$ increases for small Γ and decreases to one for large Γ, $Q(\infty, \Gamma)$ must have an odd number of critical points. Thus, if $Q(\infty, \Gamma)$ has a critical point at some $\Gamma > \Gamma_2$, it must have at least two critical points. Let Γ_6 and Γ_7 be the largest critical points smaller than Γ_3. See Fig. 6.

(8) A contradiction will be established, given the existence of Γ_6 and Γ_7, by showing that then $V(\Gamma_3) > W(\Gamma_3)$, whereas by direct computation, $V(\Gamma_3) = W(\Gamma_3)$.

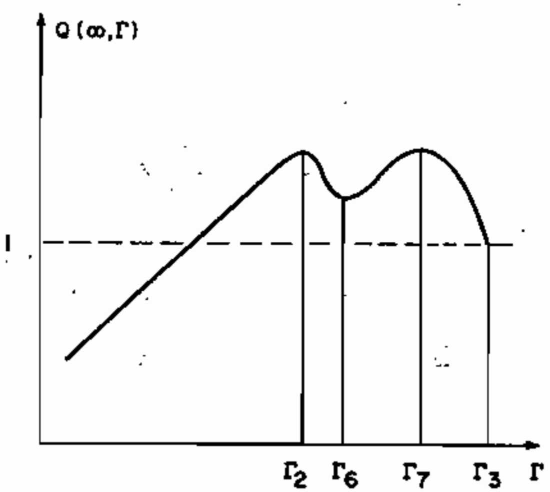

Fig. 6

The two comparison functions are defined as follows:

$$V(\Gamma) = \frac{\partial f/\partial \Gamma + \partial g/\partial \Gamma}{\partial f/\partial \Gamma + \partial h/\partial \Gamma} \tag{84}$$

and

$$W(\Gamma) = \frac{\partial^2 f/\partial \Gamma^2 + \partial^2 g/\partial \Gamma^2}{\partial^2 f/\partial \Gamma^2 + \partial^2 h/\partial \Gamma^2} \tag{85}$$

To verify step 1, differentiate (61) and substitute (84) in the result to find

$$\frac{\partial Q(\infty, \Gamma)}{\partial \Gamma} = \frac{\partial f/\partial \Gamma + \partial h/\partial \Gamma}{1 + f + h} [V(\Gamma) - Q(\infty, \Gamma)] \tag{86}$$

and note that $\partial f/\partial \Gamma$ and $\partial h/\partial \Gamma$ are negative.

We now verify step 2. Differentiation of (85) shows that

$$\text{sign} \frac{dW(\Gamma)}{d\Gamma} = \text{sign} \left[\frac{\partial^2 U}{\partial \Gamma^2} \cdot \frac{\partial^3 f}{\partial \Gamma^3} - \left(\frac{\partial^2 f}{\partial \Gamma^2} + \frac{\partial^2 g}{\partial \Gamma^2} \right) \frac{\partial^3 h}{\partial \Gamma^3} \right] \tag{87}$$

To compute (87), note that

$$\partial^2 f/\partial \Gamma^2 = e^{\alpha T_1}/\alpha \tag{88}$$

$$\partial^3 f/\partial \Gamma^3 = e^{\alpha T_1}/(I - \alpha \Gamma) \tag{89}$$

and

$$\partial^2 g/\partial \Gamma^2 = (e^{\alpha\lambda}/\alpha)\, E(\tau, [L-1]\,\tau)\, E^{-1}(-\tau, 0) \tag{90}$$

By (83), (88), and (90), $(\partial^2 f/\partial \Gamma^2 + \partial^2 g/\partial \Gamma^2)\, \partial^3 h/\partial \Gamma^3 \leqslant 0$ for all $\Gamma \geqslant 0$. By (87), $\text{sign}[(\partial^2 U/\partial \Gamma^2) \cdot (\partial^3 f/\partial \Gamma^3)] = \text{sign}\, \partial^2 U/\partial \Gamma^2$. Part II implies that $\partial^2 U/\partial \Gamma^2$ changes sign once from positive to negative. Therefore $dW/d\Gamma$ changes sign at most once from positive to negative. In particular, by part II, $\partial^2 U/\partial \Gamma^2 > 0$ if $\Gamma > \Gamma_2$, and thus $dW/d\Gamma$ can only change sign if $\Gamma \geqslant \Gamma_2$.

Step 3 follows from the equation

$$\frac{dV}{d\Gamma} = \frac{\partial^2 f/\partial \Gamma^2 + \partial^2 h/\partial \Gamma^2}{\partial f/\partial \Gamma + \partial h/\partial \Gamma} (W - V)$$

Step 4 is trivial.

Consider step 5. If $\partial Q(\infty, \Gamma)/\partial \Gamma = 0$ at some $\Gamma > \Gamma_2$, then $Q(\infty, \Gamma) < 1$ by step 4. The remaining assertions follows from the equation

$$\frac{\partial^2 Q}{\partial \Gamma^2} = \frac{\partial^2 f/\partial \Gamma^2 + \partial^2 h/\partial \Gamma^2}{1 + f + h} (W - Q)$$

which holds whenever $\partial Q/\partial \Gamma = 0$.

Steps 6 and 7 are self-explanatory.

Finally, consider step 8. Figure 7 illustrates the source of contradiction. For $\Gamma \geqslant \Gamma_7$, $V > Q > W$. Direct computation shows, however, that $V(\Gamma_3) = W(\Gamma_3)$.

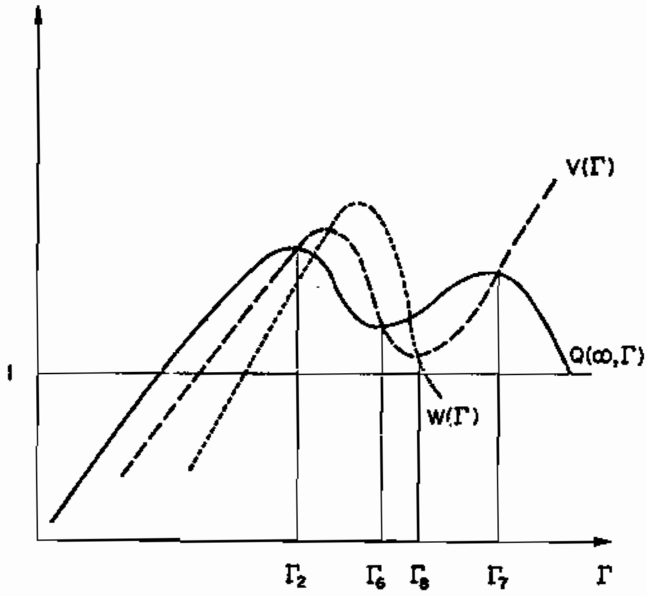

Fig. 7

Indeed, all the partial derivatives $\partial f/\partial\Gamma$, $\partial g/\partial\Gamma$, and $\partial h/\partial\Gamma$ vanish at $\Gamma = \Gamma_3$ and are negative in a left-neighborhood of $\Gamma = \Gamma_3$. Application of L'Hospital's rule to (84) and comparison with (85) proves our claim, and hence the theorem.

Proposition 3. Let $t \geqslant L\tau$. (I) If $j > k$, $y_{jk}(t)$ is a negatively accelerated, monotone-increasing function of k, where $j = 2,..., L$ and $k = 1,..., j-1$. (II) If $j < k-1$, $y_{jk}(t)$ is a negatively accelerated, monotone-decreasing function of k, where $j = 1,..., L-2$ and $k = j+2,..., L$. (III) $y_{j,j+2}(t) > y_{j,j+2}(t)$ if $Z(\lambda, \Gamma) > AE(1)\, B(\lambda, 0)$, where $Z(\lambda, \Gamma) = D(\lambda - T_1, T_1) + \Gamma C(T_1, 0)$. Note that $\partial Z(\lambda, \Gamma)/\partial\Gamma > 0$.

Proof. To prove I, we will show that $\partial y_{jk}(t)/\partial k > 0$ and $\partial^2 y_{jk}(t)/\partial k^2 > 0$. By (20), (30), (33), and (37), the denominator of $y_{jk}(t)$ is independent of k. Thus, it suffices to show $\partial f_{jk}(t)/\partial k > 0$ and $\partial^2 f_{jk}(t)/\partial k^2 > 0$ for $j > k$. Indeed by (22),

$$\partial f_{jk}(t)/\partial k = \alpha\tau f_{jk}(t) \qquad \text{and} \qquad \partial^2 f_{jk}(t)/\partial k^2 = (\alpha\tau)^2 f_{jk}(t)$$

In proving II, we again note that the denominator of $y_{jk}(t)$ is independent of k. Thus, we need only show that $\partial f_{jk}(t) < 0$ and $\partial^2 f_{jk}(t)/\partial k^2 > 0$ for $j < k-1$. This is clear by inspection of the following equations derived from (27):

$$\partial f_{jk}(t)/\partial k = -\alpha\tau AE(k-j-1)\{B(\lambda, 0) + \tfrac{1}{2}AE(2\lambda, 2[t-(k-1)\tau]\}$$
$$-A^2\tau e^{-\alpha[t-(k-1)\tau]}[e^{-\alpha(t-j\tau)} - e^{-\alpha T_2}]$$

and

$$\partial^2 f_{jk}(t)/\partial k^2 = (\alpha\tau)^2 AE(k-j-1)\{B(\lambda, 0) + \tfrac{1}{2}AE(2\lambda, 2[t-(k-1)\tau]\}$$
$$+ \alpha\tau^2 A^2 e^{-\alpha[t-(k-1)\tau]} e^{-\alpha T_2}$$

For $y_{j,j+1}(t) > y_{j,j+2}(t)$ to be satisfied, it suffices by (20) to show that $f_{j,j+1}(t) > f_{j,j+2}(t)$. Since, clearly,

$$\tfrac{1}{2}A^2E(2\lambda, 2[t - j\tau]) + \Gamma AE(t - j\tau, \lambda)$$
$$> \tfrac{1}{2}A^2E(2\lambda, 2[t - (j + 1)\tau]) + \Gamma AE(t - [j + 1]\tau, \lambda)$$

(27) and (30) show that we need only prove

$$D(\lambda - T_1, T_1) - \Gamma C(\lambda - T_1, T_1) > AE(1)\, B(\lambda, 0) - \Gamma C(\lambda, 0)$$

This inequality reduces to

$$D(\lambda - T_1, T_1) > AE(1)\, B(\lambda, 0) - \Gamma C(T_1, 0)$$

which is the desired result.

8. PARAMETRIC STUDIES

We now present the results of computer studies that were done to quantitatively illustrate the effects of varying the parameters α and Γ on the strength of primacy vs. recency. In addition, $Q(\infty, \Gamma)$ and $U(\Gamma)$ were studied for input functions not in the classes $J^{(2)}$ and $J^{(3)}$. The values of τ and λ will be held fixed at $3\pi/16$ and $\pi/8$, respectively.

Increasing Γ. As we have seen in Theorem 4, for small Γ, recency asymptotically dominates primacy. This fact is illustrated in Fig. 8, where $\alpha = 3\pi/16$, $\Gamma = 0$,

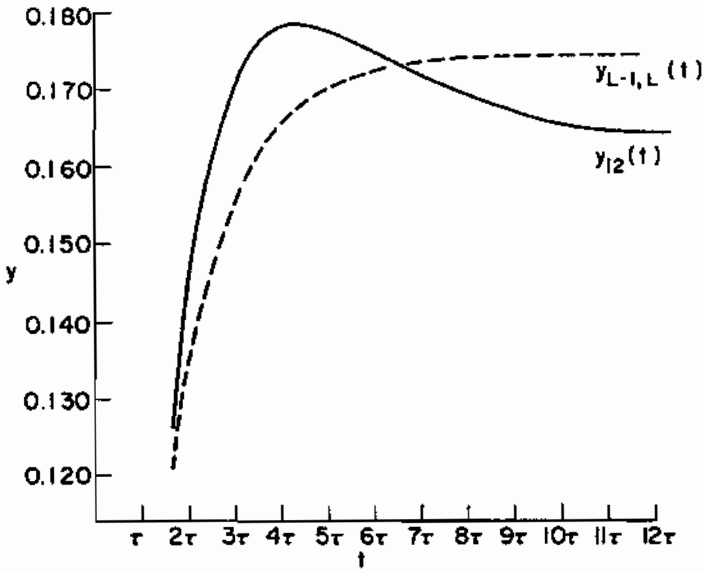

Fig. 8

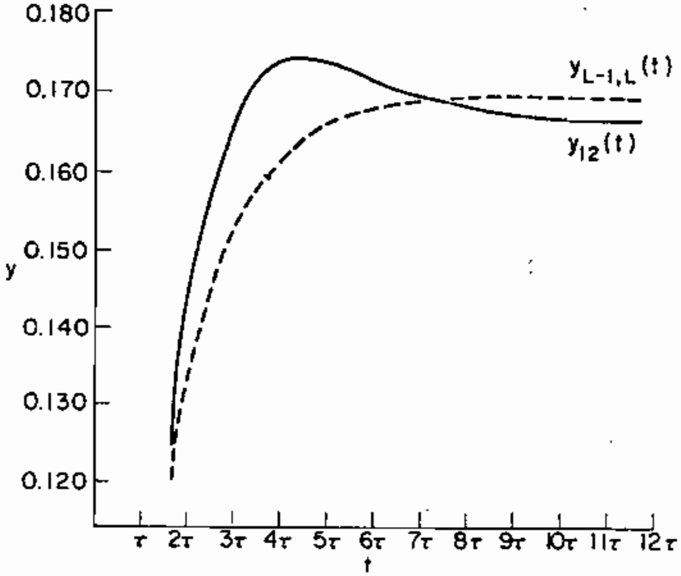

Fig. 9

and $J(t)$ is a rectangular input pulse. Then, as Γ is increased to 0.002 in Fig. 9, the gap between the asymptotic values of y_{12} and $y_{L-1,L}$ becomes smaller. Recency is losing ground to primacy due to the decreased sampling time of v_1. Finally, primacy dominates recency for $\Gamma = 0.004$, as is illustrated in Fig. 10.

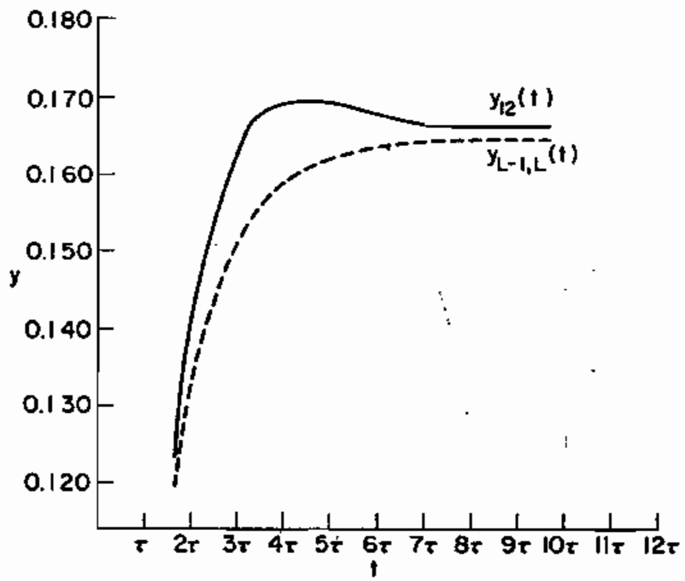

Fig. 10

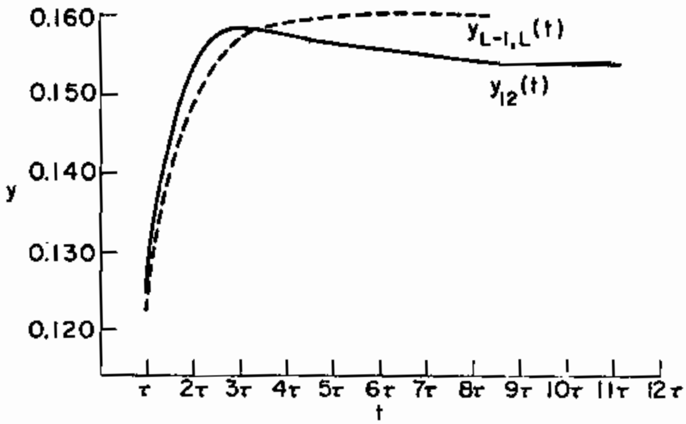

Fig. 11

Increasing α. For small α, the maximum of $y_{12}(t)$ is greater than the maximum of $y_{L-1,L}(t)$. This fact is illustrated in Fig. 8, where $\alpha = 3\pi/16$, $\Gamma = 0$, and $J(t)$ is a rectangular input pulse. By increasing α and holding Γ constant, we do not decrease the number of future list items v_1 can sample; hence, asymptotic recency will still dominate primacy for small Γ. However, for large α, the maximum of $y_{12}(t)$ is smaller than the maximum of $y_{L-1,L}(t)$. See Figs. 11 and 12, where Γ is held constant at 0 and $\alpha = 5\pi/16$ and $7\pi/16$, respectively.

Q(∞, Γ) and U(Γ). Let $J(t) = \sin(8t)$ if $0 \leqslant t \leqslant \pi/8$ and 0 otherwise. Then, $Q(\infty, \Gamma)$ and $U(\Gamma)$ are illustrated in Figs. 13 and 14, respectively. In addition, if $J(t) = (8/\pi)\, t$ for $0 \leqslant t \leqslant \pi/8$ and 0 otherwise, then $Q(\infty, \Gamma)$ and $U(\Gamma)$ are given

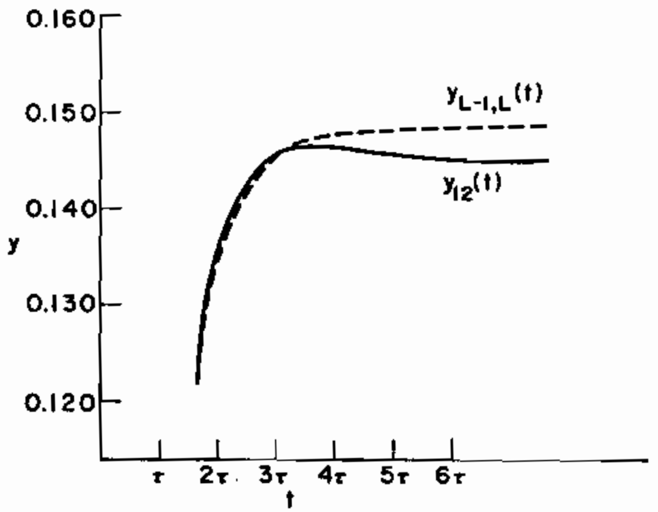

Fig. 12

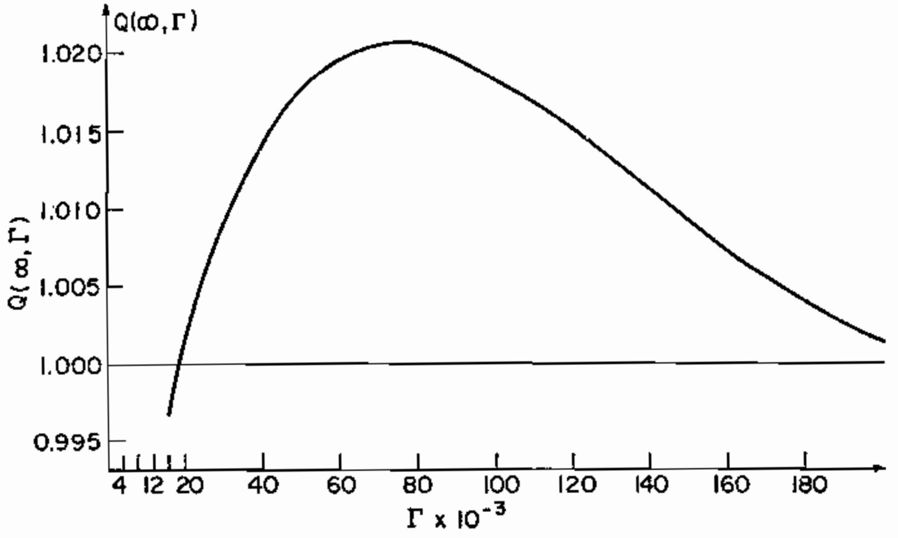
Q(∞, Γ)
1.020
1.015
1.010
1.005
1.000
0.995
Q(∞, Γ)
4 12 20 40 60 80 100 120 140 160 180
Γ x 10⁻³

Fig. 13

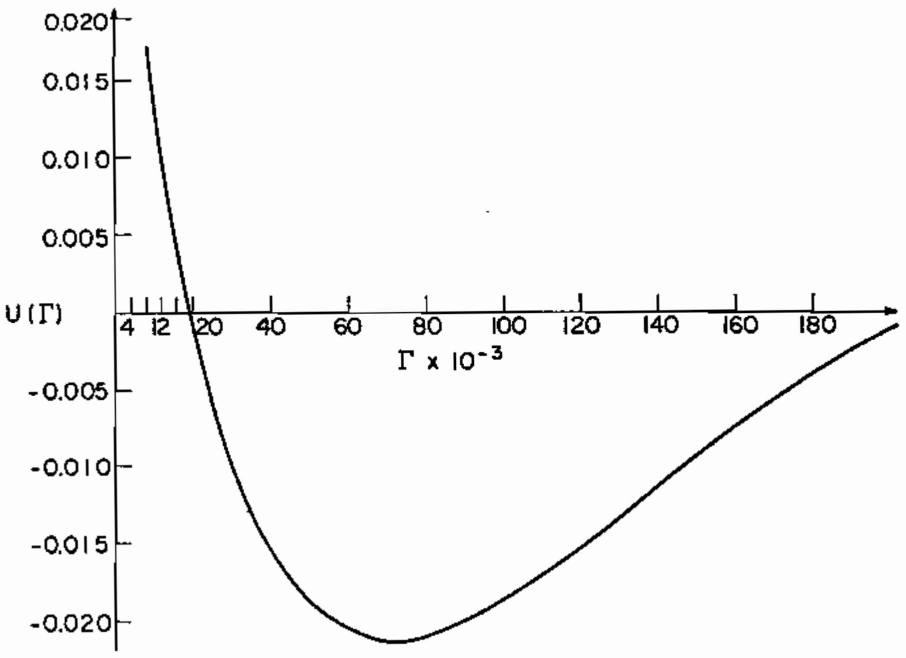
0.020
0.015
0.010
0.005
U (Γ)
-0.005
-0.010
-0.015
-0.020
4 12 20 40 60 80 100 120 140 160 180
Γ x 10⁻³

Fig. 14

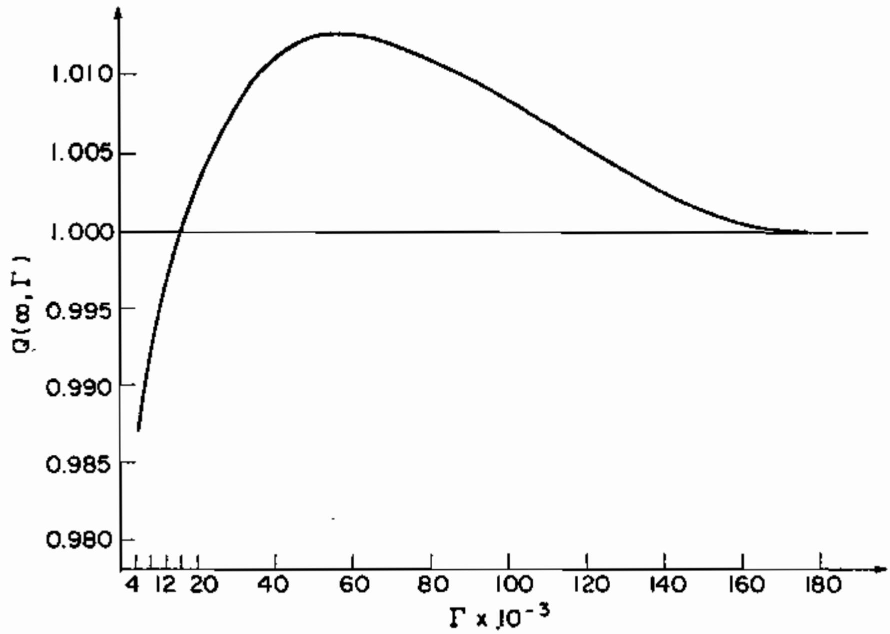
1.010
1.005
1.000
0.995
0.990
0.985
0.980
Q(8, Γ)
4 12 20 40 60 80 100 120 140 160 180
Γ x 10⁻³

Fig. 15

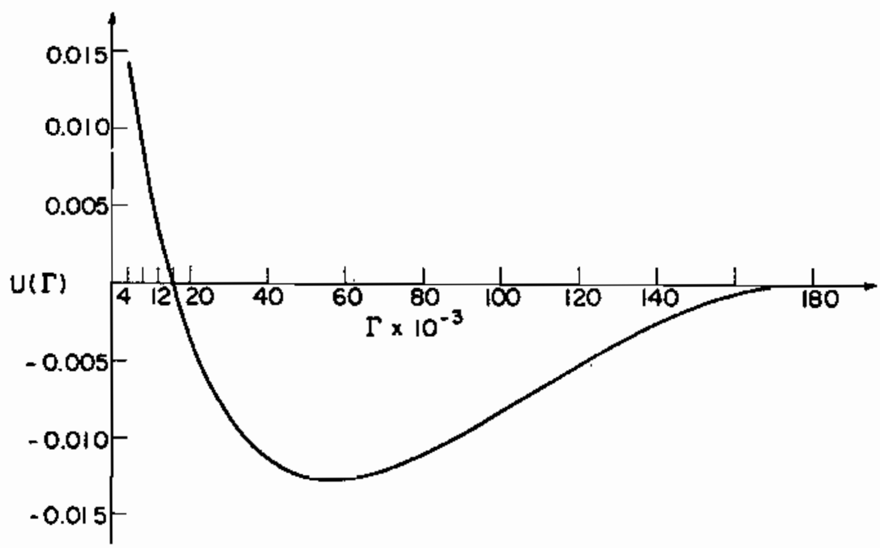
0.015
0.010
0.005
U(Γ)
- 0.005
- 0.010
- 0.015
4 12 20 40 60 80 100 120 140 180
Γ x 10⁻³

Fig. 16

by Figs. 15 and 16, respectively. These figures show that the results of Theorem 4 are true for larger classes of inputs then $J^{(2)}$ and $J^{(3)}$, possibly for all inputs satisfying conditions (b)–(d).

Chunking. These parametric studies show a very pronounced chunking effect in learning a list of length L. That is, if the L items are presented in the form of sublists (chunks) each separated by sufficiently long intertrial intervals, then these sublists are easier to learn than the original list presented in its entirety. The data on chunking were obtained by holding α and Γ constant at $3\pi/16$ and 0, respectively, while varying L. A typical set of data is the following. For $L = 5$, $y_{12}(\infty) = 0.25$; for $L = 10$, $y_{12}(\infty) = 0.125$; and for $L = 20$, $y_{12}(t) = 0.05$. Thus, if a list of length 20 is presented, we expect the associational strengths of successive list items to be on the order of 0.05. However, if this list is broken into four chunks each of length 5, the associational strengths are now on the order of 0.25. This fact shows a more rapid learning of the chunks than of the original list presented in its entirety. Otherwise expressed, "part" learning is easier than "whole" learning.

REFERENCES

1. S. Grossberg and J. Pepe, Schizophrenia: Possible dependence of associational span, bowing, and primacy vs. recency on spiking threshold, *Behavioral Sci.* **15** (4): 359–362 (1970).
2. S. Grossberg, Embedding fields. A theory of learning with physiological implications, *J. Math. Psychol.* **6**(2):209–239 (1969).
3. S. Grossberg, On the production and release of chemical transmitters and related topics in cellular control, *J. Theoret. Biol.* **22**:325–364 (1969).
4. S. Grossberg, On learning, information, lateral inhibition, and transmitters, *Math. Biosci.* **4**:255–310 (1969).
5. S. Grossberg, On the serial learning of lists, *Math. Biosci.* **4**:201–253 (1969).
6. S. Grossberg, On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks, *J. Stat. Phys.* **1**(2):319–350 (1969).
7. S. Grossberg, Some networks that can learn, remember, and reproduce any number of complicated space-time patterns (I), *J. Math. Mech.* **19**(1):53–91 (1969).
8. S. Grossberg, Some networks that can learn, remember and reproduce any number of complicated space-time patterns (II), *Studies Appl. Math.* **XLIX** (2): 135–166 (1970).
9. S. Grossberg, Neural pattern discrimination, *J. Theoret. Biol.* **27**:291–337 (1970).
10. S. Grossberg, On learning of spatiotemporal patterns by networks with ordered sensory and motor components. 1. Excitatory components of the cerebellum, *Studies Appl. Math.* XLVIII(2): 105–132 (1969).
11. C. Kornetsky and M. Eliasson, Reticular stimulation and chlorphromazine: An animal model for schizophrenic overarousal, *Science* **165** (3899):1273–1274 (1969).
12. M. I. Phillips and P. B. Bradley, Reticular stimulation and chlorpromazine, *Science* **168**(3935): 1122–1123 (1970).
13. D. O. Hebb, Drives and the CNS (conceptual nervous system), *Physiol. Rev.* **62**:243–54 (1955).
14. B. A. Maher, The shattered language of schizophrenia, *Psychology Today* **2**(6):30–33 (1968).

Printed in Belgium by the St. Catherine Press Ltd., Tempelhof 37, Bruges