Reprinted from:
DELAY AND FUNCTIONAL DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS

Academic Press, Inc., New York and London
ED.: K. SCHMITT

# PATTERN LEARNING BY FUNCTIONAL-DIFFERENTIAL NEURAL NETWORKS WITH ARBITRARY PATH WEIGHTS

Stephen Grossberg*

Massachusetts Institute of Technology

## Introduction

The Theory of Embedding Fields studies systems of nonlinear functional-differential equations which can be derived from psychological postulates and interpreted as neural networks [1]. These systems describe cross-correlated flows on signed directed graphs. They have been applied to problems in pattern discrimination, learning, memory, and recall (e.g., [1] - [10]).

The theory is derived in several stages ([1], [4], [9], [10]). Each stage exhibits the minimal systems that are compatible with a given list of psychological postulates. Successive stages refine either the dynamical equations themselves, or the synthesis of network connections, to satisfy additional postulates. This paper reviews the derivation of stage one, for completeness, and proves two general theorems about spatial pattern learning by a suitable class of networks of Embedding Field type. Weaker versions of these theorems were announced without proof in [7]. The theorems will also be interpreted psychologically and physiologically. They describe properties of learning that are invariant under broad changes in physiological and anatomical constraints. In particular, they permit a discussion of learning in an essentially arbitrary anatomy.

## Derivation of Some Networks

We will globally analyse systems of the form

*Supported in part by the Alfred P. Sloan Foundation and the Office of Naval Research (N00014-67-A-0204-0051).

$$\dot{x}_i = A_i x_i + \sum_{k\in J} B_{ki} z_{ki} + C_i(t) \tag{1}$$

$$\dot{z}_{ji} = D_{ji} z_{ji} + E_{ji} x_i,$$

where $i\in I$, $j\in J$, and $I$ and $J$ are finite but possibly arbitrarily large sets that will be subject to suitable constraints. The symbols $A_i$, $B_{ji}$, $D_{ji}$, and $E_{ji}$ denote continuous functionals, not necessarily linear, with all $B_{ji}$ and $E_{ji}$ nonnegative. The input functions $C_i$ and initial data are chosen nonnegative and continuous.

Systems of this type can be derived by considering an experimentalist $\mathcal{E}$ who interacts with a machine $M$ to teach $M$ to predict B given A by practicing AB. An alternative version of this task is described by the following experiment. A hungry dog is presented with food and thereupon salivates. A bell is rung but the dog does not salivate. Then the bell is rung just before food presentation on several learning trials. Thereafter presentation of the bell alone yields salivation. This learning process is called respondent, or Pavlovian, conditioning [11]. Food is called the unconditioned stimulus (UCS), salivation is called the unconditioned response (UCR), and the bell is called the conditioned stimulus (CS). The sensory presentation of A is analogous to a CS, the sensory presentation of B is analogous to a UCS, and the motor response B is analogous to a UCR.

Systems (1) and (2) will thus describe versions of machines $M$ capable of learning complicated patterns by respondent conditioning. The inputs $C_i(t)$ will be chosen to represent a particular experiment performed on $M$ by $\mathcal{E}$. The outputs of $M$ will be suitable functionals of the vector function $X = (x_i : i\in I)$. The simplest version of $M$ is derived below. The derivation is given in story-book form to emphasize its intuitive basis.

A) Each Letter Seems Simple

In daily speech and listening, a letter is never decomposed into two parts. To maintain close contact with experience, we assume that a single state $v_A$ in $M$ corresponds to A. In a similar fashion, let

$v_B$ correspond to B, $v_C$ to C, etc. We designate each $v_i$ by a point, or vertex.

B) Presentation Times

The times at which letters are presented to $M$ must be represented within $M$. For example, presenting A and then B with a time spacing of twenty-four hours should yield far different behavior than presentation with a time spacing of two seconds. Thus various functions of time should be associated with each vertex. To maintain contact with the "one-ness" of each letter, and to maximize the simplicity of our derivation, we let one function $x_A(t)$ be associated with $v_A$, one function $x_B(t)$ be associated with $v_B$, etc., as in Figure 1.

C) Continuous Vertex Functions

The function $x_A(t), \ldots, x_Z(t)$ will be chosen continuous, and in fact differentiable. Several reasons for this exist. The most specific reason is the following. Consider the question: What follows ABC? It is tempting to say D, but really the problem is ill-defined if the letters are presented one at a time with time spacing $w$ between successive letters. If indeed $w$ is small, say $w \cong 2$ seconds, then D might well be the correct response, but if $w \cong 24$ hours then to the sound C (="see") one can also reply "see what?" That is, as $w$ varies from small to large values, the influence of A and B on the prediction following C gradually wears off. Since $x_A(t)$ and $x_B(t)$ describe the relevance at time $t$ of A and B in $M$, we conclude that these functions also vary gradually in time.

D) Perturbations Instead of Presentations

Suppose A is never presented to $M$. Corresponding to the occurrence of "nothing" is the natural mathematical predisposition to set $x_A(t) = 0$ at all times $t$. (The equilibrium point 0 can, it turns out, be rescaled ultimately relative to the signal thresholds).

Suppose A is presented to $M$ for the first time at time $t = t_A$. Then $x_A(t)$ must be perturbed from 0 for certain $t > t_A$, or else $M$ would have no way of knowing that A occurred. We associate the occurrence of "something" with a positive deflection in the graph of $x_A$.

(The theory could also, in principle, be carried out with negative deflections.)

Shortly after A is presented, A no longer is heard by $M$. That is, $x_A(t)$ gradually returns to the value signifying no recent presentation of A, namely 0. In a similar fashion, if A is presented at times $t_A^{(1)} < t_A^{(2)} < \dots < t_A^{(N_A)}$, then we find the graph of Figure 2. The same construction holds true for all letters. In this way, we have translated the presentation of any letters A, B, C, ... in the alphabet at prescribed times into a definite sequence of perturbations of the vertex functions $x_A(t)$, $x_B(t)$, $x_C(t)$, ... .

E) Linearity

For notational convenience, we replace the alphabet A,B,C,... by any sequence $r_i$, $i = 1,2,\dots,n$, of $n$ behavioral atoms; the vertices $v_A$, $v_B$, $v_C$, ... by the vertices $v_i$, $i = 1,2,\dots,n$; and the vertex functions $x_A(t)$, $x_B(t)$, $x_C(t)$, ... by the vertex functions $x_i(t)$, $i = 1,2,\dots,n$. Now $r_i$ corresponds to $(v_i, x_i(t))$, $i = 1,2,\dots,n$.

What is the simplest way to translate Figure 2 into mathematical terms? Since we are constructing a system whose goal is to adapt with as little bias as possible to its environment, we are strongly advised to make the system as linear as possible. The simplest linear way to write Figure 2 is in terms of the equations

$$\dot{x}_i(t) = -\alpha_i x_i(t) + C_i(t) \tag{3}$$

with $\alpha_i > 0$, $x_i(0) \geq 0$, and $i = 1, 2, \dots, n$. The input $C_i(t)$ can, for example, have the form

$$C_i(t) = \sum_{k=1}^{N_i} J_i(t - t_i^{(k)}),$$

where $J_i(t)$ is some nonnegative and continuous function that is positive in an interval of the form $(0, \lambda_i)$.

F) After Learning

In order that $M$ be able to predict B given A after practicing AB, interactions between the vertices $v_i$ must exist. Suppose for example that $M$ has already learned AB, and that A is presented to $M$ at time $t_A$. We expect $M$ to respond with B after a short time interval, say at time $t = t_A+\tau_{AB}$, where $\tau_{AB}>0$. $\tau_{AB}$ is called the *reaction time* from A to B. Let us translate these expectations into graphs for the functions $x_A(t)$ and $x_B(t)$. We find Figure 3. The input $C_A(t)$ controlled by $E$ gives rise to the perturbation of $x_A(t)$. The internal mechanism of $M$ must give rise to the perturbation $x_B(t)$. In other words, after AB is learned $x_B(t)$ gets large $\tau_{AB}$ units after $x_A(t)$ gets large.

There exists a linear and continuous way to say this; namely, $v_A$ sends a linear signal to $v_B$ with time lag $\tau_{AB}$. Then (1) with $i = B$ is replaced by

$$\dot{x}_B(t) = -\alpha_B x_B(t) + C_B(t) + \beta_{AB} x_A(t-\tau_{AB}),$$

with $\beta_{AB}$ some positive constant. More generally if $r_i r_j$ has been learned we conclude that

$$\dot{x}_j(t) = -\alpha_j x_j(t) + C_j(t) + \beta_{ij} x_i(t-\tau_{ij}). \tag{4}$$

If $\beta_{ij} = 0$, then the list $r_i r_j$ cannot be learned since a signal cannot pass from $v_i$ to $v_j$.

## G) Directed Paths

The signal $\beta_{ij} x_i(t-\tau_{ij})$ from $v_i$ to $v_j$ in (4) is carried along some pathway at a finite velocity, or else the locality of the dynamics would be violated. Denote this pathway by $e_{ij}$. The pathways $e_{ij}$ and $e_{ji}$ are distinct because the lists $r_i r_j$ and $r_j r_i$ are distinct. To designate the direction of flow in $e_{ij}$, we draw $e_{ij}$ as an arrow from $v_i$ to $v_j$ whose arrowhead $N_{ij}$ touches $v_j$, as in Figure 4.

## H) Before Learning

Before any learning occurs, if A leads only to B, then learning would have already occurred. A must therefore also be able to lead to C, D, or some other letters. Thus the process of learning can be viewed as elimination of the incorrect pathways AC, AD, etc., while the correct pathway AB endures, or is strengthened.

I) Distinguishing Order

How does $M$ know that AB and not AC is being learned? By Figure 3 practicing AB means that $x_A$ and then $x_B$ become large several times. Saying A alone, or B alone, or neither A nor B should yield no learning. This can be mathematically stated most simply as follows. If AB occurs with a time spacing of $w$, then the product $x_A(t-w)x_B(t)$ is large at suitable times $t \cong t_A^{(i)}+w$, $i = 1,2,\ldots,N_A$. We therefore seek a process in $M$ that can compute products of past $x_A(v)$ values $(v<t)$ and present $x_B(t)$ values. Denote this process by $z_{AB}(t)$. Note that $z_{AB} \neq z_{BA}$.

Where in $M$ do past values of $x_A(v)$ and present values of $x_B(t)$ come together, so that $z_{AB}(t)$ can compute them? (Locality again!) By Figure 4, this happens only in the arrowhead $N_{AB}$. Thus $z_{AB}(t)$ takes place in $N_{AB}$. But then the past $x_A(v)$ value received by $N_{AB}$ at time $t$ is the signal $\beta_{AB}x_A(t - \tau_{AB})$. The most linear and continuous way to express this rule for $z_{AB}(t)$ is the following.

$$\dot{z}_{AB}(t) = -\gamma_{AB}z_{AB}(t) + \delta_{AB}x_A(t - \tau_{AB})x_B(t),$$

with $\gamma_{AB}$ a positive constant, and $\delta_{AB}$ a nonnegative constant that is positive only if $\beta_{AB}$ is positive. More generally, for $r_ir_j$ we find in $N_{ij}$ the process

$$\dot{z}_{ij}(t) = -\gamma_{ij}z_{ij}(t) + \delta_{ij}x_i(t - \tau_{ij})x_j(t). \tag{5}$$

J) Gating Outputs

The $z_{ij}(t)$ function can distinguish whether or not $r_ir_j$ is practiced. But more is desired. Namely, if $r_ir_j$ is practiced, presenting $r_i$ should yield a delayed output from $v_j$. If $r_ir_j$ is not practiced, presenting $r_i$ should not yield an output from $v_j$. And even if $r_ir_j$ is practiced, no output from $v_j$ should occur if $r_i$ is not presented. In other words, $x_j(t)$ should become large only if $x_i(t-\tau_{ij})$ *and* $z_{ij}(t)$ are large. Again a product is called for, and (4) is changed to

$$\dot{x}_j(t) = -\alpha_jx_j(t) + C_j(t) + x_i(t-\tau_{ij})\beta_{ij}z_{ij}(t). \tag{6}$$

K) Independence of Lists in First Approximation

If B is not presented to $M$, then in first approximation CA should be learnable without interference from B. (Not so in second approximation, since a signal could travel from C to B to A.) Similarly if C is not presented to $M$, then BA should be learnable without interference from C, in first approximation. Mathematically speaking, this means that all signals to each $v_j$ combine additively at $v_j$. Thus (6) becomes

$$\dot{x}_j(t) = -\alpha_j x_j(t) + C_j(t) + \Sigma_{i=1}^n x_i(t-\tau_{ij})\beta_{ij} z_{ij}(t) \tag{7}$$

The system (5) and (7) is a mathematically well-defined proposal for a learning machine that uses only such general notions as linearity, continuity, and locality, and a mathematical analysis of how a machine can learn to predict B given A on the basis of practicing AB.

L) Thresholds

One further modification of systems (5) and (7) is convenient; namely, the introduction of signal thresholds. Here we introduce this modification directly to keep background noise down. A more fundamental analysis would introduce it by first analysing the need in complex learning situations for inhibitory interactions, and then by pointing out that learning becomes difficult without signal thresholds if inhibitory interactions exist.

A possible difficulty in (5) and (7) is this. Small signals can possibly be carried round-and-round the network thereby building up background noise and interfering with the processing of behaviorally important inputs. We therefore seek to eliminate the production of signals in response to small $x_i(t)$ values, in the most linear possible way. Thresholds do this for us. Letting $[\xi]^+ = \max(\xi, 0)$, we replace (5) and (7) by

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \Sigma_{m=1}^n [x_m(t-\tau_{mi})-\Gamma_{mi}]^+ \beta_{mi} z_{mi}(t) + C_i(t) \tag{8}$$

and

$$\dot{z}_{jk}(t) = -\gamma_{jk} z_{jk}(t) + \delta_{jk}\left[x_j(t-\tau_{jk})-\Gamma_{jk}\right]^+ x_k(t), \tag{9}$$

where all $\Gamma_{jk}$ are positive thresholds, and $i,j,k = 1,2,\ldots, n$. Systems (8) and (9) complete the derivation of this paper.

## Psychophysiological Interpretation

The function $x_i(t)$ is called the $i^{th}$ *stimulus trace:* it responds to the *stimulus* $C_i(t)$. The function $z_{jk}(t)$ is called the $(j,k)^{th}$ *memory trace:* it records the pairing of successive events $r_j$ and $r_k$. Alternatively, $x_i(t)$ is called the $i^{th}$ *short-term memory trace:* it represents brief activation of the state $v_i$ either by inputs $C_i(t)$ or by signals from other states $v_j$. Similarly, $z_{jk}(t)$ is called the $(j,k)^{th}$ *long-term memory trace:* its record of past events can endure long after the short term memory traces have decayed. *Transfer* from short-term memory to long-term memory denotes the operation whereby the $z_{jk}$'s are altered by the distribution of $x_i$'s. *Activation* of short-term memories via long-term memories denotes the operation whereby signals from a given set of $v_j$'s, modulated in the pathways $e_{jk}$ by the $z_{jk}$'s, activate a given pattern of $x_k$'s.

$\Gamma_{jk}$ is the $(j,k)^{th}$ *signal threshold:* no signal is emitted by $v_j$ into $e_{jk}$ at time $t$ unless $x_j(t) > \Gamma_{jk}$. $v_j$ is said to *sample* $v_k$ at time $t$ if the signal from $v_j$ to $N_{jk}$ is positive at time $t$. The *signal strength* at $N_{jk}$ at time $t$ is defined by $B_{jk}(t) = \left[x_j(t-\tau_{jk})-\Gamma_{jk}\right]^+ \beta_{jk}$. The constant $\beta_{jk}$ is a structural parameter called the *path strength* of $e_{jk}$. The $n \times n$ matrix $\beta = \|\beta_{jk}\|$ determines which directed paths between vertices exist, and how strong they are. Otherwise expressed, $\beta$ determines the "anatomy" of connections between all vertices.

A physiological interpretation of these variables in terms of cell bodies $(v_i)$, axons $(e_{jk})$, synaptic knobs $(N_{jk})$, cell potentials $(x_i(t))$, spiking frequencies $(\propto B_{jk}(t))$, and transmitter production rates $(z_{jk}(t))$ can also be noted [1].

Mathematical analysis of system (8) and (9) shows that important properties of learning are preserved in the more general systems (1) and (2). Given the psychophysiological interpretation above, this generalization has an important physical meaning.

A) Short-Term Memory Decay

Consider the replacement of the exponential decay term $-\alpha_i x_i$ in (8) by the general decay term $A_i x_i$ in (1). For example, let $A_i(t) = -\alpha_i + f_i(t)$, where $0 \leq f_i(t) < \alpha_i$. $f(t)$ represents a "shunt" of cell potential at $v_i$ that can amplify the effects of the input $C_i(t)$ at prescribed times when $f_i(t)$ is large, as the equation $\dot{x}_i = -(\alpha_i - f_i(t)) \cdot x_i + C_i$ illustrates. These "Now Print" times will presumably occur when information important to the survival of the network is delivered; e.g., the presentation of food in a prescribed situation. The inputs $f_i(t)$ are often controlled by "nonspecific arousal" cells that are activated by biologically important events. See Figure 5a. An alternative shunt can act directly on the synaptic knobs delivering $C_i(t)$ to $v_i$, thereby replacing $C_i(t)$ by $f_i(t)C_i(t)$, as in the equation $\dot{x}_i = -\alpha_i x_i + f_i(t)C_i(t)$. See Figure 5b. An additive Now Print mechanism is for some purposes more useful than the shunting mechanism ([4], [9], [10]).

The functional $A_i(t)$ also permits noisy variations in decay parameters, as well as feedback from system variables, as in

$$A_i(t) = -\alpha \;\; -\sum_{k\in I}\int_0^t \left\{\left[x_k(v-\tau_k)-\Gamma_k\right]^+\right\} g(v)dv$$

B) Long-Term Memory Decay

The exponential decay term $-\gamma_{jk} z_{jk}$ i
wide variety of alternatives. For example,
$\Gamma \;]^+$

[ ]

$v_k$ $v_j$

Long-term memory decay can also be altered indirectly by changing the functional $B_{jk}$. For example, the choice

(10) $$B_{jk}(t) = \beta_{jk}\left[x_j(t-\tau_j)-\Gamma_j\right]^+\left(\sum_{m\in I} z_{jm}(t)\right)^{-1}$$

can have the effect of making memory perfect until new items are practiced [13]. In particular, memory can be perfect during performance trials. This is true, however, only in certain anatomies; for example, the "outstar" anatomy of Figure 6a. There exist other anatomies, such as those depicted in Figures 6b and 6c, in which perfect memory is replaced by "phase transitions in memory" or "imprinting" of memories ([14], [15]). That is, for certain choices of numerical parameters, memory of a given class of patterns decays; for the remaining parameter choices, memory is rigid. The particular class of patterns for which memory decays depends on the anatomy of the network ([1], p. 35). Prescribed alterations in network parameters can "imprint" an ongoing input pattern by transforming the system from its plastic phase to its rigid phase.

C) Signal Strength

The signal terms $\beta_{jk}[x_j(t-\tau_{jk})-\Gamma_{jk}]^+$ and $\delta_{jk}[x_j(t-\tau_{jk})-\Gamma_{jk}]^+$ in (8) and (9), respectively, can be replaced, say, by

$$B_{jk}(t) = \beta_{jk}(t)\left[x_j(t-\tau_{jk}(t))-\Gamma_{jk}(t)\right]^+$$

$$E_{jk}(t) = \delta_{jk}(t)\left[x_j(t-\sigma_{jk}(t))-\Omega_{jk}(t)\right]^+$$

which permit different, and variable, time lags, thresholds, and path strengths in the two signal strength functionals. This includes the possibility of coupling a Now Print mechanism to these functionals. Functional $E_{jk}(t)$ describes the effect of the signal from $v_j$ on the cross-correlational process within $N_{jk}$ that determines $z_{jk}(t)$. Functional $B_{jk}(t)$ describes the net signal from $v_j$ that ultimately influences $v_k$ after being processed in $N_{jk}$. It is therefore natural to physically expect that $\Gamma_{jk} \geq \Omega_{jk}$. This *local flow condition* says little more than that the signal from $v_j$ passes through $N_{jk}$ on its way to $v_k$. Such a condition is, in fact, needed to guarantee that many cells can simultaneously sample a given pattern without creating asymptotic biases in their memory. The local flow condition provides examples of systems which can learn patterns without performing them until later, but which cannot perform old patterns without also learning new patterns that are imposed during performance.

The functionals $B_{jk}$ and $E_{jk}$ permit more complicated possibilities as well. For example, *in vivo,* after a signal is generated in $e_{jk}$, it is impossible to generate another signal for a short time afterwards (absolute refractory period) and harder to generate another signal for a short time after the absolute refractory period (relative refractory period). Also, some cells emit signals in complicated bursts. All such continuous variations are, in principle, covered by our theorems, which say, that whereas such variations can influence transient motions of the system, the classification of limits and oscillatory possibilities is unchanged by them.

It is physically interesting that those terms, such as $B_{jk}$ and $E_{jk}$, which describe processes that act over a distance (such as signals flowing along $e_{jk}$) are the terms in equations (1) and (2) that permit the most nonlinear distortion without destroying learning properties. The term $x_i$ in (2) is not of this type. This term is computed in $N_{ji}$ from the value $x_i(t)$ in the contiguous vertex $v_i$.

## Local Symmetry Axes

In its final form, the theorem shows that unbiased pattern learning can occur in systems with arbitrary positive path weights $\beta_{ji}$ from $j \in J$ to $i \in I$. This is achieved by first restricting attention to systems of the form

$$\dot{x}_i = Ax_i + \sum_{k \in J} B_k z_{ki} + C_i(t) \tag{11}$$

and

$$\dot{z}_{ji} = D_j z_{ji} + E_j x_i , \tag{12}$$

where $i \in I$ and $j \in J$. That is, all functionals $A_i$, $B_{ji}$, $D_{ji}$, and $E_{ji}$ are chosen independent of $i \in I$, and the anatomy is constrained to make this possible. These constraints mean that all cells $\mathcal{B} = \{v_i : i \in I\}$ are sampled by a given cell $v_j$ in $\mathcal{A} = \{v_j : j \in J\}$ without biases due to system parameters ($B_{ji} = B_j$, $D_{ji} = D_j$, $E_{ji} = E_j$), and that the inputs to all cells $\mathcal{B}$ are averaged by their cell potentials without biases due to averaging rates ($A_i = A$). See Figure 7a.

Systems (11) and (12) allow each cell to have a different time lag, threshold, and axon weight, as in $B_j(t) = \beta_j\left[x_j(t-\tau_j)-\Gamma_j\right]^+$. Even if all cells interact, as in Figure 7b, no biases in asymptotic learning need occur due to these asymmetries in signal transfer among possibly billions of cells.

Figure 7b and 7c illustrate two extremal anatomies, the completely recurrent $(I=J)$ and completely nonrecurrent $(I\cap J = \phi)$ cases. Generalizations of Figure 7a are also possible. In these generalizations, $A$ and $B$ are replaced by sets $\{A_k\}$ and $\{B_k\}$ of subsets such that each cell in a given $B_k$ is sampled by all cells in $A_k$. One seeks the maximal subsets $B_k$ for which this decomposition exists. For some purposes, a fixed set $\{B_k\}$ is predetermined by physical considerations; e.g., each $B_k$ controls a different motor effector. It is then sometimes profitable to introduce fictitious cells into the sampling cells $A$ if some cells in $A$ sample two or more subsets $B_k$. For example, if cell $v_i$ in $A$ samples $B_1$ and $B_2$, replace $v_i$ by two cells $v_{i1}$ and $v_{i2}$ such that $v_{ij}$ samples only $B_j$, $j = 1,2$, and each $v_{ij}$ receives the same inputs, and has the same parameters and initial data, as the original cell $v_i$ had.

## Mathematical Results

The theorems discuss learning by any number of cells in $A$ of a spatial pattern presented to any number of cells in $B$. Learning is by respondent conditioning. The CS's are delivered to $A$. The UCS is the spatial pattern received by $B$. Generalization to learning an arbitrary space-time pattern is readily accomplished [12].

A *spatial pattern* delivered to $B$ is a vector function of the form $C_i(t) = \theta_i C(t)$, $i\in I$. The relative pattern weights $\theta_i\left(\theta_i \geq 0, \sum_{k\in I}\theta_k = 1\right)$ characterize the pattern. The total intensity $C(t)$ can fluctuate wildly in time. This definition of spatial pattern notes, for example, that recognition of a picture is invariant under wide fluctuations of background illumination. The inputs $C_i(t)$ need not, however, represent a pattern created by external events at peripheral network receptor sites. Any spatial pattern playing on any finite collection of cells (e.g., cells controlling motor outputs, or centrally located sensory cells of any type) can be learned and performed in the manner described by the following theorems. References [2] and [3] discuss other applications of this definition, especially to problems of pattern discrimination.

Since the relative intensity $\theta_i$ at each $v_i$ characterizes the pattern, we study the limiting and oscillatory behavior of "pattern variables"; namely, the relative stimulus traces (potentials) $X_i = x_i(\Sigma_{k\in I}x_k)^{-1}$ and the relative memory traces (transmitters) $Z_{ji} = z_{ji}(\Sigma_{k\in I}z_{jk})^{-1}$. Once behavior of these variables is established for general functionals, analysis of the "total energy" variables $x = \Sigma_{k\in I}\, x_k$ and $z_j = \Sigma_{k\in I}\, z_{jk}$ can be carried out for particular choices of functionals. Then behavior of $x_i$ and $z_{ji}$ is also known. See, for example, [8].

Theorem 1 studies what happens when *all* cells in $A$ receive their CS's "sufficiently often", given also that the UCS is presented sufficiently often at times when the CS's can practice it. This theorem is not always realistic. A given CS need not activate all the cells $A$ that are capable of sampling $B$. Theorem 2 studies the case in which an arbitrary subset of $A$ is activated "sufficiently often" by CS's. The local flow condition is needed in this situation. Proposition 1 shows what can go wrong if the local flow condition is not imposed. Theorem 1 will be stated without proof, since the same method is used as for the more difficult Theorem 2.

Theorem 1 will be expressed in terms of the function $f(S,T) = \int_S^T C\exp\left(\int_t^T A dv\right)dt$; the functions $M(i)\colon\ [0,\infty) \to J$ such that $Z_{[M(i)](t),i}(t) = \max\{Z_{ji}(t)\colon j\in J\}$ and $m(i)\colon [0,\infty) \to J$ such that $Z_{[m(i)](t),i}(t) = \min\{Z_{ji}(t); j\in J\}$; and the functional $L$ defined for every piecewise constant function $m\colon[0,\infty) \to J$ and every continuous $g$ on $[S,T)$ by

$$L(m,g;S,T] = \int_S^T E_m g\exp\left[-\int_S^t D_m dv\right]dt.$$

Theorem 1. *Suppose that*

(i) *the system is bounded;*

(ii) *each* CS *is presented sufficiently often; that is, for every* $j \in J$,

$$L[j,x;\ 0,\infty] = \infty \quad \textit{(also necessary)};$$

(iii) *the* UCS *is presented sufficiently often; that is,*

$$\int_0^\infty Cx^{-1}\,dt = \infty \quad \textit{(also necessary)};$$

*and*

(iv) *each* CS *and the* UCS *are practiced together sufficiently often; that is for some* $\epsilon > 0$ *and each* $i \in I$, *there exist increasing divergent sequences* $\{S_{in}\}$ *and* $\{T_{in}\}$ *such that*

$$\sum_{n=1}^{\infty} \frac{L[M(i),f(S_{in},\cdot);\ S_{in},S_{i,n+1}]}{\epsilon + L[M(i),x;\ S_{in},S_{i,n+1}]}$$

*and*

(16) $$\sum_{n=1}^{\infty} \frac{L[m(i),f(T_{in},\cdot);\ T_{in},T_{i,n+1}]}{\epsilon + L[m(i),x;\ T_{in},T_{i,n+1}]} = \infty$$

*Then the perfect pattern learning occurs; that is, all the limits* $Q_i = \lim_{t\to\infty} X_i(t)$ *and* $P_{ji} = \lim_{t\to\infty} Z_{ji}(t)$ *exist globally and*

(17) $$P_{ji} = Q_i = \theta_i .$$

Condition (i) can be removed, but leads to a physically implausible situation. Then the $n^{th}$ appearances of $\epsilon$ in (iv) are replaced by $z_{[M(i)](S_{in})}(S_{in})$ and $z_{[m(i)](T_{in})}(T_{in})$, respectively. A counterexample can be constructed if (iv) is violated.

Equation (17) means that the pattern is perfectly learned in the sense that a test input to *any* subset of cells $v_j$ after learning occurs can reproduce the pattern at all cells $v_i$. The act of recall can, however, destroy the memory in special cases [14].

Corollary 1. *Conditions* (ii) - (iv) *are implied by the following conditions:* (i),

(v) *for every* $j \in J$,

$$\int_0^{\infty} E_i dt =$$

(vi) *the* UCS *energy is presented, on the average, with a uniform lower bound; that is, there exist positive constants* $K_1$ *and* $K_2$ *such that for every* $T \geq 0$,

$$f(T,T+t) \geq K_1 \quad \text{if} \quad t \geq K_2$$

Theorem 2 uses the following functions. Let $N(i): [0,\infty) \to J(1)$ be defined by $Z_{[N(i)](t),i}(t) = \max \{Z_{ji}(t): j \in J(1)\}$, and $n(i): [0,\infty) \to J(1)$ be defined by $Z_{[n(i)](t),i}(t) = \min \{Z_{ji}(t): j \in J(1)\}$, where $J(1) = \{j \in J: \int_0^{\infty} B_j z_j x^{-1} dt = \infty\}$.

Theorem 2. *Again suppose that the system is bounded, the* UCS *is presented sufficiently often , and*

(vii) *those* CS's *which are performed continually are also practiced with the* UCS *sufficiently often; that is, if* $J(1) \neq \phi$, *then condition* (iv) *holds with* $M(i)$ *and* $m(i)$ *replaced by* $N(i)$ *and* $n(i)$.

*Then the potentials pick up the pattern weights and all transmitters learn the pattern at least partially; that is, all the limits* $Q_i$ *and* $P_{ji}$ *exist with* $Q_i = \theta_i$. *If, moreover, a* CS *is practiced with the* UCS *sufficiently often, then it learns the pattern perfectly; that is, if (13) holds for some* $j \in J$, *then* $P_{ji} = \theta_i$.

The analog of Corollary 1, including a suitable version of the local flow condition, is given by Corollary 2.

Corollary 2. *Conditions* (iii) *and* (vii) *are implied by conditions* (i), (vi), *and*

(viii) *a local flow condition holds; that is, for every* $j \in j$, *either*

$$\int_0^{\infty} B_j z_j x^{-1} dt = \infty \quad \textit{only if} \int_0^{\infty} E_j x \exp\left(-\int_0^t D_j dv\right) dt = \infty$$

*or*

(21 $$\int_0^\infty B_j dt = \infty \text{ only if } \int_0 E_j dt = \infty$$

*Under these circumstances, if either* $L[j,x;0,\infty] = \infty$ *or* $\int_0^\infty E_j dt = \infty$ *then* $P_{ji} = \theta_i$.

Suppose for example that $B_j(t) = \beta_j[x_j(t-\tau_j)-\Gamma_j]^+$ and $E_j(t) = \delta_j[x_j(t-\sigma_j)-\Omega_j]^+$. Then condition (21) is satisfied if $\Gamma_j \geqslant \Omega_j$. In applications, (21) is a constraint on the parameters of the system, rather than on its trajectories.

Condition (15) has the following intuitive meaning. The $n^{th}$ summand in (15) considers how much total input $C(t)$ reaches $M$ during the time interval $[S_{in}, S_{i,n+1})$. The function $f(S_{in},\cdot)$ describes the effect of averaging by functional $A$ on $C$ to yield $C$'s contribution to the total potential $x$. For fixed $j\in J$, the functional $L[j, f(S_{in},\cdot); S_{in}, S_{i,n+1}]$ describes the effect of averaging $D_j$ and $E_j$ on $f(S_{in},\cdot)$ to yield $C$'s contribution to the $j^{th}$ total transmitter $z_j$. $L[M(i),f(S_{in},\cdot);S_{in},S_{i,n+1}]$ measures the effect of $C$ on the cell $v_j$ whose relative memory trace $Z_{ji}$ has been least attracted downwards towards $\theta_i$, whenever this case occurs.

The term $L[M(i), x; S_{in}, S_{i,n+1}]$ in (15) has a similar interpretation, except that $x(t)$ replaces $f(S_{in},t)$ to express the total effect of potential on transmitter. $x(t)$ differs from $f(S_{in},t)$ due to the interaction term $\Sigma\{B_j z_{ji}:i\in I,j\in J\}$ in (11), which is also averaged by $A$ to yield a contribution to $x$. These terms tend to preserve the old patterns that are already in $M$'s memory. (See Proposition 2.) Condition (15) therefore says the following. For each $i\in I$, there exists some sequence of time intervals $[S_{in},S_{i,n+1})$, such that enough input energy $C(t)$ is presented in each $[S_{in},S_{i,n+1})$ to guarantee that, after averaging by potentials and transmitters, this energy suffices to overcome the stabilizing effect of interaction terms and thereby drive all the $Z_{ji}$'s towards the limits $\theta_i$ imposed by the new pattern.

Proposition 1 below notes that the local flow condition is not superfluous in a case of some physical interest.

Proposition 1. *Suppose* (viii) *does not hold. Partition* J *into subsets* J(2) *and* J(3) *such that*

$$J(2) = \{j: \int_0^\infty B_j dt = \infty \text{ and } \int_0^\infty E_j dt < \infty\} \neq \phi \tag{22}$$

*Suppose that the system is bounded, that* (vi) *holds, that*

(ix) *there is perfect memory until recall in* J(2)*; that is,* $D_j \geq -\gamma_j E_j$ *for some constant* $\gamma_j > 0$, $j \in J(2)$*; and that*

(x) *average performance energy in* J(2) *does not converge to zero; that is, for every* $T \geq 0$,

$$\limsup_{t\to\infty} \sum_{k\in J(2)} \int_T^t B_k \exp\left[\int_v^t A d\xi\right] dv > 0 \tag{23}$$

*Then given initial data such that* $\max(Z_{ji}(0): j\in J) \geq X_i(0) > \theta_i$ *and* $\min(Z_{ji}(0): j\in J) > \theta_i$, *even if* $Q_i$ *exists,* $Q_i \neq \theta_i$, *so that even if* $P_{ji}$ *exists and* $\int_0^\infty E_j dt = \infty$, $P_{ji} \neq \theta_i$.

Theorem 1 will be proved below. The first step of the proof is to transform (11) and (12) into equations in the pattern variables $X_i = x_i x^{-1}$ and $Z_{ji} = z_{ji} z_j^{-1}$, where $x = \Sigma_{k\in I} x_k$ and $z_j = \Sigma_{k\in I} z_{jk}$.

Lemma 1. *Suppose* $x(0) > 0$ *and* $z_j(0) > 0$, $j\in J$. *Then*

$$\dot{X}_i = \Sigma_{k\in J} F_k (Z_{ki} - X_i) + G(\theta_i - X_i \tag{24}$$

*and*

$$\dot{Z}_{ji} = H_j (X_i - Z_{ji}), \tag{25}$$

*where*

$$F_j = B_j z_j x^{-1} \tag{26}$$

$$G = Cx^{-1} \tag{27}$$

*and*

$$H_j = E_j x z_j^{-1} \tag{28}$$

The proof uses the standard equation

$$\left(\frac{f}{g}\right) = \frac{1}{g}\left(\dot{f}\right.$$

along with the equations

(29) $$\dot{x} = Ax + \Sigma_{k\in J}B_k z_k + C$$

and

(30) $$\dot{z}_j = D_j z_j + E_j x.$$

Equations (23) and (24) are then readily transformed into equations for the difference variables $X_i^{(\theta)} = X_i - \theta_i$ and $Z_{ji}^{(\theta)} = Z_{ji} - \theta_i$.

Lemma 2. *Given any spatial pattern, then*

(31) $$\dot{X}_i^{(\theta)} = \Sigma_{k\in J}F_k(Z_{ki}^{(\theta)} - X_i^{(\theta)} - GX_i^{(\theta)}$$

*and*

(32) $$\dot{Z}_{ji}^{(\theta} = H_j(X_i^{(\theta)} - Z_{ji}^{(\theta)})$$

Lemmas 1 and 2 provide information concerning the oscillations of pattern variables. In order to conveniently describe these facts, we introduce the following terminology.

Definition 1. If a system of inequalities, valid at some time $t = T$, is therefore valid at all times $t \geqslant T$, we say that the inequalities *propagate* in time.

The oscillations depend on whether or not the UCS is being presented. Hence let $U_0$ be the union of all intervals during which $C = 0$ and let $U_1$ be the union of all intervals during which $C > 0$. In both cases, $X_i$'s motion is compared with that of $Y_i = \max(Z_{ji}: j\in J)$ and $y_i = \min(Z_{ji} : j\in J)$; that is, with the "envelope" of all relative memory traces facing $v_i$. We will see that $X_i$ is a kind of nonlinear "center of mass" of these variables, and therefore attracts and is attracted towards the values in the interval $[y_i, Y_i]$. The influence of $C$ changes the configuration to which $X_i$ is attracted by also attracting $X_i$ to $\theta_i$. For $t\in U_1$, we will therefore compare $X_i^{(\theta)}$ with $Y_i^{(\theta)} = Y_i - \theta_i$ and $y_i^{(\theta)} = y_i - \theta_i$.

Proposition 2. *Given any* $t$ *in an interval of* $U_0$, *the following* *cases are exhaustive.*

I) *The inequalities* $y_i \leq X_i \leq Y_i$ *propagate with* $y_i$ *monotone increasing and* $Y_i$ *monotone decreasing. Thus the limits* $y_i(\infty)$ *and* $Y_i(\infty)$ *exist.*

II) *The inequality* $X_i > Y_i$ *either propagates with* $X_i$ *monotone decreasing and all* $Z_{ji}$ *monotone increasing, or switches into Case (I). Thus,* $Y_i$ *oscillates at most once, and* $y_i$ *is always monotone increasing. Hence either all limits exist, or* $y_i(\infty)$ *and* $Y_i(\infty)$ *exist.*

III) *If* $X_i < y_i$ *holds, then the conclusions of Case (II) hold, with* $y_i$ *replacing* $Y_i$ *and all inequalities reversed.*

These results follow by inspection of (24) and (25) given that $G = 0$ and $F_j$ and $H_j$ are nonnegative. In short, either $X_i$ is attracted to the interval $[y_i, Y_i]$ as it attracts all $Z_{ji}$, or $X_i$ is trapped between $y_i$ and $Y_i$ as they are drawn together. If, therefore, as a result of prior practice $y_i \cong X_i \cong Y_i \cong \theta_j$, then these approximations propagate, yielding perfect memory of pattern weights.

Proposition 3. *Given any* $t$ *in an interval of* $U_1$, *the following cases are exhaustive*

*The inequalities* $X_i^{(\theta)} \geq 0$ *and* $y_i^{(\theta)} \geq 0$ *propagate. If moreover*

(IVa) $X_i^{(\theta)} \leq Y_i^{(\theta)}$, *then the inequality propagates with* $Y_i^{(\theta)}$ *decreasing. Hence* $Y_i(\infty)$ *exists. If however*

(IVb) $X_i^{(\theta)} > Y_i^{(\theta)}$, *then the inequality propagates as* $X_i^{(\theta)}$ *decreases and all* $Z_{ji}^{(\theta)}$ *increase until (if ever) Case (IVa) is entered. Hence either all limits exist or Case (IVa) is entered.*

V) *The inequalities* $X_i^{(\theta)} \leq 0$ *and* $Y_i^{(\theta)} \leq 0$ *also propagate. The conclusions of Case (IV) hold with* $y_i$ *and* $Y_i$ *interchanged, and all inequalities reversed.*

VI) *The inequalities* $Y_i^{(\theta)} \geq 0 \geq y_i^{(\theta)}$ *either propagate, or Cases (IV) or (V) are entered. If moreover,*

(VIa) $Y_i^{(\theta)} \geq X_i^{(\theta)} \geq y_i^{(\theta)}$, *then this inequality propagates with* $y_i^{(\theta)}$ *monotone increasing and* $Y_i^{(\theta)}$ *monotone decreasing. Both limits* $y_i(\infty)$ *and* $Y_i(\infty)$ *thus exist. If however*

(VIb) $X_i^{(\theta)} \notin [y_i^{(\theta)}, y_i^{(\theta)}]$, *then* $X_i^{(\theta)}$ *and all* $Z_{ji}^{(\theta)}$ *are monotonic until Case (VIa) is entered. If Case (VIa) is never entered, then all limits exist.*

*In short,* $X_i$ *is attracted towards the interval* $[y_i, Y_i]$, *but also towards* $\theta_i$. *If both attractive forces steer* $X_i$ *in the same direction, then the inequalities propagate.*

*Proof of Theorem 2:* First we consider the subcases of (IV)-(VI) in which all $X_i$ and $Z_{ji}$ are monotonic for large t. Suppose for example that Case (IVb) holds. Let this Case hold at all $t \geq 0$ for convenience. Then by (31), $\dot{X}_i^{(\theta)} \leq -GX_i^{(\theta)}$, and $X_i$ decreases to the limit $Q_i \geq \theta_i$. Thus $\dot{X}_i^{(\theta)} \leq G(\theta_i - Q_i)$, or in integral form,

$$0 \leq X_i(t) \leq X_i(0) + (\theta_i - Q_i)\int_0^t G(v)dv \leq 1 + (\theta_i - Q_i)\int_0^t G(v)dv,$$

for all $t \geq 0$. In particular, if $Q_i > \theta_i$, then

$$\int_0^\infty G(v)dv \leq (Q_i - \theta_i)^{-1} < \infty$$

which contradicts (14). Thus $Q_i = \theta_i$. A similar procedure works in the other monotonic subcases.

To show that all $P_{ji}$ exist in these cases, we prove the following:

Lemma 3. *If* $Q_i$ *exists, then all* $P_{ji}$ *exist.*

*Proof:* For every $\epsilon > 0$, there exists a $T_\epsilon$ such that $t \geq T$ implies $X_i(t) \in [Q_i - \epsilon, Q_i + \epsilon]$. If $Z_{ji}(t) \notin [Q_i - \epsilon, Q_i + \epsilon]$ for some $t \geq T_\epsilon$, then $Z_{ji}(t)$ is monotonically attracted towards the interval $[Q_i - \epsilon, Q_i + \epsilon]$ until it enters it, by (25). If for some $\epsilon > 0$, $Z_{ji}(t)$ never enters $[Q_i - \epsilon, Q_i + \epsilon]$, then $Z_{ji}(t)$ is monotonic for $t \geq T_\epsilon$, and hence $P_{ji}$ exists. If for every $\epsilon > 0$, $Z_{ji}(t) \in [Q_i - \epsilon, Q_i + \epsilon]$ for all sufficiently large t, then $P_{ji} = Q_i$.

It remains in Case (IVb) to show that $P_{ji} = \theta_i$ if (13) holds. In this Case, also $Y_i^{(\theta)}(t) \leq 0$ for $t \gg 0$. Otherwise, there will exist a T such that $Y_i(T) > \theta_i$, and since $Y_i(t)$ is monotone increasing, $X_i(t) \geq Y_i(t) \geq Y_i(T) > \theta_i$ for $t \geq T$. In particular, $Q_i > \theta_i$, which is impossible.

We can therefore restrict attention to the subcase in which $X_i^{(\theta)}(t) \geq 0 \geq Y_i^{(\theta)}(t)$ for $t \gg 0$. We will assume that $\theta_i > P_{ji}$ to derive a contradiction. Since $X_i$ is monotone decreasing, and all $Z_{ji}$ are monotone increasing, (25) implies

$$\dot{Z}_{ji} \geq H_j(\theta_i - P_{ji}),$$

and thus for $t \geq T$ and $T$ sufficiently large,

$$1 \geq Z_{ji}(t) - Z_{ji}(T) \geq (\theta_i - P_{ji}) \int_T^t H_j dv$$

Consequently $\int_0^\infty H_j dv < \infty$

By (28) and (30),

$$H_j = \frac{d}{dt} \log [z_j(0) + \int_0^t E_j x e^{-\int_0^v D_j d\xi} dv].$$

Hence for every $t$

$$\int_0^t H_j dv = \log \{1 + z_j^{-1}(0) E[j, x; o, t]\}.$$

Letting $t \to \infty$ shows that $E[j, x; o, \infty] < \infty$, which contradicts (13).

Now we consider the nonmonotonic subcases of Proposition 3; namely

VII) $Y_i^{(\theta)} \geq X_i^{(\theta)} \geq 0$ *and* $y_i^{(\theta)} \geq 0$ *with* $Y_i^{(\theta)}$ *monotone decreasing for large* $t$;

VIII) *the reverse inequalities with* $Y_i^{(\theta)}$ *and* $y_i^{(\theta)}$ *interchanged; and*

IX) $X_i^{(\theta)} \in [y_i^{(\theta)}, Y_i^{(\theta)}]$ *and* $y_i^{(\theta)} \leq 0 \leq Y_i^{(\theta)}$ *with* $y_i^{(\theta)}$ *increasing and* $Y_i^{(\theta)}$ *decreasing at large values of* $t$.

Only Case VII will be explicitly considered, since Cases VIII and IX can be treated by an analogous method. First, we treat the subcase in which

$$\sum_{k \in J} \int_0^\infty F_k dt < \infty,$$

noting by (26) the relevance of $F_j$ to (20). Then, for every $\epsilon > 0$, there exists a $T_\epsilon$ such that $t \geq T_\epsilon$ implies

$$\sum_{k \in J} \int^{\infty} F_k dv \leq \epsilon/2 \tag{33}$$

By (24),

$$\dot{X}_i^{(\theta)} \leq \sum_{k \in J} F_k - G X_i^{(\theta)}$$

and thus for $t \geq T_\epsilon$,

$$0 \leq X_i^{(\theta)}(t) \leq X_i^{(\theta)}(T_\epsilon) \exp\left(-\int_{T_\epsilon}^{t} G dv\right) + \sum_{k \in J} \int_{T_\epsilon}^{t} F_k \exp\left(-\int_{v}^{t} G d\xi\right) dv,$$

which by (33) implies

$$0 \leq X_i^{(\theta)}(t) \leq \exp\left(-\int_{T_\epsilon}^{t} G dv\right) + (\epsilon/2).$$

Let $t \to \infty$ and apply (14) to conclude that $Q_i = \theta_i$. Lemma 3 now implies that all $P_{ji}$ exist.

It remains only to consider Case VII - and Cases VIII and IX analogously - if

$$\sum_{k \in J} \int_0^{\infty} F_k dv = \infty$$

Partition $J$ into two sets $J(1)$ and $J(2)$ such that $j \in J(1)$ iff

$$\int_0^{\infty} F_j dv = \infty$$

By (35) and the nonnegativity of $F_j$, $J(1) \neq \phi$, and we can define the function

$$\tilde{Y}_i^{(\theta)} = \max\{y_{ji}^{(\theta)} : j \in J(1)\}.$$

We now show that $\tilde{Y}_i^{(\theta)}(\infty) \equiv \lim_{t \to \infty} \tilde{Y}_i^{(\theta)}(t)$ exists in Case VII. To do this note by (32) that

$$\operatorname{sign}\left[\frac{d}{dt} \tilde{Y}_i^{(\theta)}(t)\right] = \operatorname{sign}\left[X_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(t)\right]$$

whenever $(d/dt)\ \tilde{Y}_i^{(\theta)}(t) \neq 0$. Thus, if there exists a $T$ such that $X_i^{(\theta)}(t) \leq \tilde{Y}_i^{(\theta)}(t)$ for all $t \geq T$, then $\tilde{Y}_i(\infty)$ exists. There cannot exist a $T$ such that $\tilde{Y}_i^{(\theta)}(t) \leq X_i^{(\theta)}(t)$ for all $t \geq T$, as we now show.

Define the functions $L^{(i)} = \sum_{k \in J(i)} F_k$, $i = 1,2$. Then by (31)

$$\dot{X}_i^{(\theta)} \leq L^{(2)} + (\tilde{Y}_i^{(\theta)} - X_i^{(\theta)})L^{(1)} - GX_i^{(\theta)}. \tag{38}$$

If $\tilde{Y}_i^{(\theta)} \leq X_i^{(\theta)}$, for $t \geq T$, then (38) implies

$$\leq L^{(2)} - GX_i^{(\theta)}$$

for $t \geq T$, and since by (37) $\tilde{Y}_i^{(\theta)}(t)$ is nondecreasing for $t \geq T$,

$$\dot{X}_i^{(\theta)} \leq L^{(2)} - \tilde{Y}_i^{(\theta)}(T)G \quad \text{for } t \geq T,$$

where we can assume that $\tilde{Y}_i^{(\theta)}(T) > 0$ without loss of generality. Integrating (38) between $T$ and $\infty$ readily yields the inequality $\int_0^\infty G(v)dv < \infty$, which contradicts (14).

Thus, either $\tilde{Y}_i^{(\theta)}(t) \geq X_i^{(\theta)}(t)$ for all $t \geq T$ and some $T$ sufficiently large, or $\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t)$ changes sign at arbitrarily large times. We use this fact to prove that, for all $t \geq T$ and $T$ sufficiently large, the inequality

$$\max\,[X_i^{(\theta)}(t),\ \tilde{Y}_i^{(\theta)}(t)] \leq \tilde{Y}_i^{(\theta)}(T) + U(T, t) \tag{39}$$

holds, where

$$U(T, t) = \int_T^t L^{(2)}(v)dv. \tag{40}$$

(39) can be used to show that $\tilde{Y}_i^{(\theta)}(\infty)$ exists in all cases. Since by definition of $J(2)$, $\int_0^\infty L^{(2)}(v)dv < \infty$, the function $U(T) \equiv \int_T^\infty L^{(2)}(v)dv$ monotonically decreases to zero as $T \to \infty$. Thus by (39) and (40), the bounded and continuous function $Y_i^{(\theta)}(t)$ is alternately monotone decreasing or increasing by an amount that approaches zero as $t \to \infty$. Hence $\tilde{Y}^{(\theta)}(\infty)$ exists.

Inequality (39) is proved as follows. It is trivial if $\tilde{Y}_i^{(\theta)}(t)$

$\geq X_i^{(\theta)}(t)$ for all $t \geq T$ and some $T$ sufficiently large. Suppose that $X_i^{(\theta)}(t) \geq \tilde{Y}_i^{(\theta)}(t)$ only in the disjoint sequence of intervals $[S_{ik}, T_{ik}]$, $k = 1,2,\ldots$ . Then $\tilde{Y}_i^{(\theta)}(S_{ik}) = X_i^{(\theta)}(S_{ik})$ for all $k = 1,2,\ldots,$ and by (38)

$$\dot{X}_i^{(\theta)}(t) \leq L^{(2)}(t), \quad t \in [S_{ik}, T_{ik}],$$

which implies

$$\tilde{Y}_i^{(\theta)}(t) \leq X_i^{(\theta)}(t) \leq \tilde{Y}_i^{(\theta)}(S_{ik}) + U(S_{ik}, t) \tag{41}$$

for $t \in [S_{ik}, T_{ik}]$. Since $X_i^{(\theta)}(t) \leq \tilde{Y}_i^{(\theta)}(t)$ with $\tilde{Y}_i^{(\theta)}(t)$ monotone nonincreasing for $t \notin \cup_{k=1}^{\infty}[S_{ik}, T_{ik}]$, (39) readily follows by pasting these cases together.

Inequalities (38) and (39) along with the existence of $\tilde{Y}_i^{(\theta)}(\infty)$ will now be used to prove that $\tilde{Y}_i(\infty) = \theta_i$, and thus by (39) that $Q_i = \theta_i$. Then the existence of all $P_{ji}$ follows by Lemma 3, and the proof is readily completed. Define the functions $M = (d/dt)\log x - A - G$ and $N = (d/dt)\log x\text{-}A$. By (26) and (29), $L^{(1)} + L^{(2)} = \sum_{k\in J} F_k = M$. (38) therefore implies

$$\dot{X}_i^{(\theta)} \leq 2L^{(2)} + M\tilde{Y}_i^{(\theta)} - NX_i^{(\theta)} \tag{42}$$

which can be expressed in integral form for any $t \geq T \geq 0$ as

$$X_i^{(\theta)}(t) \leq P_i(T, t) + Q_i(T, t) + R(T, t)$$

using the notation

$$P_i(T, t) = X_i^{(\theta)}(T)x(T)x^{-1}(t)\exp\left(\int_T^t A dv\right), \tag{44}$$

$$Q_i(T, t) = x^{-1}(t)\int_T^t \tilde{Y}_i^{(\theta)}(v)M(v)x(v)\exp\left(\int_v^t A d\xi\right) dv,$$

and

$$R(T,t) = 2x^{-1}(t)\int_T^t L^{(2)}(v)x(v)\exp\left(\int_v^t A d\xi\right) dv.$$

We now estimate each of these terms from above. First note that

This llo from which shows that Ax. and thus the

exp $\int$ Adξ

used est ding

(θ U

$\int$ Mx exp $\int$ Adξ

Not the

exp Adξ ·Ax· exp $\int$ Adξ

exp $\int$ Ad $\int$ exp $\int$ Adξ dv

al ed to estimate y ling

(θ

$\int$

Combining and yi the bas inequal ty

(θ (θ T

$\int$ exp $\int$ Adξ

Inequal ty is the bas al. llow in the

(∞) Inequality ll be app ed to

d
t

where

$$N_i(t) = H_{[N(i)](t)}(t).$$

Interpret the derivative in (52) as a left-handed derivative at times when $N(i)$ changes value. Letting $Y_i(T,t) = \tilde{Y}_i(T) - \tilde{Y}_i(t)$, we find by (50) and (52) that

$$\frac{\partial}{\partial t} \tilde{Y}_i(T,t) = N_i(\tilde{Y}_i - X_i)$$

$$\geq -N_i Y_i(T,t) + \tilde{Y}_i^{(\theta)}(T)N_i P(T,t) + N_i U(T,t)P(T,t)$$

Integrate this inequality from $t = T$ to any $t > T$. Then

$$Y_i(T,t) \geq \tilde{Y}_i^{(\theta)}(T)R_i(T,t) + S_i(T,t)$$

$$R_i(T,t) = \int_T^t P(v,T)N_i \exp\left(-\int_T^v N_i d\xi\right) dv$$

and

$$S_i(T,t) = \int_T^t U(v,T)\left[P(v,T)-3\right]N_i \exp\left(-\int_T^v N_i d\xi\right) dv.$$

In other words,

$$\tilde{Y}_i^{(\theta)}(T)\left[1-R\ (T,t)\right] \geq \tilde{Y}_i^{(\theta}\ (t) + S_i(T,t).$$

Inequality (55) will now be iterated. Let $\{T_{in}\}$ be an increasing divergent sequence. Choose $T = T_{in}$ and $t = T_{i,n+1}$ and iterate (55). (We suppose for simplicity that $T_{i1}$ is chosen sufficiently large that (55) holds for all $t \geq T_{i1}$.) Then iterating the inequality from $k = K$ to $k = n$, we find

$$(56)\quad \tilde{Y}_i^{(\theta)}(T_{iK})\Pi_{k=K}^n\left[1-R_i(T_{ik},\ T_{i,k+1})\right] \geq \tilde{Y}_i^{(\theta)}(T_{i,n+1}) + G_i(K,n),$$

where

$$(57)\quad G_i(K,n) = \sum_{k=K}^n S_i(T_{ik},\ T_{i,k+1})\Pi_{m=k+1}^n\left[1-R_i(T_{im},\ T_{i,m+1})\right].$$

($\theta$ by hypothes ;ill suff to prove tha ∞)
it al be shown tha

$$\Pi_{k} \quad -P \quad ik \quad +1$$

and tha

$$\text{lim lim sup}$$

To prov note by and tha for every

$$\frac{d}{dt} \log z_N \quad \int E_N \quad \exp \int \quad d\xi$$

al erna vely

$$\frac{d}{t} \log z_N \quad \exp \int$$

Thus in for any T and

$$\exp \int \quad \frac{N(i)(\mu)\exp\left(-\int_T^{\mu} D_{N(i)}\,dt\right)}{\int_T^{\nu} E_{N(i)} x \exp\left(-\int_T^{\nu} D_{N(i)}\,d\xi\right)}$$

and hence

$$\frac{L\left[N(i),\ f(T,\cdot);\ T,\ t\right]}{(T)\, L[N(i \quad ]}$$

ince ho

$$\sum \quad ik$$

by repl ing M and
suf to show that

$$\mathcal{U} \quad iK \quad T \quad n+1$$

fo pos constan By and

$$G_i(K,n) \leqslant \sum_{k=K}^{n} |S_i(T_{ik}, \quad T_{i,k+1})|.$$

By (54),

$$G_i(K,n) \leqslant \sum_{k=K}^{n} U(T_{ik}, \quad T_{i,k+1}) \int_{T_{ik}}^{T_{i,k+1}} |P(v,t)-3| N_i \exp\left(-\int_{T_{ik}}^{v} N_i d\xi\right) dv$$

By (51), $P(v,T) - 3$ is bounded. Moreover

$$\int_T^t N_i \exp\left(-\int_T^v N_i d\xi\right) dv = 1-\exp\left(-\int_T^t N_i dv\right) \leqslant 1.$$

These estimates complete the proof.

Corollary 2 is proved as follows:

*Proof of Corollary 2.* (14) follows from (19) and (47) by noting that, for $t \geqslant K_2$,

$$K_1 \leqslant f(T, T+t) \leqslant x(t+T) \int_T^{t+T} G dv \leqslant (\sup x) \int_T^{t+T} G dv.$$

Thus

$$\int_0^\infty G dv = \sum_{n=0}^{\infty} \int_{nK_2}^{(n+1)K_2} G dv = \infty$$

We now show that (21) implies (20). Then (20) will be used to prove the remaining assertions. It suffices to show that

(61) $$L[j, x; 0, \infty] < \infty \text{ implies } \int_0^\infty E_j dt < \infty$$

and that

(62) $$\int_0^\infty B_j dt < \infty \text{ implies } \int_0^\infty B_j z_j x^{-1} dt < \infty.$$

To prove (61), note by (30) that

$$\exp\left(-\int_T^t D_j dv\right) \geqslant z_j(T)(\sup z_j)^{-1}$$

and by (19) that $x(t) \geqslant K_1$ for $t \geqslant K_2$. To avoid trivialities, we assume that $z_j(T) > 0$ for some $T \geq K$. Then

$$L[j,\ x;\ 0,\ t] \geq R_j \int_T^t E_j dv,$$

where

$$R_j = z_j(T)(\sup z_j)^{-1} K_1 \exp\left(-\int_0^T D_j dv\right)$$

This proves (61). (62) has a similar proof.

(15) and (16) with $N(i)$ and $n(i)$ replacing $M(i)$ and $m(i)$, respectively, are proved using (18) and (19). Consider (15) for definiteness. It suffices to prove that, for any fixed $T \geq 0$ and $t$ sufficiently large,

$$\frac{L[N(i),f(T,\cdot);T,t]}{E+L[N(i),x;T,t]} \geq K_1(2 \sup x)^{-1}$$

Then define the sequence $\{S_{in}\}$ in (15) iteratively, setting $S_{in} = T$ and $S_{i,n+1} = t$ at each stage, and (15) readily follows. Consider (63). If (63) holds for $T \leq t < T + K_2$, we are done. If not, use condition (i) and (19) to conclude for $t \geq T + K_2$ that

$$\frac{L[N(i),f(T,\cdot);T,t]}{\epsilon+L[N(i),x;T,t]} \geq \frac{f_i(T)+K_1(\sup x)^{-1}h_i(T,t)}{g_i(T)+h_i(T,t)}$$

where

$$f_i(T) = L[N(i),f(T,\cdot);T,T+K_2],$$

$$g_i(T) = \epsilon + L[N(i),x;T,T+K_2],$$

and

$$h_i(T,t) = \int_{T+K_2}^t E_{N(i)} x \exp\left(-\int_T^v D_{N(i)} d\xi\right) dv.$$

By (64), it suffices to show that

$$E[N(i),\ x;\ 0,\ \infty] = \infty,$$

which will be proved by contradiction. Namely, we show that

(65) $$E[N(i), x; 0, \infty] < \infty$$

implies

(66) $$E[j, x; 0, \infty] < \infty \quad \text{for some } j \in J(1).$$

This contradicts the local flow condition (20)

Assume (65) and let Case (VII) hold for all $t \geq 0$ with $\tilde{Y}_i^{(\theta)}(0) > 0$. Then by (52), $\frac{d}{dt}\tilde{Y} \geq -N_i \tilde{Y}_i$. Thus there exists an $\eta > 0$ such that $\tilde{Y}_i^{(\theta)}(t) \geq \eta$ for all $t \geq 0$. Moreover, by (19) and (50) for any $T \geq 0$ and all $t \geq T+K_2$,

$$0 \leq X_i^{(\theta)}(t) \leq U(T) + (1-\mu)[\tilde{Y}_i^{(\theta)}(T) + U(T)],$$

where $\mu = K_1(\sup x)^{-1}$ and $U(T)$ monotonically approaches zero as $t \to \infty$. In all, there exists a $T_1 \geq 0$ and a $\nu \in (0,1)$ such that

(67) $$X_i^{(\theta)}(t) \leq (1-\nu)\tilde{Y}_i^{(\theta)}(T)$$

if $T \geq T_1$ and $t \geq T + K_2$. Since trivially,

$$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) = \tilde{Y}_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta)}(T) + \tilde{Y}_i^{(\theta)}(T) - X_i^{(\theta)}(t)$$

for any $t$ and $T$, (67) shows that

(68) $$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) \geq \nu\tilde{Y}_i^{(\theta}(T) + \tilde{Y}_i^{(\theta)}(t) - \tilde{Y}_i^{(\theta}(T)$$

if $T \geq T_1$ and $t \geq T + K_2$. (68) and the existence of $\tilde{Y}_i(\infty)$ (which follows without involving the hypothesis to be proved) imply the existence of a time $T_2$ such that

(69) $$\tilde{Y}_i^{(\theta)}(t) - X_i^{(\theta)}(t) \geq (\nu\eta/2) > 0 \quad \text{for } t \geq T_2.$$

(69) will now be shown to be impossible, thereby completing the proof.

By (69) there exists a $T_3$ such that for $t \geq T_3$,

$$\tilde{Y}_i^{(\theta)}(t) \geq \tilde{Y}_i^{(\theta)}(\infty) - (\nu\eta/8) > \tilde{Y}_i^{(\theta)}(\infty) - (3\nu\eta/8) \geq X_i^{(\theta)}(t).$$

Thus if for any $j \in J(1)$ and any $t \geq T_3$, $Z_{ji}^{(\theta)}(t) < \tilde{Y}_i^{(\theta)}(\infty) - (\nu\eta/4)$, then $Z_{ji}^{(\theta)}(t) < \tilde{Y}_i^{(\theta)}(t)$ for all $t \geq T_3$. In other words, every

$j \in J(1)$ such that $\tilde{Y}_i^{(\theta)}(t) = Z_{ji}^{(\theta)}(t)$ at any $t \geqslant T_3$ satisfies $Z_{ji}^{(\theta)}(t) - X_i^{(\theta)}(t) \geqslant \nu\eta/8$ for all $t \geqslant T_3$. By (25) $\dot{Z}_{ji}^{(\theta)}(t) \leqslant -(\nu\eta/8)H_j(W_t,t)$ for $t \geqslant T_3$, and thus (66) holds. This completes the proof.

Now we turn to the

*Proof of Proposition 1:* Suppose $Q_i$ exists. Then by Lemma 3, all $P_{ji}$ exist. Suppose $Q_i = \theta_i$. Choose $j \in J(2)$ and let Case VII hold for all $t \geqslant 0$ with $y_i^{(\theta)}(0) > 0$. Then since $\dot{Z}_{ji}^{(\theta)} \geqslant -H_j Z_{ji}^{(\theta)}$, it follows from $\int_0^\infty E_j dt < \infty$ that $P_{ji} > \theta_i$.

Now consider (24). Let $\omega_i = \min(P_{ji} - \theta_i : j \in J(2)) > 0$. Then for all sufficiently large $t$

$$\dot{X}_i^{(\theta)} \geqslant \omega_i \sum_{k \in J(2)} F_k - (F+G)X_i^{(\theta)}$$

where $F = \sum_{k \in J} F_k$. Integrating from $t = T$ to any $t \geqslant T$ yields

$$X_i^{(\theta)}(t) \geqslant P_i(t,T),$$

where

$$P_i(t,T) = \omega_i \sum_{k \in J(2)} \int_T^t F_k \exp\left[-\int_v^t (F+G)d\xi\right]dv.$$

It will suffice to show that $\limsup_{t\to\infty} P(t,T) > 0$. By (26), (27), and (29),

$$\exp\left[-\int_v^t (F+G)d\xi\right] = \frac{x(v)}{x(t)} \exp\left[\int_v^t A d\xi\right].$$

$$P_i(t,T) = \omega_i x^{-1}(t) \sum_{k \in J(2)} \int_T^t B_k z_k \exp\left[\int_v^t A d\xi\right]dv.$$

But $x$ is bounded. Moreover $z_k$ has a positive lower bound, since by condition (ix) and (19),

$$\dot{z}_k \geqslant E_k(-\gamma_k z_k + x)$$

$$\geq E_k(-\gamma_k z_k + K_1)$$

$t \geq K_2$. Thus there exists a $\lambda_i > 0$ such that

$$P(t,T) \geq \lambda_i \sum_{k \in J(2)} \int_T^t B_k \exp\left[\int_v^t A d\xi\right] dv.$$

Now apply (23) to show that $\limsup_{t\to\infty} X_i(t) > \theta_i$. In particular, $Q_i \neq \theta_i$. Moreover, if $\int_0^\infty E_j dt = \infty$, then $P_{ji} = Q_i \neq \theta_i$.

## Unbiased Learning with Arbitrary Positive Axon Weights Using Chemical Transmission and Action Potentials

Let (11) be replaced by

$$\dot{x}_i = A x_i + \sum_{k \in J} B_k \beta_{ki} z_{ki} + C_i; \tag{70}$$

that is, let the path weights $\beta_{ji}$ from $v_j$ to $v_i$ be arbitrary positive numbers. Can we transform (12) analogously so that learning and performance of spatial patterns is unimpaired? The answer is "yes".

We want the ratios $Z_{ji}^{(\beta)} = \beta_{ji} z_{ji} [\sum_{k \in I} \beta_{jk} z_{jk}]^{-1}$ to converge to $\theta_i$ after sufficient practice. This will happen if (12) is replaced by

$$\dot{z}_{ji} = D_j z_{ji} + E_j \beta_{ji}^{-1} x_i, \tag{71}$$

since letting $w_{ji} = \beta_{ji} z_{ji}$, (70) and (71) yield

$$\dot{x}_i = A x_i + \sum_{k \in J} B_k w_{ki} + C_i$$

and

$$\dot{w}_{ji} = D_j w_{ji} + E_j x_i,$$

which are again of the form (11) - (12).

Our goal could *not* be achieved by replacing (12) with

$$\dot{z}_{ji} = D_j z_{ji} + E_j \beta_{ji} x_i, \tag{72}$$

which would be the natural thing to do if we supposed that $E_j \beta_{ji}$ is determined wholly by spiking frequency. That (72) is inadmissible can be seen by transforming (70) and (72) into pattern variables. Doing this yields an infinite hierarchy of equations in the variables

$$\theta_{j_1 j_2 \ldots j_m i} = \frac{\beta_{j_1 i}\beta_{j_2 i}\ldots\beta_{j_m i}\theta_i}{\sum_{k\in I}\beta_{j_1 k}\beta_{j_2 k}\ldots\beta_{j_m k}\theta_k}$$

$$X_{j_1 j_2 \ldots j_m i} = \frac{\beta_{j_1 i}\beta_{j_2 i}\ldots\beta_{j_m i}x_i}{\sum_{k\in I}\beta_{j_1 k}\beta_{j_2 k}\ldots\beta_{j_m k}x_k}$$

$$Z_{j_1 j_2 \ldots j_r i} = \frac{\beta_{j_1 i}\beta_{j_2 i}\ldots\beta_{j_r i}z_{j_r i}}{\sum_{k\in I}\beta_{j_1 k}\beta_{j_2 k}\ldots\beta_{j_r k}z_{j_r k}}$$

where all $j_k \in J$, $m = 0,1,2,\ldots,$ and $r = 1,2,3,\ldots$ . These equations have the form

$$\dot{X}_{j_1 j_2 \ldots j_m i} = \sum_{k\in J} F_{j_1 j_2 \ldots j_m k}(Z_{j_1 j_2 \ldots j_m k i} - X_{j_1 j_2 \ldots j_m i})$$

$$+ G_{j_1 j_2 \ldots j_m}(\theta_{j_1 j_2 \ldots j_m i} - X_{j_1 j_2 \ldots j_m i})$$

$$\dot{Z}_{j_1 j_2 \ldots j_m i} = H_{j_1 j_2 \ldots j_m}(X_{j_1 j_2 \ldots j_m i} - Z_{j_1 j_2 \ldots j_m i}).$$

Note that when $X$ in (73) depends on $m$ values of $J$, $Z$ in (74) depends on $(m+1)$ values of $J$. Thus the hierarchy of equations never ends.

Suppose that we could analyse (73) and (74) and that all $X$'s and $Z$'s had limits $\theta$ which were approached with sufficient regularity that all $\dot{X}$'s and $\dot{Z}$'s approached zero as $t \to \infty$. Since all the coefficients $F$, $G$, and $H$ are nonnegative, we would expect each term on the right-hand side of (73) and (74) to also approach zero. In particular, we would find that

(75) $$P_{j_1 j_2 \ldots j_m ki} = Q_{j_1 j_2 \ldots j_m i}$$

and

$$\theta_{j_1 j_2 \ldots j_m i} = Q_{j_1 j_2 \ldots j_m i},$$

from (73), and

$$Q_{j_1 j_2 \ldots j_m i} = P_{j_1 j_2 \ldots j_m i}$$

from (74). Letting $j_1 = j_2 = \ldots = j_m = k = j$, this would mean that

$$\underbrace{\theta_{jj\ldots ji}}_{m \text{ times}} = \underbrace{Q_{jj\ldots ji}}_{m \text{ times}} = \underbrace{P_{jj\ldots ji}}_{m+1 \text{ times}} = \underbrace{Q_{jj\ldots ji}}_{m+1 \text{ times}} = \underbrace{\theta_{jj\ldots ji}}_{m+1 \text{ times}}$$

for every $m \geqslant 0$. In particular

$$\theta_i = \frac{\beta_{ji}^m \theta_i}{\sum_{k\in I} \beta_{jk}^m \theta_k}, \quad m \geqslant 0.$$

Letting $m \to \infty$ in (78) and defining $\beta_j = \max \{\beta_{ji} : i \in I\}$, we find for any $\theta_i > 0$ that

$$\theta_i = \lim_{m\to\infty} \frac{\beta_{ji}^m \theta_i}{\sum_{k\in I} \beta_{jk}^m \theta_k} = \begin{cases} 0 & \text{if } \beta_{ji} < \beta_j \\ \dfrac{\theta_i}{\Sigma\{\theta_k : \beta_j = \beta_{jk}\}} & \text{if } \beta_{ji} = \beta_j \end{cases}$$

In particular, $\Sigma\{\theta_k : \beta_j = \beta_{jk}\} = 1$, so one can at best learn patterns which are concentrated on the cells $v_i$ with $\beta_{ji} = \beta_j$. These cells have uniformly distributed path weights. If there exists a subset $\tilde{J} \subset J$ such that

$$\bigcap_{j\in\tilde{J}} \{i \in I : \beta_{ji} = \beta_j\} = \emptyset,$$

then no pattern can simultaneously be learned by all the cells $\tilde{J}$. This is a very inflexible system.

How can the $\beta_{ji}$'s in (70) and (71) be interpreted? Suppose $\beta_{ji}$ $\lambda_j R_{ji}$, where $\lambda_j > 0$ and $R_{ji}$ is the circumference of the cylindrical axon $e_{ji}$. Let the signal in $e_{ji}$ (e.g., the action potential [16]) propagate along the circumference of the axon to its synaptic knob. Let the signal disperse throughout the cross-sectional area of the knob (e.g., as ionic fluxes [16]). Let local chemical transmitter production in the knob be proportional to the local signal density. Finally, let the effect of the signal on the postsynaptic cell be proportional to the product of local signal density and local available transmitter density and the cross-sectional area of the knob.

These laws generate (70) and (71) as follows. Signal strength is proportional to $R_{ji}$, or $\beta_{ji}$. The cross-sectional area of the knob is proportional to $R_{ji}^2$. Hence signal density in the knob is proportional to $R_{ji}R_{ji}^{-2} = R_{ji}^{-1}$, or to $\beta_{ji}^{-1}$, as in (71). Thus (signal density) x (transmitter density) x (area of knob) $\cong R_{ji}^{-1}z_{ji}R_{ji}^2 = R_{ji}z_{ji} \cong \beta_{ji}z_{ji}$, as in (70).

By contrast, a mechanism whereby signals propagate throughout the cross-sectional area of the axon could not produce unbiased learning given arbitrary axon connection strengths, or at least such a mechanism is still elusive. The difficulty here is that signal strength is proportional to $R_{ji}^2$, signal density is proportional to $R_{ji}^{-2}$, and local transmitter production rate is then proportional to $R_{ji}^{-2}$. The post synaptic signal is proportional to (signal density) x (transmitter density) x (area of knob) $\cong z_{ji}$. Thus we are led to the system

$$\dot{x}_i = Ax_i + \sum_{k \in J} B_k z_{ki}$$

and

$$\dot{z}_{ji} = D_j z_{ji} + E_j \beta_{ji}^{-2}$$

which can be written as

$$\dot{x}_i = Ax_i + \sum_{k \in J} B_k \gamma_{ki} w_{ki} + C_i$$

and

$$= D_j w_{ji} + E_j \gamma_{ji} x_i$$

in terms of the variables $w_{ji} = \beta_{ji} z_{ji}$ and $\gamma_{ji} = \beta_{ji}^{-1}$. This system again yields an infinite hierarchy of equations for the pattern variables.

These observations suggest that the action potential not only guarantees faithful signal transmission over long cellular distances, as is well known, but also executes a subtle transformation of signal densities into transmitter production rates that compensates for differences in axon diameter. Note also that this transformation seems to require the chemical transmitter step. Purely electrical synapses presumably could not execute it. Thus our laws for transmitter production not only guarantee that learning occurs, but also that unbiased learning occurs, under very weak anatomical constraints.

## References

[1] Grossberg, S., Embedding Fields: Underlying Philosophy, Mathematics, and Applications to Psychology, Physiology, and Anatomy, J. of Cybernetics, 1(1971), 28.

[2] Grossberg, S., Neural Pattern Discrimination, J. of Theoretical Biology, 27(1970), 291.

[3] Grossberg, S., Neural Expectation: Cerebellar and Retinal Analogs of Learnable and Unlearned Pattern Classes, Kybernetik 10(1972), 49.

[4] Grossberg, S., On the Dynamics of Operant Conditioning, J. of Theoretical Biology, 33(1971), 225.

[5] Grossberg, S., On the Serial Learning of Lists, Math. Biosci. 4 (1969), 201.

[6] Grossberg, S. and J. Pepe, Spiking Threshold and Overarousal Effects in Serial Learning, J. of Statistical Physics, 3(1971), 95.

[7] Grossberg, S., Pavlovian Pattern Learning by Nonlinear Neural Networks, Proc. of the Natl. Acad. of Sciences, U.S.A., 68(1971), 68.

Grossberg, S., On Learning and Energy-Entropy Dependence in Recurrent and Nonrecurrent Signed Networks, J. of Statistical Physics, 1(1969), 319.

Grossberg, S., A Neural Theory of Punishment and Avoidance, I. Qualitative Theory, Math. Biosci., in press.

Grossberg, S., A Neural Theory of Punishment and Avoidance, II. Quantitative Theory, Math. Biosci., in press.

Kimble, G. A., Foundations of Conditiong and Learning, Appleton-Century-Crafts, New York, 1967, 26.

Grossberg, S., Some Networks That Can Learn, Remember, and Reproduce any Number of Complicated Patterns, II, Studies in Applied Math., XLIX(1970), 137.

Grossberg, S., A Prediction Theory for some Nonlinear Functional-Differential Equations, I, J. Math. Anal. and Applics., 21(1968), 643.

Grossberg, S., A Prediction Theory for Some Nonlinear Functional-Differential Equations, II, J. Math. Anal. and Applics., 22(1968), 490.

Grossberg, S., 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.

Ruch, T. C., H. D. Patton, J. W. Woodbury and A. L. Towe, Neurophysiology, W. B. Saunders, Philadelphia, 1961.

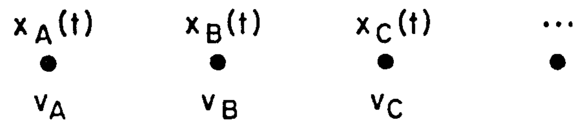
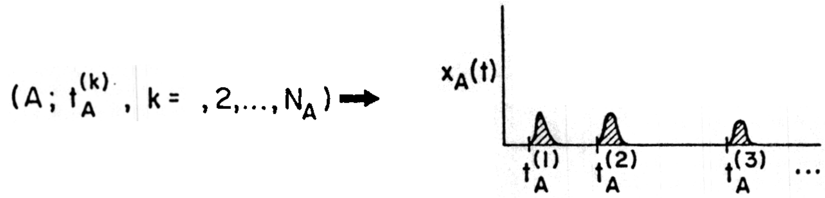
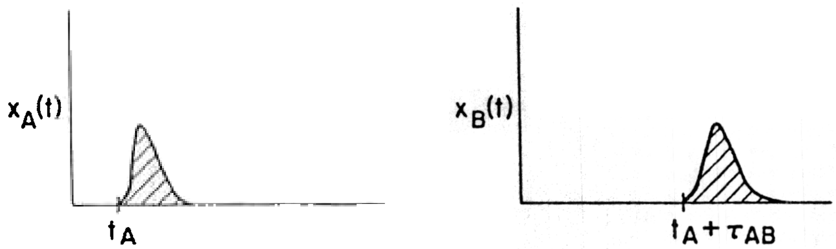



Fig. 1. Vertices and vertex functions.




2. Input presentations induce vertex perturbations.




3. Vertex translation of predicting B given A.

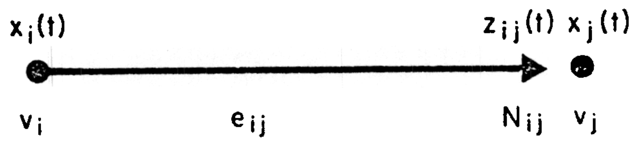



Fig. 4. Directed network and network processes

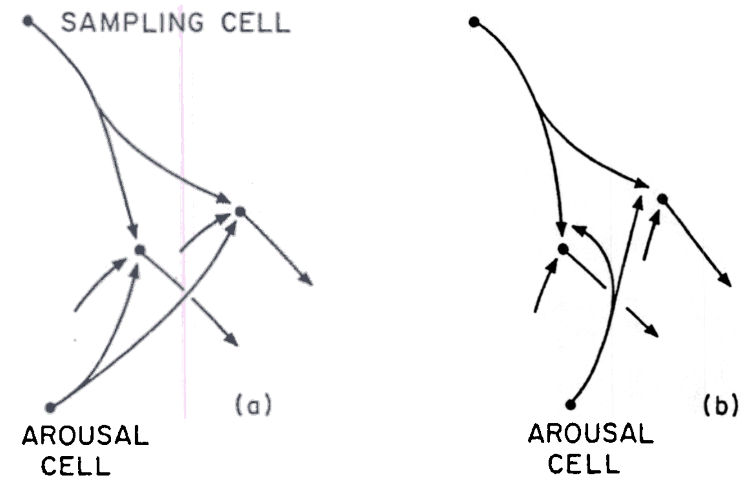



Fig. 5. Interactions between sampling and arousal cells.

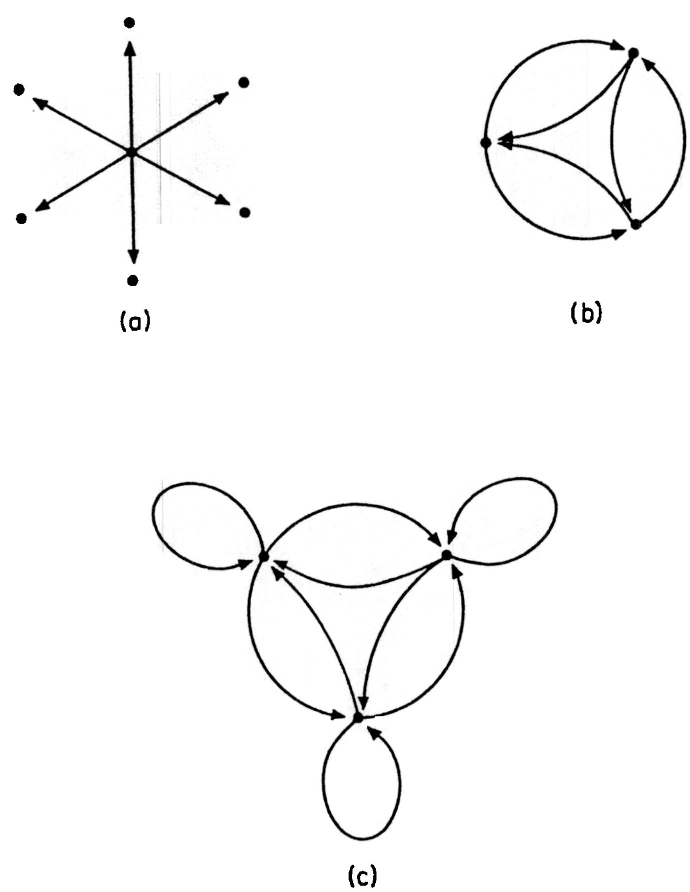



Fig. 6. Phase transitions depend on anatomy.

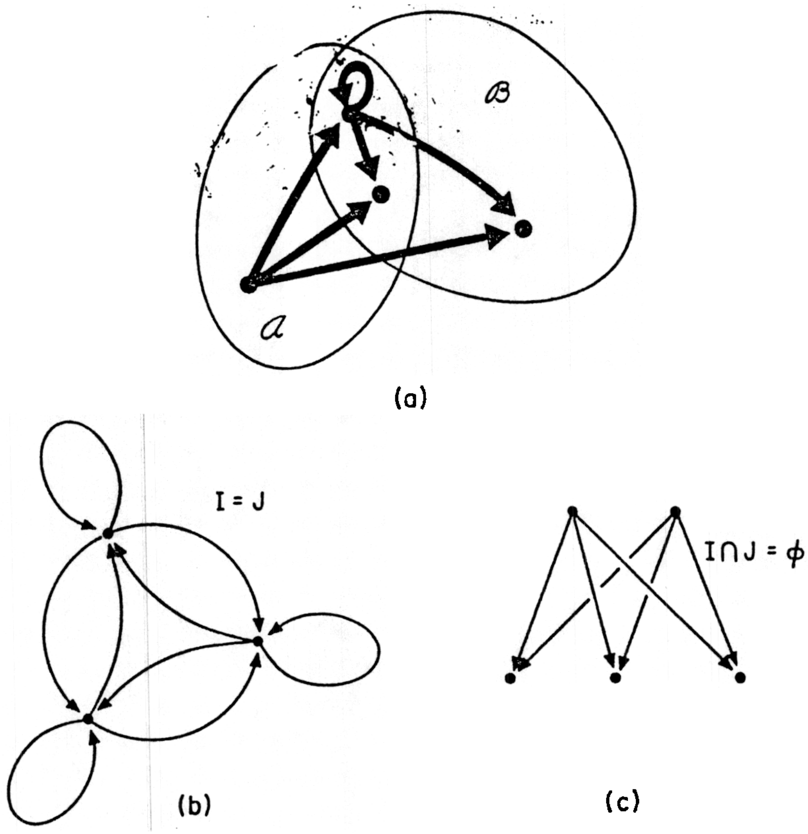



Fig. 7. Connections between sampling and sampled cells.