On Learning, Information, Lateral Inhibition, and Transmitters

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

Communicated by Richard Bellman

ABSTRACT

This article continues the derivation of a learning theory with neurophysiological implications. Equations are derived that contain formal analogs of physiological interactions between membrane potentials, spiking frequencies, transmitter production and release, and various trophic effects. These equations reduce to the Hartline-Ratliff equation for lateral inhibition in a special case. The need for inhibition in making possible perfect learning is suggested. A formal connection between contour enhancement due to lateral inhibition, the Ward-Hovland or "reminiscence" phenomenon in learning (i.e., spontaneous improvement of memory), bowing in serial verbal learning, and the information functional is described. A unified explanation of spatiotemporal masking and of the decrease of reaction time with increased learning or increased stimulus energy is given. A formal analog of the inward flow of Na^+ and the outward flow of K^+ through the membrane in response to excitatory transmitter and of the outward flow of K^+ in response to inhibitory transmitter is derived.

I. INTRODUCTION

A. Some "Neural" Properties

This article continues the program, introduced in [1], of deriving a learning theory with physiological implications from psychological postulates. Psychological postulates are herein reviewed and extended that give rise to equations that resemble such familiar neural phenomena as lateral inhibition [2], transmitter production and release [3], the production of spikes at suprathreshold membrane potential values [4].
the inward flow of Na⁺ and outward flow of K⁺ across the cell membrane in response to an excitatory transmitter [3], and the outward flow of K⁺ in response to an inhibitory transmitter [3]. The equations reduce in a special case to the Hartline-Ratliff equation for lateral inhibition in the Limulus retina [5], and provide some theoretical information about the parameters that enter into determining the empirical coefficients of that equation. The equations also contain predictions about phenomena that have not as yet been experimentally determined; for example, a binding of Na⁺, K⁺, Ca²⁺, and Mg²⁺ within the end bulb in complexes of varying strength to produce transmitter production and release rates that are sensitive to prior presynaptic and postsynaptic levels of membrane potential. Some of these facts are only sketched herein, since the article is primarily concerned with heuristic gedanken experiments that give rise to these equations. Later papers will derive the equations on a postulational basis, extend the formalism, and analyze the neural results in greater detail.

B. Lateral Inhibition and Perfect Learning

Lateral inhibition plays a useful role in making possible perfect, or deterministic, learning in the systems that we derive herein. In [1] are derived learning machines in which no manifestly inhibitory processes occur, and yet these machines can learn, remember, and recall in specific experimental situations [6-11]. Nonetheless, because some of the postulates used in this derivation are merely approximate, a severe limitation is placed on the behavior sequences that these machines can perfectly learn. The alphabet A = ABC . . . . Z can, in principle, be perfectly learned by some of the machines in [1], but the alternating sequence $B = ABACABAC$ can be learned only statistically (i.e., some machines can learn that B and C each follow A 50% of the time in B). Lateral inhibition arises in this article as one of the mechanisms used to construct machines that learn $B$ deterministically.

Statistical learning of $B$ is simpler than deterministic learning for an obvious reason. Statistical learning merely requires that a machine know the frequency with which B and C follow A. Deterministic learning requires that the machine know the subsequences of length greater than one that determine the next letter uniquely. For example, the subsequence BA uniquely determines C, and the subsequence CA uniquely determines B. Deterministic learning of $B$ thus requires that the machine be able to learn the context of letters that precede a given letter, not only the frequency with which one letter follows the next.

Inhibition is thus a suggested tool for solving the problem of deterministically learning behavior sequences in which nontrivial context effects occur. This problem arises in any behavioral language that has an alphabet (e.g., letters) that is smaller than its vocabulary (e.g., words). Only the simplest behavioral languages are not of this type, if only because the number of distinct behavioral signs that an individual can produce is so small. The universal appearance of lateral inhibition in neural systems therefore emerges as a counterpart of the universal problem of deterministically learning nontrivial contexts.

C. Lateral Inhibition and Information

We will show that lateral inhibition and the mathematical concept of information are closely related. It is well known that the information function, or entropy, of mathematical information theory can be uniquely derived from postulates that are a plausible formal realization of our intuitive concept of information [12]. The strong intuitive impression also exists that the nervous systems of animals are excellent information-processing systems. The basic question therefore arises: how does the information function “sit” in an animal’s nervous system? We find that the transformation from inputs to outputs executed by our equations behaves qualitatively in specific cases as if the mathematical information contained in the inputs were being computed by these equations. Lateral inhibition is crucial in producing these qualitative effects.

This article begins by pointing out some limitations of the systems in [1] and sketches a simple plausible way out of these limitations. We then label the mathematical variables of the new system in a natural neural way. The equations governing these variables thereby give rise to some neural expectations, which are stated along with references containing confirmatory neural data.

2. EXCITATORY EMBEDDING FIELDS

We briefly review some conclusions from [1] about our learning machines $\mathcal{M}$ as a point of departure for the present account.

A. Recognition

1. A “simple” behavioral symbol $r_i$ (such as the letter A as used in speech) is represented in $\mathcal{M}$ by a single abstract state $v_i$.
2. The presentation of $s_i$ to $\mathcal{M}$ at time $t = t_0$ is represented in $\mathcal{M}$ as an input pulse to $v_i$ with onset time $t = t_0$. More complicated behavioral symbols are represented by correspondingly more complicated input patterns.

3. A nonnegative state function $x_i(t)$ sits at $v_i$ and fluctuates in response to inputs. When all inputs are zero, $x_i(t)$ decays to zero. Increasing the input $I_i(t)$ received by $v_i$ increases $x_i(t)$.

These facts are summarized in the equations

$$\dot{x}_i(t) = -\alpha x_i(t) + I_i(t), \quad x_i(0) > 0, \quad (1)$$

for each symbol $r_i$, where $i = 1, 2, \ldots, n$. The input $I_i(t)$ is given by

$$I_i(t) = \sum_{j=1}^{N} J_{ij}(t - t_{im})$$

where $J_{ij}(t)$ is an input pulse with onset time $t = 0$, and $t_{im}$ is the $m$th onset time of $r_i$.

In the special case that $x_i(t)$ has a fixed upper bound $M_i$, (1) is replaced by

$$\dot{x}_i(t) = -\alpha x_i(t) + (M_i - x_i(t))I_i(t), \quad 0 \leq x_i(0) \leq M_i. \quad (1')$$

B. Learning

1. After a list $r_F$ of symbols has been learned by $\mathcal{M}$, a presentation of $r_i$ to $\mathcal{M}$ creates the reply $r'_i$ a short time, say, $\tau_{ij}$ time units, later. That is, an input to $v_i$ creates an output from $v_j$ a time $\tau_{ij}$ later. To accomplish this, a signal will be sent from $v_i$ to $v_j$ at a finite velocity over a pathway $e_{ij}$. $e_{ij}$ is a directed pathway because the links $r_i$ and $r_j$ are not the same. The pathway $e_{ij}$ is therefore drawn as an arrow from $v_i$ to $v_j$.

2. Before learning occurs, $\mathcal{M}$ will be able to learn several lists $r_F, r'_F, \ldots, r_{Fk}$, or else $r_j$ would already be $\mathcal{M}$’s only reply to $r_i$. Therefore, $r_i$ can send signals to all points $v_j$ that represent possible replies to $r_i$.

By Section 2, B. 1 and 2, a process $z_{ij}(t)$ will exist that controls the size of signals from $v_i$ to $v_j$ at time $t$; $z_{ij}(t)$ measures the frequency with which $r_i$ and $r_j$ have been presented consecutively to $\mathcal{M}$ in the past, and cuts off the signal from $v_i$ to $v_j$ if this frequency is small.

3. Since $z_{ij}(t)$ measures the frequency of consecutive $r_i$ and $r_j$ presentations, it takes place in $\mathcal{M}$ at a position where past $x_i(\xi)$ signals (\(\xi < t\)) and present $x_i(t)$ signals coexist. Only one such place exists in $\mathcal{M}$, namely, the arrowhead $N_{ij}$ of $e_{ij}$ at which the signal from $v_i$ to $N_{ij}$ is contiguous with $x_i(t)$ in $v_j$.

Mathematical Biosiences 4 (1969), 255-310
and

$$z_{jk}(t) = -u z_{jk}(t) + \beta y_{jk}(t - \tau) x_{jk}(t).$$

(6)

Systems (a) and (aa) are both examples of embedding fields. These fields are called excitatory because their variables, including the interaction terms $\beta x_{jk}(t - \tau) y_{jk}(t)$ of (a) and $\beta y_{jk}(t - \tau) y_{jk}(t)$ of (aa) are always nonnegative.

We now itemize some limitations of excitatory embedding fields.

3. NONLOCAL ASSOCIATIONS IN A LOCAL FLOW

Consider the interaction term $\beta x_{jk}(t - \tau) y_{jk}(t)$ from $r_{jk}$ to $r_{j}$ in (aa). This interaction has the following interpretation. Vertex $r_{jk}$ sends out a signal $\beta x_{jk}(t - \tau)$ along $e_{jk}$ at time $t - \tau$. This signal travels along $e_{jk}$ until it reaches the arrowhead $N_{jk}$ of $e_{jk}$ at time $t$. The process $y_{jk}(t)$ in $N_{jk}$ is thereupon activated and the quantity $\beta x_{jk}(t - \tau) y_{jk}(t)$ is instantaneously transmitted from $N_{jk}$ to $r_{j}$. Process $y_{jk}(t)$, in turn, is constructed from the ratio of $p_{jk} z_{jk}(t)$ and $\sum_{k} p_{jk} z_{jk}(t)$. Each $z_{jk}(t)$ cross-correlates the pulse $\beta p_{jk} z_{jk}(t - \tau)$ received from $e_{jk}$ by the arrowhead $N_{jk}$ of $e_{jk}$ at time $t$ with the value $x_{j}(t)$ of $r_{j}$ contiguous to the arrowhead at this time. The puzzle thereupon arises: for any $k \neq j$, how does the cross-correlation $z_{jk}(t)$, which sits at the arrowhead $N_{jk}$ at time $t$, make itself instantaneously felt at the arrowhead $N_{jk}$ in $y_{jk}(t)$?

This difficulty cannot be overcome without changing (aa), since no provisions are therein made for a transport of any $z_{jk}(t)$ quantity from one arrowhead to another. This change must not, however, destroy the improvements in learning achieved by replacing $p_{jk} z_{jk}(t)$ in (2) by $y_{jk}(t)$ in (4). The basic improvement is that an increase in $z_{jk}(t)$ causes not only an increase in $y_{jk}(t)$, but also a decrease in $y_{jk}(t)$, for all $i \neq j$, and conversely. That is, growth of the “associational strength” $y_{jk}(t)$ is inhibited by growth of the associative strength $y_{jk}(t)$, $i \neq j$, and conversely. The replacement of $p_{jk} z_{jk}(t)$ by $y_{jk}(t)$ is thus a way of introducing inhibition into an embedding field without sacrificing the nonnegativity of all its variables.

A heavy conceptual price is paid for this formal improvement; namely, the introduction of a “virtual” process that instantaneously leaps from arrowhead to arrowhead without any apparent geometrical mechanism underlying it. Since this price is conceptually intolerable, we must seek another way to realize the mutual inhibition of associational strengths.

We will find that essentially only one method of doing this is available to us. This method will sacrifice the nonnegativity of the $x_{j}(t)$ values to eliminate the virtual inhibitory process. As shown in [1], nonnegativity of the $x_{j}(t)$'s means that all states of $\mathcal{X}$ are observable to a psychological experimenter studying $\mathcal{X}$. Thus, observability must be sacrificed to create a dynamics of learning that corresponds in a sensible way to the geometry on which it plays.

4. LATERAL INHIBITION: LOCALITY VERSUS OBSERVABILITY

A way of preserving inhibition between associations $y_{jk}(t)$ and $y_{jk}(t)$ without requiring a virtual process is easily seen. It is clear that some process must carry this inhibition between $N_{jk}$ and $N_{jk}$. We know only one process that carries information around the network. This is the process whereby the states $r_{j}$ and $r_{j}$ send signals to one another. In other words, inhibitory signals must pass between the states $r_{j}$ and $r_{j}$. Only the process $x_{j}(t)$ can create a signal from $r_{j}$ to $r_{j}$ and only the process $x_{j}(t)$ can create a signal from $r_{j}$ to $r_{j}$. The signals with which we are familiar in (aa) are excitatory in the sense that an increase in $x_{j}(t)$ causes an increase in $x_{j}(t)$ after a suitable time lag elapses. An inhibitory signal from $r_{j}$ to $r_{j}$ has the property that an increase in $x_{j}(t)$ causes a decrease in $x_{j}(t)$ after a suitable time lag elapses.

Can such an inhibitory signal from $r_{j}$ to $r_{j}$ alter $r_{jk}(t)$ in much the same way that the transformation from $z_{jk}(t)$ to $y_{jk}(t)$ does? The answer is manifestly yes, because as (6) shows, a decrease (or increase) in $y_{jk}(t)$ causes a corresponding decrease (or increase) in $z_{jk}(t)$. Just as the virtual inhibition from $y_{jk}(t)$ to $y_{jk}(t)$ occurs instantaneously, the inhibitory signals from $r_{j}$ to $r_{j}$ must occur "rapidly" where by this we clearly mean: (1) more rapidly than excitatory signals, and (2) rapidly compared to the rate with which the correlations $z_{jk}(t)$ and $z_{jk}(t)$ change in response to changes in $x_{j}(t)$ and $x_{j}(t)$, respectively.

We conclude that rapid inhibitory signals from $r_{j}$ to $r_{j}$ can inhibit $z_{jk}(t)$ much as the transformation from $z_{jk}(t)$ to $y_{jk}(t)$ does. The great advantage of the former method of inhibiting $z_{jk}(t)$ is that all interactions can then be described by local processes; that is, by processes that move along the geometry of $\mathcal{X}$ in a sensible way. The loss of observability is a small price to pay for this advantage.
5. INHIBITION IN THE MIDDLE OF A LONG LIST

In a previous paper [8], the bowed curve of serial verbal learning was explained, using (**)9. Bowing occurs at a fixed item \( r_i \) in the middle of a long list because the associations \( y_{ij}(t) \) remain quite uniformly distributed in \( \kappa \) for a relatively long amount of time.

In terms of inhibition between states, this means that inhibition among states is maximal when items in the middle of the list occur. This fact corroborates the classical work of Hull and his associates ([13], page 516), who also postulated an "accumulation of inhibition" near a list's middle. Of course, the actual mechanisms that Hull envisaged are not the same as our own, but it is altogether remarkable that he reached this qualitative conclusion some thirty odd years ago.

The introduction of inhibition between states to replace virtual inhibition between correlations carries with it a substantial conceptual consequence. Having identified Hull's "accumulating inhibition" as inhibition between states, we will, in Sections 11-13, identify inhibition between states as lateral inhibition of the neural variety, and will be able to formally derive the Hartline-Katlin equation. A unifying conceptual bridge is hereby constructed from the seemingly diverse phenomena of serial verbal learning in humans to lateral inhibition within the Limulus retina.

6. INHIBITION AND SELF-IMPROVING MEMORY

Using the functions \( y_{ij}(t) \) instead of \( z_{ij}(t) \) in (**)9 has an important effect on the way in which \( \kappa \) remembers. The following fact has been rigorously proved [7, 9, 10]. If \( \kappa \) has learned a given task to a "moderate" degree of accuracy, then \( \kappa \)'s memory of the task improves spontaneously without further practice. Since we have replaced the \( y_{ij}(t) \)'s by inhibitory signals, we should now be able to say that inhibition between states helps to create self-improving memories. This is indeed the case.

Speaking roughly, the mechanism is as follows. First, a rapid inhibition between states eliminates behaviorally insignificant (e.g., uniformly distributed) background noise. The more slowly varying associations \( z_{ij}(t) \) then learn only from the behaviorally significant (e.g., dominant) states. The data on which \( \kappa \) constructs its memory are thus "crisper" than the data that \( \kappa \) receives. Once a significant \( z_{ij}(t) \) association is formed, it tends to create a distribution of state values that are compatible with its existence by funneling large excitatory signals from \( c_i \) to \( v_i \). Then inhibition between states "crisps" these state values and thereby causes a further increase in \( z_{ij}(t) \).

The "crisping" of associations \( z_{ij}(t) \) is a new idea concerning the way in which \( \kappa \) learns. It is, however, intimately connected with the "perceptual crisping" or "contour enhancement" due to lateral inhibition that has been a subject of experimental interest since the time of Mach [8]. Perceptual crisping has been noticed by experimentalists for some time because it leads to a crisping of neural signals and thus of behavior itself. Crisping of associations has remained unnoticed both because it is coupled so intimately to perceptual crisping effects, and because it describes a change within the end bulbs, and therefore creates no directly observable signals.

7. OBSERVABLE PROBABILITY MODELS VERSUS UNOBSERVABLE INHIBITION BETWEEN STATES

At the risk of belaboring the now obvious, we emphasize an important conceptual consequence of replacing virtual inhibition between correlations by a real inhibition between states. We perhaps all share the strong intuitive belief that we can legitimately hope to describe the observable behavior of many organisms by using some kind of probability model, and it seems perfectly natural to discuss the transition probability that the organism will move from one behavioral state to the next at any given time. We are equally aware that much of our behavior seems to be quite deterministic. For example, the very act of writing a mathematical paper has approximately probability zero in any reasonably simple probabilistic model for producing letters and numbers. How can the seemingly contradictory intuitive impressions of merely probabilistic versus deterministic behavior be reconciled in a philosophically satisfactory way?

The simple model (**) suggests a way: (***) is manifestly a deterministic model, in the sense that it describes a prediction theory for individual inputs and outputs. System (**) also contains transition probabilities of moving from a given behavioral state \( v_i \) to another state \( v_j \) at time \( t \), namely, the associational strengths \( y_{ij}(t) \). In particular, when \( y_{ij}(t) \approx 1 \), then the probabilistic transition from \( v_i \) to \( v_j \) at time \( t \) will look as deterministic as we please. In this sense, (***) is both deterministic and probabilistic at every instant of time.
The properties that define a probability distribution, namely,

(a) \( p_i(t) \geq 0 \) and (b) \( \sum_{i=1}^{n} p_i(t) = 1 \).

are, within (*), consequences of two facts: (a') observability, and (b') the need in an observable model for virtual inhibition between correlations to guarantee efficient learning. Observability, or (a'), must, however, be sacrificed to preserve the more basic property of locality. Virtual inhibition between correlations, or (b'), must then be replaced by real inhibition between states. The "impression" that the observable behavior of system (*') is probabilistic is hereby formally translated into a wholly deterministic, but not wholly observable, process in which both excitatory and inhibitory factors are admitted.

Since the basic ideas, such as locality, that led to these conclusions are so very simple and general, it seems quite possible that analogs therefore exist in many other systems in which evolutionary (or learning) trends appear. That is, the seemingly probabilistic dynamics of these processes might well be the observable epiphenomenon of a not entirely observable interaction between excitatory and inhibitory factors.

S. ENTROPY AND SUFFICIENT REASON: NONLOCAL OUTPUTS OF LOCAL STATES

One of the conceptual deficiencies of model (*) is that it does not abide by the principle of sufficient reason. By this we mean that presentation of a symbol \( r_i \) to a machine \( \mathcal{M} \) can produce large outputs from several states \( r_i \) even when no symbol \( r_i \) is a preferred response to \( r_i \).

For example, consider an outstar [9] with source \( r_i \) and border \( B = \{ i; \ i = 2, 3, \ldots, n \} \). That is, let \( p_{ij} = 1/(n-1) \), \( i = 2, 3, \ldots, n \), and let all other \( p_{ij} = 0 \). Then (*) becomes

\[
\begin{align*}
\hat{x}_i(t) &= -x_i(t) + I_i(t), \\
\hat{z}_i(t) &= -z_i(t) + \beta x_i(t - \tau) y_i(t) + I_i(t), \\
y_i(t) &= z_i(t) \left( \sum_{i=2}^{n} z_i(t) \right)^{-1},
\end{align*}
\]

and

\[
\hat{z}_i(t) = -u z_i(t) + \beta x_i(t - \tau), \quad i = 2, 3, \ldots, n.
\]

Suppose that no list \( r_i r_j \) is preferred to any other in \( \mathcal{M} \) before time \( t - \tau \). That is, all \( x_i(\xi) \) are equal and all \( y_i(\xi) = 1/(n-1) \), where \( \xi \leq t - \tau \) and \( i = 2, 3, \ldots, n \). Let \( r_i \) be presented to \( \mathcal{M} \) at time \( \xi = t - \tau \). Then \( x_i(\xi) \) becomes large shortly after time \( \xi = t - \tau \). Consequently, the signal \( \beta x_i(t) y_i(t) \) from \( r_i \) to \( r_i \) also increases. Since \( y_i(t) = 1/(n-1) \), each signal is the same. Moreover, the increase in this common signal is substantial, and in any case depends only on the number \( n - 1 \) of points in \( B \). The outputs \( x_i(t) \) from all \( r_i \) therefore increase substantially shortly after time \( \xi = t \). These large outputs occur in spite of the fact that all \( x_i(\xi) \) remain equal and all \( \gamma_i(\xi) = 1/(n-1) \) throughout this experiment on \( \mathcal{M} \). That is, a large output from \( r_i \) can occur in response to the presentation of \( r_i \) even though \( r_i \) is not a preferred response alternative to \( r_i \).

This difficulty is formally overcome in [9] by modifying the output from \( r_i \). Instead of using \( x_i(t) \) itself as the output, we use

\[
O_i^j(t) = \max \left\{ 0, x_i(t) \left( 1 - \frac{H(X_i(t), X_j(t), \ldots, X_m(t))}{\ln(n-1)} \right) \right\},
\]

\( i = 2, 3, \ldots, n \), where \( \Gamma \) is some positive response threshold.

and

\[
X_i(t) = \frac{\sum_{i=2}^{n} x_i(t)}{n-1},
\]

and

\[
H(q_1, q_2, \ldots, q_{n-1}) = \sum_{i=1}^{n-1} q_i \ln q_i
\]

for any probability distribution \( \{ q_i; \ i = 1, 2, \ldots, n-1 \} \); \( H \) is the familiar information function, or entropy, of mathematical information theory [12].

The modified output \( O_i^j(t) \) realizes the principle of sufficient reason in \( \mathcal{M} \) in the following way. Suppose that no list \( r_i r_j \) is preferred by \( \mathcal{M} \) at time \( t \); in particular, all \( x_i(t) \) are equal. Then \( O_i^j(t) = 0 \) for all \( i = 2, 3, \ldots, n \) no matter how large the common \( x_i(t) \) value is. That is, in order that a modified output \( O_i^j(t) \) from \( r_i \) occur, it must represent a preference by \( \mathcal{M} \) for the list \( r_i r_j \).

Consider now the case in which a definite preference for \( r_i r_j \) does exist within \( \mathcal{M} \) at time \( t \). Suppose in particular that \( x_i(t) \) far exceeds all \( x_j(t) \) values in size, \( j \neq 1, i \). Then all \( O_i^j(t) \) reduce essentially to the old outputs \( x_i(t) \), with the improvement that background noise is eliminated by the thresholds \( \Gamma_j \). In mathematical terms this means:

\[
x_i(t) \geq x_j(t) \iff 0, \quad \text{for all } j \neq 1, i.
\]

Mathematical Biosciences 4 (1969), 255-310
implies that
\[ O_j^T(t) = \max \{0, x_j(t) - l'_j\} . \]

Since \( x_j(t) \geq 0 \) and \( l'_j > 0 \) for all \( j \neq i, i' \),
\[ O_j^T(t) = 0, \quad j \neq i, \]
whereas
\[ O_i^T(t) = x_i(t) - l', \]
which differs from \( x_i(t) \) only by an inessential scaling factor \( l' \).

Although the modified outputs \( O_i^T(t) \) eliminate an unpleasant conceptual difficulty in a formal way, they create yet another conceptual difficulty. In order to evaluate \( O_i^T(t) \), \( \mathcal{N} \) must instantaneously bring together all \( x_j(t) \) values from \( \mathcal{B} \) to evaluate the probabilities \( X_j(t) \), must then compute the entropy of these probabilities, and must then compute \( O_i^T(t) \) itself at each state \( x_i(t) \). Yet within \( \mathcal{N} \), there exist no geometrical pathways, let alone dynamical reasons, by which these transformations can take place. Again locality has been violated. Model (**) must therefore be modified to eliminate this conceptual difficulty without losing the formal advantages of \( O_i^T(t) \).

The transformation from \( x_i(t) \) to \( O_i^T(t) \) replaces equal \( x_i(t) \)'s by zeros, but allows the \( O_i^T(t) \) value corresponding to a unique large \( x_i(t) \) value to remain uninfluenced by the other \( x_i(t) \) values except for a threshold shift. Thus the transformation from \( x_i(t) \) to \( O_i^T(t) \) acts as if the \( x_i(t) \)'s inhibit each other out of existence when they are equal, but a unique large \( x_i(t) \) value inhibits all other \( x_i(t) \) values to zero without suffering reciprocal inhibition.

We therefore conclude that both the transformation
\[ z_i(t) \rightarrow y_i(t) \]
between associations and the transformation
\[ x_i(t) \rightarrow O_i^T(t) \]
between outputs covertly describe an inhibition between states.

The remarks above suggest that the entropy \( H \) "sits" in \( \mathcal{N} \) as an inhibitory mechanism that accentuates dominant \( x_i(t) \) values and annihilates background noise to prepare these data for subsequent learning by the associations \( z_i(t) \). This role for \( H \) is certainly compatible with the intuitive idea that \( H \) plays some part in the information processing of a learning machine.

**Mathematical Biosciences 6 (1969), 245-270**

9. WHAT IS A "SIMPLE" ACT? NONLOCAL PATTERNS WITH LOCAL CONTROLS

Every "simple" behavioral symbol \( r_i \) has been placed in correspondence with a single abstract state \( s_i \) in \( \mathcal{N} \). Learning a sequence of simple symbols \( r_{i1} r_{i2} \cdots r_{i\ell} \) can be viewed as the formation of a new composite symbol that itself gradually becomes simple as our experience with it grows. Should not then the new symbol \( r_{i1} r_{i2} \cdots r_{i\ell} \) eventually correspond to a new state? If this is so, then are not the states \( r_i \) to which the original simple symbols \( r_i \) correspond also, in some sense, composite states?

Questions such as these illustrate the approximate status of the postulate that a simple symbol \( r_i \) corresponds to a single abstract state \( s_i \) in \( \mathcal{N} \). Even the production of such seemingly simple symbols as \( A \) requires the integration of very complicated muscular motions that cannot adequately be described by a single state, and are carried through the air by sound waves of great complexity. We must therefore expect to eventually find that even simple behavioral symbols are represented by aggregates of many states interacting together. In the next section we discuss an example that provides some formal reasons why this will be so. This one–many correspondence between symbols and states will gradually lose its abstract character as we derive better neural equations for the states.

Given such a one–many correspondence between symbols and states, the question why simple behavioral symbols "seem" to be simple becomes an urgent one. A partial answer is fortunately already provided by studying the way in which (**) learns the alphabet ABC \cdots Z.

Consider a machine \( \mathcal{N} \) obeying (**) that can learn the alphabet. Once the alphabet is learned, a chain of associations
\[ F_{A} \equiv F_{ABC} \equiv F_{ABD} \cdots F_{XYZ} \equiv F_{YZ} \equiv 1 \]
is formed. The alphabet seems simple to us once we know it because we can simply "rattle it off" once we choose to do so. Can \( \mathcal{N} \) "rattle off the alphabet?" The answer is manifestly yes.

Simply present \( A \) to \( \mathcal{N} \) after \( \mathcal{N} \) has learned the alphabet. Then \( x_A(t) \) grows and sends a signal only to \( v_B \). A large output \( x_A(t) \) (or \( O_A^T(t) \)) is created at \( v_B \) \( \tau \) time units later. Suppose that this output rapidly creates a feedback input to \( v_A \) through the medium surrounding \( \mathcal{N} \) (much as we "hear ourselves talk"). Then \( x_B(t) \) grows further and a signal is sent only to \( v_C \). A large output \( x_A(t) \) occurs from \( v_C \) \( \tau \) time units later, and hereby creates a rapid feedback input to \( v_A \). The process of sending signals one
LEARNING, INFORMATION, AND INHIBITION

After the alphabet is learned, each subset of it has some status as a simple symbol, and thus all the subsets constructed from A, B, and C can contribute to a prediction of D, but to differing degrees. Thus we write

\[ (A)(B)(C) \]
\[ (ABC) \]
\[ (BC) \]
\[ (AB) \]
\[ \ldots \]
\[ \Omega. \]

How much weight does each of these subsets carry in the prediction of D? The answer manifestly depends on the amount of time that elapses between presentation of successive letters to the machine [1]. If this time is large, then only C immediately precedes presentations of D, and so only (C) can influence the prediction of D. If A has been said long ago, but B and C are presented rapidly before D is, then only (B), (C), and (BC) can influence the prediction of D. We must therefore seek a way of mathematically expressing the fact that only sets of letters that were presented contingently in time to a given letter substantially influence the prediction of that letter.

There is a simple way to say at time \( t \) that B and C have recently been said. It is: the product \( x_B(t)x_C(t) \) is large. Indeed, if any set

\[ I(J) = \{ r_k: k \in J \equiv \{ j_1, j_2, \ldots, j_m \} \} \]

where \( J \) is large. We need now only find a way to say that if \( I(J) \) and \( r_j \) have been consecutively presented very often in the past, then presenting \( I(J) \) alone in the future generates an output from \( v_j \). Model (199) solves this problem for the special case that

\[ I(J) = \{ r_j \}. \]

Since the interaction corresponding to the transition \( r_i \rightarrow r_j \) is

\[ \beta x_i(t - \tau) v_j(t). \]
we suppose by analogy that the interaction of the transition \( I(J) \rightarrow r_j \) is given by
\[
\dot{z}_{J}(t) = -\alpha z_{J}(t) + \beta \rho_{J} \prod_{k \in J} \xi_k(t - \tau) \eta_J(t).
\]
It remains only to define \( \eta_J(t) \) in terms of the correlations \( z_{J_n}(t) \). Formally this can be done by analogy with (5) as
\[
\eta_J(t) = \rho_{J} \sum_{m=1}^{m=1} p_{J_m} \xi_{J_m}(t).
\]
We have hereby derived the following equations (system (***)) for an excitatory embedding field with higher-order correlations.

\[
\dot{x}_i(t) = -Lx_i(t) + \rho \sum_{J \subseteq \mathcal{K}} \prod_{k \in J} x_k(t - \tau) \eta_J(t) + I_i(t).
\]

\[
\dot{y}_i(t) = \rho_{J} \sum_{m=1}^{m=1} p_{J_m} \xi_{J_m}(t).
\]

\[
\dot{z}_{J_n}(t) = -\alpha z_{J_n}(t) + \beta \rho_{J_n} \prod_{k \in J_n} \xi_k(t - \tau) \eta_J(t).
\]

\( i = 1, 2, \ldots, n \) where \( \mathcal{K} = \{1, 2, \ldots, n\} \).

This system has the following desirable property. If a set \( I(J) \) is often paired with \( r_j \), then \( z_{J_n} \) and hence \( \eta_J \) will grow. A future occurrence of \( I(J) \) will therefore guarantee that the interaction \( \prod_{k \in J} x_k(t - \tau) \eta_J(t) \) becomes large, and thus that the output \( x_i(t) \) is large. In this sense, the set \( I(J) \) eventually acts like a simple symbol within \( \mathcal{K} \). If \( I(J) \) is often paired with \( r_j \), then every subset \( I \) of \( I(J) \) will also be paired with \( r_j \). The relative contribution of each subset \( I \) to the output \( e_j \) depends on the relative timing whereby symbols in \( I \) and \( r_j \) are presented to \( \mathcal{K} \). Consider for example the case in which
\[
p_{J_n} = p^{\mid J \mid}, \quad |J| > 1,
\]
where \( |J| \) is the number of indices in \( J \) and \( p \) is chosen so small that
\[
p \xi_m(t) \ll 1
\]
for all \( t \geq 0 \). Start \( \mathcal{K} \) in a state of "rest" and "maximal ignorance" [1], and let
\[
I(t, f) = \{ f - t, f - t + 1, f - t + 2, \ldots, f - 1 \}.
\]
Then the correlations
\[
\dot{z}_{I(t, f)}(t)
\]
will be monotone decreasing in \( t \), corresponding to the idea that symbols \( r_{J_n} \), which occur long before \( r_j \) does, cannot easily become associated with \( r_j \).

Although the system (***), it embodies some useful formal properties, it is full of conceptual difficulties. These difficulties are, however, quite informative. We now discuss some of them.

A. Reducing Higher-Order Associations to Simple Associations

Where and how are the products \( \prod_{k \in J} x_k(t) \) computed? Figure 1 partially answers this question: it introduces a compound edge \( e_{J_n} \) for every set of indices \( J_n \), in addition to the simple edges \( e_{ij} \) from \( v_i \) to \( v_j \). The signals \( x_i(t), k \in J_n \), travel from \( v_k \) along \( e_{ij} \), until they reach \( v_j \), where the product \( \rho_{J_n} \prod_{k \in J_n} x_k(t) \) travels to the compound arrowhead \( x_{J_n}(t + \tau) \) correlates it with \( x_j(t + \tau) \).

The process whereby this product is computed and then correlated within \( x_{J_n} \) is entirely mysterious in Fig. 1. We will fortunately find, in Section 19, that the process mentioned in Sections 3 and 8 can (a) compute products of states, and can thereby (b) reduce higher-order associations \( \overline{x}_{J_n}(t) \) to combinations of first-order associations \( x_{J_k}(t) \). This reduction property provides an alternative tool for "simplifying" learned behavior without creating formidable technical difficulties, as the following paragraphs show.

B. Overlapping Nonlocal Patterns Reduce Background Noise

There is too much background noise in (**). Every time \( m \) points \( e_{ij}, j \in I(J) \), are rapidly excited, \( n(2^m - 1) - m \) edges of the form \( e_{K^t} \) are activated, over subsets \( K \subseteq J \) and \( t = 1, 2, \ldots, n \). The signals
LEARNING, INFORMATION, AND INHIBITION

Output $O_i^{(0)}(t)$ has a “local” form, however, in exactly one case, namely, if only the state $e_i$ has a large $x_i(t)$ value. Then

$$O_i^{(0)}(t) = \max(x_i(t) - \Gamma_o, 0),$$

which depends only on $x_i(t)$ and a known positive threshold $\Gamma_o$. The principle of sufficient reason can now be invoked to show that $x_i(t)$ must be replaced by (7) in all output expressions. To see this, let a “local” (or “near-sighted”) observer $\mathcal{O}$ inhibit the state $e_i$; $\mathcal{O}$ cannot determine whether or not exactly one state has a large value at time $t$, if only because signals from state to state are not transmitted instantaneously. Nor can $\mathcal{O}$ distinguish the destination of one output signal emitted by $e_i$ from that of any other. In this sense, sufficient reason requires that the output function from $e_i$ be independent of the distribution of values at other states. Since locality requires that this function be (7) for one distribution of values, (7) is the output for all distributions of values. In particular, the excitatory signal received by the arrowhead $N_{ij}$ from $e_i$ at time $t$ is, instead of $\beta_i(x_i(t - \tau_{ij})p_{ij})$,

$$\beta_i \max(x_i(t - \tau_{ij}) - \Gamma_o, 0)p_{ij},$$

where we have indexed the threshold $\Gamma_o$ by $i$ as well as by $t$ to admit the possibility that signals to different arrowheads have different thresholds. Similarly, $\beta_i$ can depend on the state $e_i$.

In order to unambiguously designate that (8) represents an excitatory signal, we label the parameters in (8) with superscript +; (8) becomes

$$\beta_i^+ \max(x_i(t - \tau_{ij}) - \Gamma_o, 0)p_{ij}.$$ (9)

To shorten the writing of (9), we introduce the notation

$$[a]^+ = \max(a, 0).$$ (10)

Then (9) becomes

$$\beta_i^+[x_i(t - \tau_{ij}) - \Gamma_o]^+]p_{ij}.$$ (11)

**Remark.** Given that $x_i(t)$ is always replaced by the local form $\max(x_i(t) - \Gamma_o, 0)$ of $O_i^{(0)}(t)$, how can we possibly recapture the nonlocal properties of $O_i^{(+)}(t)$ when more than one state has a large value? We will find that these nonlocal properties are consequences of the *interactions* both excitatory and inhibitory, between the various states.

The nonlocal association $y_{ij}(t)$ of (9) will now be replaced by the local association $\tau_{ij}(t)$ of (8). The excitatory signal received by $e_j$ from $e_i$ at time $t$ is therefore, by (11),

$$\beta_j^+[x_j(t - \tau_{ij}) - \Gamma_o]^+]p_{ij}^+\tau_{ij}(t).$$ (12)
The total excitatory signal received by \( r_i \) from all states \( v_m \) at time \( t \) is

\[
J_x(t) = \sum_{m=1}^{n} \beta_{mi}[v_m(t - \tau_{mi}) - \Gamma_{mi}^-] p_{mi}^- \tau_{mi}(t).
\] (13)

The inhibitory signal received by \( r_i \) from \( v_i \) at time \( t \) is defined by analogy with (12). Two considerations determine this signal: (1) an increase in \( v_i \) causes a decrease in \( x_i \) \( \tau_{ij}^- \) time units later; and (2) no reason

\[
x_i(t - \tau_{ij}^-) \rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ \rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]

\[
\rightarrow \beta_i^+ [x_i(t - \tau_{ij}) - \Gamma_{ij}^-] p_{ij}^+ z_i(t) \rightarrow \rightarrow \rightarrow
\]
only measure whether \( z_{jk}(t) \) is above or below various prescribed equilibrium and threshold values. Once this is realized, Eq. (3)
\[
\dot{z}_{jk}(t) = -u_{jk}(t) + \beta \gamma_{jk}(t - \tau_{jk}) \eta_{jk}(t)
\]
of (\( n \)) is easily adapted to the signed case.

If we let \( Q_{jk} \) be the new equilibrium value of \( z_{jk}(t) \), the decay term \(-u_{jk}(t)\) of (3) is generalized to
\[
u_{jk} [Q_{jk} - z_{jk}(t) - \nu_{jk} [Q_{jk} - Q_{jk}]].
\]
The term \( \beta \gamma_{jk}(t - \tau_{jk}) \eta_{jk}(t - \tau_{jk}) \) in (3) designates the excitatory signal received by \( N_{jk} \) from \( x_{ij} \) at time \( t \), which is replaced by (11) in the signed case. The term \( x_{ij}(t) \) in (3) need not be compared to its zero equilibrium in the signed case. It suffices to measure whether or not \( x_{ij}(t) \) exceeds some threshold value \( \lambda_{ij} \). The term \( x_{ij}(t) \) in (3) can therefore be replaced by
\[
[x_{ij}(t) - \lambda_{ij}]
\]
in the signed case. In all, (3) is replaced by
\[
\dot{z}_{jk}(t) = \nu_{jk}[Q_{jk} - z_{jk}(t)] - \nu_{jk}[z_{jk}(t) - Q_{jk}]
\]
\[
+ \beta \gamma_{jk}(t - \tau_{jk}) \eta_{jk}(t - \tau_{jk}) \left[ x_{ij}(t - \tau_{jk}) - \lambda_{ij} \right].
\]
The numerical factor \( \gamma_{jk} \) in (20) is merely a scaling factor. The only immediately obvious constraint to be placed on \( \lambda_{ij} \) is
\[
\lambda_{ij} \geq \nu_{jk},
\]
which means heuristically that \( z_{jk}(t) \) does not grow if \( x_{ij}(t) \) is “too small.” The choice \( \nu_{jk} = \nu_{jk} = 0 \) of parameters in (20) is not forbidden.

It remains only to rescale \( z_{jk}(t) \) in (17); \( z_{jk}(t) \) is replaced by \( [z_{jk}(t) - Q_{jk}] \), where \( Q_{jk} \) satisfies
\[
Q_{jk} \geq \nu_{jk},
\]
which means that \( z_{jk}(t) \) will not grow if \( x_{ij}(t) \) is “too small.”

Then (17) becomes
\[
\beta \gamma_{jk}(t - \tau_{jk}) \left[ x_{ij}(t) - \lambda_{ij} \right] - \nu_{jk} \left[ z_{jk}(t) - Q_{jk} \right],
\]
and (18) becomes
\[
J_{jk}(t) = \sum_{a=1}^{n} \beta \gamma_{ja}(t - \tau_{ja}) \left[ x_{ja}(t) - \lambda_{ja} \right] - \nu_{jk} \left[ z_{jk}(t) - Q_{jk} \right].
\]
The equation for \( x_{ij}(t) \) is therefore
\[
\dot{x}_{ij}(t) = \beta \gamma_{ij}(t - \tau_{ij}) \left[ x_{ij}(t) - P_{ij} \right] + J_{ij}(t) - J_{ij}(t) + I_{ij}^{+}(t) + I_{ij}^{-}(t),
\]
where \( I_{ij}^{+}(t) \) and \( I_{ij}^{-}(t) \) are known excitatory and inhibitory inputs, respectively; \( J_{ij}(t) \) is defined in (24); and \( J_{ij}^{+}(t) \) is defined in (15). Equations (26) and (25) arise from (\( n \)) merely by (a) admitting inhibitory signals, and (b) introducing the more general equilibria and thresholds required by (a).

12. A NEURAL INTERPRETATION

Reference [1] contains a qualitative neural interpretation of excitatory embedding fields. We now extend this interpretation to signed fields.

A. Cell Body, Axon, End Bulb, and Synapse

Inspection of Figs. 2 and 3 readily suggests a neural interpretation of these networks. Each point \( v_{i} \) corresponds to a nerve cell body, or to a cluster of mutually “indistinguishable” nerve cell bodies. Each edge \( e_{ji} \) corresponds to the axons, or clusters of axons, leading from \( v_{i} \) to \( v_{j} \). Each arrowhead \( N_{ij} \) corresponds to the end bulbs associated with the axons of \( e_{ji} \). And the spaces between the “end bulbs” \( N_{ij} \) and the “postsynaptic cell bodies” \( v_{i} \) are synapses.

B. Membrane Potential, Spiking Frequency, and Threshold

The process \( x_{ij}(t) \) sits in the cell bodies \( v_{i} \). It is therefore plausible to identify it with the average membrane potential of these cells: \( x_{ij}(t) \) gives rise to an excitatory signal along the axons \( e_{ji} \) that equals
\[
\beta \gamma_{ij}(t) \left[ x_{ij}(t) - \lambda_{ij} \right] p_{ij}^{+},
\]
This signal cannot represent an individual neural event, because \( x_{ij}(t) \) is itself an average. It can represent the frequency of individual neural signals in \( e_{ji} \), throughout a unit time interval, if such signals exist. Indeed neural signals exist, and an individual neural signal is called a spike. Expression (26) is thus the average spiking frequency created by the cells \( v_{i} \) in the axons \( e_{ji} \) at time \( t \). This signal equals zero whenever \( x_{ij}(t) \leq \lambda_{ij} \). Thus \( \lambda_{ij} \) is the spiking threshold, if such exists, which it does, and \( P_{ij} \) is the equilibrium potential of \( x_{ij}(t) \), if such exists, which it does. Given these identifications, (26) claims that the spiking frequency equals zero until the membrane potential rises above a spiking threshold that strictly exceeds the equilibrium potential. For suprathreshold values, average spiking frequency should be linearly related to membrane potential. (In the bounded case to be discussed in Section 14, “saturation” occurs at large values of membrane potential, and thus also in the spiking frequency.)

These claims are a natural consequence of an obvious neural labeling of our psychologically derived mathematical variables. It is therefore gratifying that some neural data behave like the claims. In particular, the
existence of a process \( x_i(t) \) fluctuating at \( r_i \) is well known [2, 3, 4]. That this process gives rise to signals in \( e_{ij} \) is also well known [4]. That a threshold exists above which these signals vary linearly in \( x_i(t) \) has also been reported [5, 16, 17, 18]. Our equations are also compatible with the so-called slow potential view, since the fluctuations in the average membrane potential \( v_i \) determine changes in spiking frequency along \( e_{ij} \) that activate the potentials of the recipient cells \( r_j \) [19, 20, 21], as the next paragraphs show.

C. Transmitter Production and Release

The excitatory signal (26) reaches the end bulbs \( N_{ij} \) at time \( t + \tau_{ij} \) and thereupon activates the process \( z_{ij}(t + \tau_{ij}) \). A quantity

\[
\beta^i_j \left[ x_i(t) - \Gamma_{ij}^* \right] p_{ij} \left[ z_{ij}(t + \tau_{ij}) - \Lambda_j \right] \tag{27}
\]

is then released into the “synaptic clefts” between \( N_{ij} \) and the postsynaptic cell bodies \( r_i \), and thereupon causes a proportional change in the rate of change, namely, \( \dot{x}_i(t + \tau_{ij}) \), of the “average postsynaptic potential.” A transmission from end bulbs over the synapses to postsynaptic cells is experimentally well established [3, 22, 23]. We therefore identify (27) with the amount of excitatory transmitter released into the synaptic clefts at time \( t + \tau_{ij} \). The process \( z_{ij}(t + \tau_{ij}) \) taking place in the end bulbs \( N_{ij} \) corresponds to the transmitter production and storage process in these end bulbs at time \( t + \tau_{ij} \). The transmission law (27) then says that the average amount of excitatory transmitter released from \( N_{ij} \) into the synaptic clefts increases if either the presynaptic spiking frequency or the amount of available transmitter in \( N_{ij} \) increases.

D. Presynaptic and Postsynaptic Control of Transmitter Production

Equation (20) can now be read as an equation describing the average rate of transmitter production in \( N_{ij} \). The most important term in (20) is

\[
\beta^i_j \frac{P_{ij} - z_{ij}}{P_{ij} + z_{ij}} \left[ x_i(t) - \Lambda_j \right]. \tag{28}
\]

As this expression increases, the rate of excitatory transmitter production also increases; (28) can be increased either by increasing the spiking frequency (11)

\[
\beta^i_j p_{ij} \left[ x_i(t) - \tau_{ij} \right] - \Gamma_{ij}^* \tag{27a}
\]

of \( e_{ij} \), or by increasing the postsynaptic potential \( x_i(t) \) of \( e_{ij} \), and thereby increasing (19)

\[
\left[ x_i(t) - \Lambda_j \right]. \tag{27b}
\]

LEARNING, INFORMATION, AND INHIBITION

Both (11) and (19) must be positive in order for \( z_{ij}(t) \) to grow. This property leads to the prediction that excitatory transmitter production in \( N_{ij} \) depends on both the presynaptic spiking frequency in \( e_{ij} \) and the postsynaptic potential of \( r_i \). In particular, if the value of \( x_i(t) \) is depressed below its equilibrium potential \( P_e \) by an inhibitory input to \( r_i \), then since \( \Lambda_j > P_e \), \( z_{ij}(t) \) cannot grow. This is, moreover, just the kind of inhibition that the transformation \( z_{ij}(t) \rightarrow j_{ij}(t) \) in (28b) describes.

E. Exponential Decay of Membrane Potential

Expression (16) states that the average membrane potential decays at an exponential rate to its equilibrium potential in the absence of incoming transmissions from other cells. This fact has been experimentally reported [2].

13. THE HARTLINE-RATLIFF EQUATION

In a series of distinguished papers (e.g., [5, 24–26]), Hartline, Ratliff, and their colleagues have developed an empirical equation concerning the inhibitory interactions within the Limulus retina. This equation has the following form.

\[
r_i = e_i - \sum_{j=1}^{n} K_{ij} \left[ r_j - r_i^0 \right] \tag{29}
\]

where \( i = 1, 2, \ldots, n \); \( e_i \) is the frequency of impulses within the \( i \)th ommatidial axon under a fixed light source in the absence of inhibition from nearby cells. When inhibitory contributions are not negligible, \( e_i \) is reduced to \( r_i \), which is the net frequency of discharge of impulses in the \( i \)th axon due to the combined effects of the light source and the inhibition from nearby cells. The \( K_{ij} \) are “inhibitory coefficients,” and the \( r_i^0 \) are “threshold frequencies.” We now show that the Hartline-Ratliff equation can be formally derived from (25). We carry out the derivation in the simplest possible way to clearly delineate the main ideas. In particular, various geometrical complexities of the Limulus retinal network that can, in principle, be discussed using (25) will be ignored.

All interactions in (29) are inhibitory. Thus we set all \( f_{ij}^* = 0 \), ignoring the possibility that excitatory links activate the inhibition indirectly. Only the light sources perturb the retina, and so all \( I_i(t) \equiv 0 \). Equation (25) becomes

\[
x_i(t) = \frac{\beta^i_j [P_i - x_i(t)] - \Gamma_{ij}^* [x_i(t) - P_i]}{J_i^* + I_i(t)}. \tag{30}
\]
The light source is stationary in time. That is,
\[ I^+_i(t) = I^+_i = \text{constant}. \]
Equation (29) describes the retina's steady-state response to this source.
In other words,
\[ x_i(t) = 0 \quad \text{and} \quad x_i(t) = x_i = \text{constant}, \]
for all \( i. \) Two cases now arise.

Case 1. Only one ommatidium, say \( r_j, \) receives light. Therefore \( r_j \)
receives no inhibition from other \( r_i, \) and since \( x_i \) is excited above its
equilibrium value by the light,
\[ x_i(P_i - x_i) = 0 \quad \text{and} \quad x_i(x_i(t) - P_i) = x_i(x_i(t) - P_i). \]
Equation (30) becomes
\[ 0 = -x_i(x_i(t) - P_i) + I^+_i, \]
or
\[ x_i = P_i + \frac{1}{\alpha_i} I^+_i. \quad (31) \]
The spiking frequency along the ommatidium axon is, as in (26), of the form
\[ e_i = \mu_i^+ [x_i - \Gamma_i^+], \quad (32) \]
where \( \mu_i^+ \) plays the role of a composite coefficient \( \beta_i^+ \rho_i^+ \), and \( \Gamma_i^+ > P_i. \)
By (31),
\[ e_i = \mu_i^+ \left[ P_i - \Gamma_i^+ + \frac{1}{\alpha_i} I^+_i \right], \]
and choosing the light \( I^+_i \) sufficiently intense to make \( e_i \) suprathreshold,
\[ e_i = \mu_i^+ \left( P_i - \Gamma_i^+ + \frac{1}{\alpha_i} I^+_i \right). \quad (33) \]

Case 2. Let several ommatidia receive intense light. Then not all
\( J_i^+ = 0. \) Since all \( x_m(t) \) are constant, (15) becomes
\[ J_i^+ = \sum_{m=1}^{n} \beta_m [x_m(t) - \Gamma_m^+] \rho_m, \]
and thus for any index \( i \) such that \( x_i \geq P_i, \) (25) becomes
\[ 0 = -\alpha_i (x_i - P_i) + I^+_i - \sum_{m=1}^{n} \beta_m (x_m - \Gamma_m^+) \rho_m. \]
Rearranging terms yields
\[ x_i = P_i + \frac{1}{\alpha_i} I^+_i - \frac{1}{\alpha_i} \sum_{m=1}^{n} \beta_m (x_m - \Gamma_m^+) \rho_m. \quad (34) \]
Let us now suppose that our assumptions on \( I_j^r \) and \( \rho_j^r \) are not so unreasonable as to entirely invalidate the formulas (38) and (39): (38) then shows that the inhibitory coefficients \( K_{il} \) are composites of the rate parameters \( \rho_j^r \), \( \mu_j^r \), \( \rho_j^r \), and \( \alpha_j^r \), and of the inhibitory path weight \( \gamma_j^r \). If the shape and size of the retinal cells (say, at the retina's border, or due to its elliptical symmetry) vary systematically with retinal position, then the rate parameters need not be independent of \( i \) and \( j \). In this case, measurements of \( K_{il} \) might well give a systematically distorted view of the path weights \( \gamma_j^r \).

A curious fact emerges if we interpret \( \gamma_j^r \) as a threshold spiking frequency, as (29) bids. Then \( \gamma_j^r > 0 \) and (39) implies
\[
\Gamma_j > \Gamma_i^r.
\]
(40)

This inequality is plausible, for example, if the ommatidium axon must start firing before inhibitory interactions can set in.

14. BOUNDED SIGNED EMBEDDING FIELDS

The system of (20) and (25) has the property that the values \( x_i(t) \) can become arbitrarily large if the inputs \( I_j^r(t) \) are chosen sufficiently large. It is in some ways physically more plausible to assume that \( x_i(t) \) has fixed bounds that cannot be exceeded under any circumstances, much as we find in (2') and (3'). We therefore derive below the natural adaptations of (2') and (3') to the signed case.

Suppose that \( x_i(t) \) has a fixed maximum \( M_i \) and a fixed minimum \( m_i \). The excitatory term due to excitatory signals \( J^r_i(t) \) from states \( e^r \), and to other excitatory sources \( I^r_i(t) \) can be written down by direct analogy with (2') in the form
\[
x^r_i (M_i - x_i(t)) J^r_i(t) + I^r_i(t),
\]
(41)
with initial data that satisfy \( x_i \leq M_i \). The inhibitory term due to inhibitory signals \( J^r_i(t) \) from states \( e^r \), and to other inhibitory sources \( I^r_i(t) \) can be written down by analogy with (41) in the form
\[
-x^r_i (x_i(t) - m_i) J^r_i(t) + I^r_i(t),
\]
(42)
with initial data subject to the constraint \( x_i \geq m_i \). It remains only to determine the decay term for \( x_i(t) \). One possibility is to simply use the decay term of (16). Another possibility arises by noting in (41) and (42) that \( x_i(t) \) is compared with its extrema \( M_i \) and \( m_i \), rather than with its equilibrium value \( P_i \). We therefore single out for particular attention the following equation for \( x_i(t) \):
\[
x_i(t) = x_i(M_i - x_i(t)) J^r_i(t) + I^r_i(t) \]
\[
- x_i^r (x_i(t) - m_i) J^r_i(t) + I^r_i(t).
\]
(43)
with decay term
\[
x^r_i (M_i - x_i(t)) - x^r_i (x_i(t) - m_i).
\]
(44)
and initial data subject to the constraints \( m_i \leq x_i \leq M_i \). The inequalities \( m_i \leq x_i \leq M_i \) are then satisfied for all time, since \( x_i(t) = m_i \) implies \( x_i(t) \geq 0 \), and \( x_i(t) = M_i \) implies \( x_i(t) \geq 0 \).

The equilibrium \( P_i \) of \( x_i(t) \) in (43) is defined by the value of \( x_i \) that satisfies (43) when \( \dot{x}_i(t) \) and all inputs \( J_i^r(t) \), \( I_i^r(t) \), \( J^r_i(t) \), and \( I^r_i(t) \) are zero. Under these circumstances, (43) becomes
\[
0 = x^r_i (M_i - P_i) - x^r_i (P_i - m_i),
\]
or
\[
P_i = \frac{x^r_i (M_i + x^r_i (M_i - m_i))}{x^r_i + x^r_i}.
\]
(45)

A similar argument holds if \( z_{jk}(t) \) has the fixed maximum \( M_{jk} \). Then the term (28)
\[
\beta_{jk}^r x^r_{jk} [x_{jk}(t) - \gamma_{jk}] [x_{jk}(t) - \gamma_{jk}]
\]
in (20) is premultiplied by \( (M_{jk} - z_{jk}(t)) \) and the initial data of \( z_{jk} \) are constrained by the inequality \( z_{jk} \leq M_{jk} \). If \( z_{jk}(t) \) also has a fixed minimum \( m_{jk} \), then the initial data satisfy \( z_{jk} \geq m_{jk} \), and the decay term of (20) either remains unchanged or is replaced by an expression of the form
\[
\frac{1}{\beta_{jk}^r} (M_{jk} - z_{jk}(t)) - \frac{1}{\beta_{jk}^r} z_{jk}(t) - m_{jk})
\]
by analogy with (44). In the latter case,
\[
\dot{z}_{jk}(t) = - u_{jk}(z_{jk}(t) - m_{jk}) + (M_{jk} - z_{jk}(t)) \times (u_{jk} + \beta_{jk}^r) \gamma_{jk} [x_{jk}(t) - \gamma_{jk}] [x_{jk}(t) - \gamma_{jk}]
\]
(46)
The possibility that \( u_{jk} = \gamma_{jk} = 0 \) is not ruled out.

15. INFORMATION AND LATERAL INHIBITION

This section introduces a signed embedding field \( z_{jk} \) that illustrates the connection between information and lateral inhibition that was conjectured in Section 3. For simplicity we consider a signed field that is a direct analog of an outstar [9]. The smallest (excitatory) outstar is pictured in
Fig. 4. Section 3 shows that the transformation $z_i \rightarrow y_1$ in an excitatory outstar can be replaced by mutual inhibitory signals between the states $r_2$ and $r_3$ in the outstar's border. We therefore introduce the signed outstar of Fig. 5. By analogy with the case of excitatory outstars, we call $r_1$ the excitatory source, $V^+ = \{r_2, r_3\}$ the excitatory border, and $V^- = \{r_4, r_5\}$ the inhibitory border of $\mathcal{M}$. The inhibitory signal from $r_3$ to $r_5$ is created in the following way: $r_3$ sends an excitatory signal to $r_1$, and $r_4$ thereupon sends an inhibitory signal to $r_5$. Similarly, $r_3$ inhibits $r_5$ by sending an excitatory signal to $r_6$, which thereupon sends an inhibitory signal to $r_2$. Just as in an excitatory outstar, only the set of points $V_o = \{r_4, r_5, r_6\}$ receives inputs from an experimentalist $E$, and only these points send outputs back to $E$: that is,

$$I^+_i(v_i) = I^+_i(0), \quad i = 4, 5.$$  

Set $V_o$ is therefore called the set of observable states of $\mathcal{M}$. The set $V^+_o = \{r_4, r_5\}$ merely mediates inhibitory signals between the observable states and is therefore called the set of unobservable states of $\mathcal{M}$. We also suppose that all inputs from $E$ to $\mathcal{M}$ are excitatory, by analogy with the case of excitatory outstars; that is,

$$I^+_i(v_i) = 0, \quad i = 1, 2, 3.$$  

The parameters of $\mathcal{M}$ are chosen in a "homogeneous" way. By this we mean two things. First, we think of the points in $V^+_o$ and $V^-_o$ as belonging to (possibly) two distinct "cell types." Mathematically speaking, this means that all "local" parameters of points in $V^+_o$ and (separately) of points in $V^-_o$ are the same (e.g., $\alpha_i = \beta_i = \gamma_i$ and $\delta_i = \rho_i$). Second, we impose a "cylindrical symmetry" on $\mathcal{M}$ by requiring that all parameters remain unchanged when points in $V^+_o$ are interchanged and then corresponding points in $V^-_o$ are interchanged (e.g., $\alpha_i = \beta_i = \gamma_i$ and $\delta_i = \rho_i$). This condition says that the geometry of $\mathcal{M}$ does not bias the learning of either $r_i$ or $r_j$. In all, the following constraints on parameters are imposed.

a. Rate constants

$$\tilde{z}_i = z_i, \quad \tilde{y}_i = y_i, \quad \tilde{\alpha}_i = \alpha_i, \quad \tilde{\beta}_i = \beta_i, \quad \tilde{\gamma}_i = \gamma_i, \quad \tilde{\delta}_i = \delta_i, \quad i = 1, 2, 3,$$

and

$$\tilde{z}_i = z_i, \quad \tilde{y}_i = y_i, \quad \tilde{\alpha}_i = \alpha_i, \quad \tilde{\beta}_i = \beta_i, \quad \tilde{\gamma}_i = \gamma_i, \quad \tilde{\delta}_i = \delta_i.$$  

b. Interaction coefficients

$$p_{12} = p_{13}, \quad p_{24} = p_{25}, \quad p_{34} = p_{35}, \quad p_{45} = p_{46},$$

and

$$\gamma_{12} = \gamma_{13}, \quad \gamma_{24} = \gamma_{25}, \quad \gamma_{34} = \gamma_{35}.$$  

All other coefficients $p_{ij}$ and $\gamma_{ij}$ are zero. It therefore suffices to specify the following thresholds, time lags, and equilibrium constants.

c. Thresholds

$$\Gamma_{12} = \Gamma_{13}, \quad \Gamma_{24} = \Gamma_{25}, \quad \Gamma_{34} = \Gamma_{35}, \quad \Omega_{12} = \Omega_{13}, \quad \Omega_{24} = \Omega_{34},$$

and

$$\Lambda_2 = \Lambda_3, \quad \Lambda_4 = \Lambda_5.$$  

d. Time lags

$$\tau_{12} = \tau_{13}, \quad \tau_{24} = \tau_{25}, \quad \tau_{34} = \tau_{35}, \quad \tau_{45} = \tau_{56}.$$  

e. Equilibrium constants

$$P_1 = P_2 = P_3, \quad P_4 = P_5,$$

and

$$Q_{12} = Q_{13}, \quad Q_{24} = Q_{25}.$$  

---

Mathematical Biosciences 4 (1969), 255-310
LEARNING, INFORMATION, AND INHIBITION

value, and in particular below the threshold $\Gamma_\delta > P_2$. Thus by (51), $O_2(t) \equiv 0$, as expected. Moreover, since $\Gamma_{\delta \gamma} > P_2$, $r_2$ never sends an inhibitory signal to $r_1$ via $r_\gamma$, so that $O_2(t)$ is indeed never influenced by inhibition from $r_\gamma$.

Notice that these conclusions do not depend on the numerical values of the parameters, but only on their “homogeneity.”

B. Equal Border Values

Given an excitatory outsize with a common border value $x_b(t) = x_b(t)$, no matter how large, then $O_2^{(\delta)}(t) = 0$. This fact is translated in a signed outsize as follows: equal reciprocal inhibition between $r_2$ and $r_\gamma$ keeps $x_{\gamma}(t)$ and $x_b(t)$ so small that $O_2(t) \equiv 0$. The following thought experiment shows how this inhibition takes place.

Let $\mathcal{K}$ be in equilibrium until time $t = 0$, and suppose that equal inputs $I_2^{(\delta)}(t)$ and $I_\gamma^{(\delta)}(t)$ are received by $r_2$ and $r_\gamma$ thereafter. By homogeneity, $x_b(t) \equiv x_b(t)$ for all $t \geq 0$, and in particular $O_2(t) \equiv 0$. We wish to choose the parameters of $\mathcal{K}$ to guarantee that $O_2(t) \equiv 0$, which by (51) is the same as

$$x_{\gamma}(t) < \Gamma_\delta, \quad t \geq 0. \tag{52}$$

If $I_2^{(\delta)}(t)$ has very small values, this is easily done. By “very small,” we mean the following. Consider the thought experiment of Section 15. A, in which only $I_2^{(\delta)}(t)$ is positive: $I_2^{(\delta)}(t)$ is very small if $\Gamma_{\delta \gamma}$ is larger than the maximum of $x_b(t)$ in that experiment. The only nontrivial case arises when $I_2^{(\delta)}(t)$ is sufficiently large to drive $x_b(t)$ above the fixed threshold $\Gamma_\delta$ at some time. Suppose that such an $I_2^{(\delta)}(t)$ occurs in the present experiment, along with an equal $I_\gamma^{(\delta)}(t)$. Inequality (52) can then be achieved for all $t \geq 0$ only if inhibitory signals between $r_2$ and $r_\gamma$ take effect before the inputs drive $x_b(t)$ above $\Gamma_\delta$, Various constraints must be imposed on the parameters of $\mathcal{K}$ to guarantee that this occurs. In order to conveniently discuss these constraints, let

$$T_1 = \text{minimum time needed for } I_2^{(\delta)}(t) \text{ to drive } x_b(t) \text{ above } \Gamma_\delta \text{ if } I_\gamma^{(\delta)}(t) = 0,$$

and

$$T_2 = \text{minimum time needed for } I_2^{(\delta)}(t) \text{ to create an excitatory signal in } r_\gamma,$$

An inhibitory signal from $r_\gamma$ will arrive at $r_2$ before $x_b(t)$ exceeds $\Gamma_\delta$ if and only if

$$T_1 + \tau_{\delta 1} + T_2 + \tau_{\delta \gamma} < T_1. \tag{53}$$

Mathematische Zeitschriften 4 (1969), 255-310
Since the expression

\[ T_s = T_a + \tau_{2a} + T_b + \tau_{3b} \]

is the total time needed for a signal to be created in \( r_3 \) and to reach \( r_2 \), (53) implies

\[ T_{s2} < \Gamma_{2c}^- \]  \hspace{0.5cm} (54)

Otherwise, no signal could leave \( r_3 \) until \( x_2(t) \) reached \( \Gamma_{2c}^- \), by homogeneity between \( r_2 \) and \( r_3 \).

Inequality (53) cannot be guaranteed for an arbitrary input \( I_2^*(t) \). By (54), it suffices to show that there exists an \( I_2^*(t) \) such that

\[ \tau_{2a} + \tau_{3b} > T_s \]  \hspace{0.5cm} (55)

given fixed parameters for \( \kappa \). This is readily done by letting \( \tau_{2a} \) equal a constant \( \Gamma_{2c}^- \) for all \( t \geq 0 \), and then choosing this constant so large that (55) is satisfied. The following properties of (48) show that this can always be done.

Before an inhibitory signal reaches \( r_2 \) from \( r_3 \), (48) implies

\[ \frac{dx_2}{dt} \geq D_{12}(t) + I_2 \]  \hspace{0.5cm} (56)

since

\[ \beta_1^* \frac{D_2}{p_2^*} [x_3(t) - \Gamma_{1c}^-] \]

is nonnegative, and

\[ \beta_1^* \frac{D_2}{p_2^*} [x_3(t) - \Gamma_{1c}^-] \]

equals zero. Since \( \kappa \) starts out in equilibrium and no inhibition has reached \( r_2 \) in (56), (56) implies

\[ \frac{dx_2}{dt} \geq \alpha_2^* (P_1 - x_2(t)) + I_2^* \]

which readily yields

\[ x_2(t) \geq \alpha_2^* (P_1 - x_2(t)) + I_2^* \]

where

\[ X_2(t) = P_1 + \frac{I_2^*}{\alpha_2^*} [1 - \exp(-x_2(t))] \]

If \( I_2^* \) is chosen greater than \( \alpha_2^* (P_1 - \Gamma_{1c}^-) \), then \( X_2(t) \) achieves the value \( \Gamma_{1c}^- \) at time

\[ T_s = \frac{1}{\alpha_2^*} \log \left[ 1 - \frac{\alpha_2^* (P_1 - \Gamma_{1c}^-)}{I_2^*} \right] \]

and \( X_2(t) > \Gamma_{1c}^- \) for all \( t > T_s \). Since \( \lim_{t \to \infty} x_2(t) = 0 \), \( x_2(t) \) can be made to exceed any fixed threshold \( \Gamma_{1c}^- \) in an arbitrarily short time simply by increasing \( I_2^* \). In fact, it suffices to choose any \( I_2^* \) such that

\[ I_2^* > \alpha_2^* (\Gamma_{1c}^- - P_1) \]

and

\[ \tau_{2a} + \tau_{3b} > \frac{1}{\alpha_2^*} \log \left[ 1 - \frac{\alpha_2^* (\Gamma_{1c}^- - P_1)}{I_2^*} \right] \]

to violate (53).

A necessary condition for (53) to hold is thus that there exists a fixed finite number \( N_{20}^{(s)} \) such that

\[ I_2^*(t) \leq N_{20}^{(s)} < \infty \]  \hspace{0.5cm} (57)

for all \( t \geq 0 \) and all inputs \( I_2^* \) that ever perturb \( r_2 \). Inequality (57) is always fulfilled if \( I_2^*(t) \) is a signal created by a point whose state function has a finite maximum, as in Section 14.

It can also readily be seen that (57) must be supplemented by a condition of the form

\[ \frac{d}{dt} I_2^*(t) \leq N_{20}^{(s)} \leq \infty \]  \hspace{0.5cm} (58)

for all \( t \geq 0 \) and all inputs \( I_2^* \) that ever perturb \( r_2 \). Figure 6 illustrates the need for (58); it describes two different inputs \( I_{22}^*(t) \) and \( I_{23}^*(t) \) delivered to \( r_3 \) in two identical copies \( \kappa_2 \) and \( \kappa_3 \), respectively, of the signed outster \( \kappa \). The input \( I_{22}^*(t) \) to \( \kappa_2 \) is a rectangular pulse taking on the value \( N_{20}^{(s)} \) for all \( t \geq 0 \), and thus satisfies (57). The parameters in \( \kappa_2 \) are chosen to keep \( x_2(t) \leq \Gamma_{2c}^- \) for all \( t \geq 0 \), but in such a way that \( T_{22}/T_{21} \approx 1 \). That is, the inhibitory signals arrive just before \( x_2(t) \) reaches \( \Gamma_{2c}^- \) and just barely manage to keep \( x_2(t) \) below \( \Gamma_{2c}^- \).

\( \kappa_3 \) has the same parameters as \( \kappa_2 \), but the inhibitory signal created in \( \kappa_3 \) at time \( T_{23} \) is smaller than the signal created in \( \kappa_2 \) at time \( T_{22} \) because \( \delta < N_{20}^{(s)} \). This inhibitory signal can be made as small as we please by choosing \( \delta \) so that \( x_2(t) \geq \Gamma_{2c}^- \) for \( t \equiv T_{23} \), as (49) shows. \( \epsilon \) can be chosen in such a way that the inhibitory signal arrives when \( I_{22}^*(t) = N_{20}^{(s)} \), and thus, by (48), when \( x_2(t) \) is growing rapidly. The inhibitory signal is
often too weak to overcome the large excitatory signal, and thus the
inequality \( x_2(t) \leq \Gamma_{2}^+ \) is violated for some \( t \geq 0 \). This difficulty arises
because the small inhibitory signal created by \( r_1 \) at a given time is not
large enough to overcome the large excitatory input that exists at \( r_1 \) when
it arrives after transit through \( c_{21}^- \) and \( c_{22}^- \). Inequality (58) overcomes this
difficulty by guaranteeing that the early values of \( I_{2}^-(t) \), which create
inhibitory signals, are not too much smaller than the later values of
\( I_{2}^+(t) \) with which the inhibitory signals compete; (58) holds whenever
\( I_{2}^+(t) \) is the output signal of a bounded state.

For any particular choice of bounds \( N_1^{[0]} \) and \( N_2^{[0]} \) on admissible inputs,
our task is to find parameters of \( \lambda \) that create and deliver inhibitory
signals of sufficient strength and with sufficient rapidity to overcome the
effects of prolonged excitation. As \( N_1^{[0]} \) and \( N_2^{[0]} \) are allowed to increase,
the parameters \( T_{2} \), \( \tau_{12}^- \), \( T_{3} \), and \( \tau_{13}^- \) must be chosen smaller. \( T_{2} \) and \( T_{3} \) can
be decreased, for example, by decreasing the differences \( \Gamma_{22}^- - \Gamma_{1} \) and
\( \Gamma_{33}^- - \Gamma_{2} \) between thresholds and equilibrium values. To guarantee that a
sufficiently strong inhibitory signal arrives at \( r_2 \) from \( r_1 \), we can amplify
the signal as much as we please by choosing arbitrarily large values of
\( c_{22}^- \). This extra degree of freedom in amplifying inhibitory signals is a
major advantage of using two cell types \( V_{0} \) and \( V_{1} \).

C. Learning

We now show that large values at the states need not cause any changes
in the associations unless these values represent a learning experiment.

For example, let \( \lambda_{i} \) be in equilibrium until time \( t = 0 \). Let any
admissible input \( I_{1}^+(t) \) occur at \( r_1 \) (i.e., \( A \) is said to \( \lambda_{i} \)). Equal excitatory
signals are sent to \( r_1 \) and \( r_2 \), but since \( z_{12} \) and \( z_{13} \) begin in equilibrium, the
signals reaching \( r_2 \) and \( r_3 \) are zero, by (48) and (49), and \( z_{12} \) and \( z_{13} \) never
leave equilibrium. In short, saying \( A \) teaches us no list \( AB \) or \( AC \), so the
associations do not change.

More generally, let \( z_{12}(t) = z_{13}(t) > \Omega_{12}^- \) until time \( t = 0 \). Let \( A \) occur once again. Then equal and positive signals reach \( r_2 \) and \( r_3 \) from \( r_1 \),
but these signals create inhibitory signals between \( r_2 \) and \( r_3 \) that keep
\( x_2(t) \) and \( x_3(t) \) small. In fact, by choosing \( \lambda_{2} \) sufficiently large, we can
guarantee, as in the previous section, that \( x_2(t) \) and \( x_3(t) \) never exceed
\( \Lambda_{2} \). Then

\[
[x_2(t) - \Lambda_{2}^+] = [x_3(t) - \Lambda_{2}^+] = 0,
\]
and by (49),

\[
z_{12}(t) = z_{13}(t) < 0,
\]

for all \( t \geq 0 \). Even a large signal to \( A \) alone need not make the associations
grow.

These examples show, parenthetically, that by choosing sufficiently
large initial \( z_{12}(0) \) and \( z_{13}(0) \) values, even a small input to \( r_1 \) can create
signals at \( r_2 \) and \( r_3 \) that violate (57) and (58) for fixed \( N_1^{[0]} \) and \( N_2^{[0]} \). This
difficulty need not arise if \( z_{12}(t) \) and \( z_{13}(t) \) obey bounded equations.

Now let \( \lambda \) be in equilibrium until time \( t = 0 \) and present only the letter
\( B \) to \( \lambda \); that is, only \( I_{2}^+(t) \) is ever positive. Then only \( x_2(t) \) ever exceeds
a threshold, and in particular,

\[
[x_2(t) - \Lambda_{2}^+] = 0, \quad t \geq 0;
\]

\( z_{12}(t) \) and \( z_{13}(t) \) once again remain at equilibrium for all \( t \geq 0 \).

Finally let \( A \) be presented to \( \lambda \) at time \( t_{1} \), and let \( B \) be presented to
\( \lambda \) at time \( t_{2} \), \( t_{1} > t_{2} \). That is, inputs occur at \( r_1 \) and \( r_2 \) with a time
separation of \( t_{1} - t_{2} \). If \( t_{1} \) is so much greater than \( t_{2} \) that the signal
created by \( I_{2}^+(t) \) from \( r_1 \) to \( r_2 \) traverses \( c_{12} \) and decays before \( I_{2}^+(t) \) becomes
positive, then once again no learning occurs, as we see by a straightforward
application of the preceding remarks. If, however, this signal arrives at \( \lambda \) as \( I_{2}^+(t) \) becomes positive, then both

\[
[x_2(t) - \Lambda_{2}^+] = [x_3(t) - \Lambda_{2}^+] = 0
\]

will be positive, and so \( z_{12}(t) \), and only \( z_{12}(t) \), grows.

6. LEARNING DECREASES REACTION TIME

Daily life amply illustrates that familiar behavior sequences can often
be emitted more rapidly than unfamiliar ones. This also happens in our
machines, as we illustrate in the simplest possible case of Fig. 7. Let the
equations of this machine \( \lambda \) be unbounded, for simplicity. Then

\[
\dot{x}_{1}(t) = D_{11}(t) + I_{1}(t),
\]

\[
\dot{x}_{2}(t) = D_{22}(t) + I_{2}(t) + \beta_{1} \Gamma_{22}^+ [x_{2}(t) - \Omega_{12}^-],
\]

\[
\dot{z}_{12}(t) = D_{122}(t) + \beta_{1} \Gamma_{22}^+ [x_{2}(t) - \Omega_{12}^-][x_{2}(t) - \Lambda_{2}^+],
\]

\[
I_{1}(t) = D_{122}(t) + \Gamma_{22}^+ [x_{2}(t) - \Omega_{12}^-][x_{2}(t) - \Lambda_{2}^+].
\]

Fig. 7

Mathematical Biosciences 4 (1969), 255-310
and
\[ O_2(t) = \mu_2 \overline{X}(x_2(t) - \Gamma_k^+). \] (62)

Let \( r_1 \) and \( r_2 \) be in equilibrium until time \( t = 0 \); that is, \( x_i(t) = P_i, \ i = 1, 2 \). Suppose that \( \tau_{12}(0) > \Omega_{12} \), so that signals from \( r_2 \) reach \( r_1 \). Let a positive input \( I_1^+(t) \) perturb \( r_1 \) at time \( t = 0 \), and let \( I_2^-(t) = 0 \). The reaction time \( \tau = \tau(I_1^+) \) of \( \mathcal{M}_1 \) to the input \( I_1^+ \) is defined as the total time that elapses from the onset of \( I_1^+ \) at time \( t = 0 \) until the output \( O_2(t) \) from \( r_2 \) becomes positive for the first time, which by (62) is
\[ \tau = \inf \{ t; x_2(t) > \Gamma_k^+ \}. \]

\( \tau \) is the sum of three factors:
\[ \tau = T_1 + \tau_{12} + T_2, \]
where \( T_1 = \) minimum time needed for \( I_1^+ \) to drive \( x_1(t) \) above the threshold \( \Gamma_{1,2}^- \) and to thereby create a signal in \( e_1^+ \); \( \tau_{12} \) = time required for the signal to traverse \( e_1^+ \); and \( T_2 = \) minimum time needed for the signal from \( e_1^+ \) to \( r_2 \) to drive \( x_2(t) \) above the threshold \( \Gamma_k^+ \) after the signal reaches \( r_2 \) at time \( T_1 + \tau_{12} \). That is,
\[ T_1 = \inf \{ t; x_1(t) > \Gamma_{1,2}^- \} \]
and
\[ T_2 = \inf \{ t - T_1 - \tau_{12}; x_2(t) > \Gamma_k^+ \}. \]

Whereas \( \tau_{12} \) is constant, \( T_1 \) is a functional of \( I_1^+ \), and \( T_2 \) is a functional of \( I_2^- \) and \( \tau_{12} \). It often suffices to approximate \( z_{12}(t) \) by its initial value \( z_{12}(0) \), since \( T_2 \) depends on the values of the slowly fluctuating \( z_{12}(t) \) only in the time interval \( [0, T_1 + \tau_{12} + T_2] \). We therefore suppose for simplicity that \( z_{12}(t) \equiv z_{12}(0) \) in all of the following remarks.

The qualitative behavior of \( T_1 \) and \( T_2 \) is easily found in the special case that \( I_1^+(t) = I_2^- \) = constant, \( t \geq 0 \). Then (59) readily shows that \( T_1 \) is a monotone decreasing function of \( I_2^- \), since \( x_1(t) \) is a monotone increasing function of \( I_1^+ \) for all \( t \geq 0 \). The signal from \( r_2 \) to \( r_1 \) is therefore also monotone increasing in \( I_2^- \), and thus, by (60), \( T_2 \) is a monotone decreasing function of \( I_2^- \). In all, both \( T_1 \) and \( T_2 \) are monotone decreasing functions of \( I_2^- \), so that also the reaction time \( \tau \) decreases as the input (or "energy") increases.

To study the effects of learning on reaction time, we need merely consider several copies \( \mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_n \) of the machine \( \mathcal{M}_1 \), each with identical initial data and input \( I_1^+(t) = I_2^-, \) and differing only in that \( \mathcal{M}_{i+1} \) has a larger \( z_{12}(0) \) value than \( \mathcal{M}_i \) has. That is, \( \mathcal{M}_{i+1} \) knows the transition from \( r_1 \) to \( r_2 \) better than \( \mathcal{M}_i \) does. Each of the machines \( \mathcal{M}_i \) has the same \( x_i(t) \) function, since they all have the same initial data and inputs. In particular, \( T_1 \) is the same in all the machines. But the signal from \( r_1 \) to \( r_2 \) is larger in \( \mathcal{M}_{i+1} \) than in \( \mathcal{M}_i \), as (60) shows, because \( z_{12}(0) \) is larger in \( \mathcal{M}_{i+1} \) than in \( \mathcal{M}_i \). Thus, \( x_2(t) \) grows faster in \( \mathcal{M}_{i+1} \) than in \( \mathcal{M}_i \), so that \( T_2 \) is smaller in \( \mathcal{M}_{i+1} \) than in \( \mathcal{M}_i \). Since \( \tau = T_1 + \tau_{12} + T_2 \), \( \tau \) is also smaller in \( \mathcal{M}_{i+1} \) than in \( \mathcal{M}_i \). That is, learning decreases the reaction time.

The dependence of \( \tau \) on \( I_1^+(t) \) and \( z_{12}(t) \) is easily found in the special case that \( I_1^+(t) = l = \) constant, \( z_{12}(t) \equiv z_{12}(0) \), and \( z_1^- = z_k^+ = z \). Then \( \tau \) is the smallest root of the equation
\[ x_2(t) = \Gamma_k^+ \] (63)
subjected to the constraint
\[ x_2(t) > 0. \] (64)

Since
\[ T_1 = \left\{ \begin{array}{ll} \infty & \text{if } l < z(\Gamma_{1,2}^- - P_1) \\ \frac{1}{2} \log \left[ \frac{I}{l - z(\Gamma_{1,2}^- - P_1)} \right] & \text{if } l > z(\Gamma_{1,2}^- - P_1) \end{array} \right. \]
\[ \tau = \infty \text{ unless } l > z(\Gamma_{1,2}^- - P_1). \] If \( l > z(\Gamma_{1,2}^- - P_1) \), then by (59) and (60), (63) implies the equation
\[ x_A = B e^z \]
where
\[ x = r - z_{12}, \]
\[ A = \frac{1}{z} \left( \log \left[ \frac{l}{l - z(\Gamma_{1,2}^- - P_1)} \right] \right), \]
and
\[ B = \frac{1}{z} \left( \frac{P_1 - L_{12}^- - \frac{z(\Gamma_{1,2}^- - P_2)}{P_2}}{P_1 P_2 \beta_1 \beta_2 (z_{12}(0) - \Omega_{12})} \right). \] (64) implies
\[ \tau > \tau_{12}^+ - \frac{1}{z} \left[ \frac{1}{l} \log \left[ \frac{l}{l - z(\Gamma_{1,2}^- - P_1)} \right] \right] \frac{z(\Gamma_{1,2}^- - P_2)}{P_2} + \frac{z(\Gamma_{1,2}^- - P_2)}{P_2 \beta_1 \beta_2 (z_{12}(0) - \Omega_{12})}. \]

These equations for the reaction time of \( \mathcal{M} \) in terms of \( l \) and \( z_{12}(0) \) must, of course, be modified in more realistic situations, where many components of the type illustrated in Fig. 7 mutually excite and inhibit one another before a peripheral input can give rise to a peripheral output. The next section describes some of the additional possibilities that arise as a result of inhibitory interactions between points.

---

Mathematical Biosciences 4 (1969), 255-310
17. Spatiotemporal Masking

The speeding up of reaction time due to prior learning helps to eliminate response interference, to spontaneously improve the memory of prior learning, and to create context effects.

Consider, for example, the homogeneous negative outset \( \mathcal{M} \) of Fig. 5. Suppose that the sequence \( r_{F2} \) has been substantially better learned than \( r_{F3} \) before time \( t = 0 \); that is, \( z_{130} \gg z_{230} \). Let all \( x_{i}(t) = F_{i} \) for \( i \in \{-F_{1}, 0, \} \), and suppose that a recall trial occurs at time \( t = 0 \); that is, \( I_{i}^{+}(t) \) becomes positive in an interval of the form \((0, z_{i})\), and all other inputs are identically zero. To avoid trivialities, suppose \( I_{i}^{+}(t) \) is sufficiently large to guarantee that \( x_{1} \) and \( x_{2} \) eventually exceed the thresholds \( \Gamma_{1}^{+} \) and \( \Gamma_{2}^{+} \), respectively.

By the homogeneity of \( \mathcal{M} \), \( I_{i}^{+}(t) \) creates equal signals from \( r_{1} \) along the excitatory edges \( e_{12}^{+} \) and \( e_{13}^{-} \) to the arrowheads \( N_{12} \) and \( N_{13} \). Since \( z_{130} \gg z_{230} \), the input from \( N_{13} \) to \( r_{2} \) created by this signal is substantially larger than the input from \( N_{12} \) to \( r_{3} \). Therefore, \( x_{2}(t) \) grows at a faster rate than \( x_{3}(t) \) does, and reaches larger values. In particular, \( x_{2}(t) \) generates an inhibitory signal from \( r_{2} \) to \( r_{3} \) before \( x_{3}(t) \) can generate an inhibitory signal from \( r_{3} \) to \( r_{2} \). If \( z_{230} \) sufficiently exceeds \( z_{130} \), then this inhibitory signal from \( r_{2} \) to \( r_{3} \) can reach \( r_{3} \) in force before \( x_{3}(t) \) generates an inhibitory signal to \( r_{2} \) and can keep \( x_{2}(t) \) below the inhibitory threshold \( \Gamma_{2}^{+} \). Under these circumstances, \( O_{2}(t) \equiv 0 \) since the inhibitory threshold \( \Gamma_{2}^{+} \) at \( r_{2} \) is smaller than the output threshold \( \Gamma_{2}^{+} \).

In short, speeding up the reaction time of well-learned behavior sequences inhibits the output from lesser learned sequences to subthreshold values.

The same argument shows that for \( z_{130} \gg z_{230} \) sufficiently large, \( x_{2}(t) \) can be kept below \( \Lambda_{2}^{+} \), and thus \( z_{230} \leqslant 0 \) for all \( t \geqslant 0 \), whereas there exist times \( t \) for which

\[
[x_{2}(t - \tau_{23}) - \Gamma_{23}^{-} x_{2}(t) + \Lambda_{2}^{+}] > 0
\]

during which \( z_{23}(t) \) grows.

In short, speeding up the reaction time of well-learned behavior sequences tends to preserve \( \mathcal{M} \)'s memory of these sequences.

These arguments do not require that \( z_{130} \) be at equilibrium, but only that \( z_{130} \gg z_{230} \). In other words, some memory of an \( r_{F2} \) transition can exist within \( \mathcal{M} \), but this memory never gives rise to an observable output signal \( O_{2}(t) \) if \( z_{130} \gg z_{230} \), because \( x_{2}(t) \) grows faster than \( x_{3}(t) \) grows, and can therefore inhibit \( x_{3}(t) \) to small values before \( O_{2}(t) \) becomes positive. Since a \( z_{23}(t) \) memory exists, but never appears in \( \mathcal{M} \)'s event behavior, we say that the inhibitory signal from \( r_{2} \) to \( r_{3} \) spatially masks the \( z_{23}(0) \) memory by cutting off the output from \( r_{2} \). Since this masking depends on the relative timing of the input signals received by \( r_{2} \) and \( r_{3} \), it is really more proper to call the masking process spatiotemporal masking.

The foregoing example of spatiotemporal masking depends on unequal \( z_{ij}(0) \) values. Section 16 suggests the existence of another form of masking due to unequal inputs \( I_{i}^{+}(t) \). Consider, for example, the machine \( \mathcal{M} \) of Fig. 8. Choose the parameters and the initial data of \( \mathcal{M} \) "homogeneously"; for example, \( x_{1}^{-} = x_{2}^{-} \), \( \rho_{1}^{+} = \rho_{2}^{+} \), and \( \rho_{1}^{-} = \rho_{2}^{-} \). In particular, let \( x_{1}(t) = x_{2}(t) = \rho_{1}^{-}, t \leqslant 0 \). For simplicity, let the inputs \( I_{i}^{+}(t) \) and \( I_{2}^{+}(t) \) have the graphs given in Fig. 9. That is, \( I_{i}^{+}(t) \) is a rectangular

\[
\begin{array}{c}
I_{i}^{+}(t) \\
\end{array}
\]

\[
\begin{array}{c}
\text{Mathematical Biosciences 4 (1969), 355-310}
\end{array}
\]
input pulse with onset time \( t = 0 \), and \( I_2(t) \) is a rectangular input pulse with onset time \( t = \epsilon > 0 \).

Although \( I_2(t) \) perturbs \( x_2(t) \) before \( I_1(t) \) does, we can often find a time lag \( \epsilon \) and an \( I_2^{*} \) that is sufficiently larger than \( I_1^{*} \) that the output \( O_3(t) = 0 \) for all \( t \geq 0 \), whereas \( O_3(t) \) becomes positive for some \( t \). That is, the input \( I_1(t) \) masks the earlier input \( I_2(t) \). The reasoning behind this conclusion is the same as in the case of \( z_{2h}(0) \neq z_{2h}(0) \).

Since
\[
x_1(t) = O_1(t) + I_1(t), \quad i = 1, 2.
\]
with \( x_2(t) = P_1 \) for all \( t \leq 0 \), we find
\[
x_1(t) = P_1 + \exp(-z_{1}(t)) \exp(z_{1}(t)) I_1(t) \, dt, \quad i = 1, 2.
\]
In particular,
\[
x_1(t) = \frac{I_1}{z_{1}(t)} [1 - \exp(-z_{1}(t))], \quad t \geq 0,
\]
whereas
\[
x_2(t) = \begin{cases} P_1 & \text{if } 0 \leq t \leq \epsilon, \\ P_1 + \frac{I_2}{z_{2}(t)} [1 - \exp(-z_{2}(t - \epsilon))], & \epsilon \leq t. \end{cases}
\]
Letting \( T_1^{*} \) be the minimum time needed for \( x_1(t) \) to generate an excitatory signal in \( e_1^{*} \), we have
\[
T_1^{*} = \begin{cases} \infty & \text{if } I_1 \leq z_{1}(\Gamma_{12} - P_1), \\ \frac{1}{z_{1}(t)} \log \frac{I_2}{I_{1}} & \text{if } I_1 > z_{1}(\Gamma_{12} - P_1). \end{cases}
\]
and
\[
T_2^{*} = \begin{cases} \infty & \text{if } I_2 \leq z_{2}(\Gamma_{12} - P_1), \\ \left( \epsilon + \frac{1}{z_{2}(t)} \log \frac{I_2}{I_{1}} \right) & \text{if } I_2 > z_{2}(\Gamma_{12} - P_1). \end{cases}
\]
x creates a signal in \( e_2 \) before \( x_4 \) can create a signal in \( e_2^{*} \) if and only if
\[
T_2^{*} > T_1^{*},
\]
which by (65) and (66) is equivalent to
\[
\frac{I_2}{I_1} \frac{z_{2}(\Gamma_{12} - P_1)}{z_{1}(\Gamma_{12} - P_1)} > \exp(\alpha_1), \quad (67)
\]
given also that \( I_1^{*} > z_{1}(\Gamma_{12} - P_1) \). Although \( r_1 \) receives an input before \( r_2 \) does, \( r_1 \) gives rise to a signal before \( r_3 \) does if \( I_2^{*} \) is so much larger than \( I_1^{*} \) that (67) is satisfied.

If (67) is satisfied, then \( x_4(t) \) begins growing before \( x_5(t) \) does, and \( r_3 \) sends out an inhibitory signal to \( r_2 \) before \( r_3 \) sends an inhibitory signal to \( r_4 \). If \( I_2^{*} \) is sufficiently large, the inhibitory signal from \( r_3 \) to \( r_2 \) can keep \( r_3 \) below the thresholds \( \Gamma_2^{*} \) and \( \Gamma_3^{*} \). Consequently, \( O_3(t) = 0 \), and \( x_4(t) \), which receives no inhibition from \( r_3 \), creates a positive \( O_3(t) \) signal for some \( t \geq 0 \). Again spatiotemporal masking has occurred.

Exciting the single point \( r_3 \) a great deal can also create a masking effect, which we also call a "warm-up," "practice," or "context" effect. This happens because exciting \( r_2 \) more than \( r_1 \) creates the inequality \( z_{2h}(t) \geq z_{3h}(t) \). If equal inputs are then presented to \( r_1 \) and \( r_2 \), equal outputs \( O_3(t) \) and \( O_3(t) \) will not be created, because the inhibitory signal from \( r_1 \) to \( r_2 \) will be larger than the inhibitory signal from \( r_2 \) to \( r_1 \).

The two ways of achieving masking in a homogeneous machine, namely, through inhomogeneous choices of inputs or of associations, can be mixed in more complicated machines to achieve some remarkably subtle effects. Masking can also be guaranteed by constructing machines whose geometry is not homogeneous.

18. REMEMBRANCES OF EVENTS LONG SUPPRESSED

We now briefly discuss an important case of spatiotemporal masking due to both the choice of inputs and of associations. Memories of experiences (say) from childhood that never come into consciousness in our customary adult environment can be triggered suddenly and with remarkable clarity by a fragment of our childhood environment. How can memories be stored with such clarity over such long time intervals without ever influencing our overt behavior or our conscious thoughts in a new environment? Spatiotemporal masking suggests a way, simply by pointing out that a learned transition whose output is inhibited by one distribution of inputs can nonetheless be strongly facilitated by a different distribution of inputs.

We will consider a very simple and highly idealized machine to illustrate this fact. Consider the machine \( A_6 \) of Fig. 10, which differs from the signed outstar of Fig. 9 only by the addition of a point \( r_6 \), an excitatory edge \( e_6 \), and an end bulb \( N_{62} \) impinging on \( r_6 \). We can apply all of the remarks concerning learned spatiotemporal masking from Section 17 to

the signed outstar part of \(\mathcal{M}\), which we denote by \(\mathcal{M}^\circ\). The "new environment" of \(\mathcal{M}\) consists of outputs only to \(\mathcal{M}^\circ\), whereas the "old environment" also permits inputs to \(r_0\). Observable learned transitions within \(\mathcal{M}\) can occur along \(\mathcal{E}_{13}\), \(\mathcal{E}_{15}\), and \(\mathcal{E}_{65}\). The inputs that activate these transitions are therefore concentrated at \(r_1\) and \(r_6\). In the new environment, only inputs to \(r_1\) can activate a transition, and since \(z_{1\tau} \gg z_{13}\), \(x_{1\tau}(t)\) is always inhibited by \(x_{13}(t)\) in the new environment, and only the output from \(O_3(t)\) ever becomes positive. In the old environment, inputs to both \(r_1\) and \(r_6\) can become positive. In particular, if \(I_1^+\) and \(I_6^+\) are simultaneously large, then even though the input from \(r_1\) to \(r_6\) exceeds the input from \(r_1\) to \(r_3\) because \(z_{1\tau} \gg z_{13}\), the total input from \(r_1\) and \(r_6\) to \(r_3\) can easily exceed the input received by \(r_2\) from \(r_1\). Therefore, \(x_{13}(t)\) can inhibit \(x_{1\tau}(t)\), and we can easily guarantee that only the output from \(r_2\) becomes positive. Thus the transition along \(\mathcal{E}_{13}\), which is not strong enough to create any output whatsoever in the new environment, can nonetheless overcome the stronger transition along \(\mathcal{E}_{1\tau}\) in the old environment with the help of the transition along \(\mathcal{E}_{13}\). Once this qualitative point is clearly understood, the reader can easily construct for himself more complicated and realistic instances of spatial-temporal masking due to a mixture of learned transitions and a fluctuating input environment.

19. REDUCING HIGHER-ORDER ASSOCIATIONS TO SIMPLE ASSOCIATIONS

Section 10 suggested that a signed embedding field can compute products \(\prod_{\alpha \in J} x_{\alpha}(t)\) of states even for large sets \(J\) of indices without introducing higher-order associations. We now introduce the simplest signed field that can (approximately) accomplish this.
Since the maximum input received by \( r_3 \) from either \( r_1 \) or \( r_2 \) is
\[
K_{13} = \sum p_i p_i^\prime M_i - \Gamma_{13} M_{13} - \Omega_{13},
\]
(74) implies that the maximum input received by \( r_3 \) from \( r_1 \) and \( r_2 \) is 2\( K_{13} \).

To guarantee that \( \tau_{31} \) responds only to the joint excitation of \( r_1 \) and \( r_2 \), choose the threshold \( \Gamma_{31} \) controlling the excitatory signal from \( r_3 \) to \( r_1 \), sufficiently large that a rectangular input to \( r_3 \) of size \( 2K_{13} \) cannot drive \( x_3(t) \) above \( \Gamma_{31} \), but sufficiently small that a rectangular input of size \( 2K_{13} \) can drive \( x_3(t) \) above \( \Gamma_{31} \). Any choice of \( \Gamma_{31} \) such that
\[
\frac{\gamma_{31} M_{31} \gamma_{31} \frac{K_{13}}{\gamma_{31}} + \frac{\gamma_{31} M_{31} \gamma_{31} \frac{K_{13}}{\gamma_{31}}}{\gamma_{31} \gamma_{31} + \gamma_{31} \gamma_{31} + 2K_{13}}}{\gamma_{31} \gamma_{31} + \gamma_{31} \gamma_{31} + 2K_{13}} < \Gamma_{31}
\]
and
\[
\Gamma_{31} < \frac{\gamma_{31} M_{31} \gamma_{31} \frac{K_{13}}{\gamma_{31}} + \frac{\gamma_{31} M_{31} \gamma_{31} \frac{K_{13}}{\gamma_{31}}}{\gamma_{31} \gamma_{31} + \gamma_{31} \gamma_{31} + 2K_{13}}}{\gamma_{31} \gamma_{31} + \gamma_{31} \gamma_{31} + 2K_{13}}
\]
accomplishes this goal, as (69) readily shows. For such a \( \Gamma_{31} \), \( r_3 \) does not send a signal to \( r_1 \) unless both \( r_1 \) and \( r_2 \) send signals to \( r_3 \). By (72), \( \tau_{31}(t) \) cannot grow unless both \( \tau_{31} \) and \( \tau_{31} \) are large at approximately time \( \tau = \tau_{31} + \tau_{31} - T_1 - T_2 \) where \( T_1 \) is the minimum time needed for \( x_3(t) \) and \( x_3(t) \) to exceed the threshold \( \Gamma_{31}^{+} \) and \( T_2 \) is the minimum time needed for \( x_3(t) \) to exceed the threshold \( \Gamma_{31}^{+} \) after the signals from \( x_{13} \) and \( x_{13} \) reach \( r_3 \). It is in this sense that \( \tau_{31}(t) \) computes the product
\[
\tau(t - \tau_{31} - \tau_{31} - T_1 - T_2) x_3(t) \tau(t - \tau_{31} - \tau_{31} - T_1 - T_2) x_3(t)
\]
In order that the argument above hold, \( \tau_3(t - \tau_{31} - T_2) \) and \( \tau_3(t - \tau_{31} - T_2) \) must also have large values due to frequent prior excitation of \( r_1 \) and \( r_2 \). A machine in which this warm-up effect does not occur can be constructed simply by eliminating the \( [\tau_{31} - \Omega_{31}] \) terms from (69), \( i = 1, 2 \). The foregoing argument also holds in an unbounded signed field whose inputs are subjected to (57), just so as \( \tau_{31} \) and \( \tau_{31} \) are not too large.

"Products" of the three or more components can be computed by a simple extension of the idea given above. Computing \( x_3(t) x_3(t) x_3(t) \) requires only a bounded signed field \( \mathcal{M} \) of the form pictured in Fig. 12, where we have chosen the time lags to satisfy \( \tau_{31} = \tau_{31} \) and \( \tau_{31} = \tau_{31} + \tau_{31} \). The thresholds \( \Gamma_{31} \) and \( \Gamma_{31} \) are chosen such that \( r_3 \) sends a signal to \( r_3 \) only if both \( r_1 \) and \( r_2 \) send signals to \( r_3 \), and \( r_3 \) sends a signal to \( r_5 \) only if both \( r_2 \) and \( r_3 \) send a signal to \( r_3 \). Thus \( r_3 \) sends a signal to \( r_5 \) only if all the values \( x_3(t) \), \( x_3(t) \), and \( x_3(t) \) are large at a common prior time. The size of \( x_3(t) \) therefore measures the frequency with which all these values have been large in the past.

---

Fig. 12

When a list (say, \( \tau_{13} \tau_{13} \)) is presented to a machine \( \mathcal{M} \) with time lag \( w \) between successive symbols, \( x_3(t) \) becomes large \( w \) time units before \( x_3(t) \) does, and \( x_3(t) \) becomes large \( w \) time units before \( x_3(t) \) does, so that it is desirable that \( \mathcal{M} \) compute the product
\[
x_3(t) x_3(t + w) x_3(t + 2w)
\]
rather than \( x_3(t) x_3(t) x_3(t) \). This can be done by simply choosing the time lags \( \tau_{31} \) of \( \mathcal{M} \) differently. For example, it suffices to let
\[
\tau_{31} = \tau_{31} = \tau_{31} = \tau_{31}
\]
and to choose \( w = \tau_{31} \), as the reader can easily check.

The idea of summing quantities and then truncating them by a threshold is not new, although it arises in our machines as a consequence of more fundamental considerations rather than as an end in itself. The introduction of correlations that measure these operations and control the size of future signals using these measurements is new, however, and opens up qualitatively new possibilities for the use of this old idea.

20. THE EXCITATORY-INHIBITORY DUALISM

Daily experience abounds in examples of a dualism between positive versus negative, or excitatory versus inhibitory, factors. This dualism is mathematically represented in our number system by such classes as positive versus negative numbers; in the physical description of nature in such examples as particles versus antiparticles; in our philosophies as such concepts as the yang versus the yin; and in our ethics in such values as good versus evil. A special case of an excitatory versus inhibitory dualism is also visible in the excitatory versus inhibitory interactions of a signed embedding field. We will treat this dualism as a fundamental principle for deepening our understanding of signed fields, rather than merely as an amusing formal consequence of our previous gedanken
experiments. This point of view seems inevitable if we ever wish to bring
the concepts stated herein into harmony with the results of other disciplines.

Invoking excitatory–inhibitory dualism (or EID) in the present context
has far-reaching concrete consequences as well as philosophical appeal.
It will, for example, lead us by simple formal manipulations to equations
that can, in a natural way, be interpreted to imply

(a) the existence of inhibitory as well as excitatory transmitter sub-

stances,

(b) the existence of two quantities that are formal caricatures of Na⁺

and K⁺ concentrations inside the cell membrane,

(c) the creation by excitatory transmitters of an inward flow of Na⁺

that induces an outward flow of K⁺ at suprathreshold values, whereas

only an outward flow of K⁺ is created by an inhibitory transmitter,

(d) the existence of two quantities that are formal caricatures of Ca²⁺

and Mg²⁺ concentrations inside the cell membrane,

(e) the binding of Na⁺, K⁺, Ca²⁺, Mg²⁺ within the end bulb in

complexes of varying strength to produce transmitter production and

release rates that are sensitive to prior presynaptic and postsynaptic levels

of membrane potential, and related facts and predictions.

Some qualitative insight also emerges concerning such fundamental
problems as the way in which a nerve cell’s functions in learning determines
its shape, and the way in which a nerve cell “knows” how much it must
produce to meet extracellular demands upon it. This article merely intro-
duces some of the formal machinery needed to derive these results and
insights. Later papers will investigate these and related topics in greater
detail.

21. THE COUPLING OF FORMAL Na⁺ AND K⁺ TO FORMAL
EXCITATORY AND INHIBITORY TRANSMITTERS

This section uses EID to show how the simplest features of (a)–(c) in
the preceding section can be derived. A bounded signed embedding field,
as in Section 14, will always be considered for definiteness.

Our derivation begins with the observation that a “sin of omission”
ocurred when we passed from (8) to (9) by adjoining + superscripts to
the parameters. Should not + superscripts be adjoined to the processes
xᵢ and zᵢ, as well? The answer surely is yes, since xᵢ and zᵢ must grow
in order to create a signal from vᵢ to vᵢ, and this property imparts to xᵢ and

\[
\gamma^- γ⁻(xᵢ(t) - mᵢ)
\]

an “excitatory polarity.” We therefore seek an equation for the
bounded variable xᵢ⁺.

We find this equation, at least formally, in (43) if we merely adjoin an
extra superscript + to all expressions therein. Then (43) becomes

\[
\dot{x}ᵢ^+(t) = \gamma^+(Mᵢ - xᵢ^+(t))(xᵢ^+(t)^- + Jᵢ^+(t) + Jᵢ^-(t)) - \gamma^- xᵢ^-(t) - mᵢ + \gamma^- γ⁻(xᵢ(t) - mᵢ)
\]

(75)

It remains only to determine the input functions Jᵢ^+(t) and Jᵢ^-(t); Jᵢ^+(t)
is readily found by adjoining sufficiently many + superscripts to Jᵢ(t) in
(24). Then

\[
Jᵢ^+(t) = \sum αᵢ [xᵢ(t) - \tauᵢ] - \tauᵢ) + \sum αᵢ [xᵢ(t) - \Omegaᵢ)]
\]

(76)

Input Jᵢ^+(t) is determined from Jᵢ^-(t) by invoking EID. This we do
in the most heuristic way possible to keep the physical meaning of the
results clear.

Compare the excitatory part

\[
\gamma^+(Mᵢ - xᵢ(t))
\]

of the decay term

\[
xᵢ^+ γ⁻(Mᵢ - xᵢ(t) - mᵢ)
\]

of (75) with its inhibitory part

\[
\gamma^-(xᵢ(t) - mᵢ)
\]

(77)

(78)

In (77), xᵢ⁺(t) is compared with its maximum Mᶠ, whereas in (78) xᵢ⁻(t)
is compared with its minimum mᵢ. We illustrate this situation in Fig. 13 in
a suggestive way: the figure shows that xᵢ⁺(t) has an excitatory polarity
in the excitatory part (77), which we denote by †, and an inhibitory polarity
in the inhibitory part (78), which we denote by ‡. Guided by this fact,
we invoke EID by requiring that all expressions in $J_i^{-}(t)$ have the opposite polarity to that of the corresponding expression in $J_i^{+}(t)$. For example, the expression $[x_{mi}^{+}(t - \tau_{mi}^{-}) - \Gamma_{mi}^{-}]^\ast$ in $J_i^{+}(t)$ corresponds to the expression $[\Gamma_{mi}^{+} - x_{mi}^{-}(t - \tau_{mi}^{+})]^\ast$ in $J_i^{-}(t)$. In all, we find, by reversing polarities in (76), that

$$J_i^{-}(t) = \sum_{m=1}^{n} \beta_{mi}^{-} [\Gamma_{mi}^{+} - x_{mi}^{-}(t - \tau_{mi}^{+})]^\ast \rho_{mi}^{-} [\Gamma_{mi}^{+} - z_{mi}^{+}(t)]. \quad (79)$$

Given the existence of a process $x_i(t)$. EID requires that there also exists a process $\chi_i(t)$, whose equation follows from (75) by changing the + superscripts corresponding to the + in $x_i$ to − superscripts. Then (75) becomes

$$\dot{x}_i^-(t) = x_i^-(M_i - x_i(t))(\gamma_i^- + J_i^-(t) + I_i^-(t))$$
$$- x_i^-(x_i(t) - \eta_i^+)(\gamma_i^+ + J_i^+(t) + I_i^+(t)); \quad (80)$$

$J_i^{-}(t)$ and $J_i^{+}(t)$ are also found by changing + superscripts corresponding to $x_i$ to − superscripts in (76) and (79). We find

$$J_i^{+}(t) = \sum_{m=1}^{n} \beta_{mi}^{+} [x_{mi}^{+}(t - \tau_{mi}^{-}) - \Gamma_{mi}^{-}]^\ast \rho_{mi}^{+} [x_{mi}^{+}(t) - \Omega_{mi}^{-}] \quad (81)$$

and

$$J_i^{-}(t) = \sum_{m=1}^{n} \beta_{mi}^{-} [\Gamma_{mi}^{+} - x_{mi}^{-}(t - \tau_{mi}^{+})]^\ast \rho_{mi}^{-} [\Gamma_{mi}^{+} - z_{mi}^{+}(t)]. \quad (82)$$

Equations (75) and (80) satisfy EID formally, and therefore seem to describe a complete symmetry between excitatory and inhibitory processes. Actually, as we will now show, these equations do not give rise to a sensible learning process unless the coefficients $\beta_{mi}^{+}, \beta_{mi}^{-}, \rho_{mi}^{+},$ and $\rho_{mi}^{-}$ are constrained in an asymmetric way. This “symmetry breaking” within the formal symmetry of (75) and (80) is needed to guarantee the “evolutionary trend” with an excitatory bias that we call “learning” in these systems. (Is it possible that symmetry breaking of ostensibly symmetric systems in various other physical disciplines occurs to guarantee analogous evolutionary trends?)

The need for symmetry breaking is clearly seen by considering $J_i^{-}(t)$ in (79). $x_{mi}^{+}(t)$ creates an inhibitory signal along $e_{mi}$ whenever $\Gamma_{mi}^{+} > x_{mi}^{+}(t)$ and $\rho_{mi}^{+} > 0$. Since $\Gamma_{mi}^{+} > P_{mi}^{+}$, an inhibitory signal is created whenever $x_{mi}^{+}(t)$ is at equilibrium and $\rho_{mi}^{+} > 0$. This conclusion is absurd, since we have introduced the threshold $\Gamma_{mi}^{+}$ to guarantee that only large $x_{mi}^{+}(t)$ values create signals. Inhibitory signals at equilibrium would ultimately wash away all learned associations. To eliminate this catastrophic possibility, we let

$$\beta_{mi}^{-} = 0, \quad (83)$$

and thus $J_i^{-}(t) \equiv 0$. By analogy, we also let $I_i^{-}(t) \equiv 0$. Equation (75) then becomes

$$\dot{x}_i(t) = x_i^{+} x_i^{-} (M_i - x_i(t))(\gamma_i^+ - J_i^{-}(t) + I_i^{-}(t))$$
$$- x_i^{+} x_i^{-} (x_i(t) - \eta_i^+)(\gamma_i^- + J_i^{+}(t) + I_i^{+}(t)); \quad (84)$$

Equation (84) still obeys EID formally, but the constraint (83) gives the equation an asymmetric appearance.

Since no actual inhibitory inputs are permitted in (84), the very concept of inhibition can only be salvaged if inhibitory inputs are permitted in (80). As in all our previous discussions, we must require that an inhibitory signal from $e_{mi}$ into $e_{mi}$ at time $t$ is created by an increase in $x_{mi}(t)$ above its inhibitory threshold $\Gamma_{mi}^{-}$. and therefore has the form

$$\beta_{mi}^{-} [x_{mi}(t) - \Gamma_{mi}^{-}] \quad (85)$$

One of the two output expressions

$$\beta_{mi}^{+} [x_{mi}(t) - \Gamma_{mi}^{+}] \quad (86)$$

or

$$\beta_{mi}^{+} [\Gamma_{mi}^{-} - x_{mi}(t)] \quad (87)$$

in $J_{mi}^{+}(t)$ and $J_{mi}^{-}(t)$, respectively, must therefore equal (85), for every $m = 1, 2, \ldots, n$. Suppose (86) equals (85). Then, as in (84), we must impose the constraint $\beta_{mi}^{+} = 0$ for every $m = 1, 2, \ldots, n$. Thus $I_i^{-}(t) \equiv 0$, and by analogy, $\dot{I}_i^{-}(t) \equiv 0$. $x_i^{-}(t)$ and $x_i^+(t)$ then obey equations of identical form in which no inhibitory signals occur. Inhibitory signals are only possible if (85) equals (87). Then

$$\beta_{mi}^{+} [x_{mi}(t) - \Gamma_{mi}^{+}]^\ast = \beta_{mi}^{-} [\Gamma_{mi}^{-} - x_{mi}(t)]^\ast \quad (88)$$

and

$$I_i^{-}(t) \equiv 0, \quad i, m = 1, 2, \ldots, n \quad (89)$$

and $I_i^{-}(t) \equiv 0, \quad i, m = 1, 2, \ldots, n$. Equation (80) becomes

$$\dot{x}_i(t) = x_i^{-} (M_i - x_i(t))(\gamma_i^{-} - J_i^{-}(t) + I_i^{-}(t)); \quad (90)$$

The important relation (88) implies that whenever $x_{mi}(t)$ has a suprathreshold value, then

$$\beta_{mi}^{+} x_{mi}(t) + \beta_{mi}^{-} x_{mi}(t) = \beta_{mi}^{+} \Gamma_{mi}^{+} + \beta_{mi}^{-} \Gamma_{mi}^{-} \quad (91)$$

is a constant.
That is, $x_m^-(t)$ and $x_m^+(t)$ are linearly, but antagonistically, coupled when $x_m^-(t)$ is driven to suprathreshold values by the excitatory input $J_m^+(t)$ in (84). The coupling (88) cannot be valid when $x_m^+(t)$ is driven to small values by the inhibitory input $J_m^-(t)$ in (90); otherwise, we must accept the absurd conclusion that a large inhibitory input to $v_m$ creates a large excitatory output from $v_m$. The coupling (88) is therefore broken by an inhibitory input signal. This situation can be formally expressed in terms of the function

$$
\chi(w) = \begin{cases} 
1 & \text{if } w > 0 \\
0 & \text{if } w \leq 0.
\end{cases}
$$

Then we replace (88) by

$$
\chi(x_m^-(t) - \Gamma_m^-)\beta_m^-[x_m^-(t) - \Gamma_m^-] - \beta_m^+[1 - \chi(x_m^-(t))] = 0,
$$

(92)

so that the coupling (88) holds only if

$$
\chi(x_m^-(t) - \Gamma_m^-) = 1,
$$

which is the same as saying that $x_m^-(t)$ has been driven to suprathreshold values by an excitatory input.

The foregoing simple manipulations can be summarized in the following way. The equation (43) contains a latent symmetry between its excitatory and inhibitory interactions. The effort to make this symmetry explicit shows that it describes a process $x_m^+(t)$ with positive polarity and a process $x_m^-(t)$ with negative polarity, along with four possible correlational (or associational) processes $z_m^+/z_m^-(t)$, $z_m^-/z_m^+(t)$, and $x_m^{+/+}(t)$. The four new processes $x_m^+(t)$, $x_m^-/z_m^+(t)$, $z_m^-/z_m^+(t)$, and $z_m^-/z_m^+(t)$ can be thought of as new formal degrees of freedom that must be coupled to the old variables $x_m^+(t)$ and $x_m^-/z_m^+(t)$ in such a way that the dynamics of learning (or the evolutionary trends) in our original psychologically derived equations are not lost. The mechanism needed for learning is, however, manifestly asymmetric with respect to excitatory and inhibitory interactions. We hereby find the equations (84), (90), and (92), which exhibit a symmetric coupling that must nonetheless sometimes be broken. In this sense, the modest symmetry breaking within the very special situation of (84), (90), and (92) reconciles, at least formally, two pervasive but not manifestly compatible tendencies within nature: namely, (i) the creation of systems whose excitatory and inhibitory interactions are as symmetric as possible, and (ii) the creation of systems that can benefit from their experience, and can thereby evolve to ever more efficient and elaborate levels of organization. We will presently see that this symmetry breaking seems to occur in vivo in the response of intracellular $Na^+$ and $K^+$ fluxes to excitatory and inhibitory transmitter substances.

Before establishing this fact, we remark that another antagonistic coupling is suggested by EID. Certainly the associations $z_{mi}(t)$ that occur in the expression $J_{mi}^-(t)$ should have the notation $z_{mi}^-(t)$. By (82) we therefore find that

$$
p_m^+[z_m^-(t) + z_{mi}^-] = p_m^-[z_m^-(t) + z_{mi}^-(t)],
$$

(93)

which means that at suprathreshold values of $z_{mi}^-(t)$.

$$
p_m^+[z_m^-(t) + z_{mi}^-(t)] = \text{constant},
$$

(94)

EID has hereby created the following expectations.

1. Two processes $x_m^+(t)$ and $x_m^-(t)$ exist in every cluster of "cell bodies" $v_m$.

2. These processes are antagonistically coupled at suprathreshold values, such that:

3. An excitatory signal to $v_m$ at time $t$ causes an increase in $x_m^+(t)$ within $v_m$, which in turn at suprathreshold values causes a decrease in $x_m^-(t)$ within $v_m$.

4. An inhibitory signal to $v_m$ at time $t$ merely causes a decrease in $x_m^-(t)$ within $v_m$.

5. A process $z_{mi}^-(t)$ occurs in every excitatory "end bulb" cluster $N_{mi}^+$ and contributes to the excitatory signal discussed in property 3.

6. A process $z_{mi}^+(t)$ occurs in every inhibitory end bulb cluster $N_{mi}^-$ and contributes to the inhibitory signal discussed in property 4.

Properties 1–6 suggest an obvious neural interpretation of the quantities $x_m^+(t)$, $x_m^-(t)$, $z_{mi}^+(t)$, and $z_{mi}^-(t)$. We introduce this interpretation herein in a qualitative way because our formal work must still be extended to achieve a quantitative connection. Let

$$\begin{align*}
x_m^+(t) &= \text{amount of } Na^+ \text{ inside the } v_m \text{ "cell membrane" at time } t, \\
x_m^-(t) &= \text{amount of } K^+ \text{ inside the } v_m \text{ "cell membrane" at time } t, \\
z_{mi}^+(t) &= \text{amount of excitatory transmitter inside the end bulbs } N_{mi}^+, \\
z_{mi}^-(t) &= \text{amount of inhibitory transmitter inside the end bulbs } N_{mi}^-.
\end{align*}
$$

The conclusions 1–6 now read:

1'. Each cell body cluster $v_m$ contains significant amounts of $Na^+$ and $K^+$,
2'. An antagonistic coupling between the amounts of Na⁺ and K⁺ in $e_i$ exists at suprathreshold values, such that:

3'. Excitatory transmitter causes an inward flow of Na⁺, which in turn at suprathreshold values causes an outward flow of K⁺;

4'. Inhibitory transmitter merely causes an outward flow of K⁺.

Moreover, the process by which the signal $[x^+_i(t) - I^+_i]^{c_j}$ traverses the axons $e^+_i$ and $e^-_i$ now reads:

5'. The spikes along $e^+_i$ are due to an inward flow of Na⁺ coupled to an outward flow of K⁺.

All of these phenomena have been experimentally reported [3, 4, 27-30]. A number of other interesting phenomena, some new, sit in the equations waiting to be interpreted. This will be done in a later article, which will also describe the equations for $z^-_m(t)$ and $z^+_m(t)$.

It is certain that our formalism is at best a rough description of neural events. In fact, later papers will show how to extend the formalism considerably in a rational way. Nonetheless it is gratifying that such highly nontrivial experimental properties as 1'-5' should have formal analogs in a theoretical picture that is derived in a simple and rather inevitable way from such basic principles as locality, principle of sufficient reason, and excitatory–inhibitory dualism, along with an elementary analysis of what we mathematically mean by learning. If the leap from mathematical to neural variables is accepted, then the results 1'-5' are consequences of these principles, and in this sense, we know "why" 1'-5' occur and how these experimental properties contribute to learning. Most important, some of the basic facts about Na⁺ and K⁺ fluxes, which heretofore have been thought of as part of the repetitive, and therefore stationary, responses of nerves to input signals are now implicated in the dynamics of neural learning, which is a nonstationary phenomenon.

Remark. Equations (84), (90), and (92) cannot possibly be in their final form. We can see this by analogy with (**), in which two kinds of transformations occur: "differential" transformations that describe the rate of change of a process using a differential equation, and "algebraic" transformations such as $z^+_i(t) = y^+_i$, which describe a very fast process in an approximate way. Equations (84) and (90) are examples of differential transformations, whereas (92) is a very fast process approximately described. Equation (92) manifestly describes a coupling of $x^+_i(t)$ and $x^-_i(t)$ within the cell bodies $v_i$, but thus far $v_i$ has no "cellular interior" in which to study the details of the coupling. Each point must be blown up into an extended cell body before (92) can be replaced by a differential transformation. How this can be done will be shown in a later paper, where dendritic and other effects of cell shape will also be discussed. Clearly a first step in this modification is to weaken the coupling between $x^+_i(t)$ and $x^-_i(t)$ in (90) and (92), so that an outward flux of K⁺ can be created by a linear mixture of inward excitatory transmitter-induced Na⁺ fluxes and of directly applied inhibitory transmitter at different spatial loci on the blow-up boundary of $v_i$. Otherwise (90) and (92) are in general incompatible as they stand.

ACKNOWLEDGMENTS

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-0294-0016).

REFERENCES