

*Some Networks That Can Learn, Remember,
and Reproduce Any Number of Complicated
Space-Time Patterns, I*

STEPHEN GROSSBERG

Offprint from

JOURNAL OF MATHEMATICS AND MECHANICS

Vol. 19, No. 1, July, 1969, pp. 53–91

Some Networks That Can Learn, Remember, and Reproduce Any Number of Complicated Space-Time Patterns, I

STEPHEN GROSSBERG*

Communicated by GIAN-CARLO ROTA

	Page
1. Introduction.....	53
2. Network equations.....	54
3. Cross-correlated flows on signed networks.....	55
4. Outstars.....	57
5. Learning a single spatial pattern of arbitrary complexity.....	59
6. Respondant conditioning.....	61
7. Complete graph with loops: Pattern completion and the rigidity-plasticity continuum.....	62
8. Approximating any continuous space-time pattern by a series of spatial patterns.....	64
9. A better approximation: Signal thresholds eliminate source noise.....	67
10. Simultaneous storage of patterns and multimodal learning.....	80
11. Diffuse arousal inputs.....	82
12. Feedback inhibition from control vertices.....	84
13. A hierarchy of facilitatory and incompatible behavioral acts.....	85
14. Novelty and habituation.....	88
15. Higher-order controls.....	88
16. "On"- "Off" performance of complicated reflexes and "paying attention".....	88
17. Alternative systems.....	90

1. Introduction. This paper describes some networks \mathfrak{N} that can learn, simultaneously remember, and individually reproduce on demand any number

* The preparation of this work was supported in part by the National Science Foundation (GP 9003) and the Office of Naval Research (N00014-67-A-0204-0016).

of spatiotemporal patterns (*e.g.*, “motor sequences”) of essentially arbitrary complexity. Because these networks are *embedding fields*, their behavior can be psychologically, neurophysiologically, and anatomically interpreted ([1], [2], [3], [4]). The network properties include the following.

- (a) “Practice makes perfect”.
 - (b) Memory of each pattern is essentially perfect if no competing experimental practice is imposed.
 - (c) New patterns can be learned without at all destroying the memory of old patterns.
 - (d) All errors can be corrected.
 - (e) No “subject-induced” overt or covert practice is needed to ensure perfect memory.
 - (f) Given a moderate amount of practice, memory spontaneously improves (*i.e.*, “reminiscence” occurs).
 - (g) Memory is not destroyed by recall trials.
 - (h) Learning occurs by a mixture of respondent and operant conditioning, the operant effects including nonspecific arousal inputs in response to “novel” stimuli, and induced blocking of incoming inputs by inhibitory signals, leading to “habituation” of repeated inputs. Both respondent and operant factors are unified into a single comprehensive learning mechanism.
 - (i) Only one “control neuron” is needed to activate reproduction of an entire space-time pattern.
 - (j) The time needed to begin recall of a pattern can be made as small as we please, and is independent of pattern complexity.
 - (k) The network is insensitive to wild “behaviorally irrelevant” oscillations of inputs.
 - (l) The network dynamics, though nonlinear, can be analysed globally.
- Networks that perform any number of complicated “reflex acts” (*e.g.*, “walking”, “clasping”, “sniffing”) will also be constructed, as a special case of the learning networks. These “reflex” networks also satisfy (i)–(l) above.

2. Network equations. The networks we will use were derived in [2] and [3], and are reviewed in [1]. They are a special case of the following equations.

$$(1) \quad \dot{x}_i(t) = -\alpha_i x_i + \sum_{m=1}^n \beta_m [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ y_{mi}(t) - \sum_{m=1}^n [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ q_{mi} + I_i(t),$$

$$(2) \quad y_{ik}(t) = p_{ik} z_{ik}(t) \left[\sum_{m=1}^n p_{im} z_{im}(t) \right]^{-1}, \quad \text{and}$$

(*)

$$(3) \quad \dot{z}_{ik}(t) = \begin{cases} -u_{ik} z_{ik}(t) + v_{ik} [x_i(t - \tau_{ik}) - \Gamma_{ik}]^+ x_k(t), & \text{if } p_{ik} > 0 \\ 0, & \text{if } p_{ik} = 0, \end{cases}$$

for $i, j, k = 1, 2, \dots, n$ where

$$(4) \quad [\eta]^+ = \max(\eta, 0)$$

for any real number η . The parameters, initial data, and inputs of (*) satisfy the following constraints.

(I) *Parameters:*

- (1) All constant parameters are nonnegative; e.g., $\alpha_i, \beta_m, \Gamma_{mi}, p_{ik}$.
- (2) The $n \times n$ matrix $P = ||p_{ik}||$ is *semi-stochastic*; i.e., $p_{ik} \geq 0$ and $\sum_{m=1}^n p_{im} = 0$ or 1 .
- (3) v_{jk} is positive only if p_{jk} is positive.
- (4) All time lags τ_{jk} are positive.
- (5) $p_{m,q_{mi}} = 0$ for all $i, m = 1, 2, \dots, n$.

(II) *Initial Data.* All initial data of $x_i(v)$ and $z_{jk}(v)$ for $v \leq 0$ is nonnegative and continuous. Moreover we suppose for convenience that $z_{jk}(0) > 0$ if and only if $p_{jk} > 0$.

(III) *Inputs.* All inputs $I_i(t)$ are bounded, nonnegative, and continuous for $t \geq 0$ and vanish for $t < 0$.

When we say henceforth that parameters, initial data, or inputs are chosen "arbitrarily", we will always mean "arbitrarily subject to (I)-(III)".

3. **Cross-correlated flows on signed networks.** The equations (1)-(3) describe a cross-correlated flow on a signed network \mathfrak{N} . Since variants of this flow have been previously described in several places ([1]-[13]), the following summary will be brief.

Let a finite directed graph $G = (V, E)$ be given with vertices $V = \{v_i : i = 1, 2, \dots, n\}$ and directed edges $E = \{e_{jk} : j, k = 1, 2, \dots, n\}$. e_{jk} is drawn as an arrow facing from the point v_j whose arrowhead N_{jk} touches the point v_k . $x_i(t)$ describes a process at v_i , whereas $z_{jk}(t)$ and $y_{jk}(t)$ describe processes at N_{jk} . At every time $t - \tau_{mi}$, the value $x_m(t - \tau_{mi})$ at v_m creates an excitatory signal of size

$$(5) \quad \beta_m [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ p_{mi}.$$

In particular, the signal (5) is positive only if the *path weight* p_{mi} is positive and if $x_m(t - \tau_{mi})$ is greater than the signal threshold Γ_{mi} . The signal (5) travels at a finite velocity along e_{mi} and reaches the arrowhead N_{mi} at time t . It thereupon interacts with the processes z_{mi} and y_{mi} , and a signal of size

$$(6) \quad \beta_m [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ y_{mi}(t)$$

is released from N_{mi} and reaches v_i , where it perturbs x_i . All signals from the various v_m combine additively at v_i , yielding the second term on the right hand side of (1). $x_i(t)$ also decays exponentially at a rate α_i , and is perturbed by the input $I_i(t)$ which is under the control of an experimentalist or other external environmental factors.

An inhibitory signal of size

$$(7) \quad [x_m(t - \tau_{mi}) - \Gamma_{mi}]^+ q_{mi}$$

leaves v_m at time $t - \tau_{mi}$. By (I5) of Section 2, either (5) or (7) is identically zero.

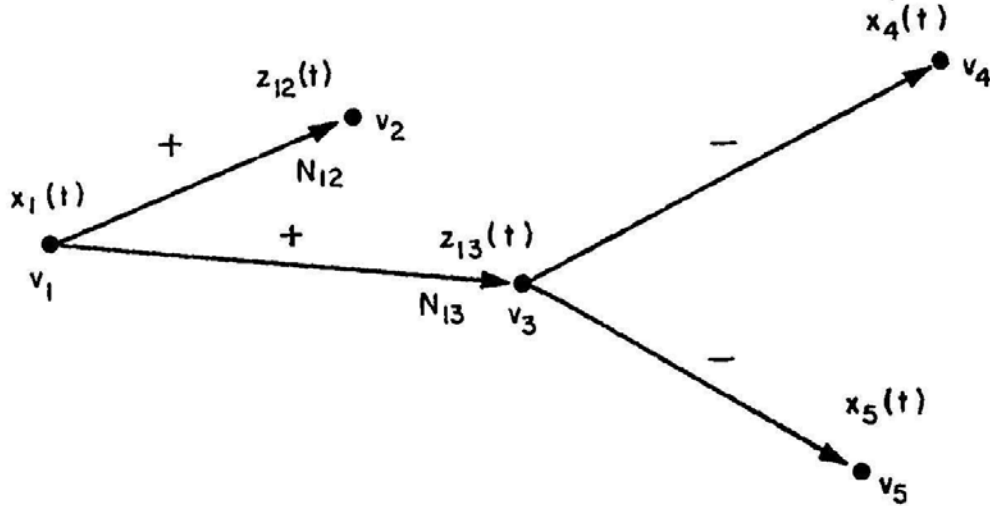


FIGURE 1.

$z_{mi}(t)$ sits in N_{mi} and *cross-correlates* the signal (5) received by N_{mi} at time t with the contiguous value $x_i(t)$ of v_i ; hence the condition (I3). $z_{mi}(t)$ also decays exponentially at the rate u_{ik} , which can be zero in special cases (see [5]).

$y_{mi}(t)$, rather than $z_{mi}(t)$, controls the excitatory signal that reaches v_i from v_m . y_{mi} is a ratio of cross-correlators, as in (2). This ratio can be interpreted as a kind of "competition between associations" or as "lateral inhibition coupled to cross-correlators", and is helpful in some form to make perfect learning possible ([1], [2], [3]). [13] describes systems where z_{mi} replaces y_{mi} .

Each of the dynamical variables in (*) has a mathematical, psychological, and neural name. Thus

$$\begin{aligned} x_i &= i^{\text{th}} \text{ vertex function, or} \\ &= i^{\text{th}} \text{ stimulus trace, or} \\ &= i^{\text{th}} \text{ average membrane potential,} \end{aligned}$$

and

$$\begin{aligned} y_{jk} &= (j, k)^{\text{th}} \text{ edge (or interaction) function, or} \\ &= \text{associational strength from } v_j \text{ to } v_k, \text{ or} \\ &= \text{average activity of transmitter control process in } N_{jk}. \end{aligned}$$

The network components also have qualitative anatomical labels. Thus

$$\begin{aligned} v_i &= i^{\text{th}} \text{ cell body cluster,} \\ e_{jk} &= \text{cluster of axons from } v_j \text{ to } v_k, \text{ and} \\ N_{jk} &= \text{cluster of synaptic knobs at terminal ends of } e_{jk} \text{ axons.} \end{aligned}$$

Each choice of the matrices $P = \|p_{ik}\|$ and $Q = \|q_{ik}\|$ defines a different "anatomy" for a network \mathfrak{N} by picking out the directed paths $v_i \rightarrow v_k$ over which signals can be transmitted and the relative strengths of these signals. Varying P and Q can dramatically change the qualitative properties of learning, memory, and recall in a network ([5]–[13]). The task of this paper is, given (1)–(3), mainly to choose an "anatomy" (P, Q) that can accomplish the paper's stated task. This will not be the only anatomy that can do the job, but it will certainly be a very simple one, and its deficiencies, where they exist, must be remedied by passing to more elaborate anatomies; see [12] for an example of a more realistic anatomy.

4. Outstars. An especially simple example of (*) is given by the equations

$$(8) \quad \dot{x}_1(t) = -\alpha x_1(t) + I_1(t),$$

$$(9) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta x_1(t - \tau) y_{1i}(t) + I_i(t),$$

(**)

$$(10) \quad y_{1i}(t) = z_{1i}(t) \left[\sum_{m=2}^n z_{1m}(t) \right]^{-1},$$

and

$$(11) \quad \dot{z}_{1i}(t) = -u z_{1i}(t) + \beta x_1(t - \tau) x_i(t),$$

where $i = 2, 3, \dots, n$. (**) is characterized by the following parameter choices in (*).

- a) $\alpha_i = \alpha,$
- b) $\beta_i = \beta,$
- c) $u_{jk} = u,$
- d) $v_{jk} = \beta,$
- e) $\Gamma_{jk} = q_{jk} = 0,$ and

$$(12) \quad \text{f) } p_{ij} = \begin{cases} 0, & \text{unless } i = 1 \\ 0, & \text{if } i = j = 1 \\ \frac{1}{n-1}, & \text{if } i = 1 \neq j. \end{cases}$$

The probabilistic graph characterized by (12) is drawn in Figure 2.

Hence the network $\mathfrak{N}^{(1)}$ which obeys (**) is called an *outstar* with *source vertex* v_1 , *sink vertices* $v_i, i \neq 1$, and *border* $B_n = \{v_i : i = 2, \dots, n\}$. We will construct the network of this paper using several copies of suitably modified outstars.

The dynamics of $\mathfrak{N}^{(1)}$ have been studied in [8] and are reviewed in [6]. These references describe the following theorem, which discusses an infinite sequence $G^{(1)}, G^{(2)}, \dots, G^{(N)}, \dots$ of outstars. Each outstar represents a "learning subject" who is matched in prior learning experience (*i.e.*, initial data) with all other subjects. $G^{(N)}$ differs from $G^{(N-1)}$ only by receiving more practice than

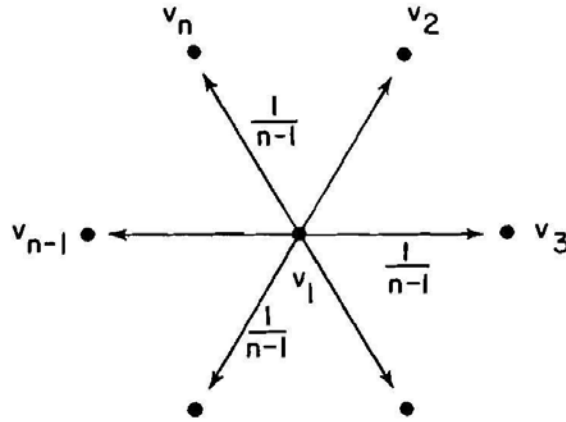


FIGURE 2.

$G^{(N-1)}$, and by perhaps being subjected to a different sequence of recall trials. The theorem means heuristically that in an outstar,

- 1) "practice makes perfect",
- 2) an isolated outstar suffers no memory loss,
- 3) an isolated outstar remembers without overtly practicing,
- 4) after moderate amounts of practice, the memory of an outstar spontaneously improves,
- 5) the act of making a correct recall can be repeated as many times as one wishes without destroying the outstar's memory, and
- 6) all errors can be corrected, although the retraining time might be lengthened due to response interference.

The theorem discusses the probabilities $y_{1i}(t)$ of (10) and the correspondingly defined probabilities

$$(13) \quad X_i(t) = \frac{x_i(t)}{\sum_{m=2}^n x_m(t)},$$

$i = 2, \dots, n$. We denote the y_{1i} , X_i , and I_i functions of $G^{(N)}$ by $y_{1i}^{(N)}$, $X_i^{(N)}$, and $I_i^{(N)}$ respectively.

Theorem 1. Let $G^{(1)}, \dots, G^{(N)}, \dots$, be any sequence of outstars with equal but otherwise arbitrary initial data. Let $I_i^{(N)}$ be any inputs of the form

$$(14) \quad I_1^{(N)}(t) = I_1(t)\chi(t - U_1(N)) + h_N(t),$$

and

$$(15) \quad I_i^{(N)}(t) = \theta_i I_i(t)\chi(t - U(N)),$$

$i = 2, \dots, n$, where

- a) $\{\theta_i : i = 2, \dots, n\}$ is a fixed but arbitrary probability distribution (i.e., $\theta_i \geq 0$ and $\sum_{i=2}^n \theta_i = 1$);

b) $U_1(N)$ and $U(N)$ are any nonnegative and strictly increasing functions of the integer $N \geq 1$;

c) the inputs $I_i(t)$ and $I(t)$ are constrained only by the existence of positive constants c and T_0 such that

$$(16) \quad \int_0^t e^{-\alpha(t-v)} I_1(v) dv \geq c$$

and

$$(17) \quad \int_0^t e^{-\alpha(t-v)} I(v) dv \geq c,$$

for $t \geq T_0$;

d) the input $h_N(t)$ is constrained only by being zero until $t > U(N)$; and

$$e) \quad x(t) = \begin{cases} 1, & \text{if } t < 0 \\ 0, & \text{if } t \geq 0. \end{cases}$$

Then

A) for every $N \geq 1$, the limits $Q_i^{(N)} = \lim_{t \rightarrow \infty} X_i^{(N)}(t)$ and $P_{1i}^{(N)} = \lim_{t \rightarrow \infty} y_{1i}^{(N)}(t)$ exist and are equal,

B) for every $N \geq 1$ and $t \geq U(N)$, $X_i^{(N)}(t)$ and $y_{1i}^{(N)}(t)$ are monotonic and are contained in the interval $[m_i^{(N)}, M_i^{(N)}]$, where

$$m_i^{(N)} = \min (X_i^{(N)}(U(N)), y_{1i}^{(N)}(U(N))),$$

$$M_i^{(N)} = \max (X_i^{(N)}(U(N)), y_{1i}^{(N)}(U(N))),$$

and

$$(18) \quad \lim_{N \rightarrow \infty} m_i^{(N)} = \lim_{N \rightarrow \infty} M_i^{(N)} = \theta_i.$$

In particular, by (A) and (B),

$$\lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} X_i^{(N)}(t) = \lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} y_{1i}^{(N)}(t) = \theta_i;$$

$i = 2, \dots, n$.

C) for every $N \geq 1$ and $i \neq 1$, the functions

$$\dot{y}_{1i}^{(N)}, f_i^{(N)} = y_{1i}^{(N)} - X_i^{(N)}, \quad \text{and} \quad g_i^{(N)} = X_i^{(N)} - \theta_i,$$

change sign at most once and not at all if $f_i^{(N)}(0)g_i^{(N)}(0) \geq 0$. Moreover $f_i^{(N)}(0)g_i^{(N)}(0) > 0$ implies that $f_i^{(N)}(t)g_i^{(N)}(t) > 0$ and that $y_{1i}^{(N)}(t)$ is monotonic for all $t \geq 0$.

Remark. In [S], each $G^{(N)}$ is denoted by $G^{(N,f)}$ and Theorem 1 is Theorem 5(f).

5. Learning a single spatial pattern of arbitrary complexity. Theorem 1 allows us to use an outstar to learn a spatial pattern of arbitrary complexity.

Suppose for example that we wish $\mathfrak{N}^{(1)}$ to learn the spatial pattern "A". First we must specify the level of spatial discrimination that is desired. If, for example, A is to be presented within a square region \mathcal{R} , then we can arrange the $n - 1$ border vertices B_n of $\mathfrak{N}^{(1)}$ in a rectangular grid spread over \mathcal{R} . The larger n is taken, the greater will be $\mathfrak{N}^{(1)}$'s ability to discriminate fine spatial details in the pattern. Since Theorem 1 holds for any value of $n = 2, 3, \dots$, any prescribed level of spatial discrimination can be guaranteed. Figure 3 depicts the case $n = 10$.

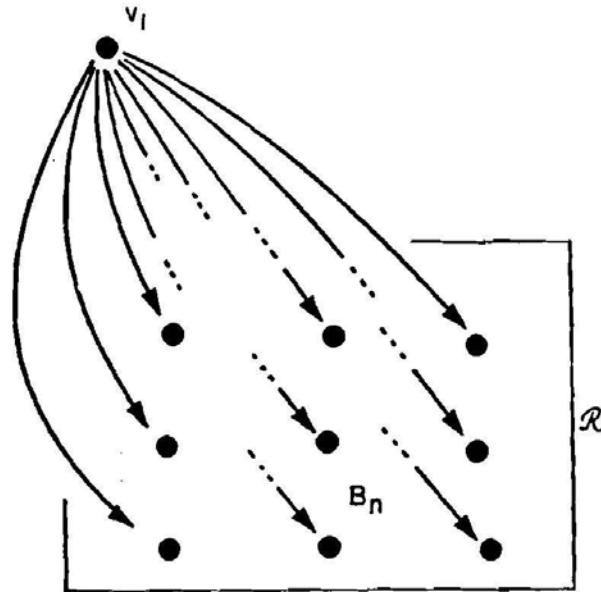


FIGURE 3.

The "A-ness" of the pattern "A" does not depend on the absolute blackness of its lines, but only on their relative blackness as compared to the surround. A *spatial pattern* (in "black", "white", and "shades of gray") is therefore defined as an input function

$$(19) \quad I_i(t) = \theta_i J(t), \quad i = 2, \dots, n,$$

delivered to B_n , where the nonnegative number θ_i specifies the relative blackness of the portion of the pattern that is seen by v_i . The θ_i 's can clearly be normalized to form a probability distribution without loss of generality, and then $J(t)$ specifies the total intensity of the pattern at time t . The pattern "A" is the same whether or not we view it in steady light or in flickering light, within substantial physiological limits. $J(t)$ can thus oscillate quite wildly without changing the pattern described by the θ_i 's. Henceforth we therefore speak of the spatial pattern $\theta^{(n)} = \{\theta_i : i = 2, \dots, n\}$.

The inputs $I_i^{(N)}(t)$, $i \neq 1$, in (15) are a spatial pattern $\theta^{(n)}$ with total intensity function

$$J(t) = I(t)\chi(t - U(N)).$$

In other words, presentation of the pattern ceases at time $t = U(N)$. Theorem 1 guarantees that if the spatial pattern $\theta^{(n)}$ is presented sufficiently often (*i.e.*, $N \rightarrow \infty$), and if the source vertex v_1 is also perturbed sufficiently often, then a later perturbation of v_1 *alone* will recreate the spatial pattern $\theta^{(n)}$ on the grid.

6. Respondant conditioning. The above learning paradigm can readily be interpreted as a form of respondant conditioning. Consider Figure 4.

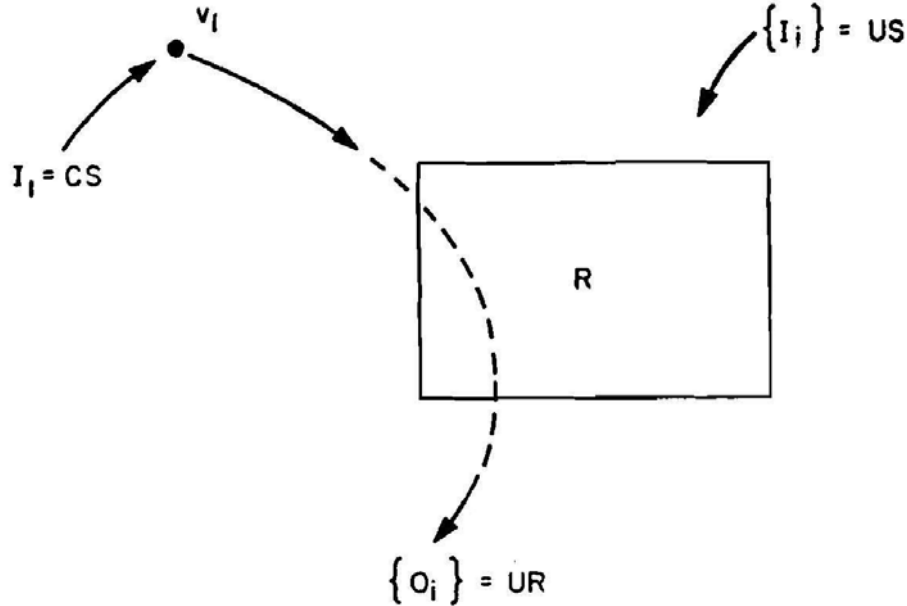


FIGURE 4.

Figure 4 emphasizes that inputs $I_i(t)$ reaching border vertices v_i give rise to outputs $O_i(t)$ that are proportional to $x_i(t)$, $i = 2, \dots, n$. Suppose moreover that all inputs $I_i(t)$ have equalled zero for a long time prior to $t = t_0$. Then by (8) and (9) we readily find that

$$(20) \quad x_i(t_0) \cong 0, \quad i = 1, 2, \dots, n$$

(See [8], p. 673). Now let the spatial pattern $\theta^{(n)}$ perturb B_n at times $t \geq t_0$. Then by (8) and (9),

$$x_1(t) \cong 0, \quad t \geq t_0,$$

and thus

$$(21) \quad \begin{aligned} \dot{x}_i(t) &\cong -\alpha x_i(t) + I_i(t) \\ &= -\alpha x_i(t) + \theta_i J(t), \end{aligned}$$

$i = 2, \dots, n$, where $J(t)$ becomes positive after $t = t_0$. By (20) and (21),

$$x_i(t) \cong \theta_i e^{-\alpha t} \int_{t_0}^t e^{\alpha v} J(v) dv,$$

$i = 2, \dots, n$, ([8], pp. 675–677), and thus

$$(22) \quad \frac{x_i(t)}{x_i(t)} \cong \frac{\theta_i}{\theta_i},$$

for $i, j = 2, \dots, n$, and $t > t_0$. In other words, the spatial pattern input to B_n creates the same spatial pattern output from B_n . We can therefore call the input pattern $I_i(t)$ an *unconditioned stimulus* (US) to B_n , and the output pattern $O_i(t)$ an *unconditioned response* (UR) from B_n [14]. Learning in an outstar has the effect of producing the unconditioned response from B_n , given an input to v_i alone, by previously pairing inputs to v_i with unconditioned stimuli to B_n . If we call the input to v_i a *conditioned stimulus* (CS), this learning procedure falls easily within the paradigm of respondent conditioning [14].

7. Complete graph with loops: pattern completion and the rigidity–plasticity continuum. [7] and [9] describe a different network that can also learn a spatial pattern of arbitrary complexity, namely a *complete graph with loops*, in which every $p_{ij} = 1/n$, all parameters are independent of their indices, and all thresholds and inhibitory signals equal zero. Learning in the complete graph with loops has some interesting properties that are not found in an outstar, such as the following ones.

a) *Pattern completion.* After the pattern is learned, even a single “speck of light” shined anywhere on the grid will suffice to reproduce the entire spatial pattern on the grid. The complete graph with loops is, in fact, the smallest embedding field that can do this.

b) *Path rigidity vs. plasticity.* Qualitative properties of memory in a complete graph with loops depend on whether or not the number $\sigma(\tau)$ is positive, where

$$(23) \quad \sigma(\tau) = u + 2s(\tau)$$

and $s(\tau)$ is the supremum of the real parts of the roots s of the equation

$$(24) \quad s + \alpha - \beta e^{-\tau s} = 0.$$

If $\sigma(\tau) > 0$, then

$$(25) \quad \lim_{t \rightarrow \infty} y_{ij}(t) = \lim_{t \rightarrow \infty} \tilde{X}_i(t), \quad \text{for all } i \text{ and } j,$$

where

$$\tilde{X}_i(t) = \frac{x_i(t)}{\sum_{m=1}^n x_m(t)}.$$

In other words, the associations converge to a spatial pattern in the absence of practice, and in fact to the spatial pattern which is “closest” to the graph’s values y_{ij} and \tilde{X}_i , when practice ceases. If $\sigma(\tau) < 0$, this is false. For example, if $\tau = 0$, then

$$(26) \quad |y_{ik}(t) - y_{ik}(0)| \leq 2 \log \left(1 + \frac{k_j}{|\sigma(0)|} \right),$$

where k_j depends on initial data. Thus by choosing $\sigma(\tau) < 0$ and $|\sigma(\tau)|$ sufficiently large, the associations $y_{ik}(t)$ can be made to remember anything arbitrarily well. In this sense, the network is "plastic" for $\sigma(\tau) > 0$ and "rigid" for $\sigma(\tau) < 0$.

Suppose also that $\alpha > \beta$. This condition means heuristically that large outputs from the graph occur only in response to inputs, for any time lag $\tau \geq 0$. In this case—which is the only case of physical interest—[9] shows that $\sigma(\tau)$ is a monotone increasing function of $\tau \geq 0$. Hence if a way existed to gradually decrease the time lag τ as a function of time, then the sign of $\sigma(\tau)$ could be made to shift from positive to negative values if $\sigma(0) < 0$. Since

$$(27) \quad \sigma(0) = u + 2(\beta - \alpha)$$

this can happen only if

$$0 \leq u < 2(\alpha - \beta).$$

Supposing that the length of the edges remains fixed through time, decreasing τ means increasing the signal velocity along the edges. Thus an increase in signal velocity can take the network from a "plastic" memory phase to a "rigid" memory phase capable of better preserving old learning. Since also the best learning speeds approximate τ [2], increasing signal velocity also lets the network learn best at faster presentation rates.

In vivo, a standard method of increasing axonal signal velocity is to *myelinate* the axon [15]. Hence an abstract "myelination" process, if coupled to the dynamics of learning in individual "cells" of a complete graph with loops, can help to speed up learning and to better preserve learned patterns in the cellular ensemble. A similar phenomenon occurs in complete graphs without loops (*i.e.*, $p_{ij} = 1/(n-1)$ if $i \neq j$, and $p_{ii} = 0$), at least in the case $\tau = 0$ and $n = 3$ [10]. Here if $\sigma(0) > 0$, the graph forgets *everything* it has learned in the absence of overt practice, whereas if $\sigma(0) < 0$ the graph can be made to remember arbitrarily well by choosing $|\sigma(0)|$ sufficiently large. Suppose, for example, that the excitation parameter β and the decay parameter α are equal, but are chosen with any positive value. Then by (27), any choice $u > 0$ of the correlational decay parameter, no matter how small, guarantees that $\sigma(0) > 0$ and hence that the graph will eventually forget everything that it has learned. Such a graph can remain an "unbiased" input filter for all time if inputs are presented at a sufficiently slow rate compared to the decay parameter $\sigma(0) > 0$, but can remember arbitrarily well if $\sigma(0) < 0$ and $|\sigma(0)|$ is taken arbitrarily large.

Neither pattern completion nor dependence on the sign of $\sigma(\tau)$ occur in an outstar, which exhibits no pattern completion because $p_{ij} = 0$ whenever v_i is in B_n , and whose memory has the same qualitative properties for all $\tau \geq 0$. Nonetheless, an outstar has a vital property which complete graphs lack. We will find that the outstar associations $y_{1i}(t)$ will not learn patterns in the border

B_n during times t for which the source function $x_1(t - \tau)$ is zero. Hence an outstar can be made to "sample" the patterns playing on B_n at prescribed times t ; namely, those times for which $x_1(t - \tau)$ is large. This cannot be done by complete graphs because perturbation of any vertex of a complete graph can indirectly perturb all other vertices via signals along the edges, as in pattern completion.

8. Approximating any continuous space-time pattern by a series of spatial patterns. A space-time pattern delivered to a grid of vertices v_i is created by an arbitrary array of input functions $I_i(t)$, $i = 2, \dots, n$. If the pattern weights

$$\theta_i(t) = \frac{I_i(t)}{\sum_{m=2}^n I_m(t)},$$

$i = 2, \dots, n$, vary "sufficiently" slowly, however, then we can approximate the space-time pattern by a series of spatial patterns in the following way.

Suppose for example that each function $\theta_i(t)$ varies very slowly in intervals of length ξ , where if some $\theta_i(t)$ is a rapidly varying function, then ξ must be chosen sufficiently small. We wish to construct a network that samples the space-time pattern briefly every ξ time units, and thereby learns this pattern as the series of spatial patterns

$$(28) \quad \theta^{(n)}(k\xi), \quad k = 1, 2, \dots,$$

where at every fixed time t , $\theta^{(n)}(t)$ is the spatial pattern

$$(29) \quad \{\theta_i(t) : i = 2, \dots, n\}.$$

If we can construct a network that samples briefly every ξ time units for *any* fixed positive value of ξ , then our approximation of (29) by (28) can be made arbitrarily good, by continuity of the inputs $I_i(t)$.

A plausible proposal for doing this is readily suggested. Suppose that (29) is presented over a time interval of length T , and that the sampling interval ξ is prescribed. Let

$$(30) \quad K(\xi, T) = \left[\frac{T}{\xi} \right] + 1,$$

where $[w]$ is the greatest integer less than w for every real number w . $K(\xi, T)$ is the number of spatial patterns by which the space-time pattern will be successively approximated. Let $K(\xi, T)$ outstars

$$\mathfrak{N}_1^{(1)}, \mathfrak{N}_2^{(1)}, \dots, \mathfrak{N}_{K(\xi, T)}^{(1)}$$

be given, and denote the vertex v_i and functions X_i , $y_{i,i}$, and I_i of $\mathfrak{N}_k^{(1)}$ by $v_{k,i}$, $X_{k,i}$, $y_{k,i}$, and $I_{k,i}$, respectively. Suppose that

$$v_{k,i} = v_{m,i}$$

for every $k, m = 1, 2, \dots, K(\xi, T)$, and $i = 2, \dots, n$; that is, each $\mathfrak{N}_k^{(1)}$ has the

same border. Let $v_{k,1}$ be perturbed by an input pulse of duration less than ξ at a time ξ units after $v_{k-1,1}$ is perturbed. See Figure 5.

Figure 5 depicts the series of outstars $\mathfrak{N}_k^{(1)}$ whose source vertices are excited successively every ξ time units by "axon collaterals" of the edge leading from v_0 . The space-time pattern is presented to the common border B_n of these outstars by an independent source of inputs and the border gives rise to outputs $O_i(t)$ as before. If the k^{th} outstar could learn the k^{th} spatial approximation to the space-time pattern, then a later input to v_0 alone would recreate the entire space-time pattern at B_n by successively activating each source $v_{k,1}$, $k = 1, 2, \dots, K(\xi, T)$. A single control vertex v_0 could hereby activate an arbitrarily complicated space-time pattern!

The equations describing the network of Figure 5 are

$$(31) \quad \dot{x}_0(t) = -\alpha_0 x_0(t) + I_0(t),$$

$$(32) \quad \dot{x}_{k,1}(t) = -\alpha_1 x_{k,1}(t) + \beta_0 x_0(t - k\xi),$$

$$(33) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{K(\xi, T)} x_{k,1}(t - \tau) y_{k,1,i}(t) + I_i(t),$$

(***)

$$(34) \quad y_{k,1,i}(t) = z_{k,1,i}(t) \left[\sum_{m=2}^n z_{k,1,m}(t) \right]^{-1},$$

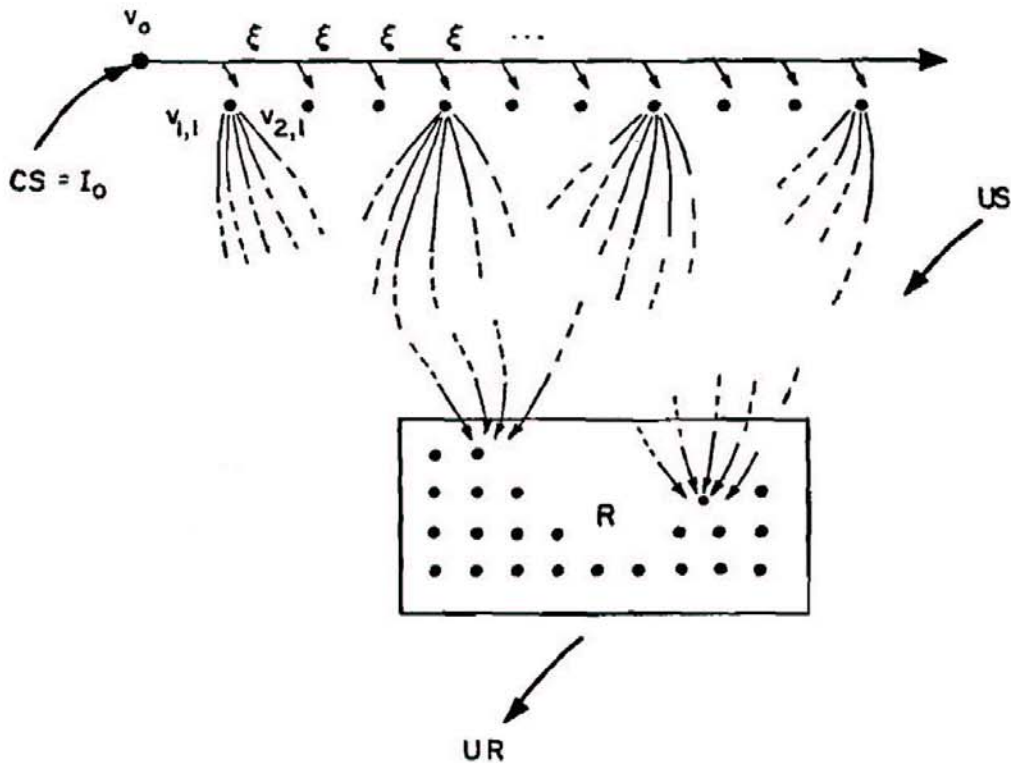


FIGURE 5.

and

$$(35) \quad \dot{z}_{k,1i}(t) = -uz_{k,1i}(t) + \gamma x_{k,1}(t - \tau)x_i(t),$$

where $x_i(t)$ is the common value of all $x_{k,i}(t)$, $k = 1, 2, \dots, K(\xi, T)$, and $i = 2, 3, \dots, n$. (***) is called an *outstar avalanche* by analogy with Ramón y Cajal's description of parallel fiber conduction in the cerebellum as "avalanche" conduction ([16], p. 196).

Alternatively, (***) can be replaced by the following system of equations:

$$(31) \quad \dot{x}_0(t) = -\alpha_0 x_0(t) + I_0(t),$$

$$(32) \quad \dot{x}_{k,1}(t) = -\alpha_1 x_{k,1}(t) + \beta_0 x_0(t - k\xi),$$

$$(33') \quad \dot{x}_{k,i}(t) = -\alpha x_{k,i}(t) + \beta x_{k,1}(t - \tau)y_{k,1i}(t) + I_i(t),$$

(***)

$$(34) \quad y_{k,1i}(t) = z_{k,1i}(t) \left[\sum_{m=2}^n z_{k,1m}(t) \right]^{-1},$$

and

$$(35') \quad \dot{z}_{k,1i}(t) = -uz_{k,1i}(t) + \gamma x_{k,1}(t - \tau)x_{k,i}(t).$$

In this system, each $\mathfrak{N}_k^{(1)}$ has a separate grid, and the same input $I_i(t)$ is delivered to the i^{th} border vertex $v_{k,i}$ of each $\mathfrak{N}_k^{(1)}$ via axon collaterals from the input source. The i^{th} output from this avalanche is given by

$$O_i(t) = \delta \sum_{k=1}^{K(\xi, T)} x_{k,i}(t).$$

The following qualitative argument shows that the source vertex $v_{k,1}$ in the k^{th} outstar $\mathfrak{N}_k^{(1)}$ can be activated primarily in a time interval $[k\xi, k\xi + \lambda]$ of length less than ξ , and therefore $\mathfrak{N}_k^{(1)}$ learns primarily the k^{th} spatial approximation to the space-time pattern reaching B_n at time $k\xi + \tau$.

First note that $\mathfrak{N}_k^{(1)}$ cannot learn from the border inputs $I_i(t)$ in intervals of time $[t_0, T_0]$ for which $x_1(t - \tau) \cong 0$, since then by (35),

$$\dot{z}_{k,1i}(t) \cong -uz_{k,1i}(t),$$

or

$$z_{k,1i}(t) \cong e^{-u(t-t_0)} z_{k,1i}(t_0),$$

and thus by (34)

$$y_{k,1i}(t) \cong y_{k,1i}(t_0),$$

whence no new learning has occurred. Supposing that x_0 and each $x_{k,1}$ is initially zero, (32) implies that

$$\begin{aligned} x_{k,1}(t) &= \beta_0 e^{-\alpha_1 t} \int_0^t e^{\alpha_1 v} x_0(v - k\xi) dv \\ &= \begin{cases} 0, & t \leq k\xi, \\ \beta_0 e^{-\alpha_1(t-k\xi)} \int_0^{t-k\xi} e^{\alpha_1 v} x_0(v) dv, & t \geq k\xi. \end{cases} \end{aligned}$$

Thus if $x_0(v)$ is large only in an interval of length $w < \xi$, then

$$(36) \quad x_{k,1}(t) \cong \beta_0 e^{-\alpha_1(t-k\xi)} \Lambda$$

for $t \geq w + k\xi$, where

$$\Lambda = \int_0^w e^{\alpha_1 v} x_0(v) dv.$$

If moreover the decay rate α_1 is large, (36) guarantees that

$$x_{k,1}(t) \cong 0$$

for $t \geq k\xi + \lambda$, where $w < \lambda < \xi$. If this can be achieved, then $\mathfrak{N}_k^{(1)}$ will learn only the inputs reaching B_n within the interval $[k\xi + \tau, k\xi + \tau + \lambda]$; i.e., only the k^{th} spatial approximation to the space-time pattern.

It remains only to check that $x_0(v)$ can be made large in an interval of length $w < \xi$. By (31),

$$(37) \quad x_0(t) = e^{-\alpha_0 t} \int_0^t e^{\alpha_0 v} I_0(v) dv,$$

so it suffices to make the duration of $I_0(v)$ less than λ and to choose the decay rate α_0 large.

This argument shows that a sequence of outstars can, at least crudely, learn a space-time pattern as a sequence of spatial patterns.

9. A better approximation: Signal thresholds eliminate source noise. The accuracy of learning by an outstar avalanche is limited by the fact that each $x_{k,1}(t) > 0$ for all $t \geq t_0$ if $x_{k,1}(t_0) > 0$, $k = 1, 2, \dots, K(\xi, T)$. Thus an outstar $\mathfrak{N}_k^{(1)}$ can never wholly stop learning from the $(k+m)^{\text{th}}$ spatial approximations to the space-time pattern, $m = 1, 2, \dots, K(\xi, T) - k$. The influence of the $(k+m)^{\text{th}}$ approximation will decrease as the source $x_{k,1}(t)$ exponentially decays to ever smaller values, but some background interference can never be eliminated.

This difficulty can be readily overcome. We need merely guarantee that very small source values $x_{k,1}(t)$ never create signals to border vertices. Thus we replace (***) by the following system of equations.

$$(38) \quad \dot{x}_0(t) = -\alpha_0 x_0(t) + I_0(t),$$

$$(39) \quad \dot{x}_{k,1}(t) = -\alpha_1 x_{k,1}(t) + \beta_0 [x_0(t - k\xi) - \Gamma_0]^+,$$

$$(40) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta \sum_{k=1}^{K(\xi, T)} [x_{k,1}(t - \tau) - \Gamma_1]^+ y_{k,1,i}(t) + I_i(t).$$

$$(41) \quad y_{k,1,i}(t) = z_{k,1,i}(t) \left[\sum_{m=2}^n z_{k,1,m}(t) \right]^{-1},$$

(†)

and

$$(42) \quad \dot{z}_{k,1,i}(t) = -uz_{k,1,i}(t) + \gamma[x_{k,1}(t - \tau) - \Gamma_1]^+ x_i(t),$$

where the signal thresholds Γ_0 and Γ_1 are *positive*, $k = 1, 2, \dots, K(\xi, T)$, and $i = 2, \dots, n$. The system (‡) eliminates transmission of background noise from v_0 to each $v_{k,1}$ and from $v_{k,1}$ to each v_i , $i \neq 1$, by cutting off signals at the *positive* signal thresholds Γ_0 and Γ_1 , respectively.

Alternatively the following system with $K(\xi, T)$ separate grids receiving axon collaterals from a common source can be used.

$$(38) \quad \dot{x}_0(t) = -\alpha_0 x_0(t) + I_0(t),$$

$$(39) \quad \dot{x}_{k,1}(t) = -\alpha_1 x_{k,1}(t) + \beta_0 [x_0(t - k\xi) - \Gamma_0]^+,$$

$$(40) \quad \dot{x}_{k,i}(t) = -\alpha x_{k,i}(t) + \beta [x_{k,1}(t - \tau) - \Gamma_1]^+ y_{k,1,i}(t) + I_i(t),$$

(‡)

$$(41) \quad y_{k,1,i}(t) = z_{k,1,i}(t) \left[\sum_{m=2}^n z_{k,1,m}(t) \right]^{-1},$$

and

$$(42') \quad \dot{z}_{k,1,i}(t) = -uz_{k,1,i}(t) + \gamma [x_{k,1}(t - \tau) - \Gamma_1]^+ x_{k,i}(t),$$

where the i^{th} output from this avalanche is

$$O_i(t) = \delta \sum_{k=1}^{K(\xi, T)} [x_{k,i}(t) - \Gamma]^+,$$

with Γ a small, but possibly positive, output threshold.

By (38) and (39), it is clear that an input of finite duration at v_0 will create a signal $\beta [x_{k,1}(t - \tau) - \Gamma_1]^+$ of finite duration in $e_{k,1,i}$. (‡) and (‡) are constructed from $K(\xi, T)$ outstars with positive threshold Γ_1 (*i.e.*, Γ -outstars), where below we write Γ instead of Γ_1 for simplicity.

$$(43) \quad \dot{x}_1(t) = -\alpha_1 x_1(t) + I_1(t),$$

$$(44) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta [x_1(t - \tau) - \Gamma]^+ y_{1,i}(t) + I_i(t),$$

(‡ ‡)

$$(45) \quad y_{1,i}(t) = z_{1,i}(t) \left[\sum_{m=2}^n z_{1,m}(t) \right]^{-1},$$

$$(46) \quad \dot{z}_{1,i}(t) = -uz_{1,i}(t) + \gamma [x_1(t - \tau) - \Gamma]^+ x_i(t),$$

$i = 2, \dots, n$. We must check that Theorem 1 holds in some form for (‡‡). In fact the following theorem holds, in which we assume to avoid trivialities that the sums $x^{(1)}(t) = \sum_{m=2}^n x_m(t)$ and $z^{(1)}(t) = \sum_{m=2}^n z_{1,m}(t)$ are positive at $t = 0$.

Theorem 2. Let $G^{(1)}, G^{(2)}, \dots, G^{(N)}, \dots$ be any sequence of Γ -outstars with

equal but otherwise arbitrary initial data. Let $I_i^{(N)}$ be any inputs of the form

$$(47) \quad I_i^{(N)}(t) = I_1(t)\chi(t - U_1(N)) + h_N(t),$$

and

$$(48) \quad I_i^{(N)}(t) = \theta_i I(t)\chi(t - U(N)),$$

$i = 2, \dots, n$, where

- a) $\{\theta_i : i = 2, \dots, n\}$ is a fixed but arbitrary probability distribution;
- b) $U_1(N)$ and $U(N)$ are any nonnegative and strictly increasing functions of $N \geq 1$;
- c) the inputs $I_1(t)$ and $I(t)$ are constrained only by the existence of positive constants c and T_0 such that

$$(49) \quad \int_0^t e^{-\alpha(t-\xi)} \left[\int_0^\xi e^{-\alpha(\xi-v)} I_1(v) dv - \Gamma \right]^+ dv \geq c,$$

and

$$(50) \quad \int_0^t e^{-\alpha(t-v)} I(v) dv \geq c,$$

for $t \geq T_0$;

- d) the input $h_N(t)$ is constrained only by being zero until $t > U(N)$.

Then

- A) for every $N \geq 1$, the limits $Q_i^{(N)}$ and $P_{1_i}^{(N)}$ exist,
- B) for every $N \geq 1$ and $t \geq U(N)$, $X_i^{(N)}(t)$ and $y_{1_i}^{(N)}(t)$ are monotonic and are contained in the interval $[m_i^{(N)}, M_i^{(N)}]$, where

$$(51) \quad \lim_{N \rightarrow \infty} m_i^{(N)} = \lim_{N \rightarrow \infty} M_i^{(N)} = \theta_i.$$

In particular, by (A) and (B),

$$\lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} X_i^{(N)}(t) = \lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} y_{1_i}^{(N)}(t) = \theta_i;$$

- C) for every $N \geq 1$ and $i \neq 1$, the functions $\dot{y}_{1_i}^{(N)}$, $f_i^{(N)}$, and $g_i^{(N)}$ change sign at most once and not at all if $f_i^{(N)}(0)g_i^{(N)}(0) \geq 0$. Moreover, $f_i^{(N)}(0)g_i^{(N)}(0) > 0$ implies that $f_i^{(N)}(t)g_i^{(N)}(t) > 0$ and that $y_{1_i}^{(N)}(t)$ is monotonic for all $t \geq 0$.

Remark. If $\Gamma = 0$, then condition (49) is weaker than condition (16). If $\Gamma > 0$, then the limits $Q_i^{(N)}$ and $P_{1_i}^{(N)}$ are not generally equal.

Proof. The proof of Theorem 2 uses methods developed in [8] and [9]. The proof is divided into four sections. Section (I) transforms the nonlinear Γ -outstar into a linear system. Section (II) studies memory and recall in a Γ -outstar. Section (III) studies learning trials in a Γ -outstar that are not cut-off after a finite practice interval. Section (IV) cuts off learning after a finite practice interval and pastes the results of (II) onto the end of this interval.

(I) A Γ -outstar becomes a linear system in terms of the probability distributions $y = \{y_{1i} : i = 2, \dots, n\}$ and $X = \{X_i : i = 2, \dots, n\}$.

Lemma 1. *The source function x_1 and the sums $x^{(1)}$ and $z^{(1)}$ depend on time only through the known inputs I_1 and I .*

Proof. The assertion is obvious for x_1 by (43). Summing (44) and (45) over $i \neq 1$ yields

$$(52) \quad \dot{x}^{(1)} = -\alpha x^{(1)} + \beta[x_1(t - \tau) - \Gamma]^+ + I$$

and

$$(53) \quad \dot{z}^{(1)} = -uz^{(1)} + \gamma[x_1(t - \tau) - \Gamma]^+ x^{(1)}.$$

Integrating (43), (52), and (53) completes the proof.

Lemma 2. *The probability distributions y and X satisfy the following linear system of equations.*

$$(54) \quad \dot{X}_i = A_i(y_{1i} - X_i) + B_i(\theta_i - X_i)$$

and

$$(55) \quad \dot{y}_{1i} = C_i(X_i - y_{1i}),$$

where

$$A_i(t) = \frac{\beta[x_1(t - \tau) - \Gamma]^+}{x^{(1)}(t)},$$

$$B_i(t) = \frac{I(t)}{x^{(1)}(t)},$$

and

$$C_i(t) = \gamma[x_1(t - \tau) - \Gamma]^+ \frac{x^{(1)}(t)}{z^{(1)}(t)},$$

and are hence known functions of time by Lemma 1.

Proof. Since $X_i = x_i/x^{(1)}$,

$$\dot{X}_i = \frac{1}{x^{(1)}} \left(\dot{x}_i - x_i \frac{\dot{x}^{(1)}}{x^{(1)}} \right),$$

which yields (54) after substituting (44) and (52), cancelling and rearranging terms. Since $y_{1i} = z_{1i}/z^{(1)}$,

$$\dot{y}_{1i} = \frac{1}{z^{(1)}} \left(\dot{z}_{1i} - z_{1i} \frac{\dot{z}^{(1)}}{z^{(1)}} \right),$$

which yields (55) after substituting (45) and (53), cancelling and rearranging terms.

(II) Memory and recall experiments in $[t_0, \infty)$ are characterized by choosing $I_i(t) = 0$ for $t \geq t_0$ and $i = 2, \dots, n$.

Lemma 3. Let $(\ddagger\ddagger)$ be given with arbitrary initial data in $[t_0 - \tau, t_0]$, border inputs $I_i(t) = 0$ for $t \geq t_0$ and $i = 2, \dots, n$, and any input $I_1(t)$. If $x_1(t - \tau) \leq \Gamma$ for t in an interval $[t_1, t_2]$ such that $t_0 < t_1 < t_2$, then $y_{1i}(t)$ and $X_i(t)$ are constant in $[t_1, t_2]$. If $x_1(t - \tau) > \Gamma$ for t in $[t_1, t_2]$, then $y_{1i}(t)$ and $X_i(t)$ are monotonic in opposite senses with $|y_{1i}(t) - X_i(t)|$ monotone decreasing. Thus the limits

$$Q_i = \lim_{t \rightarrow \infty} X_i(t)$$

and

$$P_{1i} = \lim_{t \rightarrow \infty} y_{1i}(t)$$

exist, and $X_i(t)$ and $y_{1i}(t)$ lie in the interval $[m_i, M_i]$ for $t \geq t_0$, where $m_i = \min(X_i(t_0), y_{1i}(t_0))$ and $M_i = \max(X_i(t_0), y_{1i}(t_0))$.

Proof. By hypothesis, $I(t) = 0$ for $t \geq t_0$. Hence (54) becomes

$$(56) \quad \dot{X}_i = A_1(y_{1i} - X_i).$$

The proof follows by inspection of (55) and (56) using the facts that A_1 and C_1 are nonnegative and continuous, and are positive or zero according as $x_1(t - \tau) > \Gamma$ or $x_1(t - \tau) \leq \Gamma$. (See [8], Theorem 1).

(III) Another change of variables is needed to study learning in $(\ddagger\ddagger)$, namely from y_{1i} and X_i to $f_i = y_{1i} - X_i$ and $g_i = X_i - \theta_i$. Throughout this section, (49) and (50) will be assumed to hold.

Lemma 4. f_i and g_i satisfy the equations

$$(57) \quad \dot{f}_i = -D_1 f_i + B_1 g_i$$

and

$$(58) \quad \dot{g}_i = -B_1 g_i + A_1 f_i$$

where $D_1 = A_1 + C_1$.

Proof. Rewrite (54) and (55) in terms of f_i and g_i .

The oscillations of f_i and g_i can be studied using the following lemma, in which $f = f_i$, $g = g_i$, $a = -D_1$, $b = B_1$, $c = A_1$, and $d = -B_1$.

Lemma 5. Let the functions f and g satisfy the differential equations

$$\dot{f} = af + bg$$

$$\dot{g} = cf + dg,$$

where a, b, c , and d are continuous functions and the off-diagonal coefficients b and c are nonnegative. Then f and g change sign at most once and not at all if $f(0)g(0) \geq 0$. Moreover $f(0)g(0) > 0$ implies $f(t)g(t) > 0$ for all $t \geq 0$.

Lemma 5 is proved in [8], p. 663–664. By Lemma 5, two cases arise for large values of t . Either

A) $f_i(t)g_i(t) < 0$ for all large t ,

or

B) $f_i(t)g_i(t) \geq 0$ for all large t .

If (A) holds, then by (57) and (58), f_i and g_i are monotonic in opposite senses for all large t . Thus the limits $\lim_{t \rightarrow \infty} f_i(t)$ and $\lim_{t \rightarrow \infty} g_i(t)$ exist, and hence the limits $Q_i = \lim_{t \rightarrow \infty} X_i(t)$ and $P_{1i} = \lim_{t \rightarrow \infty} y_{1i}(t)$ exist.

The existence of P_{1i} in both cases (A) and (B) follows by Lemma 5. (55) implies

$$\dot{y}_{1i} = C_1 f_i.$$

Since $C_1 \geq 0$, \dot{y}_{1i} changes sign at most once and not at all if $f_i(0)g_i(0) \geq 0$. In particular, $y_{1i}(t)$ is monotonic for large t , whence P_{1i} exists, $i = 2, \dots, n$.

Proving the existence of $Q_i = \lim_{t \rightarrow \infty} X_i(t)$ in Case (B) and the equation $P_{1i} = Q_i = \theta_i$ in both cases requires the following estimates on $x^{(1)}$ and $z^{(1)}$.

Lemma 6. *If the inputs I_1 and I satisfy (49) and (50), then $x^{(1)}(t)$ and $z^{(1)}(t)$ are bounded from above and below by positive constants.*

Proof. Integrating (52) yields

$$(59) \quad x^{(1)}(t) = e^{-\alpha t} \left[x^{(1)}(0) + \int_0^t e^{\alpha v} (\beta [x_1(v - \tau) - \Gamma]^+ + I(v)) dv \right],$$

and thus

$$x^{(1)}(t) \geq \int_0^t e^{-\alpha(t-v)} I(v) dv,$$

which by (50) yields

$$x^{(1)}(t) \geq c \quad \text{for } t \geq T_0.$$

Since also $x^{(1)}(t)$ is positive and continuous in $[0, T_0]$, $x^{(1)}(t)$ has a positive lower bound in $[0, \infty)$.

Integrating (53) yields

$$z^{(1)}(t) = e^{-ut} \left[z^{(1)}(0) + \gamma \int_0^t e^{uv} [x_1(v - \tau) - \Gamma]^+ x^{(1)}(v) dv \right],$$

which by (59) implies

$$\begin{aligned} z^{(1)}(t) &\geq \gamma c \int_{T_0}^t e^{-u(t-v)} [x_1(v - \tau) - \Gamma]^+ dv \\ &= \gamma c e^{u\tau} \int_{T_0 - \tau}^{t - \tau} e^{-u(t-v)} [x_1(v) - \Gamma]^+ dv \end{aligned}$$

for $t \geq T_0 + \tau$. For any $w \geq 0$ and $s \geq w + T_0$, integrating (43) yields

$$\int_w^s e^{-u(t-v)} [x_1(v) - \Gamma]^+ dv \geq \int_w^s e^{-u(t-v)} \left[\int_0^v e^{-\alpha_1(v-\xi)} I_1(\xi) d\xi - \Gamma \right]^+ dv \geq c - We^{-uw}$$

where

$$W = \int_0^w e^{uw} \left[\int_0^v e^{-\alpha_1(v-\xi)} I_1(\xi) d\xi - \Gamma \right]^+ dv.$$

Thus letting $w = T_0 - \tau$ and $s = t - \tau$,

$$z^{(1)}(t) \geq \frac{\gamma c^2 e^{u\tau}}{2} \quad \text{for } t \geq \tau + \frac{1}{u} \log \frac{2W}{c}.$$

Hence $z^{(1)}(t)$ has a positive lower bound in $[0, \infty)$.

Upper bounds for $x^{(1)}$ and $z^{(1)}$ readily follow from the boundedness of I_1 and I .

These estimates can be used to show that $\tilde{y}_{1i}(t)$ is bounded. (At points where a two-sided derivative does not exist because of the threshold cut-off of $x_1(t - \tau)$ by Γ_1 in (55), a one-sided derivative is intended.) By (54) and (55),

$$\begin{aligned} |\tilde{y}_{1i}| &= |\dot{C}_1(X_i - y_{1i}) + C_1(\dot{X}_i - \dot{y}_{1i})| \\ &\leq 2 |\dot{C}_1| + C_1(|\dot{X}_i| + |\dot{y}_{1i}|) \\ &\leq 2[|\dot{C}_1| + C_1(A_1 + B_1 + C_1)]. \end{aligned}$$

Lemma 6 and the boundedness of I imply the boundedness of A_1, B_1, C_1 , and \dot{C}_1 , and hence the boundedness of \tilde{y}_{1i} .

The boundedness of \tilde{y}_{1i} and the existence of P_{1i} imply that

$$(60) \quad \lim_{t \rightarrow \infty} \dot{y}_{1i}(t) = 0$$

([8], Lemma 4). Hence by (55),

$$(61) \quad \lim_{t \rightarrow \infty} C_1(t)(X_i(t) - y_{1i}(t)) = 0.$$

In case (A), (61) implies that

$$(Q_i - P_{1i}) \lim_{t \rightarrow \infty} C_1(t) = 0.$$

Thus either $Q_i = P_{1i}$ or $\lim_{t \rightarrow \infty} C_1(t) = 0$. If $\lim_{t \rightarrow \infty} C_1(t) = 0$, then

$$(62) \quad \lim_{t \rightarrow \infty} [x_1(t) - \Gamma]^+ = 0$$

by Lemma 6. But (62) implies

$$\lim_{t \rightarrow \infty} \int_0^t e^{-u(t-v)} [x_1(v) - \Gamma]^+ dv = 0,$$

which contradicts (49). Hence $Q_i = P_{1i}$ in case (A).

We now prove that $Q_i = P_{1i}$ implies $Q_i = \theta_i$. Suppose not, and in particular let $Q_i > \theta_i$. Then there exists a T such that $t \geq T$ implies

$$\theta_i - X_i(t) \leq -\frac{1}{2}(Q_i - \theta_i).$$

We will also show that there exists a constant K_1 and a positive constant K_2 such that

$$(63) \quad \int_0^t B_1(v) dv \geq K_1 + K_2 t$$

for all $t \geq 0$. Thus for any $S \geq T$ and $t \geq S$,

$$\begin{aligned} \int_s^t B_1(v)(\theta_i - X_i(v)) dv &\leq -\frac{1}{2}(Q_i - \theta_i) \int_s^t B_1(v) dv \\ &\leq \frac{1}{2}(Q_i - \theta_i) \int_0^S B_1(v) dv - \frac{1}{2}(Q_i - \theta_i)(K_1 + K_2 t). \end{aligned}$$

Now define $h_i \equiv A_1(y_{1i} - X_i)$ and note that since A_1 is bounded and $P_{1i} = Q_i$, there exists an S such that

$$h_i(t) \leq \frac{1}{4}(Q_i - \theta_i)K_2$$

for $t \geq S$. For such an S , integrating (54) from S to any $t \geq S$ yields

$$(64) \quad X_i(t) \leq G_i(S) - \frac{1}{4}(Q_i - \theta_i)K_2 t$$

where

$$\begin{aligned} G_i(S) &= X_i(S) + \frac{1}{2}(Q_i - \theta_i) \int_0^S B_1(v) dv \\ &\quad - \frac{1}{2}K_1(Q_i - \theta_i) - \frac{1}{4}(Q_i - \theta_i)S, \end{aligned}$$

and thus by (64),

$$-\infty = \lim_{t \rightarrow \infty} X_i(t) \geq 0.$$

This contradiction shows $Q_i \leq \theta_i$. The contradiction

$$\infty = \lim_{t \rightarrow \infty} X_i(t) \leq 1$$

similarly follows if $Q_i < \theta_i$.

It remains only to prove (63). Denote the finite upper bound of $x^{(1)}(t)$ by M . Then by (50), for any $\epsilon > 0$ and $t \geq T_\epsilon$,

$$x^{(1)}(t) \leq e^{-\alpha t} \left[\epsilon + \frac{M}{c} \int_0^t e^{\alpha v} I(v) dv \right],$$

or

$$B_1(t) \geq \frac{c}{M} \frac{d}{dt} \log \left[\epsilon + \frac{M}{c} \int_0^t e^{\alpha v} I(v) dv \right],$$

which yields for $t \geq T_0$,

$$\int_0^t B_1(v) dv \geq L \log \left[\epsilon + \frac{M}{c} \int_0^t e^{\alpha v} I(v) dv \right],$$

where

$$L^{-1} = \frac{M}{c} \log \left[\epsilon + \frac{M}{c} \int_0^{T_0} e^{\alpha v} I(v) dv \right].$$

By (50),

$$\begin{aligned} \int_0^t B_1(v) dv &\geq L \log (\epsilon + M e^{\alpha t}) \\ &\geq L \log M + L \alpha t, \end{aligned}$$

which completes the proof by setting $K_1 = L \log M$ and, $K_2 = L \alpha$, and noting that L , and thus K_2 , can be made positive by choosing ϵ sufficiently large.

It remains only to consider Case (B). Suppose for specificity that $f_i(t) \geq 0$ and $g_i(t) \geq 0$ for $t \geq T_1$. (The case $f_i(t) \leq 0$ and $g_i(t) \leq 0$ can be similarly treated.) Then $y_{1i}(t) \geq X_i(t) \geq \theta_i$ and $y_{1i}(t)$ is monotone decreasing for $t \geq T_1$. The equalities $\theta_i = Q_i = P_{1i}$, therefore hold if $P_{1i} = \theta_i$. It remains only to consider the case $P_{1i} > \theta_i$.

Let $X_i^{(\theta)} = X_i - \theta_i$ and $y_i^{(\theta)} = y_{1i} - \theta_i$. Then (54) and (55) become

$$(65) \quad \dot{X}_i^{(\theta)} = A_1(y_i^{(\theta)} - X_i^{(\theta)}) - B_1 X_i^{(\theta)}$$

and

$$(66) \quad \dot{y}_i^{(\theta)} = C_1(X_i^{(\theta)} - y_i^{(\theta)}),$$

where $y_i^{(\theta)}(t) \geq 0$ for $t \geq T_1$.

Lemma 7. Suppose $y_i^{(\theta)}(t) \geq X_i^{(\theta)}(t) \geq 0$ for $t \geq T_1$, where we can choose $T_1 \geq T_0$ without loss of generality. Then there exists a $\mu \in (0, 1)$ and a $T_2 = T_2(\mu)$ such that

$$(67) \quad X_i^{(\theta)}(t) \leq (1 - \mu) y_i^{(\theta)}(t - T_2).$$

Proof. Integrating (65) in $[T, t]$ yields

$$(68) \quad X_i^{(\theta)}(t) = U_i^{(\theta)}(t, T) + V_i^{(\theta)}(t, T),$$

where

$$(69) \quad U_i^{(\theta)}(t, T) = X_i^{(\theta)}(T) Z^{-1}(t, T),$$

$$(70) \quad V_i^{(\theta)}(t, T) = Z^{-1}(t, T) \int_T^t y_i^{(\theta)} A_1 Z(v, T) dv,$$

and

$$(71) \quad Z(t, T) = \exp \left[\int_T^t (A_1 + B_1) dw \right].$$

Since $X_i^{(\theta)}(T) \leq y_i^{(\theta)}(T)$ for $T \geq T_1$, (69) implies

$$(72) \quad U_i^{(\theta)}(t, T) \leq y_i^{(\theta)}(T)Z^{-1}(t, T).$$

(71) is evaluated as follows.

$$\begin{aligned} A_1 + B_1 &= \frac{\beta[x_1(t - \tau) - \Gamma]^+ + I}{x^{(1)}} \\ &= \frac{d}{dt} \log x^{(1)} + \alpha, \end{aligned}$$

and thus

$$Z(t, T) = \frac{x^{(1)}(t)e^{\alpha t}}{x^{(1)}(T)e^{\alpha T}}.$$

(70) now becomes

$$V_i^{(\theta)}(t, T) = \frac{1}{x^{(1)}(t)e^{\alpha t}} \int_{\tau}^t y_i^{(\theta)} A_1 x^{(1)} e^{\alpha v} dv,$$

and since $y_i^{(\theta)}(v) \leq y_i^{(\theta)}(T)$ for $v \geq T$,

$$(73) \quad V^{(\theta)}(t, T) \leq y_i^{(\theta)}(T)R(t, T),$$

where

$$R(t, T) = \frac{1}{x^{(1)}(t)e^{\alpha t}} \int_{\tau}^t A_1 x^{(1)} e^{\alpha v} dv.$$

Since $A_1 = \beta[x_1(t - \tau) - \Gamma]^+ / x^{(1)}$,

$$R(t, T) = \frac{1}{x^{(1)}(t)e^{\alpha t}} \int_{\tau}^t \beta[x_1(v - \tau) - \Gamma]^+ e^{\alpha v} dv,$$

and since

$$\begin{aligned} \beta[x_1(v - \tau) - \Gamma]^+ e^{\alpha v} &= e^{\alpha v}(\dot{x}^{(1)} + \alpha x^{(1)} - I) \\ &= \frac{d}{dv} (x^{(1)} e^{\alpha v}) - I e^{\alpha v}, \end{aligned}$$

$$(74) \quad R(t, T) = 1 - Z^{-1}(t, T) - \frac{1}{x^{(1)}(t)e^{\alpha t}} \int_{\tau}^t I e^{\alpha v} dv.$$

Combining (68), (72), (73), and (74) yields

$$(75) \quad X_i^{(\theta)}(t) \leq y_i^{(\theta)}(T)P(t, T),$$

where

$$P(t, T) = 1 - \frac{1}{x^{(1)}(t)e^{\alpha t}} \int_{\tau}^t I e^{\alpha v} dv \quad (> 0).$$

By Lemma 6, there exists a positive constant M such that

$$P(t, T) \leq 1 - M \int_T^t e^{-\alpha(t-v)} I(v) dv.$$

Since

$$c \leq \left(\int_0^T + \int_T^t \right) e^{-\alpha(t-v)} I(v) dv \leq \frac{I}{\alpha}$$

for $t \geq T \geq T_0$, where $I = \sup \{I(t) : t \geq 0\} (< \infty)$,

$$\begin{aligned} \int_T^t e^{-\alpha(t-v)} I(v) dv &\geq c - \int_0^T e^{-\alpha(t-v)} I(v) dv \\ &\geq c - \frac{I}{\alpha} e^{-\alpha(t-T)}. \end{aligned}$$

Thus there exists a T_2 such that

$$\int_T^t e^{-\alpha(t-v)} I(v) dv \geq \frac{c}{2}$$

for $t \geq T + T_2$, and

$$(76) \quad P(t, T) \leq 1 - \mu, \quad \mu = \frac{1}{2}cM,$$

which along with (75) completes the proof.

Lemma 7 will now be used to draw a contradiction if $P_{1i} > \theta_i$ in Case (B). Since y_{1i} decreases monotonically to P_{1i} , there exists a $T_3 \geq T_1$ such that

$$y_{1i}(t - T_2) - y_{1i}(t) \leq \frac{\mu}{2} (P_{1i} - \theta_i)$$

for $t \geq T_4 + T_5$. Thus for $t \geq T_2 + T_3$,

$$\begin{aligned} y_{1i}(t) - X_i(t) &= (y_{1i}(t) - y_{1i}(t - T_2)) + (y_{1i}^{(\theta)}(t - T_2) - X_i^{(\theta)}(t)) \\ &\geq -\frac{\mu}{2} (P_{1i} - \theta_i) + y_{1i}^{(\theta)}(t - T_2) - X_i^{(\theta)}(t). \end{aligned}$$

which by (67) yields

$$\begin{aligned} y_{1i}(t) - X_i(t) &\geq -\frac{\mu}{2} (P_{1i} - \theta_i) + \mu(y_{1i}(t - T_2) - \theta_i) \\ &\geq \frac{\mu}{2} (P_{1i} - \theta_i) \quad (> 0). \end{aligned}$$

Thus by (61), $\lim_{t \rightarrow \infty} C_1(t) = 0$, which contradicts (49). Hence $P_{1i} = Q_i = \theta_i$ in all cases.

(IV) To complete the proof, we must study the effects of cutting off the learning interval at a finite time that is followed by recall experiments (inputs to v_1

alone) and/or memory intervals free from inputs. Thus we consider Γ -outstars with inputs given by (47) and (48).

By (II), $y_{1i}^{(N)}(t)$ and $X_i^{(N)}(t)$ are contained in $[m_i^{(N)}, M_i^{(N)}]$ for $t \geq U(N)$. Two cases now arise.

Case 1. $U(N) \leq U_1(N)$. By (III), $Q_i = P_{1i} = \theta_i$. Since

$$X_i^{(N)}(t) = X_i(t) \quad \text{and} \quad y_{1i}^{(N)}(t) = y_{1i}(t)$$

for $t \in [0, U(N)]$,

$$(77) \quad \lim_{N \rightarrow \infty} X_i^{(N)}(U(N)) = \lim_{N \rightarrow \infty} y_{1i}^{(N)}(U(N)) = \theta_i,$$

and thus

$$\lim_{N \rightarrow \infty} m_i^{(N)} = \lim_{N \rightarrow \infty} M_i^{(N)} = \theta_i.$$

In particular,

$$(78) \quad \lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} X_i^{(N)}(t) = \lim_{N \rightarrow \infty} \lim_{t \rightarrow \infty} y_{1i}^{(N)}(t) = \theta_i.$$

Case 2. $U(N) > U_1(N)$. In this case we will find that $X_i^{(N)}(U(N))$ and $y_{1i}^{(N)}(U(N))$ lie in $[m_{i\theta}^{(N)}, M_{i\theta}^{(N)}]$, where

$$m_{i\theta}^{(N)} = \min \{ \theta_i, X_i^{(N)}(U_1(N)), y_{1i}^{(N)}(U_1(N)) \}$$

and

$$M_{i\theta}^{(N)} = \max \{ \theta_i, X_i^{(N)}(U_1(N)), y_{1i}^{(N)}(U_1(N)) \}.$$

Then by (III),

$$X_i^{(N)}(U_1(N)) = X_i(U_1(N)) \quad \text{and} \quad y_{1i}^{(N)}(U_1(N)) = y_{1i}(U_1(N)),$$

and since $Q_i = P_{1i} = \theta_i$,

$$\lim_{N \rightarrow \infty} X_i^{(N)}(U_1(N)) = \lim_{N \rightarrow \infty} y_{1i}^{(N)}(U_1(N)) = \theta_i,$$

whence (77) and thus (78) follow.

For $t \in [U_1(N), U(N)]$,

$$\dot{X}_i^{(N)} = A_i^{(N)}(y_{1i}^{(N)} - X_i^{(N)}) + B_i^{(N)}(\theta_i - X_i^{(N)})$$

and

$$\dot{y}_{1i}^{(N)} = C_i^{(N)}(X_i^{(N)} - y_{1i}^{(N)}).$$

Suppose $y_{1i}^{(N)}(U_1(N)) \geq X_i^{(N)}(U_1(N)) \geq \theta_i$. Then $y_{1i}^{(N)}(t) \geq X_i^{(N)}(t) \geq \theta_i$ and $y_{1i}^{(N)}(t)$ is monotone nonincreasing for $t \geq U_1(N)$. Suppose $y_{1i}^{(N)}(U_1(N)) \geq X_i^{(N)}(U_1(N)) < \theta_i$. Then $y_{1i}^{(N)}(t)$ is monotone nonincreasing and $X_i^{(N)}(t)$ is monotone nondecreasing until $X_i^{(N)}(t) = \theta_i$ at time $t = T$, after which $y_{1i}^{(N)}(t) \geq X_i^{(N)}(t) \geq \theta_i$ and $y_{1i}^{(N)}(t)$ is monotone nonincreasing for $t \geq T$. In both cases

$X_i^{(N)}(U(N))$ and $y_i^{(N)}(U(N))$ lie in $[m_{i\theta}^{(N)}, M_{i\theta}^{(N)}]$. The two remaining cases can be similarly treated.

The oscillatory behavior of $\dot{y}_{i\theta}^{(N)}$, $f_i^{(N)}$ and $g_i^{(N)}$ readily follows from Lemmas 3, 4, and 5, and the above argument. Theorem 2 is hereby proved.

Theorem 2 holds in the special case when I_1 and I are constructed from sequences of input pulses, where an *input pulse* is a nonnegative continuous function that is positive in a finite interval.

Corollary 1. *Let*

$$I_1(t) = \sum_{k=1}^{\infty} J_1(t - t_1(k))$$

and

$$I(t) = \sum_{k=1}^{\infty} J(t - t(k)),$$

where J_1 and J are input pulses that are positive in $(0, \lambda_1)$ and $(0, \lambda)$, respectively,

$$\sup_t \int_0^t e^{-\alpha_1(t-v)} J_1(v) dv > \Gamma,$$

and the sequences $\{t_1(k): k \geq 1\}$ and $\{t(k): k \geq 1\}$ satisfy

$$\epsilon_1 \leq t_1(k+1) - t_1(k) \leq \epsilon_2$$

and

$$\delta_1 \leq t(k+1) - t(k) \leq \delta_2$$

for some positive numbers ϵ_1 , ϵ_2 , δ_1 , and δ_2 . Theorem 2 holds for any such choice of I_1 and I .

The proof is obvious, amounting merely to showing that the exponentially weighted sum of an input pulse that is iterated with bounded spacing eventually has positive upper and lower bounds.

Corollary 1 shows that a single Γ -outstar will learn a spatial pattern perfectly if its source vertex v_1 receives inputs from a v_0 which is perturbed by a sequence of sufficiently intense input pulses with bounded spacing.

Corollary 2. *Let*

$$I_1(t) = \left[\sum_{k=1}^{\infty} \int_0^t e^{-\alpha_0(t-v)} J_0(v - t_0(k)) dv - \Gamma_0 \right]^+$$

and

$$I(t) = \sum_{k=1}^{\infty} J(t - t(k)),$$

where J_0 and J are input pulses that are positive in $(0, \lambda_0)$ and $(0, \lambda)$, respectively,

$$\sup \int_0^t e^{-\alpha_1(t-\tau)} \left[\int_0^\tau e^{-\alpha_0(\tau-\xi)} J_0(\xi) d\xi - \Gamma_0 \right]^+ dv > \Gamma,$$

and the sequences $\{t_o(k) : k \geq 1\}$ and $\{t(k) : k \geq 1\}$ satisfy

$$\epsilon_1 \leq t_o(k+1) - t_o(k) \leq \epsilon_2$$

and

$$\delta_1 \leq t(k+1) - t(k) \leq \delta_2$$

for some positive numbers ϵ_1 , ϵ_2 , δ_1 , and δ_2 . Theorem 2 holds for any such choice of I_1 and I .

Let $I_1, I_2, \dots, I_n, \dots$ be the nonoverlapping closed intervals of time in which $x_1(t - \tau) \leq \Gamma_1$, and let $J_1, J_2, \dots, J_n, \dots$ be the complementary open intervals on which $x_1(t - \tau) > \Gamma_1$. (55) shows that no learning occurs in the Γ -outstar for $t \in \bigcup_{n=1}^{\infty} I_n$, since then $\dot{y}_{1i}(t) = 0$. Corollary 2 shows, nonetheless, that the Γ -outstar can learn any spatial pattern perfectly for $t \in \bigcup_{n=1}^{\infty} J_n$, if it is driven by a series of intense input pulses at v_o .

To learn a series of spatial patterns in $K(\xi, T)$ Γ -outstars, it remains only to guarantee that the source functions $x_{k,1}(t - \tau)$ of successive outstars exceed Γ_1 in successive, nonoverlapping time intervals, and that these time intervals occur when the k^{th} spatial approximation to the space-time pattern is arriving at B_n . The first condition is readily achieved by choosing J_o so that $\xi > T$, where

$$T = \sup \left\{ t : \int_0^t e^{-\alpha_1(t-\tau)} \left[\int_0^\tau e^{-\alpha_0(\tau-\xi)} J_o(\xi) d\xi - \Gamma_0 \right]^+ dv > \Gamma_1 \right\}.$$

The second condition requires that when the $(k+1)^{\text{th}}$ spatial approximation arrives at B_n , effects of the k former spatial approximations and of inputs from $v_{m,1}$, $m = 1, 2, \dots, k$, shall have substantially decayed. This condition is implemented by increasing the decay rate α in B_n , or by decreasing

$$\sup_t \max_i |\dot{\theta}_i(t)|$$

in intervals of length ξ , or by decreasing T/ξ , etc. The decay rate of previous spatial approximants relative to the arrival speed of new spatial approximants is the rate-limiting factor determining accuracy of learning in the Γ -outstar avalanche. Another way of improving this accuracy is mentioned in Section 12, and uses an inhibitory feedback mechanism. Or independent grids can be used.

10. Simultaneous storage of patterns and multimodal learning. A network that can learn any number of space-time patterns on any number of finite grids can now readily be constructed from Γ -outstar avalanches. Just one control vertex $v_{m,o}$ is needed to activate the m^{th} independent pattern. See Figure 6.

The avalanche with control $v_{1,o}$ can learn from both grids ("modalities") G_1 and G_2 . The time lags $\tau_1^{(1)}$ and $\tau_1^{(2)}$ for signals to travel from the source vertex

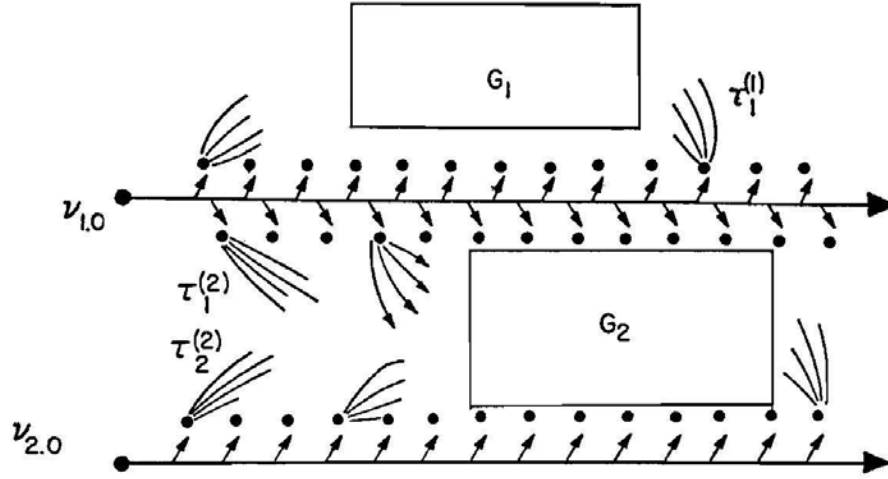


FIGURE 6.

$v_{1,k,1}^{(1)}$ to G_1 and from $v_{1,k,1}^{(2)}$ to G_2 , respectively, can be different to help synchronize the learning of patterns in modalities which filter inputs at different rates. The avalanche with control $v_{2,o}$ also learns from G_2 . Clearly the two avalanches will not interfere with each other if their controls $v_{1,o}$ and $v_{2,o}$ are perturbed at widely spaced times. For completeness, we list the equations of any number of avalanches learning from any number of grids.

$$(79) \quad \dot{x}_{m,o}(t) = -\alpha_{m,o}x_{m,o}(t) + I_{m,o}(t),$$

$$(80) \quad \dot{x}_{mk,1}^{(r)}(t) = -\alpha_{m1}^{(r)}x_{mk,1}^{(r)}(t) + \beta_{m1}^{(r)}[x_{m,o}(t - t_m^{(r)} - k\xi_m^{(r)}) - \Gamma_{m1}^{(r)}]^+,$$

$$(81) \quad \begin{aligned} \dot{x}_i^{(r)}(t) &= -\alpha^{(r)}x_i^{(r)}(t) \\ &+ \sum_{m \in M_r} \sum_{k=1}^{k_{mr}} \beta_{m1}^{(r)}[x_{mk,1}^{(r)}(t - \tau_m^{(r)}) - \Gamma_{m1}^{(r)}]^+ y_{mk,1,i}^{(r)}(t) + I_i^{(r)}(t), \end{aligned}$$

$$(82) \quad y_{mk,1,i}^{(r)}(t) = z_{mk,1,i}^{(r)}(t) \left[\sum_{p=2}^{n_r} z_{mk,1,p}^{(r)}(t) \right]^{-1},$$

and

$$(83) \quad \dot{z}_{mk,1,i}^{(r)}(t) = -u_m^{(r)}z_{mk,1,i}^{(r)}(t) + \gamma_{m1}^{(r)}[x_{mk,1}^{(r)}(t - \tau_m^{(r)}) - \Gamma_{m1}^{(r)}]^+ x_i^{(r)}(t),$$

where $x_{m,o}$ is the control vertex function of the m^{th} avalanche, $x_{mk,1}^{(r)}$ is the source vertex function of the k^{th} spatial approximant in the m^{th} avalanche leading to the r^{th} grid, $x_i^{(r)}$ is the i^{th} border vertex function of the r^{th} grid, $y_{mk,1,i}^{(r)}$ is the $(1, i)^{\text{th}}$ associational strength from the m^{th} avalanche to the r^{th} grid, etc. The avalanches indexed by $m \in M_r$ perturb the r^{th} grid, which has n_r vertices. The m^{th} avalanche has $k_{mr} = K(\xi_m^{(r)}, T_m)$ spatial approximants leading to the r^{th} grid. The relative onset time $t_m^{(r)}$ of the m^{th} avalanche to the r^{th} grid, and the time lags $\xi_m^{(r)}$ between successive spatial approximants can vary with (r, m) . If a given Γ -outstar sends signals only to a proper subset of a grid's vertices, then we partition the grid into two parts and consider each grid as a grid in its own right.

An alternative to (79)-(83) is given by (79), (80), (82),

$$(81') \quad \dot{x}_{mk,i}^{(r)}(t) = -\alpha_m^{(r)} x_{mk,i}^{(r)}(t) + \beta_{mi}^{(r)} [x_{mk,i}^{(r)}(t - \tau_m^{(r)}) - \Gamma_{mi}^{(r)}]^+ y_{mk,i}^{(r)}(t) + I_i^{(r)}(t),$$

and

$$(83') \quad \dot{z}_{mk,i}^{(r)}(t) = -u_m^{(r)} z_{mk,i}^{(r)}(t) + \gamma_{mi}^{(r)} [x_{mk,i}^{(r)}(t - \tau_m^{(r)}) - \Gamma_{mi}^{(r)}]^+ x_{mk,i}^{(r)}(t).$$

In this case, each Γ -outstar $(\mathfrak{N}^{(1)})_{mk}^{(r)}$ in the m^{th} avalanche has its own grid, which is perturbed by the inputs $I_i^{(r)}$, $i = 2, \dots, n_r$, via axon collaterals from the r^{th} input source. The output to the r^{th} output sink is given by

$$O_i^{(r)}(t) = \delta^{(r)} \sum_{m \in M_r} \sum_{k=1}^{k_{mr}} [x_{mk,i}^{(r)}(t) - \Gamma_m^{(r)}]^+,$$

where the $\Gamma_m^{(r)}$ are small, but possibly positive, signal thresholds.

11. Diffuse arousal inputs. Several presentations, or trials, of a space-time pattern are often required before the pattern can be well learned by a Γ -outstar avalanche. Suppose that on the i^{th} trial, the first spatial approximation to the pattern arrives S_i time units after the control vertex v_o emits a signal. Unless all S_i are approximately equal, a given outstar $\mathfrak{N}_k^{(1)}$ in the avalanche will learn different spatial approximations on different trials, and the avalanche will never learn any one space-time pattern well. A way must be found, therefore, to guarantee that v_o transmits a signal *only* if the relative timing of v_o signal and B_n input is approximately the same on all trials. This problem does not, of course, arise when B_n receives a single spatial pattern throughout the time interval of length T during which the signal from v_o is active.

Our goal can be reached in either of two ways. The first way supposes that two sources of inputs to v_o exist: a "conditioned stimulus", or CS, to v_o alone, such as we have previously discussed, and a "diffuse arousal input", or DAI, which is controlled by the input sources that create the pattern at B_n . We can easily guarantee that

a) x_o will exceed Γ_o , and therefore transmit a signal, only if *both* the CS and the DAI arrive at v_o almost simultaneously, and

b) the DAI arrives at v_o a prescribed time η *before* the "unconditioned stimulus" arrives at B_n . Thus the term "arousal" in DAI means that a control vertex perturbed by a CS and a DAI readies its avalanche to be able to learn from the inputs reaching the grid a short time later. η is chosen so that a signal from v_o can activate the "synaptic knobs" of the first spatial approximant in the avalanche before the space-time pattern reaches the grid.

Given (a) and (b), we must also require

c) the input to v_o on a recall trial is larger than the CS to v_o on a learning trial, or else no output from B_n could ever occur during recall.

The DAI must also have the following properties.

d) *Every* edge producing an input to B_n must send a DAI, either directly or indirectly, to *every* control vertex $v_{m,o}$. Otherwise there would exist patterns playing on a small number of grid vertices which could never produce a DAI, and control vertices $v_{m,o}$ which could never send signals. Thus the term “diffuse” in DAI refers to a widespread spatial dispersion of this input to control vertices.

e) If the parameters of two avalanches differ, then the parameters of the DAI must differ commensurately. For example, let two avalanches with control vertices $v_{1,o}$ and $v_{2,o}$ be given such that $\Gamma_{1,o} \gg \Gamma_{2,o}$. If the same DAI perturbs both avalanches, then by choosing a DAI such that $\Gamma_{1,o} > x_{1,o} = x_{2,o} > \Gamma_{2,o}$, (a) will be satisfied in the first avalanche and not in the second. The DAI perturbing $v_{2,o}$ must therefore be chosen smaller than the DAI perturbing $v_{1,o}$.

Consider a single avalanche with control v_o and border vertices v_i . The simplest choice of the DAI to v_o created by the input source perturbing v_i is

$$\tilde{I}_{i,o}(t) = \omega I_i(t + \eta),$$

and the total DAI is

$$\begin{aligned} \tilde{I}_o(t) &= \sum_{i=2}^n \tilde{I}_{i,o}(t) \\ (84) \quad &= \omega \sum_{i=2}^n I_i(t + \eta), \end{aligned}$$

where ω and η are positive. In other words, v_o receives a suitable multiple of the total input received by the border η time units later. η is chosen to satisfy (b) and ω is chosen to satisfy (a). By (a), neither a CS I_o nor a DAI \tilde{I}_o alone can exceed Γ_o , but a suitable intense combination $I_o + \tilde{I}_o$ can. That is,

$$(85) \quad \tilde{I}_o \equiv \sup_t \int_0^t e^{-\alpha_o(t-v)} I_o(v) dv < \Gamma_o$$

and

$$(86) \quad \omega \sup_t \sum_{i=2}^n \int_0^t e^{-\alpha_o(t-v)} I_i(v + \eta) dv < \Gamma_o,$$

whereas

$$(87) \quad \sup_t \int_0^t e^{-\alpha_o(t-v)} \left[I_o(v) + \omega \sum_{i=2}^n I_i(v + \eta) \right] dv > \Gamma_o$$

for sufficiently intense inputs. Other things equal, ω must decrease as n increases to satisfy (86). Suppose that n is large. Since an input to just one border vertex v_i is a spatial pattern, an intense input to such a vertex, summing as $\omega I_i(t + \eta)$ with the CS $I_o(t)$ at v_o , ought to satisfy (87). Since ω is very small for large n , $\omega I_i(t + \eta)$ will be small as well, and thus \tilde{I}_o in (85) will lie close to Γ_o for large n . The individual DAI's $\omega I_i(t + \eta)$ are therefore small “subliminal” contributions to the total input at v_o which suffice to drive x_o to suprathreshold values. See Figure 7.

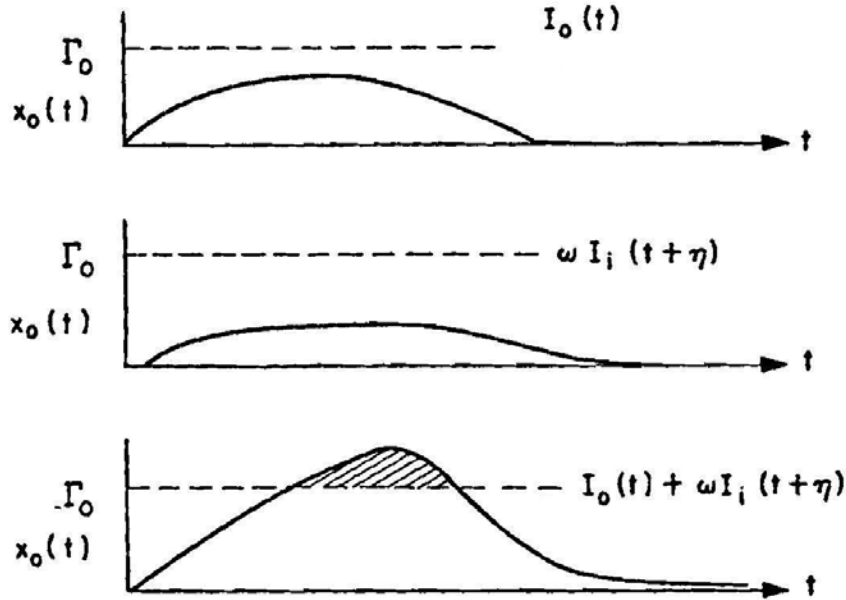


FIGURE 7.

A second way to accomplish the same goals is to let v_0 be inhibited by a “diffuse suppression input”, or DSI, which is, in turn, inhibited by the input source that perturbs B_n , η time units before B_n is perturbed. Many statements about summing excitatory influences to exceed threshold can be replaced by statements concerning inhibition of inhibiting influences (“disinhibition”) that had prevented thresholds from being reached.

12. Feedback inhibition from control vertices. Once v_0 emits a brief signal due to summation of a CS and a DAI, no new signals can be allowed to leave v_0 until all $K(\xi, T)$ outstars in the avalanche have been excited by the signal. Otherwise, the second signal from v_0 would again cause each Γ -outstar in the avalanche to learn from more than one spatial approximation to the pattern. Thus the DAI sent towards v_0 by later spatial approximations to the pattern at B_n must never reach v_0 . The signal from v_0 itself must therefore inhibit these later inputs. A release from inhibition will automatically occur after the signal from v_0 has excited the entire avalanche. See Figure 8.

Figure 8 interpolates a vertex v_{+1} between the DAI input and v_0 , which is inhibited by a feedback input from v_{-1} . The “+” signs designate “excitatory” edges and the “-” signs designate “inhibitory” edges. This particular version of feedback inhibition of the DAI by v_0 satisfies equations of the form

$$(88) \quad \dot{x}_0(t) = -\alpha_0 x_0(t) + I_0(t) + \beta_{+1}[x_{+1}(t - \tau_{+1}) - \Gamma_{+1}]^+,$$

$$(89) \quad \dot{x}_{-1}(t) = -\alpha_{-1} x_{-1}(t) + \beta_0[x_0(t - \xi_0) - \Gamma_0]^+,$$

and

$$(90) \quad \dot{x}_{+1}(t) = -\alpha_{+1} x_{+1}(t) - \beta_{-1}[x_{-1}(t - \xi_{-1}) - \Gamma_{-1}]^+ + \omega \sum_{i=2}^n I_i(t + \eta),$$

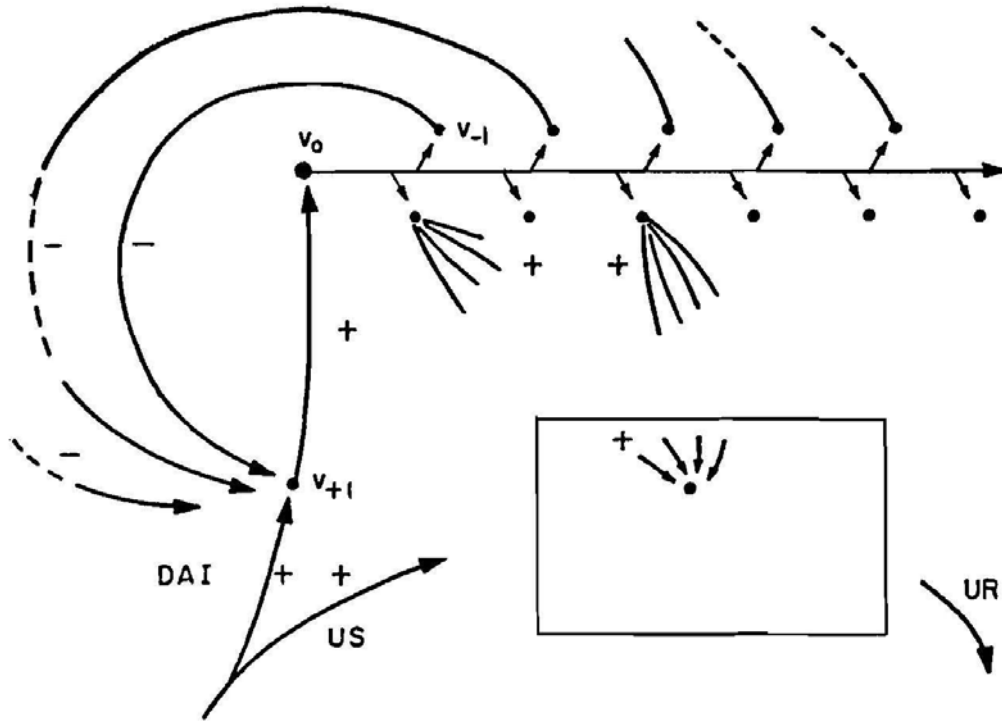


FIGURE 8.

where ω and η must be adjusted to compensate for the additional parameters and time lags in (88)–(90). The reader can readily construct improvisations on this theme, say without the inhibitory internode v_{-1} , and can modify the conditions (85)–(87) accordingly.

A very interesting phenomenon occurs if Figure 8 is replaced by Figure 9. Then the DAI input is inhibited periodically and, for suitable parameter choices, never reaches a size needed to create a second signal from v_0 , until the space-time pattern has been wholly delivered. But also the pattern input to B_n is periodically inhibited, and the period of successive inhibitory signals can be chosen equal to ξ . These inhibitory signals chop up the incoming space-time pattern into spatial approximations to the pattern with a time spacing of ξ units between successive approximations, and thereby prepare the pattern for learning by successive Γ -outstars in the avalanche. This chopping procedure enables the avalanche to learn spatial approximations to rapidly varying $\theta_i(t)$ given relatively small values of α , since it decreases the length of time intervals during which large inputs reach B_n . During a recall trial, the periodic inhibitory signals do not affect the grid. Thus the small decay parameter α allows the successively activated spatial approximants to blend smoothly on the grid, and to thereby produce a more smoothly modulated output through time than is encapsulated in the remembered successive spatial approximations of the avalanche.

13. A hierarchy of facilitatory and incompatible behavioral acts. Suppose that two avalanches perturb the same grid and have each learned different

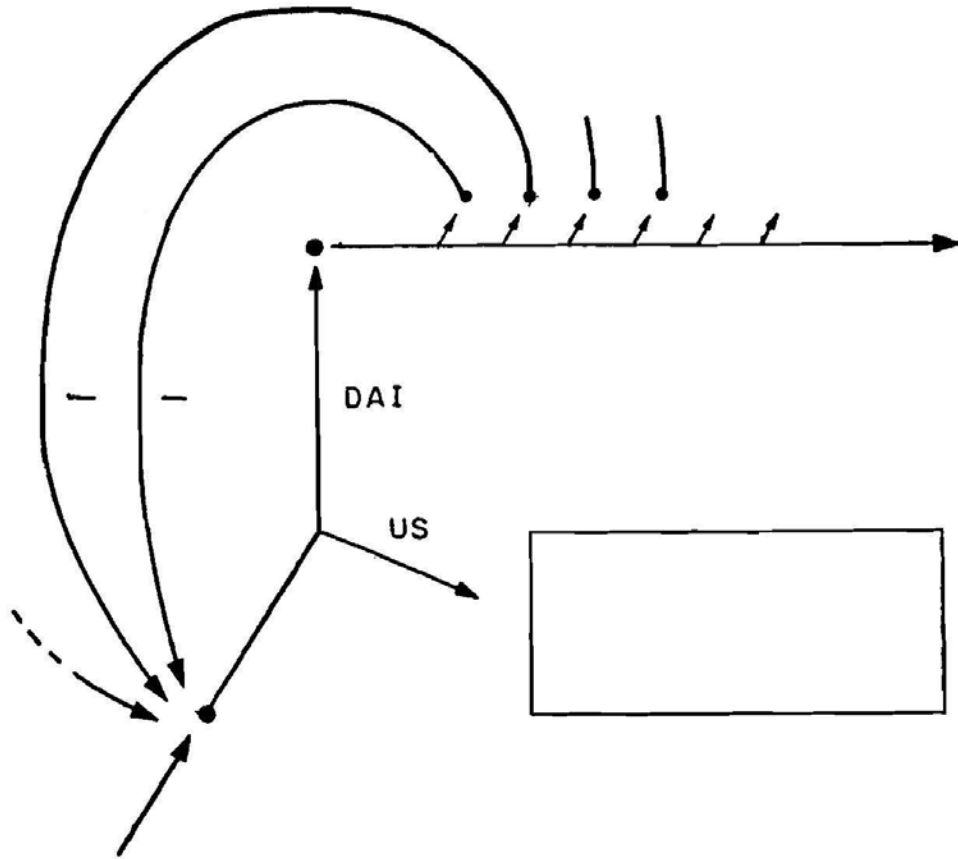


FIGURE 9.

space-time patterns. These avalanches must be prevented from thereafter perturbing the common grid simultaneously. Otherwise, the grid output will be a noisy mixture of both patterns, and the memory of each avalanche will be destroyed. This can easily be guaranteed by supposing either that the control vertices $v_{1,0}$ and $v_{2,0}$ of the two avalanches mutually inhibit each other, or that the CS to one control vertex sends an inhibitory axon collateral to the other. See Figure 10.

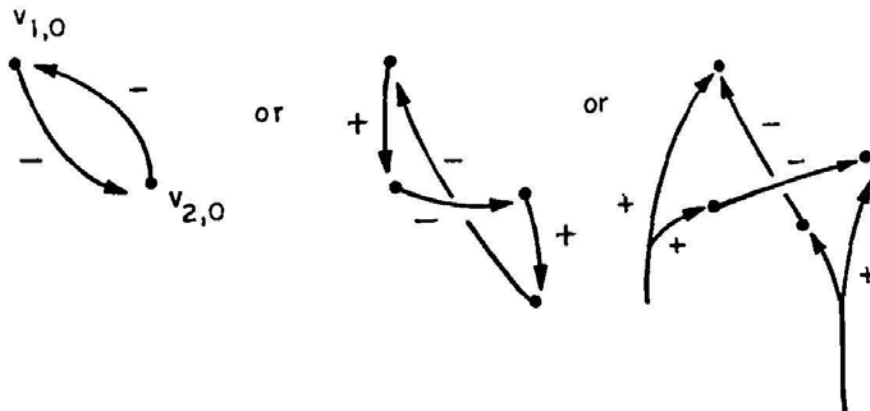


FIGURE 10.

A similar competition between control vertices occurs in a *signed outstar* ([3], Section 15).

The mechanisms of Figure 9 and 10 taken together have the following effect. Suppose that no inputs activate $v_{1,o}$ or $v_{2,o}$ for some time, so that $x_{1,o}$ and $x_{2,o}$ are approximately zero. Then let $v_{1,o}$ receive a large simultaneous CS and DAI (and much larger than the CS + DAI received by $v_{2,o}$). By Figure 10, $v_{1,o}$ inhibits $v_{2,o}$ and simultaneously sends a signal to its avalanche. By Figure 9, this signal can be made to cut off the DAI's reaching $v_{2,o}$ until the first avalanche has been played out. This mechanism guarantees that at most one avalanche will learn from, or perform upon, the grid at any time. The avalanches $v_{1,o}$ and $v_{2,o}$ are said to control "incompatible" behavioral acts, because the performance of one interferes with the performance of the other. Usage of the same grid can be thought of in a behavioral context as control of the same muscle groups, or way-stations to the same muscle groups.

Avalanches that control *different* grids need not mutually inhibit each other, since performances of space-time patterns on different grids (*i.e.*, by different "muscles") need not interfere with one another (one can talk as one walks). The spatial distribution of inhibitory connections between control vertices determines the degree of "incompatibility" between the behavioral acts controlled by the vertices. If for example, $v_{1,o}$ inhibits $v_{2,o}$ much more strongly than $v_{2,o}$ inhibits $v_{1,o}$, then $v_{1,o}$ is the *dominant* control of the pair, since it will activate its avalanche and inhibit $v_{2,o}$ if both receive equal inputs. In any given network, a *hierarchy* of incompatible controls playing on common grids will be determined by the relative strengths of mutual inhibition between the control vertices.

In a similar fashion, controls operating over different grids can facilitate one another by contributing to each other's DAI. For example, suppose that $v_{1,o}$ and $v_{2,o}$ perform upon different grids, and $v_{1,o}$ sends a small excitatory signal to $v_{2,o}$ whenever $x_{1,o}$ reaches suprathreshold values. Then $v_{2,o}$ will be able to activate its avalanche if it simultaneously receives a US. For example let $v_{1,o}$ control "sniffing" in response to a DAI caused by "hunger" and the US caused by the smell of "food", and let $v_{2,o}$ control "salivation". The above remarks suggest that \mathfrak{N} will start "salivating" after "sniffing" begins with less provocation than before "sniffing" begins. In this sense, prior search for food has "lowered \mathfrak{N} 's salivation threshold." Actually "sniffing" has raised $x_{2,o}$ closer to its signal threshold. This example is admittedly quite naive, but it can be extended to achieve rather realistic effects.

In short, the geometry of the signed graph with fixed excitatory and inhibitory path weights between \mathfrak{N} 's control vertices determines hierarchies of incompatible, facilitatory, and merely unrelated output behaviors by \mathfrak{N} . The avalanches within \mathfrak{N} that succeed in learning or performing at any time will be determined by the spatial distribution of all conditioned, unconditioned, and diffuse arousal stimuli to \mathfrak{N} , along with the feedback excitatory and inhibitory signals thereby created. The geometry of the signed graph can automatically forbid the simultaneous occurrence of incompatible behaviors and can favor the

performance of those behaviors which are most compatible with \mathfrak{N} 's internal states at any time.

14. **Novelty and habituation.** A "novel stimulus" in this setting has a clear meaning. It is an input that succeeds in activating some avalanche. Essentially all other inputs are too weak, or will be suppressed by feedback inhibition, at some stage in their development. All inputs to a grid give rise to a DAI, but the DAI will be effective in producing learning only if some avalanche is "ready" to take heed of it, because of prior preparation by a CS—that is, by other inputs that determine the network's "psychological set" at the given time. Even if a spatial pattern is novel at a given time, repeating it will cause "habituation"—that is, once an avalanche is activated, it will suppress the DAI of the repeated pattern.

15. **Higher-order controls.** The above remarks can be extended in many directions. For example, a given set of control vertices $v_{k,o}$ can be the border vertices of a Γ -outstar $\mathfrak{N}_o^{(1)}$ whose input to $v_{k,o}$ is the DAI of $v_{k,o}$. See Figure 11.

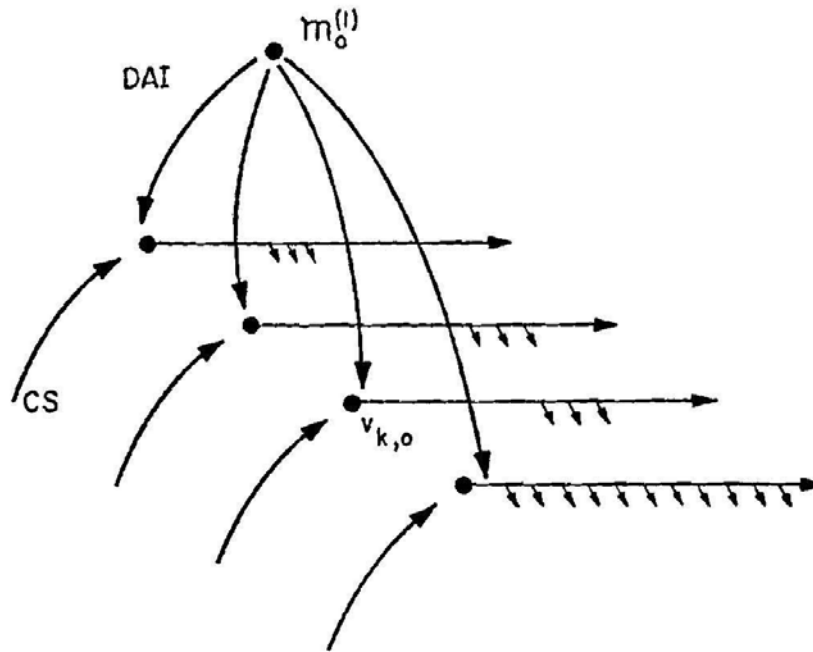


FIGURE 11.

After several trials, the associational strengths $y_{o,1k}$ of $\mathfrak{N}_o^{(1)}$ can learn any *spatial pattern* of CS's playing on the control vertices $v_{k,o}$. A later recall input to just the control vertex of $\mathfrak{N}_o^{(1)}$ can therefore activate a spatial pattern of space-time patterns.

16. "On" - "Off" performance of complicated reflexes and "paying attention". The equations for a Γ -outstar avalanche can be modified so that no

learning occurs, and an arbitrarily prescribed "reflex pattern" is produced whenever the control is activated. Merely replace (83) by the equation

$$(91) \quad \dot{z}_{mk,i}^{(r)}(t) = 0;$$

that is, choose the parameters $u_{ik} = v_{ik} = 0$ in (3). Then the system (79)–(82), and (91) can reproduce any number of reflexes, parameterized by m , on any number of grids parameterized by r . The m^{th} reflex to the r^{th} grid will be determined by the sequence $\{y_m^{(r)}(k)\}$ of spatial approximants with

$$(92) \quad y_m^{(r)}(k) = (y_{mk,12}^{(r)}(0), y_{mk,13}^{(r)}(0), \dots, y_{mk,in}^{(r)}),$$

$k = 1, 2, \dots, K(\xi_m^{(r)}, T_m)$, where T_m is the duration of the m^{th} reflex. Speaking heuristically, activating the single "control neuron" $v_{m,o}$ can reproduce a very complicated prescribed sequence of inputs to any collection of muscle groups, and can thereby produce very complicated motor behavior. It is, in fact, well known that a single neuron can activate complicated reflex acts, say in insects ([17], p. 8).

If the reflex is repetitive, as in walking, then we let the control $v_{m,o}$ excite itself via an axon collateral, thereby forming a loop, as in Figure 12.

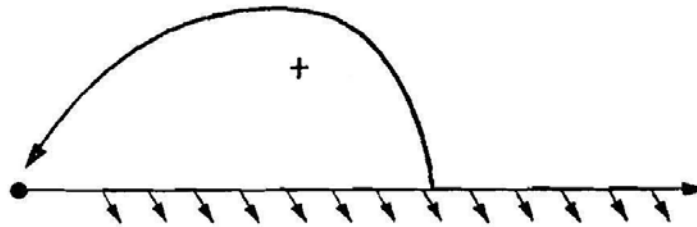


FIGURE 12.

Each total activation of the loop corresponds to a single cycle of the reflex. One can easily choose the parameters of the loop collateral such that its input alone will not activate another cycle of the avalanche. In this case, supplementary inputs must bring $x_{m,o}$ to suprathreshold values—for example, DAI's due to \mathfrak{N} 's motion through the environment. If the parameters are chosen so that the loop collateral can activate $x_{m,o}$ alone, then "walking" will continue until \mathfrak{N} sends an inhibitory signal to $v_{m,o}$ and thereby breaks the cycle. In this case, "walking" can be viewed as a simple "on"–"off" phenomenon: once the control $v_{m,o}$ is turned on, the cyclic performance of the motor pattern described by (92) can be carried out *automatically* until \mathfrak{N} shuts the reflex off with an inhibitory input. During the interim period, \mathfrak{N} need never worry about the "walking" process, albeit the pattern (92) can be a very complicated one. \mathfrak{N} can "pay attention" to other matters entirely.

A simpler case is that of a "clasp" or "clutch" reflex; for example, a hand closed tightly around an object for a long time. Here, each successive spatial approximant $y_m^{(r)}(k)$, in (92) can be chosen the same—*i.e.*, is an iterated spatial pattern activating the clutching muscle groups. This clutch pattern can be continued

indefinitely by choosing the feedback loop in Figure 12 so that the time lag needed for the feedback loop to re-excite v_0 is approximately equal to the time lag needed to finish up one cycle of the clutch reflex. Thus the second cycle can begin just as the cycle ends, unless an inhibitory signal v_0 drives x_0 to subthreshold values. Sets of outstar avalanches can, *in principle*, perform much more complicated patterns, such as learning any number of pieces of piano music. Of course, they will perform the pieces without any artistic subtlety.

17. Alternative systems. Part II of this paper [13] will study learning of space-time patterns by avalanches constructed from the following Γ -outstar.

$$(92) \quad \dot{x}_1(t) = -\alpha x_1(t) + I_1(t),$$

$$(93) \quad \dot{x}_i(t) = -\alpha x_i(t) + \beta[x_1(t - \tau) - \Gamma_1]^+ z_{1i}(t) + I_i(t),$$

and

$$(94) \quad \dot{z}_{1i}(t) = -u z_{1i}(t) + \gamma[x_1(t - \tau) - \Gamma_1]^+ x_i(t),$$

$i = 2, \dots, n$. This Γ -outstar does not use the ratios $y_{1i}(t)$. It is therefore more plausible physically, but also its memory decays exponentially at the rate u . The ratios

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

nonetheless are still remembered essentially perfectly in the absence of practice.

Part II will also discuss avalanches constructed from a Γ -outstar satisfying (92), (93) and

$$(95) \quad \dot{z}_{1i}(t) = [-u z_{1i}(t) + \gamma x_i(t)][x_1(t - \tau) - \Gamma_1]^+.$$

This system has perfect memory except during recall trials, and is also plausible physically. During recall trials, this system undergoes "extinction" of memory. (95) replaces the decay rate u in (94) by

$$(96) \quad u[x_1(t - \tau) - \Gamma_1]^+.$$

(96) couples the decay of transmitter producing activity to the presynaptic spiking frequency. Were this coupling to exist *in vivo*, it would presumably be carried out by the increase in Na^+ and the decrease in K^+ within N_{1i} , that occurs during the action potential.

Actually Part II will discuss a large class of closely related equations that include (92)–(94) and (92), (95), (96) as special cases.

REFERENCES

- [1] S. GROSSBERG, Some physiological and biochemical consequences of psychological postulates, *Proc. Natl. Acad. Sci.*, **60** (1968) 758–765.
- [2] S. GROSSBERG, Embedding fields: A theory of learning with physiological implications, *J. Math. Psych.*, **6** (1969).

- [3] S. GROSSBERG, On learning, information, lateral inhibition, and transmitters, *Math. Biosci.*, 4 (1969).
- [4] S. GROSSBERG, On the production and release of chemical transmitters and related topics in cellular control, *J. Theoret. Biol.*, 22 (1969) 325-364.
- [5] S. GROSSBERG, On the serial learning of lists, *Math. Biosci.*, 4 (1969).
- [6] S. GROSSBERG, Nonlinear difference-differential equations in prediction and learning theory, *Proc. Natl. Acad. Sci.*, 58 (1967) 1329-1334.
- [7] S. GROSSBERG, Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity, *Proc. Natl. Acad. Sci.*, 59 (1968) 368-372.
- [8] S. GROSSBERG, A prediction theory for some nonlinear functional-differential equations, I-Learning of lists, *J. Math. Anal. and Applics.*, 21 (1968) 643-694.
- [9] S. GROSSBERG, A prediction theory for some nonlinear functional-differential equations, II-Learning of Patterns, *J. Math. Anal. and Applics.*, 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. Eqns.*, 5 (1969) 531-563.
- [11] S. GROSSBERG, On the variational systems of some nonlinear difference-differential equations, *J. Diff. Eqns.*, in press.
- [12] S. GROSSBERG, On learning of spatiotemporal patterns by networks with ordered sensory and motor components, I-Excitatory components of the cerebellum, *Studies in Applied Math*, 48 (1969).
- [13] S. GROSSBERG, Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II, submitted to *SIAM J. Appl. Math.*
- [14] R. B. LIVINGSTON, Brain mechanisms in conditioning and learning, *Neurosciences Research Symposium Summaries*, Vol. 2 (ed., SCHMITT, F. O., et al.), M. I. T. Press, Cambridge, 1967.
- [15] S. R. KOREY (ed.), *The Biology of Myelin*, Hoeber-Harper, New York, 1959.
- [16] E. C. CROSBY, T. HUMPHREY, & E. W. LAUER, *Correlative Anatomy of the Nervous System*, The Macmillan Company, New York, 1962.
- [17] V. G. DETHIER, *Physiology of Insect Senses*, Methuen and Company, London, 1963.

Massachusetts Institute of Technology
Date communicated: NOVEMBER 26, 1968