Some Developmental and Attentional Biases in the Contrast Enhancement and Short Term Memory of Recurrent Neural Networks

Stephen Grossberg† and Daniel Levine

Department of Mathematics, Massachusetts Institute of Technology Cambridge, Massachusetts 02139, *U.S.A.*

This paper studies the global dynamics of neurons, or neuron populations, in a recurrent on-center off-surround anatomy undergoing nonlinear shunting interactions. In such an anatomy, a given population excites itself and inhibits other populations. The interactions are defined by multiplicative mass action laws. Grossberg (1973) studied the case in which all populations have the same weight (or total number of unit cell sites). Here the effect of an arbitrary distribution of population weights is studied; each set of populations with equal weight is called a subfield. Possible causes of variable population weights are developmental biases (e.g., which feature detectors are represented in a field), attentional changes (e.g., which features are relevant at any time), and statistical errors in network design. Such factors can bias the total field towards accentuating or suppressing in short-term memory a given subfield of sensory features. In particular, a mechanism is noted for suppressing the activity of populations whose trigger features are infrequently experienced by the network. These variables interact with the recurrent on-center off-surround interactions, that have previously been shown capable of contrast enhancing significant input information, sustaining this information in short-term memory, adapting the field's total activity while producing multistable equilibrium points of this activity, suppressing noise, and preventing saturation of population response even to input patterns whose intensities are high.

1. Introduction

The recurrent on-center off-surround anatomy is found in many different neural structures. For instance, in hippocampus (Anderson, Gross, Lomo & Sveen, 1969) the main cell type, the pyramidal cell, emits axon collaterals to interneurons. Some of these collaterals or interneurons feed back excitatory

signals to nearby pyramidal cells. Other interneurons scatter inhibitory feedback signals over a broad area. Networks with variants of the same structure are found in other brain areas, such as neocortex (Stefanis, 1969) and cerebellum (Eccles, Ito & Szentagothai, 1967). Grossberg (1973) investigated how sensory pattern processing is improved in certain neural networks by introducing surround inhibition and by making the network recurrent or reverberating, rather than non-recurrent, or feed-forward. That paper described mathematical results relevant to these issues for the subclass of on-center off-surround networks where there are no structural biases in favor of one cell population over others, and the present paper extends those results to networks where a particular kind of bias exists.

The model studied in both papers emphasizes the properties of interacting populations of cell sites. These populations are general constructs that can be interpreted, for example, as populations of small membrane patches on individual cells or else as populations of whole cells. We assume that cell sites in a given population are distributed in such a fashion that their interactions are spatially random within each population and between population pairs. The model can therefore be defined by mass action laws governing the average potentials of network populations. Our excitatory and inhibitory interactions are of multiplicative or shunting type (Hodgkin, 1964; Sperling, 1970; Sperling & Sondhi, 1968), which occur (say) in passive membranes.

Denoting the average activity (e.g., potential) at time t of the ith population v_i by $x_i(t)$, $i = 1, 2, \ldots, n$, we will study how averages are transformed through time by recurrent on-center off-surround interactions (Fig. 1); that is, each population excites itself and inhibits other populations via the systems of equations

$$\dot{x}_i = -Ax_i + (B_i - x_i)f(x_i) - x_i \sum_{k \neq i} f(x_k) + I_i, \tag{1}$$

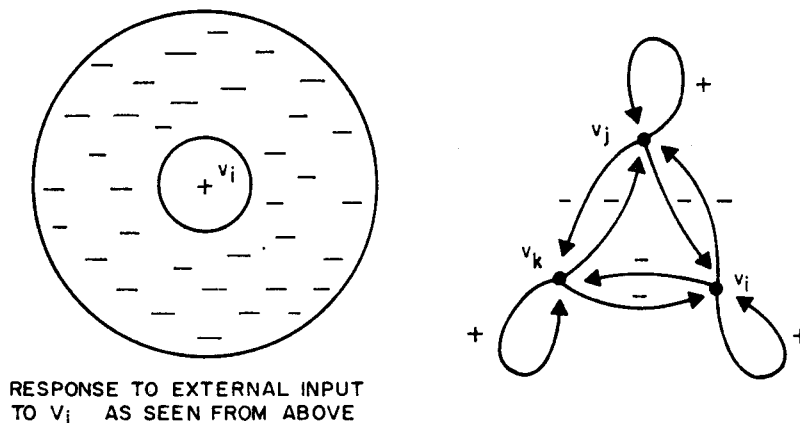RESPONSE TO EXTERNAL INPUT
TO V_i AS SEEN FROM ABOVE

Fig. 1. Recurrent on-center off-surround anatomy.

where $i = 1, 2, \ldots, n$, B_i is the "weight", i.e., maximum possible activity, of the ith population v_i of the network, and $x_i (0 \leq x_i \leq B_i)$ is the mean activity of v_i, interpretable either as a voltage or a number of active sites. Our system differs from Grossberg (1973) only in that our B_i may be any positive values, whereas in the earlier work all the B_i were equal. This generalization introduces both new physical properties and new mathematical difficulties. Four effects determine the behavior of the ith population v_i: (1) exponential decay, via the term $-Ax_i$; (2) shunting self-excitation, via the term $(B_i - x_i)f(x_i)$; (3) shunting inhibition from other populations, via the term

$$-x_i \sum_{k \neq i} f(x_k);$$

and (4) externally applied inputs, via the term I_i. The function $f(w)$ describes the mean output signal of a given population as a function of its mean activity w. In vivo, $f(w)$ is often a sigmoid function of w (Kernell, 1965a,b; Rall, 1955), and such a function arises from integrating a Gaussian, Cauchy, or other similar distribution of thresholds within a population. Grossberg (1973) showed that sigmoids are useful for the effective processing of signals in the presence of noise.

Why are the inhibitory surround and the reverberation needed at all? To motivate these interactions, consider what goes wrong in a system without interactions whose responses $x_i(t)$ to non-negative inputs $I_i(t)$ have the following natural properties: (1) linearity; (2) boundedness, say by B; (3) decay to equilibrium point, say 0, after inputs cease. Then

$$\dot{x}_i = -Ax_i + (B - x_i)I_i(t), \tag{2}$$

with $0 \leq x_i(0) \leq B$. Suppose that the relative sizes of the responses $x_i(t)$ code the relative importance of the information contained in the inputs $I_i(t)$. For example, if different populations v_i are excited by different features in a visual scene (e.g., colors, lines, edges, disparities), then the relative intensities $\theta_i = I_i(\Sigma_k I_k)^{-1}$ of the inputs give a measure of how much of that feature is in the scene. However, the equilibrium values of the responses $x_i(t)$, defined by $\dot{x}_i = 0$, satisfy

$$x_i = \frac{B\theta_i I}{A + \theta_i I} \tag{3}$$

where $I = \Sigma_k I_k$. Thus $x_i \rightarrow B$ as I becomes large, and the measure of relative importance is lost due to saturation. If this system contains noise, then the responses will not accurately measure relative importance when I is small. Hence the system is inadequate both at low and high total input intensities. Grossberg (1973) noted that an off-surround prevents this

T.B.   23

problem, since if

$$\dot{x}_i = -Ax_i + (B-x_i)I_i - x_i \sum_{k \neq i} I_k, \tag{4}$$

then the equilibrium value of x_i is

$$x_i = \theta_i \frac{BI}{A+I}, \tag{5}$$

which is proportional to θ_i no matter how large I is chosen. The off-surround hereby introduces a type of adaptation to overall levels of input activity by automatically changing the gain of the system.

In system (4), the constant B is the number of excitable sites in each population v_i. Since x_i is the mean number of active sites, $B-x_i$ is the mean number of inactive sites. Thus term $(B-x_i)I_i$ says that inactive sites are activated at a rate proportional to the number of inactive sites times the excitatory input intensity. Term

$$-x_i \sum_{k \neq i} I_k$$

says that active sites are inactivated at a rate proportional to the number of active sites times the total inhibitory input intensity. This is the meaning of "mass action" in this context.

The above example uses a nonrecurrent, or feed-forward, anatomy. A recurrent, or reverberating, anatomy is preferable when a network capable of short term memory (STM) is needed. Such a network can reverberate a pattern of activity distributed over cell populations for an indefinite interval of time. This reverberation can also be switched off rapidly by inhibitory inputs if a new pattern is delivered by external sources; the decay rates of individual cells can be large after the excitatory reverberating loop is broken by inhibition, even if the reverberation through an active excitatory loop is long lived. In a psychological context, the use of reverberation as a mechanism of STM has been suggested by Hebb (1949) and more recently by Estes (1972) and Grossberg (1971). For example, from operant conditioning experiments, one is led to seek reverberatory processes that can maintain in short term storage internal representations of sequences of external events until later rewards or punishments occur and transfer the memory of these sequences to long term storage (Grossberg, 1971).

Two main themes emerge in the discussion of reverberating networks. The first is: how does the reverberation change the distribution of activity across populations through time? In particular, how does the network suppress noise, or behaviorally irrelevant inputs, yet store behaviorally significant data in STM? This theme focusses on ways in which the *relative* sizes of population activities are transformed. The second theme is concerned with fluctuations in *total* network activity through time. In particular, when does

the total activity converge rapidly to zero, so that no stable reverberation gets established? A deeper issue is illustrated by equations (1) and (2). In equation (1), the *maximum* of the total activity

$$x(t) = \sum_{k=1}^{n} x_k(t)$$

when m populations are active is mB, since each active population has a maximum activity of B. In equation (2), the maximum of the total activity is B, since

$$x(t) = \frac{BI}{A+I},$$

which converges to B as $I \to \infty$. Thus the maximum total activity is independent of the number of active populations. This result from nonrecurrent networks has an analog in recurrent networks. Grossberg (1972c, 1973) has shown in the equal weight case, that if a persistent reverberation is established, then $x(t)$ will converge to a unique positive limit point as $t \to \infty$, for suitable choices of the signal function $f(w)$. It is also possible to find signal functions for which $x(t)$ converges to any one of a discrete set of limit points ("multi-stable equilibrium"), or even to a continuum of limit points. The situation in which one, or at most a few, limit points of total activity exist is typical. In all these cases, there is an upper bound on values of the equilibrium points that is independent of the number of populations in the network. This property is called *normalization*.

This latter property has important applications. As indicated above, it suggests a mechanism for establishing adaptation to fluctuations in total input intensity, as occurs in vision (Grossberg, 1970, 1972c). In this context, analogs of brightness and color constancy and contrast phenomena also arise. Second, it permits the construction of stable decision rules in spite of fluctuations in the number and intensity of active populations. This latter application is illustrated by the following example. A study of reinforcement mechanisms (Grossberg, 1972a,b) suggests that signals from external cue representations $(V_1 \to V_2)$ combine with signals from internal cues (i.e., drive inputs) at sites (V_2) which are capable of releasing signals $(V_2 \to V_3)$ that supply incentive motivation further downstream in the network (V_3) (see Fig. 2). The motivational signals should be released only if a large external cue input *and* a large internal drive input arrive at the motivational source cells V_2. The number and intensity of active external cue representations in V_1 can, however, vary between wide limits through time. If the total signal released by these populations also varied between wide limits, then the following dilemma would occur. Suppose that an external cue that excites only a small number of features (e.g., a pure tone) can release
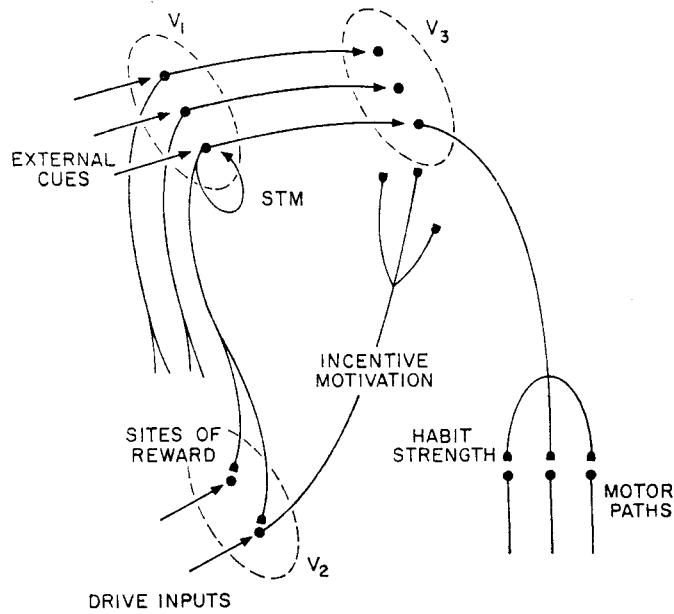
FIG. 2. Interaction of rewarded cues and drives to elicit motivation.

motivation in the presence of a suitable drive. Then an external cue that excites many features (e.g., a costumed figure) can release motivation independent of the drive level, because it controls a total signal to the motivational center that is many times larger than the signal produced by the pure tone. This dilemma is prevented by requiring that the total signal from the field of external representations has an upper bound that is independent of the number and intensity of active populations. An on-center off-surround network that joins these populations guarantees this formal property.

Third, this mechanism has been used to analyse some phenomena known from experiments on attention and discrimination learning (Grossberg, 1975a,b), such as overshadowing and behavioral contrast. It also sheds some light on how to construct position codes for motor control; for example, it suggests how to determine a fixed asymptotic position for eye movements in spite of fluctuations in target luminance (Grossberg, 1973, 1975a,b). In all of these examples, a total activity variable is, at least partially, conserved, whereas the distribution of activity across populations controls the relevant behavior.

An important theme about total activity regulation was analysed in the equal weight case: if the network can reverberate patterns imposed by behaviorally relevant inputs, then what prevents the network from reverberating behaviorally irrelevant activity levels, such as noise? Grossberg (1973) showed that a proper choice of the signal function $f(w)$ overcomes this dilemma. Below we will therefore classify our remarks according to how

the signal function is chosen. Given each choice of signal function, we will compare and contrast the global behavior of the equal weight (EW) case with that of the arbitrary weight (AW) case to see what new properties emerge in the more general situation.

Before presenting this classification, we will note some of the ways in which the AW case can arise. All properties of the EW case are due to interactions within the field of populations. They are effects of a mathematical marriage between the nonlinear shunt, the nonlinear signal function, and the recurrent on-center off-surround interaction pattern. The AW case can sometimes be interpreted as due to interactions originating from another field of cell populations (see Fig. 3). In Fig. 3, the AW field is designated
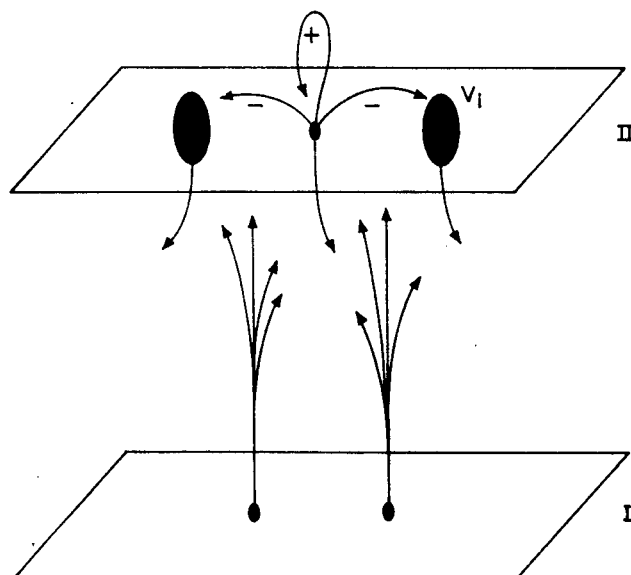

Fig. 3. Interaction of input paths to produce selective response profiles of variable weight.

by *II*. Population weight is indicated by the size of the population. The vertical arrows from *I* to *II* indicate a filtering process that combines outputs from cells in *I* to determine the response selectivity of cells in *II* (e.g., feature detector properties). See Grossberg (1970, 1972c) for a rigorous analysis of such a filtering process. Then B_i indicates the number of sites in *II* that can be excited by a particular configuration of inputs from *I*. Thus the AW case can arise from "vertical", or "external", interactions between fields, whereas the EW case describes the purely "horizontal", or "internal", effects of intrafield interactions.

Obviously not all possible features are equally represented in every field of neuron populations; various experiments have shown that the early

experience of animals can influence which feature detectors will develop. For example, depriving infant cats of monocular or binocular inputs will create a deficit of visual cortical cells that are sensitive to these features (Wiesel & Hubel, 1963, 1965). If kittens are raised viewing elongated patterns of one orientation, then the distribution of orientations of receptive fields in cortical cells is strongly biased towards that orientation, and discrimination of patterns at right angles to this orientation suffers (Blakemore & Cooper, 1970; Hirsch & Spinelli, 1970). Moreover, in a normal mature organism, the subfield of color detectors might dominate the subfield of orientation detectors (Honig, 1970). A relatively large B_i value in Fig. 3 indicates that the feature capable of exciting the ith population have a relatively large representation within the field II.

There also exists the possibility that transient shifts in performance variables can cause slow changes in B_i; for example, arousal inputs ("attentional" inputs) that increase the B_i parameters of a given subfield of cells by changing the maximal excitability of receptive populations of cell sites. Alternatively, attentional inputs could shunt the excitability of interactions within the field, either presynaptically or at interneuronal sites. Indeed the EW system

$$\dot{y}_i = -Ay_i + (B - y_i)f(\lambda_i y_i) - y_i \sum_{k \neq i} f(\lambda_k y_k) + I_i$$

is transformed into system (1) by the substitutions $x_i = \lambda_1 y_i$ and $B_i = B\lambda_i$. The "tuning" coefficients λ_i control the strength of shunting inputs to the interaction pathways of the field. Such changes in the excitability of populations would be slowly varying in time compared to the fluctuation rates of activities $x_i(t)$ in order for our model to hold.

Finally, there is the case in which nonuniformly distributed B_i parameters are due to statistical irregularities in the synthesis of an EW field. Our results then amount to statements about the stability of the EW case.

The classification of EW versus AW properties is listed below. This discussion describes how the reverberation transforms the network response to a briefly delivered pattern of inputs.

(A) CASE 1. LINEAR SIGNAL FUNCTION: $f(w) = Cw$

(i) *EW case*

Every initial pattern is preserved perfectly by this reverberation; that is, the relative activity functions

$$X_i(t) = x_i(t) \left[\sum_{k=1}^{n} x_k(t) \right]^{-1}$$

are constant. Moreover the total activity

$$x(t) = \sum_{k=1}^{n} x_k(t)$$

is normalized: the limit

$$x(\infty) = \lim_{t \to \infty} x(t)$$

exists, and equals zero or a positive constant E. Unfortunately, $x(\infty)$ has the same value *independent of the initial data*. For example, if the system does not reverberate noise (i.e., $x(0) \cong 0$ and $x(\infty) = 0$), then the system does not reverberate important inputs (i.e., $x(0) \gg 0$ and $x(\infty) = 0$); or if the system does reverberate important inputs (i.e., $x(0) \gg 0$ and $x(\infty) = E$), then it also amplifies and reverberates noise (i.e., $x(0) \cong 0$ and $x(\infty) = E$). Such a system cannot distinguish between unimportant and important inputs to be stored in STM.

This system therefore has a good and a bad property. The good property is that it can store any pattern without distortion, the bad one is that it amplifies noise. Case III below begins to show how to eliminate the bad property using a nonlinear signal function.

(ii) AW case

A form of contrast enhancement arises in this case. Label the populations so that $B_1 \leq B_2 \leq \ldots \leq B_n$. Only populations v_i such that $B_i = B_n$ can have their activities x_i reverberated in STM for an indefinite interval of time. The activities of all other populations converge to zero (see Fig. 4). Within the maximal subfield $\mathcal{F}_{\text{max}} = \{v_i : B_i = B_n\}$, the relative activities do not change. Thus the unequal weights B_i induce a form of contrast enhancement that is determined not by the activities $x_i(t)$, but by the field structure. Within the surviving subfield $\mathcal{F}_{\text{max}}$, the effect is just as in the EW case: any pattern's relative weights are preserved, and noise is enhanced as vigorously as signals.

Computer data shows that the enhancement of $\mathcal{F}_{\text{max}}$ activity occurs at a speed which depends on the differences between subfield weights. The enhancement effect is stable with respect to perturbations in the weights, in the sense that if the B_i are all close together, then the functions $X_i(t)$ will change very slowly. This is also true in all the cases below: small AW differences due to statistical imperfections in network structure cause slow changes in a prescribed direction beyond the transformations found in the EW case.

If differences in the B_i are large and due to developmental biases, then we note an interesting fact, which becomes increasingly important as $f(w)$ is chosen more realistically in the subsequent cases. Suppose, for definiteness, that the populations v_i represent feature detectors. Before an animal is born,
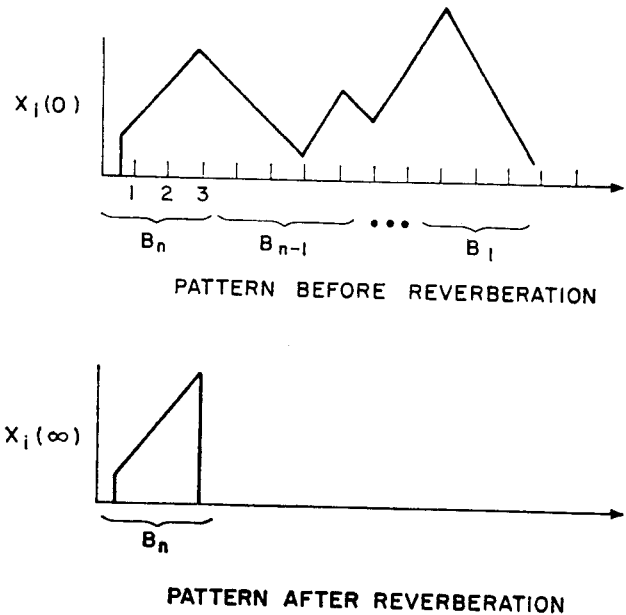
S. GROSSBERG AND D. LEVINE

FIG. 4. Transformation of input pattern by linear signal function.

it cannot know the relative frequencies with which all possible features will occur in its experiences. Its developmental mechanism must, however, be able to deal with all the features that might probably arise, even though some will turn out to be more important than others. The populations corresponding to infrequently occurring features produce a kind of un avoidable structural noise: they are there because they might have been needed. This example shows that structural noise can be totally suppressed, or masked, in the sense that activity in populations with small B_i can be annihilated by activity in populations with large B_i. A similar conclusion holds if the B_i are made large in a given time interval by attentional inputs. Then activity in unattended subfields can be suppressed by activity in attended subfields, no matter how large the initial activities are in the un- attended subfields. [Of course, whether a population will be attended to *in vivo* is not independent of its initial activity; see Grossberg (1975) for a study of some attentional mechanisms in which "vertical" feedback loops between fields are used to regulate attentiveness.]

(B) CASE II. SLOWER-THAN-LINEAR SIGNAL FUNCTION

(i) EW case

The reverberation either dies out or is normalized. If it is normalized, the limiting distribution of pattern weights is uniform; i.e., $X_i(\infty) = 1/n$, $i = 1, 2, \ldots, n$. Thus a population that is subjected to noise will ultimately

have the same weight as a population that receives a large signal. This situation is even more pathological than the linear case, since there, in the presence of signals, populations which receive only noise will maintain a small relative weight.

(ii) AW case

The reverberation again either dies out or is normalized. The limiting pattern weights are unifrom within each subfield, and the ordering of the values $X_i(\infty)$ is the same as the ordering of the B_i's (see Fig. 5). Thus any
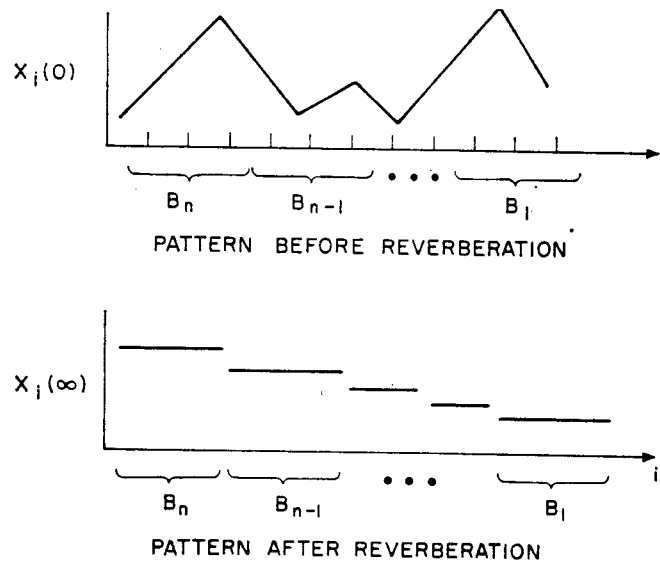

FIG. 5. Transformation of input pattern by slower-than-linear signal function.

noise, no matter how small, that impinges on a population of higher weight B_i will ultimately dominate even the largest signal to a population of lower weight. Examples show that the activities of populations of lower weight may or may not be completely suppressed. The slower-than-linear signal function is therefore unsatisfactory.

(C) CASE III. FASTER-THAN-LINEAR SIGNAL FUNCTION

(i) EW case

The main problem posed by the previous two cases is to suppress noise, or at least to prevent noise amplification. This occurs in the present case. Consider Fig. 6, where since $f(w)$ grows faster-than-linearly, the function $g(w) \equiv w^{-1}f(w)$ is strictly increasing. The parameters E_i are successive
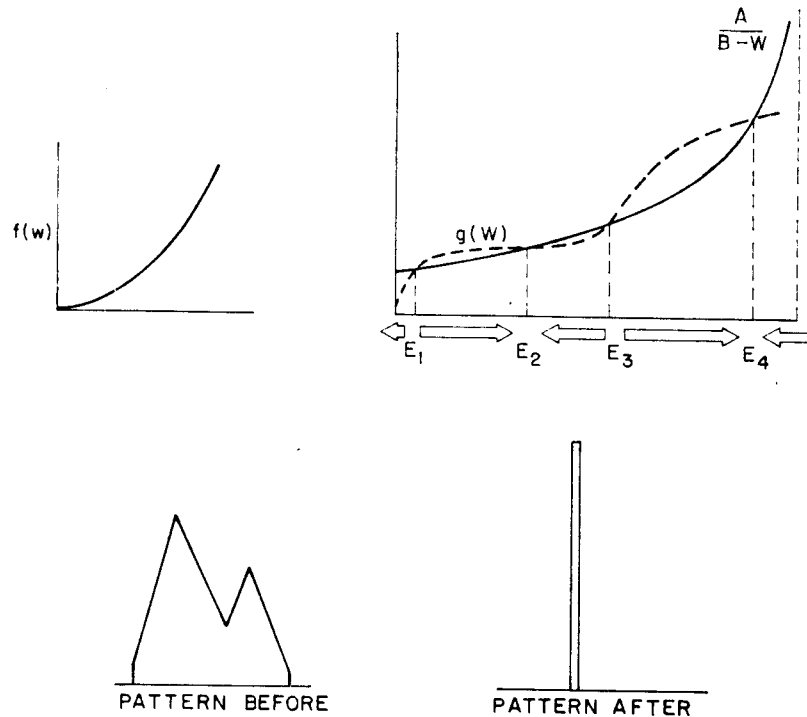
FIG. 6. Multiple limit points and transformation of input pattern by faster-than-linear signal function.

roots of the equation $g(w) = A(B-w)^{-1}$. If $x(0) < E_1$, then $x(\infty) = 0$; the value E_1 defines the level below which (total) initial activity is treated as noise and therefore suppressed. All initial values $x(0) > E_1$ lead to storage in STM. The values E_2, E_4, ..., E_{2m}, etc. are stable equilibrium points of total activity. There can be any number of these points if $g(w)$ is suitably defined; $g(w)$ measures how $f(w)$ deviates from linearity at various activity levels w. In particular, if $g(w)$ is convex, say because $f(w)$ is a sigmoid function of w, then there is a unique value E_2 to which $x(t)$ converges when reverberation in STM occurs.

The property of noise suppression implies that the pattern weights are not preserved. In fact, only those populations whose initial activities are maximal are reverberated in STM; the activities of all other populations are suppressed. In particular, if one population initially is more active than any other population, then the reverberation "chooses" this population for storage in STM, no matter how many other populations were initially active. Thus a dramatic form of contrast enhancement occurs in this case; it is the countrepart at large initial activity levels of the noise suppression that is desired at small initial activity levels.

(ii) AW case

After the reverberation reaches equilibrium, all activity is restricted to a single subfield (except for some unstable configurations). This subfield is not necessarily the one having the largest B_i value, since there is a competition between large initial activity levels (produced by particular features, etc.) and large B_i values (produced in development, by slowly varying attentional shifts, etc.); a subfield with larger initial activity and larger B_i will suppress a subfield with smaller values of these parameters, but a subfield with sufficiently large initial activity can suppress a subfield with smaller initial activity but larger B_i. Within the subfield that is stored in STM, there can be multiple stable equilibrium values, as in the EW case. Also within this subfield, only the populations with maximal initial values are stored in STM.

The AW case therefore exhibits two contrast enhancement mechanisms operating in parallel. First, only one subfield survives in STM; which one it will be is determined by an interaction between initial data $x_j(0)$ and the structural parameters B_i. Second, within this subfield, the signal function operating through the on-center off-surround anatomy chooses the populations with maximal initial activity.

The faster-than-linear case suppresses too much of a pattern in order to suppress noise. We seek a way to preserve the property of noise suppression without suppressing also all but the maxima of an input pattern. Cases I and III suggest a way.

(D) CASE IV. FASTER-THAN-LINEAR LEVELLING OFF TO LINEAR

(i) EW case

All populations whose initial activity falls below a prescribed threshold level will be quenched by the reverberation. The pattern of activity of all initially supra-threshold populations will be contrast enhanced and stored in STM (see Fig. 7). Speaking heuristically, the reason for this phenomenon
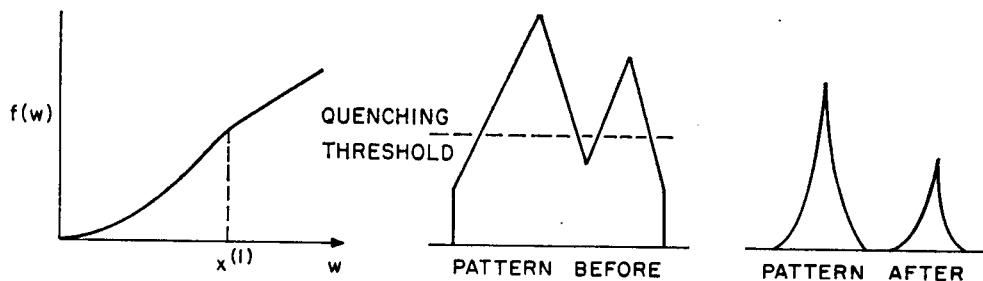

FIG. 7. Existence of quenching threshold using hybrid signal function.

is as follows. Suppose that an initial pattern of activity falls within the faster-than-linear region of $f(w)$. By Case III, contrast enhancement will begin. Were there no other region of $f(w)$, only the pattern maxima would survive. However, normalization of the total activity $x(t)$ also takes place, and can carry the pattern activities into the linear range. By Case I, *any* pattern in the linear range will be preserved by the reverberation; in particular, the partially contrast-enhanced pattern will be preserved. Thus the linear range terminates the contrast enhancement procedure which, if uninterrupted, would annihilate all but the pattern maxima.

(ii) AW case

The EW properties of contrast enhancement hold within a subfield. As with the faster-than-linear signal function, activities corresponding to only one subfield are stored in STM. Because of the linear range of the function, a bias in favor of the subfield with maximal B_i exists. However, if initial activities of all populations with maximal B_i are sufficiently small, and if initial activities of some populations with a non-maximal B_i are sufficiently large to be driven into the linear range by normalization of total activity, then a subfield with non-maximal B_i can win out asymptotically. Hence a partial contrast enhancement can occur of activities in only one subfield, which tends to be the "developmentally" or "attentionally" preferred subfield, but may not be if the initial activities of that subfield are sufficiently small.

(E) CASE V. SIGMOID SIGNAL FUNCTION

(i) EW case

In vivo, signal functions always have finite maxima at large activities. By leveling off the signal function of Case IV, one finds a sigmoid $f(w)$ (see Fig. 8). By Case II, we know that at activities corresponding to the slower-than-linear range of $f(w)$, the pattern will be uniformized. This unfortunate property can be eliminated by a suitable choice of parameters even if the maximal activity level B_i falls within the slower-than-linear
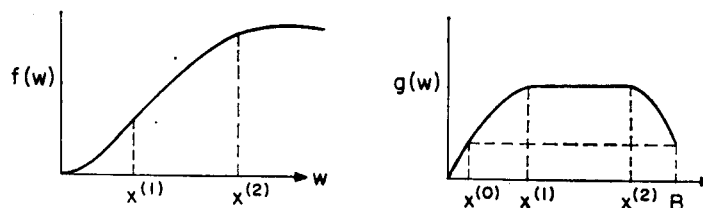

FIG. 8. Sigmoid signal functions.

range of $f(w)$. In pathological cases, however, uniformization of field activity can occur if the initial activities are sufficiently high.

Given that $f(w)$ must level off at high activity levels, the design of a sigmoid $f(w)$ requires only that $f(w)$ grow faster-than-linearly at small activity levels: since $f(w)$ is continuous, it must become approximately linear at intermediate activity levels. The width and slope of the linear region will determine how well nonmaximal initial activities get stored in STM.

A sigmoid $f(w)$ can be constructed in many ways; for example, if the signal thresholds within a population are Gaussianly distributed around a mean value, then the mean signal generated by the population will be a sigmoid function of mean population activity. The Gaussian distribution of thresholds can be viewed as a random imperfection in the structural design of the network. Yet the sigmoid $f(w)$ that results therefrom can suppress a crucial random imperfection in the functional operation of the network, namely noise. Here is a case in which randomness on one level helps to overcome randomness on another level. Randomness becomes productive in this context because part of the network design is not random, namely the on-center off-surround anatomy. Grossberg (1970, 1972c) describes an analogous example, in which a random choice of thresholds produces a wide range of efficient pattern discriminators, because the random parameter is again organized by genetically programmed inhibitory mechanisms.

(ii) AW case

Again the EW behavior is relativized to a subfield. The uniformizing effect of the slower-than-linear region of $f(w)$ can again be eliminated by a suitable choice of parameters. If no constraints are put on the parameters, however, it is possible for activities corresponding to different B_i values to be stored in STM, even though some of these activities fall in the faster-than-linear range of $f(w)$. This effect is due to an interaction of these activities with activity levels of other populations that fall in the slower-than-linear range of $f(w)$.

If all of the activities fall in the slower-than-linear range, then the results of Case II hold; namely, activity levels in STM have the same ordering as the B_i. For example, if the populations sharing a given B_i code features of a given peripheral class (e.g., colors, orientations, sounds, etc.) then over-arousal of the network can yield "hallucinations" in which the class with highest B_i value will be dominant. Another form of hallucination would be generated by any mechanism that lowers the quenching threshold of the network, so that noise, arousal, or other non-cue inputs can bootstrap population activity into STM.

The above statements are about asymptotic behavior of the system for various signal functions $f(w)$. They implicitly assume that all the activities $x_i(t)$ tend to limits as $t \to \infty$, rather than oscillating indefinitely. The existence of limits for general non-negative, continuous, and nondecreasing $f(w)$ has not yet been proved, although limits have existed in all of our computer studies. Existence of limits has however, been proved for a wide class of signal functions that includes realistic cases, such as certain sigmoid functions. All of our limit theorems constrain oscillations of the function

$$F(w) = \frac{wf'(w)}{f(w)}.$$

Since

$$F(w) = \frac{d \log f(w)}{d \log w},$$

these constraints are "multiplicative convexity" conditions on $f(w)$.

2. Mathematical Results

This section lists results whose proofs are given in the Appendix. The following notation will be used to describe these results. Let

$$x(t) = \sum_{k=1}^{n} x_k(t)$$

be the total activity, $X_i(t) = x_i(t)x^{-1}(t)$ be the ith pattern variable, $M(t) = \max_k X_k(t)$, $m(t) = \min_k X_k(t)$, $g(w) = w^{-1}f(w)$, $g_i(t) = g(x_i(t))$, and

$$H(t) = \sum_{k=1}^{n} f(x_k(t)).$$

By convention, we will order the subscripts B_i such that $B_1 \leq B_2 \leq \ldots \leq B_n$. The following definitions will also be useful.

(1) The reverberation is *persistent* if some $x_i(t)$ does not approach 0 as $t \to \infty$.

(2) A *subfield* is a maximal collection of populations v_j having the same B_i value.

(3) The limiting distribution $\{x_i(\infty)\}$ is *fair* if the limits $x_i(\infty)$ are proportional to $x_i(0)$.

(4) The limiting distribution is *uniform* if all $x_i(\infty)$ are equal.

(5) The limiting distribution is *locally uniform* if the reverberation is persistent at more than one population, but not all populations, and all nonzero $x_i(\infty)$ are equal.

(6) The limiting distribution is *0–1* if only one $x_i(\infty)$ is positive.

(7) The limiting distribution is *uniformized* if $M(t)$ is decreasing and $m(t)$ is increasing.

(8) The limiting distribution is *contrast enhanced* if $M(t)$ is increasing and $m(t)$ is decreasing.

(9) The limiting distribution exhibits *quenching* at population v_i if $x_i(0) > 0$ but $x_i(\infty) = 0$.

(10) *Normalization* occurs if $x(\infty)$ exists and is bounded above by a constant that is independent of n.

(11) An inequality *propagates* if its truth at time $t = T$ implies its truth at all times $t \geqslant T$.

These definitions have obvious extensions to analogous concepts within a subfield. For example, the limiting distribution is contrast enhanced within the subfield $\{v_j: B_j = B_i\}$ if

$$M^{(i)}(t) = \max \left\{ x_j(t) \left(\sum_{B_k = B_i} x_k \right)^{-1} : B_j = B_i \right\}$$

is increasing and

$$m^{(i)}(t) = \min \left\{ x_j(t) \left(\sum_{B_k = B_i} x_k \right)^{-1} : B_j = B_i \right\}$$

is decreasing.

We always assume below that all $x_i(0)$ are positive, since if any $x_i(0) = 0$, then v_i receives only inhibition; thereafter $x_i(t) = 0$ for all $t \geqslant 0$, and v_i can be deleted from the network without loss of generality. Our first theorems will discuss conditions under which a limiting distribution exists. Later theorems discuss the distribution of limiting values given a particular choice of signal function.

(A) EXISTENCE OF LIMITS

In all of our computer studies, limits have always existed given various choices of increasing $f(w)$. Theorem 1 shows that this might be generally true.

Theorem 1. If n = 2, then given any non-negative, continuous, and strictly increasing $f(w)$, all limits exist and there is at most one oscillation in x_1 or x_2.

For general n, weaker results have been proved, but these include physically interesting cases that are stated below in order of increasing difficulty.

Theorem 2. If $f(w)$ is differentiable, and

$$F(w) = \frac{wf'(w)}{f(w)}$$

is either monotone nondecreasing or monotone nonincreasing, then all limits exist.

Remark. $F(w)$ = constant if and only if $f(w) = aw^b$. Hence the condition on $F(w)$ tests deviations from power law signal functions.

Corollary 1. Limits exist if

$$f(w) = \sum_{k=0}^{\infty} a_k w^k$$

where all $a_k \geqslant 0$ and

$$\sum_{k=0}^{\infty} a_k B_n^k < \infty.$$

Theorem 3 (f *faster-than-linear becoming linear*). Let $f(w) = wg(w)$, where either $g(w)$ is strictly increasing, or is strictly increasing for $w < x^{(1)}$ and equals a constant C for $w \geqslant x^{(1)}$, for some $x^{(1)} \leq B_n$. Also let $F(w)$ be strictly increasing for $0 \leq w \leq \hat{x}$ and strictly decreasing for $\hat{x} < w \leq B_n$. Then all limits exist.

Remark. The physical meaning of the condition on oscillations of $F(w)$ only becomes apparent in the corollary to Theorem 4, which shows that this condition is satisfied for an important class of random factors that determine sigmoid signal functions.

Theorem 4 (f *Sigmoid*). Let $f(w) = wg(w)$, where $g(w)$ is strictly increasing for $0 \leq w \leq x^{(1)}$, equal to C for $x^{(1)} \leq w \leq x^{(2)}$, and strictly decreasing for $x^{(2)} \leq w \leq B_n$. Let $F(w)$ be strictly increasing for $0 \leq w \leq \hat{x}$ and strictly decreasing for $\hat{x} < w \leq B_n$. Define $x^{(0)}$ so that $0 \leq x^{(0)} \leq x^{(1)}$ and $g(x^{(0)}) = B_n^{-1} f(B_1)$, and suppose

$$x^{(0)} + x^{(2)} \geqslant B_n - AC^{-1} \tag{6}$$

Then all limits exist.

Remark. Condition (6) generalizes a similar condition that was applied in Grossberg (1973) to the EW case. This condition shows that activities can enter the slower-than-linear range without causing persistent "seizure".

Corollary 2. Suppose $f(w)$ is a sigmoid function such that

$$f(w) = \int_0^w p(v)\, dv,$$

where $p(v)$ is a positive, differentiable function such that

$$\lim_{v \to \infty} vp(v) = 0.$$

Let

$$P(w) = 1 + \frac{wp'(w)}{p(w)}$$

have one maximum and change sign at most once from non-negative to negative as w increases. Then all limits exist.

Remark. $f(w)$ can be rewritten as

$$f(w) = \int_0^\infty \theta(w-v)p(v) \, dv$$

where $\theta(y) = 1$ if $y \geqslant 0$ and $\theta(y) = 0$ if $y < 0$; that is, $\theta(w-v)$ equals 1 if and only if the mean activity w exceeds the threshold value v. Given this interpretation, $p(v)$ is the distribution of sites having threshold v, and $f(w)$ adds up unit signal strengths weighted by their density within the population. Corollary 2 includes such typical cases as Gaussian or Cauchy distributions. In the Gaussian case, $p(v) = \sigma^{-1} \exp\left[-\sigma^{-2}(v-a)^2\right]$ with $a > 0$ and $\sigma > 0$. Here $P(w) = 1 - \sigma^{-2}w(w-a)$. In the Cauchy case, $p(v) = \sigma[(v-a)^2 + \sigma]^{-1}$. Here $P(w) = \{1 - 2w(w-a)[(w-a)^2 + \sigma^2]^{-1}\}$. It is also obvious in both cases that $p(v) > 0$ and that

$$\lim_{v \to \infty} vp(v) = 0.$$

It has not yet been decided whether a constraint on oscillations of $F(w)$ is essential. Indeed the following theorem holds.

Theorem 5 (f *sigmoid, EW case*). Let $f(w)$ be defined as in Theorem 4. Then in the EW case (i.e., all $B_i = B_n$), all limits exist.

Similarly, there are other results, suggested by the EW case, that are compatible with the existence of limits under weaker constraints. For example, in the EW case, Grossberg (1973) demonstrated the existence of limits for any $f(w)$ whose $g(w)$ is either monotone nonincreasing or monotone nondecreasing. In the analogous AW case, it can be shown that there do not exist any periodic solutions, and hence there are no limit cycles.

Theorem 6. If $g(w)$ is either monotone nonincreasing or monotone nondecreasing, then the only periodic solutions are identically constant.

(B) DISTRIBUTION OF LIMITS

Theorem 7 (*linear* f). Let $f(w) = Cw$, $C > 0$. If the reverberation persists, then the limiting distribution is fair within the B_n-subfield and is quenched elsewhere. The reverberation persists if and only if $CB_n > A$ and is then normalized with $x(\infty) = B_n - AC^{-1}$. Indeed, for every i,

$$x_i(t) = \frac{x_i(0)e^{(CB_i - A)t}}{1 + C \sum_{k=1}^{n} x_k(0)H_k(t)}, \tag{7}$$

where

$$H_k(t) = \begin{cases} \dfrac{e^{(CB_k - A)t} - 1}{CB_k - A} & \text{if } CB_k \neq A \\[3mm] t & \text{if } CB_k = A. \end{cases}$$

Theorem 8 (slower-than-linear f). Let $f(w) = wg(w)$, where $g(w)$ is strictly decreasing. The limiting distribution is uniform within each subfield, and $x_i(\infty) \geqslant x_j(\infty)$ if and only if $B_i \geqslant B_j$. The reverberation persists if $B_1 g(0) > A$, and a given x_i is quenched unless $B_i g(0) \geqslant A$. There is a unique limiting distribution corresponding to every set of unquenched $x_i(\infty)$.

Theorem 9 (faster-than-linear f). Let $f(w) = wg(w)$, where $g(w)$ is strictly increasing. If the reverberation is persistent, then the limiting distribution is 0–1 or locally uniform within each subfield. It is 0–1 in a given subfield unless there is more than one maximal initial value in that subfield. If moreover $g(w)$ is differentiable and $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is a critical point of (1) with more than one $\xi_i > 0$, then ξ is unstable. In fact, if m of the ξ_i's are positive, then ξ can be approached only within an $(n - m + 1)$-dimensional manifold in R^n. Thus only the 0–1 distribution is stable.

If all limits exist and there are m populations with nonzero limits sharing the same B_i value and none at other B_i values, then all nonzero $x_i(\infty)$ equal xm^{-1}, where x satisfies

$$g\left(\frac{x}{m}\right) = A(B - x)^{-1}.$$

If $g(w)$ is convex, then if $g(0) \geqslant AB^{-1}$, there is a unique positive solution, whereas if $g(0) < AB^{-1}$, there are two positive solutions; the smaller is unstable, whereas, the larger solution is stable.

Remarks. Theorem 9 describes a tendency for only one subfield to reverberate, and moreover for only the maximal initial data in that subfield to reverberate. The system, in short, tries to make a choice. The theorem does not say which population(s) will be stored in STM. Populations with a larger B_i have an advantage, but this advantage can be offset by populations with smaller B_i but larger initial data. Indeed, if $F(w)$ is monotone nondecreasing, then the proof of Theorem 2 shows that $B_i g_i$ can only overtake $B_j g_j$ if $x_i \geqslant x_j$. If also $g(w)$ is monotone increasing and $B_i \neq B_j$, then $B_i < B_j$. Thus there is a tendency for populations with smaller B_i to have maximal $B_i g_i$ at arbitrarily large times. Such a population has a monotone increasing activity x_i, and therefore cannot be quenched. On the other hand,

given *any* signal function, if $B_i > B_j$, then the inequality $x_i \geq x_j$ propagates. The latter statement demonstrates that a nonuniform choice of B_i can strengthen propagating inequalities in a direction that favors the populatons with the largest B_i; this property does not however, altogether prevent other populations from becoming dominant, as the example about $B_i g_i$ overtaking $B_j g_j$ illustrates. The theorem below summarizes sharper criteria of this type. These criteria can be expressed in terms of the functions

$$M_{ij} = \max \{w/g^{-1}(B_i B_j^{-1} g(w)): 0 \leq w \leq B_i\}$$

and

$$m_{ij} = \min \{w/g^{-1}(B_i B_j^{-1} g(w)): 0 \leq w \leq B_i\}.$$

Note that if $B_i > B_j$ then $M_{ij} < 1$. Note also that $m_{ij} = 0$ if and only if $g(0) > 0$.

Theorem 10 (*faster-than-linear* f). Let $f(w) = wg(w)$, where $g(w)$ is strictly increasing. Suppose $B_i > B_j$ and $x_i(0) < x_j(0)$. If moreover $x_i(0) < m_{ij}x_j(0)$, then this inequality propagates, so that if limits exist, then $x_i(\infty) < x_j(\infty)$. If however $x_i(0) > M_{ij}x_j(0)$, then this inequality propagates and if limits exist, then $x_i(\infty) > x_j(\infty) = 0$. If $B_i g(0) > B_j g(B_j)$, then eventually $x_i > x_j$ and $x_j(\infty) = 0$. In general, if limits exist and $x_i(0) \neq 0 \neq x_j(0)$, then $x_i(\infty) > x_j(\infty)$.

If $x_i(0) < m_i$, where $g(m_i) = AB_i^{-1}$, then x_i decreases to zero, whereas if $g(0) \geq AB^{-1}$, then the reverberation is persistent.

Theorem 11 (f *faster-than-linear becoming linear*). Let $f(w) = wg(w)$, where $g(w)$ is strictly increasing for $0 \leq w \leq x^{(1)}$, and equal to a constant C for $x^{(1)} \leq w \leq B_n$. Then the limiting distribution is contrast enhanced within each subfield. Define m_{ij} and M_{ij} as above except that $g^{-1}(C)$ can be any value between $x^{(1)}$ and B_n. Then if $B_i > B_j$, the inequalities $x_i < m_{ij}x_j$ and $x_i > M_{ij}x_j$ propagate.

Remarks. As in the faster-than-linear case, the system tends to reverberate only in one subfield, and not necessarily the subfield with maximal B_i. More than one population in that subfield can reverberate if their activities fall in the linear range of f. Note in the proof of Theorem 9 that the coefficients d_{ii} vanish if $g'(\xi_i) = 0$; thus that instability result does not hold for critical points that fall in the linear range of $f(w)$. Also note that m_{ij} is smaller in the present case than in the faster-than-linear case, thereby suggesting a greater difficulty in the present case of reverberating a subfield of non-maximal B_i. Theorem 4 shows that these results also hold for suitable sigmoid signal functions.

In the sigmoid case, it is possible for activities in more than one subfield to reverberate in a stable fashion. An example in two dimensions arises when $A = 6$, $B_1 = 32/3$, $B_2 = 16$, and g is the integral of the following function

$$g'(w) = \tfrac{3}{8} - \tfrac{1}{16}w \text{ for } 0 \leq w \leq 4,$$
$$= \tfrac{33}{32}w - 4 \text{ for } 4 \leq w \leq \tfrac{9}{2},$$
$$= \tfrac{45}{8} - \tfrac{9}{8}w \text{ for } \tfrac{9}{2} \leq w \leq 5,$$
$$= -\tfrac{3}{8}w^2 \text{ for } 5 \leq w \leq 6,$$

with the values for $6 \leq w \leq 16$ irrelevant (as long as $g < 0$ and $G = F - 1 > -1$ on that interval). For these parameters, $x_1 = 6$, $x_2 = 4$ is a critical point, and the Jacobian of (1) at that point is

$$\begin{pmatrix} -2 & -3 \\ -15 & -\tfrac{67}{2} \end{pmatrix},$$

which has two eigenvalues of negative real part. One of the variables reverberates in the slower-than-linear part of f, and the other in the faster-than-linear part.

This is not surprising because even if condition (6) holds, one x_i at a critical point could have $g' < 0$. That condition would, by the proof of Theorem 7, contribute one eigenvalue with negative real part to the Jacobian of (1), and Theorem 7 shows that another negative eigenvalue can arise even if all $g' \geqslant 0$. Thus, if condition (6) holds, there can be as many as two eigenvalues of the Jacobian with negative real part. In the above two-dimensional example, (6) does not hold.

3. Discussion

This paper analyses the dynamics of certain reverberating on-center off-surround networks undergoing shunting interactions. The main results study effects of varying the number of cell sites B_i, or the tuning parameters λ_i, that correspond to different populations v_i. In cases where a given population responds only to prescribed features, a large B_i value describes a large number of cell sites that code for these features; a large λ_i value can describe an attentional shift to a particular class of features. The signal function that translates average population activity into average signal size dramatically influences how these nonuniform aspects of network architecture will manifest themselves in the network's response to input patterns. Certain

sigmoid functions are desirable because they do not amplify noise, yet also do not suppress all the graded intensity information in an input pattern. Even in networks whose B_i and λ_i are uniformly chosen, a sigmoid signal function is capable of contrast enhancing an input pattern, quenching inputs that fall below a prescribed threshold, and adapting the network's total activity in STM. Because the quenching threshold exists, any nonspecific input that amplifies an input pattern will tend to store more of its features in STM, whereas a nonspecific input that suppresses the pattern will help the network choose that feature whose population activity remains above the threshold. This property is important in discussing network responses to novel patterns, if the novelty of the pattern triggers nonspecific arousal. Grossberg (1975a) shows how the quenching threshold, along with the related property of total activity adaptation, can be used to explain related data about overshadowing and behavioral contrast. It is therefore important to establish experimentally whether a quenching threshold and adaptation exist in recurrent sensory fields *in vivo*.

Given a sigmoid signal function in a field with nonuniform B_i and/or λ_i, the features in a given input pattern can trigger a complicated tug-of-war in the network in which three feature properties struggle to determine which features will be stored in STM. Developmentally selected features compete with features that are attended to, or salient. The physical parameters of each feature, such as input energy or density, compete also. Our theorems show how these different parameters struggle until a consensus is established leading to STM storage of those features which have the strongest balance of energetic, attentional, and developmental factors.

A related paper (Grossberg, 1975b) shows how populations v_i can develop their selective responses to particular features. The reverberating network is an important part of the developmental mechanism. Features which do not occur frequently during a critical developmental period end up having small B_i values in the adult network. Our theorems show how the activity of these infrequently experienced feature detectors can be totally masked by that of other feature detectors, thereby preventing the mere existence of many feature detectors from producing overwhelming network noise. Nonetheless, when the infrequent features do occur, they can be stored in STM.

Other papers in this series analyse networks in which the on-center and off-surround coefficients depend on distance, or the excitatory and inhibitory populations respond at different rates (Ellias & Grossberg, 1976; Levine & Grossberg, 1976). These papers show how basic network designs can yield visual illusions (neutralization tilt after-effect, angle expansion), and describe other mechanisms whereby graded intensity information in an input pattern can be stored in STM.

Appendix

Proof of Theorem 1. Divide the (x_1, x_2)-plane into four regions: (I) $= \{(x_1, x_2): \dot{x}_1 > 0, \dot{x}_2 > 0\}$, (II) $= \{(x_1, x_2): \dot{x}_1 > 0, \dot{x}_2 < 0\}$, (III) $= \{(x_1, x_2): \dot{x}_1 < 0, \dot{x}_2 < 0\}$, and (IV) $= \{(x_1, x_2): \dot{x}_1 < 0, \dot{x}_2 > 0\}$. We will prove that a trajectory can never leave (II) or (IV). This suffices to prove the theorem, since a trajectory either always stays in the same region, or passes at most once from (I) or (III) into (II) or (IV); that is, at most one oscillation occurs in either x_1 or x_2. Once this passage occurs, both x_1 and x_2 are thereafter monotonic and bounded continuous functions; hence their limits exist.

The proofs for (II) and for (IV) are the same; only (II) will be explicitly considered. Write equation (1) in the form

$$\dot{x}_1 = x_1[U_1(x_1) - f(x_2)] \tag{8}$$

$$\dot{x}_2 = x_2[U_2(x_2) - f(x_1)], \tag{9}$$

where $U_i(x_i) = (B_i - x_i)g(x_i) - A$, $i = 1,2$. To establish a contradiction, let (a,b) be a point where a trajectory goes from (II) to (III). [see Fig.9(a)].
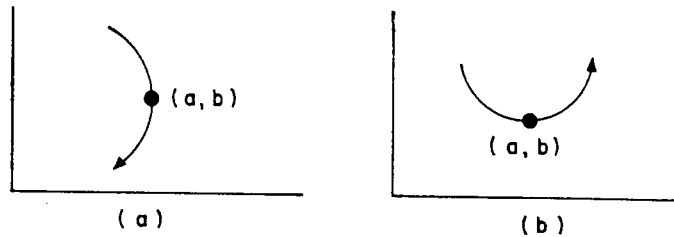

FIG. 9. Transition points between regions.

There exist positive numbers ε, δ_1, δ_2 such that $(a-\varepsilon, b+\delta_1)$ is in (II), whereas $(a-\varepsilon, b-\delta_2)$ is in (III). Thus $\dot{x}_1 > 0$ at $(a-\varepsilon, b+\delta_1)$ and $\dot{x}_1 < 0$ at $(a-\varepsilon, b-\delta_2)$. By (9),

$$(a-\varepsilon)[U_1(a-\varepsilon) - f(b+\delta_1)] > 0 > (a-\varepsilon)[U_1(a-\varepsilon) - f(b-\delta_2)],$$

so $f(b-\delta_2) > f(b+\delta_1)$, which violates the monotonicity of $f(w)$. A similar argument, using equation (10), shows that a trajectory in (II) cannot move into (I). [see Fig. 9(b)]. This completes the proof.

Proof of Theorem 2. Assume that $F(w)$ is monotone nondecreasing; a similar proof holds in the case of nonincreasing $F(w)$. Four main lemmas form the proof's core.

Lemma 1. If $B_i \geqslant B_j$, then the inequality $x_i \geqslant x_j$ propagates. Hence the x_i's are in a fixed order for all sufficiently large times.

Proof. Recalling the notation $g_k(t) = g(x_k(t))$ and

$$H(t) = \sum_{k=1}^{n} f(x_k(t)),$$

rewrite (1) in the form

$$\dot{x}_i = x_i(B_i g_i - A - H) \tag{10}$$

At a time when $x_i = x_j$, $\dot{x}_i = x_j(B_i g_i - A - H) \geq \dot{x}_j$.

Lemma 2. Let f be differentiable. If some nonzero limit $x_i(\infty)$ exists, then all limits exist.

Proof. A routine computation shows that $\ddot{x}_i$ is bounded. Since also $x_i(\infty)$ exists, it follows that $\dot{x}_i(\infty)$ exists and equals zero. Since $x_i(\infty) \neq 0$, (10) implies that $H(\infty)$ exists and that $B_i g_i(\infty) = A + H(\infty)$. By (10), for any j,

$$\dot{x}_j = B_j x_j[g(x_j) - K_j(t)], \tag{11}$$

where $K_j(t) = B_j^{-1}(A + H(t))$. Let

$$L_j = \lim_{t \to \infty} K_j(t) > 0.$$

For every $\varepsilon > 0$, there exists a T_ε such that $t \geq T_\varepsilon$ implies $|K_j(t) - L_j| \leq \varepsilon$. Hence if there exists an $\varepsilon > 0$ such that $|g(x_j(t)) - L_j| > \varepsilon$ for some $t \geq T_\varepsilon$, then either of two possibilities occurs: (1) $g(x_j(t))$ approaches the limit L_j as $t \to \infty$, or (2) $|g(x_j(t)) - L_j| < \varepsilon$ for all sufficiently large t. Fig. 10 illustrates the two possibilities. In Fig. 10(a), $g(x_j(t_1)) > L_j + \varepsilon \geq K_j(t_1)$. By (11), $\dot{x}_j(t_1) > 0$. Since $g(w)$ is increasing for all $w > L_j + \varepsilon$, $x_j(t)$ is monotone increasing for all $t \geq t_1$, and $x_j(\infty)$ exists. In Fig. 10(b), again $g(x_j(t_1)) > L_j + \varepsilon \geq K_j(t_1)$. Hence $x_j(t)$ grows for $t \geq t_1$ until $g(x_j(t))$ enters the interval $[L_j - \varepsilon, L_j + \varepsilon]$, where $K_j(t)$ is trapped. By (11), $g(x_j(t))$ will then also be trapped in $[L_j - \varepsilon, L_j + \varepsilon]$. Consequently, either there is no $\varepsilon > 0$ such that $|g(x_j(t)) - L_j| > \varepsilon$ for some $t \geq T_\varepsilon$—and this implies that
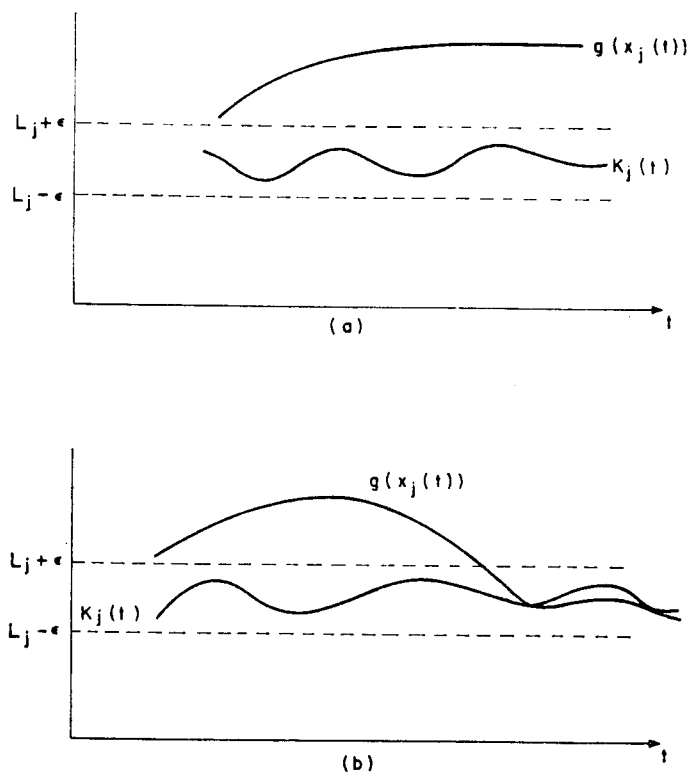
$$\lim_{t \to \infty} g(x_j(t))$$

exists—or for *every* such $\varepsilon > 0$. either

$$\lim_{t \to \infty} g(x_j(t))$$

exists or $|g(x_j(t)) - L_j| < \varepsilon$ for arbitrarily large times. In all cases, therefore,

$$\lim_{t \to \infty} g(x_j(t)) = L_j.$$

Lemma 3. Let $B_i g_i(t)$ overtake $B_j g_j(t)$ at time $t = \tau$. If $F(w)$ is monotone nondecreasing, then at times $t < \tau$ that are arbitrarily close to τ, either (I) $\dot{x}_1 > 0$, $\dot{x}_j > 0$, and $x_i > x_j$, or (II) $\dot{x}_i < 0$, $\dot{x}_j < 0$, and $x_i < x_j$.

FIG. 10. Oscillation of $g(x_j(t))$ relative to $K_j(t)$.

Proof. By hypothesis, for t arbitrarily close to τ and $t < \tau$,

$$B_i g(x_i(t)) < B_j g(x_j(t)) \tag{12}$$

and

$$B_i \frac{d}{dt} g(x_i(t)) > B_j \frac{d}{dt} g(x_j(t)). \tag{13}$$

From (12) and (13) we find

$$\frac{d}{dt} \log g(x_i(t)) > \frac{d}{dt} \log g(x_j(t)),$$

which is the same as

$$\frac{\dot{x}_i(t) g'(x_i(t))}{g(x_i(t))} > \frac{\dot{x}_j(t) g'(x_j(t))}{g(x_j(t))}. \tag{14}$$

By (10) and (12),

$$\frac{\dot{x}_i(t)}{x_i(t)} < \frac{\dot{x}_j(t)}{x_j(t)}. \tag{15}$$

Suppose $\dot{x}_i(t) > 0$ for definiteness. By (15), $\dot{x}_j(t) > 0$ as well. Thus by

(14) and (15),

$$\frac{x_i(t)g'(x_i(t))}{g(x_i(t))} > \frac{x_j(t)g'(x_j(t))}{g(x_j(t))}. \tag{16}$$

The function

$$G(w) = \frac{wg'(w)}{g(w)}$$

is readily seen to equal $F(w) - 1$, and is therefore monotone nondecreasing. From (16) now follows that $x_i(t) > x_j(t)$. The three conditions $\dot{x}_i(t) > 0$, $\dot{x}_j(t) > 0$, $x_i(t) > x_j(t)$ characterize Case (I). Case (II) occurs if $\dot{x}_i(t) < 0$. The two cases are clearly exhaustive.

Lemma 4. Let $g(w)$ be differentiable. Then the inequality

$$\max_k B_k g_k \geqslant A + H \tag{17}$$

propagates. If $\max_k B_k g_k = A + H$ at $t = \tau$, then

$$\frac{d}{dt}[\max_k B_k g_k - A - H] > 0$$

unless all $B_j g_j = A + H$.

Proof. Suppose that at $t = \tau$

$$B_i g_i = \max_k B_k g_k = A + H \tag{18}$$

It suffices to show that

$$\frac{d}{dt}[B_i g_i(\tau) - A - H(\tau)] \geq 0$$

and that

$$\frac{d}{dt}[B_i g_i(\tau) - A - H(\tau)] > 0$$

unless all $B_j g_j(\tau) = A + H(\tau)$. Obviously

$$\frac{d}{dt}[B_i g_i(\tau) - A - H(\tau)] = B_i g_i'(\tau)\dot{x}_i(\tau) - \sum_k f_k'(\tau)\dot{x}_k(\tau),$$

where by (10) and (18), $\dot{x}_i(\tau) = 0$ and

$$\dot{x}_j(\tau) = x_j(\tau)[B_j g_j(\tau) - A - H(\tau)] \qquad \leq x_j(\tau)[B_i g_i(\tau) - A - H(\tau)] \leq 0.$$

Since $f'(w) > 0$, the proof is complete.

The proof of Theorem 2 is completed as follows. By Lemma 4, the function $R \equiv \max_k B_k g_k - (A + H)$ changes sign at most once. If $R \leq 0$ for all $t \leq 0$, then by (10), all $\dot{x}_i \leq 0$ for all $t \geqslant 0$, and thus all limits exist. Otherwise, there exists a $t = T$ such that $R(t) \geqslant 0$ for all $t \geqslant T$. Thus at times $t \geqslant T$

when $B_i g_i = \max_k B_k g_k$, (10) and Lemma 4 show that $\dot{x}_i > 0$ except at isolated points in time. If $B_j g_j$ overtakes $B_i g_i$ at such a $\tau > T$, then Case (I) of Lemma 3 shows that $x_j(t) > x_i(t)$ for some t arbitrarily close to τ. That is, every change in the index i such that $B_i g_i = \max_k B_k g_k$ can only increase the corresponding x_i. By Lemma 1, the x_i's are eventually in a fixed order. There are consequently at most finitely many such changes in the index i, so that for all sufficiently large t, there exists an i such that $B_i g_i = \max_k B_k g_k$. For this index i, $\dot{x}_i \geqslant 0$ at all large times. Thus $x_i(\infty)$ exists and is positive. By Lemma 2, all limits exist.

Proof of Corollary 1. A routine computation shows that $F'(w)$ has the same sign as

$$\sum_{k,\,m=0}^{\infty} m(m-k)a_k a_m w^{k+m}.$$

Let $l = k+m$ and rewrite the series as

$$\sum_{l=0}^{\infty} b_l w^l,$$

where

$$b_l = \sum_{r=0}^{[l/2]} (l-2r)^2 a_r a_{l-r}$$

and $[l/2]$ is the largest integer $\leq l/2$. Since all $a_k \geqslant 0$, $F(w)$ is monotone nondecreasing.

Proof of Theorem 3. First we consider the case of $g(w)$ strictly increasing. Again our strategy is to show that beyond a fixed time, there exists an i such that $B_i g_i = \max_k B_k g_k$. Then x_i is monotone increasing, the limit $x_i(\infty) > 0$ exists, and hence by Lemma 2, all limits exist. We will show that there are at most finitely many changes in the index defining the maximal $B_i g_i$.

By Lemmas 1 and 4, we can assume without loss of generality that the x_i's are in invariant order and that $\max_k B_k g_k \geqslant A+F$ at all times. Also by Lemma 4, since the inequality $\max_k B_k g_k \geqslant A+H$ propagates and equality can hold only at isolated points, the maximal $B_i g_i$ is strictly increasing.

Let $i > j$; i.e., $B_i > B_j$. Consider the curve $C_{ij} = \{(x_i,x_j): B_i g(x_i) = B_j g(x_j)\}$. Along C_{ij}, $x_i < x_j$. Suppose that there is a point $(U_{ij}, V_{ij}) \varepsilon C_{ij}$ such that

$$F(U_{ij}) = F(V_{ij}). \tag{19}$$

Concerning such a point, the following Lemma holds.

Lemma 5. Let

$$\frac{d}{ds_{ij}}$$

denote differentiation with respect to any parameter increasing along C_{ij}. Then for any $(x_i, x_j) \varepsilon C_{ij}$,

$$\frac{d}{ds_{ij}}(x_i x_j^{-1}) > 0 \text{ if } x_i < U_{ij}, \ x_j < V_{ij}$$

$$= 0 \text{ if } x_i = U_{ij}, \ x_j = V_{ij}$$

$$< 0 \text{ if } x_i > U_{ij}, \ x_j > V_{ij} \tag{20}$$

Proof. First we prove that for any other point $(U, V) \varepsilon C_{ij}$,

$$U > U_{ij} \text{ implies } V > V_{ij} \text{ and } F(U) > F(V)$$

and

$$U < U_{ij} \text{ implies } V < V_{ij} \text{ and } F(U) < F(V). \tag{21}$$

Since $U_{ij} < V_{ij}$ and $F(w)$ has a unique maximum, say at $\hat{x} > 0$, it follows that $U_{ij} < \hat{x} < V_{ij}$. Since $g(w)$ is strictly increasing, given any other point $(U, V) \varepsilon C_{ij}$, it follows that $(U - U_{ij})(V - V_{ij}) > 0$. Suppose $U > U_{ij}$. If also $U < \hat{x}$, then $U_{ij} < U < \hat{x} < V_{ij} < V$, so that $F(U) > F(U_{ij}) = F(V_{ij}) > F(V)$. If instead $U > \hat{x}$, then since $U < V$ with U and V both in the decreasing range of $F(w)$, $F(U) > F(V)$. A similar argument proves the reverse inequalities if $U < U_{ij}$.

We now differentiate the identity $B_i g(x_i) = B_j g(x_j)$ along the curve C_{ij} Thus

$$B_i g'(x_i) \frac{dx_i}{ds_{ij}} = B_j g'(x_j) \frac{dx_j}{ds_{ij}},$$

or

$$\frac{g'(x_i)}{g(x_i)} \frac{dx_i}{ds_{ij}} = \frac{g'(x_j)}{g(x_j)} \frac{dx_j}{ds_{ij}}. \tag{22}$$

Along C_{ij},

$$\frac{d}{ds_{ij}}(x_i x_j^{-1})$$

has the sign of

$$\dot{x}_j \frac{dx_i}{ds_{ij}} - \dot{x}_i \frac{dx_j}{ds_{ij}},$$

which by (22) has the sign of $G(x_j) - G(x_i)$, where

$$G(w) = \frac{w g'(w)}{g(w)}.$$

Since $G(w) = F(w) - 1$,

$$\frac{d}{ds_{ij}}(x_i x_j^{-1})$$

has the sign of $F(x_j) - F(x_i)$. Letting $x_i = U$ and $x_j = V$ in (21) yields the conclusion of the lemma.

If there is no point (U_{ij}, V_{ij}) that satisfies (19), than a point (U_{ij}, V_{ij}) that satisfies (21) can still be found: choose (U_{ij}, V_{ij}) to be an endpoint of the curve C_{ij}. In all cases, therefore, for every ordered pair (i,j) with $i > j$, there exists a point $(U_{ij}, V_{ij})\varepsilon C_{ij}$ such that (20) holds.

We now define $\gamma_{ij} = B_i g(U_{ij})$ and let $\gamma = B_i g_i(U)$ for every $(U,V)\varepsilon C_{ij}$. By (20), the function $H_{ij} = x_i x_j^{-1}$ is increasing along C_{ij} for $\gamma > \gamma_{ij}$ and decreasing for $\gamma < \gamma_{ij}$. Likewise, H_{jk} is increasing for $\gamma < \gamma_{jk}$ and decreasing for $\gamma > \gamma_{jk}$. Since $H_{ik} = H_{ij}H_{jk}$, H_{ik} is increasing for $\gamma < \min(\gamma_{ij},\gamma_{jk})$ and decreasing for $\gamma > \max(\gamma_{ij},\gamma_{jk})$. By (20), H_{ik} reaches its maximum at γ_{ik}. Thus

$$\min(\gamma_{ij},\gamma_{jk}) < \gamma_{ik} < \max(\gamma_{ij},\gamma_{jk}). \tag{23}$$

Using this fact, and the following combinatorial lemma, we will show that there cannot be an infinite number of changes in which $B_i g_i$ is maximal.

Lemma 6. For all pairs of integers i,j such that $1 \le j < i \le n$, let γ_{ij} be a real number. Define $\gamma_{ji} = \gamma_{ij}$ and suppose that $i > j > k$ implies that (23) holds. Let $i_1, i_2, \ldots, i_r, i_{r+1} = i_1$ be integers between 1 and n. Let

$$Y = \max_{i_j < i_{j+1}} \gamma_{i_j i_{j+1}} \quad \text{and} \quad Z = \min_{i_j > i_{j+1}} \gamma_{i_j i_{j+1}}.$$

Then $Y \ge Z$.

Proof. The following pictorial convention will be used. Let the integers $i_1, i_2, \ldots, i_r$ be the vertices of a polygon P, arranged in clockwise order, as in Fig. 11, with $i_{r+1} = i_1$. Each edge $e_{i_j i_{j+1}}$ of P is assigned the value $\gamma_{i_j i_{j+1}}$. Thus we can identify $e_{i_j i_{j+1}}$ with $e_{i_{j+1} i_j}$, since $\gamma_{i_j i_{j+1}} = \gamma_{i_{j+1} i_j}$. If $i_j < i_{j+1}$, we call $e_{i_j i_{j+1}}$ an *upward* edge; if $ji > i_{j+1}$, we call $e_{i_j i_{j+1}}$ a *downward* edge. The inequality $M \ge m$ means that some upward edge has a larger γ than some downward edge.

The proof proceeds by induction. For $n = 2$, $M = m$. For $n = 3$, either $i_1 < i_2 < i_3$ or $i_1 > i_2 > i_3$. In the former case, $Y = \max(\gamma_{i_1 i_2}, \gamma_{i_2 i_3})$ and $Z = \gamma_{i_3 i_1} = \gamma_{i_1 i_3}$; clearly $Y \ge Z$. The latter case can be similarly treated.
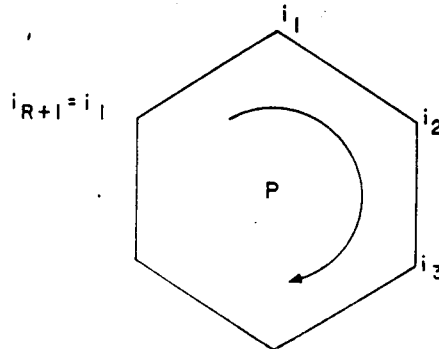

FIG. 11. Polygonal path between indices.

Suppose that the lemma holds for $n = r \geqslant 3$, and let P be an $(r+1)$-gon. Then there exists some diagonal through the polygon's vertices which is an upward edge, say between vertices i_l and $i_m (i_l > i_m)$; see Fig. 12. In Fig. 12, $e_{i_l i_m}$ is an upward edge of P_1 and $e_{i_m i_l}$ is a downward edge of P_2. P_1 satisfies the inductive hypothesis, so some upward edge P_1 has larger γ than some downward edge of P_1. That downward edge of P_1 must be a downward edge of P. That upward edge of P_1 must either be an upward edge of P or else $e_{i_l i_m}$. In the first case, the lemma holds for P. In the second case, $\gamma_{i_l i_m} \geqslant \gamma_{i_j i_{j+1}}$ for $e_{i_j i_{j+1}}$ a downward edge of P_1, and thus of P. Likewise, some downward edge of P_2 must have a smaller γ than some upward edge of P_2. By the same reasoning, either the lemma holds for P or $\gamma_{i_l i_m} \leqslant \gamma_{i_k i_{k+1}}$, where $e_{i_k i_{k+1}}$ is a downward edge of P_2, and thus of P.
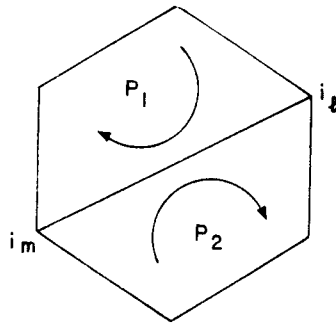

FIG. 12. Oriented subpolygons.

The only remaining case is where

$$\gamma_{i_j i_{j+1}} \leqslant \gamma_{i_l i_m} \leqslant \gamma_{i_k i_{k+1}},$$

so again an upward edge of P has a γ at least as large as the γ's of downward edges of P.

To complete the proof of the theorem, we argue by contradiction. Suppose that there exists an infinite sequence of times $t_1 < t_2 < \ldots$ when one $B_i g_i$ overtakes another $B_j g_j$ as maximal. Since there are only finitely many indices i, the indices of the maximal $B_i g_i$ reoccur infinitely often. Suppose that $B_{i_{k+1}} g_{i_{k+1}}$ overtakes $B_{i_k} g_{i_k}$ at $t = t_k$. Then there are infinitely many subsequences $t_j < t_{j+1} < \ldots t_{j+p} < t_{j+p+1}$ such that all the indices $i_j, i_{j+1}, \ldots, i_{j+p}$ are distinct, but $i_{j+p+1} = i_j$. Consider such a subsequence; i.e., let the successive subscripts of the maximal $B_{i_k} g_{i_k}$ go around a polygon. By the proof of Lemma 3, if $i_{k+1} > i_k$ and $B_{i_{k+1}} g_{i_{k+1}}$ overtakes $B_{i_k} g_{i_k}$ at $t = t_k$, then $F(x_{i_{k+1}}(t_k)) \geqslant F(x_{i_k}(t_k))$. By (21), it follows that $x_{i_{k+1}}(t_k) \geqslant U_{i_{k+1} i_k}$ and since $g(w)$ is monotone increasing, $i_{k+1} > i_k$ implies that

$$\max_i B_i g_i(t_k) = B_{i_{k+1}} g_{i_{k+1}}(t_k) \geqslant \gamma_{i_{k+1} i_k} = \gamma_{i_k i_{k+1}} \qquad (24a)$$

Similarly, if $i_{k+1} < i_k$, then

$$\max_l B_l g_l(t_k) \leq \gamma_{i_k i_{k+1}}. \qquad (24b)$$

Suppose that between times t_j and t_{j+p+1}, $\max_l B_l g_l$ never crosses a value γ_{ij}. Then (24a) and (24b) hold at all times between t_j and t_{j+p+1}. Thus on the interval $[t_j, t_{j+p+1}]$,

$$\max_{i_k < i_{k+1}} \gamma_{i_k i_{k+1}} \leq \max_l B_l g_l \leq \min_{i_k > i_{k+1}} \gamma_{i_k i_{k+1}}. \qquad (25)$$

By Lemma 6,

$$\max_{i_k < i_{k+1}} \gamma_{i_k i_{k+1}} \geq \min_{i_k > i_{k+1}} \gamma_{i_k i_{k+1}},$$

which is possible only if

$$\max_l B_l g_l = \max_{i_k < i_{k+1}} \gamma_{i_k i_{k+1}} = \min_{i_k > i_{k+1}} \gamma_{i_k i_{k+1}} \text{ on } [t_j, t_{j+p+1}].$$

But $\max_l B_l g_l$ is strictly increasing. By this contradiction, during each cycle of subscripts $i_j, \ldots, i_{j+p}$, $\max_l B_l g_l$ must cross some γ_{ij}. There are only finitely many γ_{ij} and $\max_l B_l g_l$ is increasing. Hence there are only finitely many cycles of subscripts. This contradiction proves the theorem for the case of $g(w)$ strictly increasing.

Now suppose g is strictly increasing for $w \leq x^{(1)}$ and constant on $w \geq x^{(1)}$. Then the above proof is still valid with two minor modifications.

(1) Define U_{ij}, V_{ij} as before but relative to the portion of the curve C_{ij} corresponding to x_i and $x_j \leq x^{(1)}$. Suppose $B_i > B_j$. At a time when $B_i g_i = B_j g_j$, then $g_i < g_j$, and so $x_i < x_j$ and $x_i \leq x^{(1)}$. If also $x_j \leq x^{(1)}$, then g is increasing at x_i and x_j, and so (24) still holds. If however $x_j \geq x^{(1)}$, then $x_j \geq V_{ij}$ and $B_i g_i = B_j g_j \geq B_j g(V_{ij}) = B_i g(U_{ij})$, so that $x_i \geq U_{ij}$. Thus $B_j g_j$ cannot overtake $B_i g_i$ since $(B_j g_j)^{\cdot} = 0$ and $(B_i g_i)^{\cdot} \geq 0$. Again (24) holds.

(2) The function $\max_k B_k g_k$ is no longer strictly increasing, but it is increasing whenever $B_i g_i$ overtakes $B_j g_j$ as maximal, so (25) is again impossible.

Proof of Theorem 4. By (10) and the fact that $B_1 \leq \ldots \leq B_n$, with $B_1 < B_n$, it is readily shown that

$$x < Cx[B_n - AC^{-1} - x].$$

Hence $\dot{x} < 0$ until $x \leq B_n - AC^{-1}$, and this latter inequality propagates. We consider the system after this inequality sets in. If ever some $x_i(t) > x^{(2)}$, than all other $x_j(t) < x^{(0)}$. As in Lemma 4, we assume without loss of generality that $\max_k B_k g_k \geq A + F$. Suppose for the i such that $x_i > x^{(2)}$ $B_i g_i = \max_k B_k g_k$ at $t = T$. Since $g(w)$ is increasing on $[0, x^{(0)}]$,

$$B_j g_j < B_j g(x^{(0)}) = B_j B_n^{-1} f(B_1) \leq f(B_1)$$

but since $g(w)$ is decreasing on $[x_i(T), B_i]$,

$$f(B_1) \leq f(B_i) \equiv B_i g(B_i) \leq B_i g_i.$$

In all, $B_j g_j < B_i g_i$. Thus no other $B_j g_j$ can overtake $B_i g_i$ unless x_i falls below $x^{(2)}$, which cannot happen while $B_i g_i$ is maximal since $B_i g_i \geq A + F$ and so $\dot{x}_i \geq 0$. Thus $x_i \geq 0$ for $t \geq T$, so x_i approaches a positive limit, which by Lemma 2 completes the proof. If the x_i's corresponding to $\max_k B_k g_k$ never exceed $x^{(2)}$, then the proof for $g(w)$ nondecreasing in Theorem 3 still holds.

Proof of Corollary 2. By Theorem 4, it suffices to show that $F(w)$ is uni-modal. If $F'(w) = 0$, then it readily follows that

$$f(w)f'(w)(P(w) - F(w)) = 0.$$

By hypothesis, $f'(w) > 0$, and $f(w) > 0$ unless $w = 0$. Consequently, $P(w) = F(w)$ at any critical point of $F(w)$.

Using the hypothesized unimodality of $P(w)$, it will suffice to prove that $F(w)$ is increasing at small values of w, and that $F(\infty) = 0$. For suppose that w_1 and w_2 are the smallest two positive critical points of $F(w)$. Since $F(w)$ is increasing for small w, the point w_1 is a relative maximum of $F(w)$, and the point w_2 is a relative minimum of $F(w)$. However $F(w)$ is positive and $F(\infty) = 0$. Thus there exists a third critical point, say w_3, of $F(w)$, which is a relative maximum of $F(w)$. But $F(w_i) = P(w_i)$, $i = 1, 2, 3$ which violates the inimodality of $P(w)$. Hence $F(w)$ has only one critical point, which is a maximum.

To show that $F(w)$ is increasing at small w, note by L'Hôpital's rule that

$$\lim_{w \to 0} F = \lim_{w \to 0} \frac{f' + wf''}{f'} = \frac{p(0)}{p(0)} = 1,$$

whereas since $f(w)$ is sigmoid, $g'(w) > 0$ for small w, and thus at such values of w, $F(w) = 1 + G(w) > 1$.

That $F(\infty) = 0$ is an obvious consequence of

$$\lim_{w \to \infty} wp(w) = 0.$$

Proof of Theorem 5. By Lemma 4, the inequality $\max_k B_k g_k \geq A + H$ propagates. If $\max_k B_k g_k \leq A + H$ at arbitrarily large times, then all x_i are decreasing and all limits exist.

Without loss of generality, assume $\max_k B_k g_k \geq A + H$ for all $t \geq 0$. By (10) and Lemma 4, since all $B_i = B_n$, the x_i whose g_i is maximal will always be nondecreasing. Also, by Lemma 1, we can assume that the x_i's are in invariant order for all $t \geq 0$.

Let g_i across g_j at time τ. Then $x_i(\tau) \neq x_j(\tau)$, since equality propagates in the EW case. Since $g(x_i(\tau)) = g(x_j(\tau))$, only three cases are possible:

1. $0 \leq x_i(\tau) < x^{(1)}, x^{(2)} < x_j(\tau) \leq B_n$;
2. $0 \leq x_j(\tau) < x^{(1)}, x^{(2)} < x_i(\tau) \leq B_n$;
3. $x^{(1)} \leq x_i(\tau) \leq x^{(2)}, x^{(1)} \leq x_j(\tau) \leq x^{(2)}$.

We will show that only Case 1 can arise.

Consider Case 2. Clearly $F(w) = 1 + G(w) > 1$ for $0 \leq w < x^{(1)}$ and is < 1 for $x^{(2)} < w \leq B_n$. Hence $F(x_i(\tau)) < F(x_j(\tau))$, but by the proof of Lemma 3, $F(x_i(\tau)) \geq F(x_j(\tau))$. Case 3 is impossible because g_i and g_j cannot cross each other at a time when x_i and x_j are in the constant range of $g(w)$. Consequently, only Case 1 can arise, and so $x_i < x_j$ whenever $B_i g_i$ crosses $B_j g_j$. Since the x_i's are in invariant order, there can be at most $n-1$ changes in which $B_i g_i$ is maximal. Thus some $B_i g_i = \max_k B_k k_k$ for all sufficiently large t; by (10), x_i is then increasing, consequently has a positive limit, and so by Lemma 2, all limits exist.

Proof of Theorem 6. First suppose $g(w)$ is monotone nondecreasing. Consider a nontrivial periodic solution of (1). By Lemma 4, at all sufficiently large times, $\max_k B_k g_k \geq A + H$, since otherwise all x_i are decreasing and thus not periodic. In a time interval such that $B_i g_i = \max_k B_k g_k$, $\dot{x}_i > 0$ except at isolated points, by (10) and Lemma 4. Thus $B_i g_i$ is nondecreasing. At points where some $B_j g_j$ crosses $B_i g_i$, the function $\max_k B_k g_k$ is continuous, hence it is nondecreasing for all $t \geq 0$. But $\max_k B_k g_k$ is a periodic function, and is therefore constant. If $B_i g_i = \max_k B_k g_k$ over a given time interval, then this must happen in a constant interval of g, since otherwise $\dot{x}_i > 0$ and thus $B_i g_i$ is increasing except at isolated points. If x_i is in a constant interval of g, however, $B_j g_j$ can only overtake, and not cross, $B_i g_i$ since $\max_k B_k g_k$ is constant and $B_i g_i$ cannot decrease immediately after such a time, by (10). Hence there are no crossings, and thus there is an index i such that $B_i g_i = \max_k B_k g_k$ for all time. This x_i is nondecreasing, and is hence not periodic.

The case of $g(w)$ nonincreasing can be similarly treated using the fact that inequality $\min_k B_k g_k \leq A + H$ propagates.

Proof of Theorem 7. Since $F(w) = C$, Theorem 2 implies that all limits $x_i(\infty)$ exist. By (10), the limits $x_i(\infty)$ also exist, and must equal 0, or else x_i will be unbounded, which is impossible. In particular, if $x_i(\infty) \neq 0 \neq x_j(\infty)$, then by (10),

$$B_i g(x_i(\infty)) = A + F(\infty) = B_j g(x_j(\infty)), \tag{26}$$

and since $g(w) \equiv C$, $B_i = B_j$. In other words, all nonzero limits are in the same subfield. We show that this subfield corresponds to the maximal B_i

by contradition. Let $B_i < B_j$ and $x_i(\infty) > 0 = x_j(\infty)$; in particular, $x_i(\infty)x_j^{-1}(\infty) = \infty$. By (10), however,

$$\frac{d}{dt} \log\left(\frac{x_i}{x_j}\right) = B_i g_i - B_j g_j = C(B_i - B_j) < 0, \tag{27}$$

which contradicts $x_i(\infty)x_j^{-1}(\infty) = \infty$.

The limiting distribution is fair within the B_n-subfield because $B_i = B_j$ implies, by (27), that

$$\frac{d}{dt} \log\left(\frac{x_i}{x_j}\right) \equiv 0.$$

The reverberation persists if $CBn > A$ because, by (10),

$$\dot{x}(\infty) = x(\infty)[CB_n - A - Cx(\infty)],$$

which also shows that $x(\infty) = \max(B_n - AC^{-1}, 0)$. To prove (7), note by (27) that for every i,

$$x_i(t) = x_i(0)x_n^{-1}(0)x_n(t)e^{C(B_i - B_n)t}. \tag{28}$$

By (10) and (28),

$$\dot{x}_n - \mu x_n + v(t)x_n^2 = 0 \tag{29}$$

where $\mu = CB_n - A$ and

$$v(t) = Cx_n^{-1}(0) \sum_{k=1}^{n} x_k(0)\, e^{C(B_k - B_n)t}.$$

The Riccati equation (29) can be transformed by the standard substitution

$$x_n = v^{-1}\frac{d}{dt}\log y$$

into the equation

$$\ddot{y} = g(t)\dot{y}, \quad g(t) = \mu + \frac{d}{dt}\log v(t).$$

Solving this equation shows that

$$x_n = \frac{1}{v}\frac{d}{dt}\log\left[1 + \left(\frac{d}{dt}\log y(0)\right)\int_0^t \exp\left(\int_0^v g\, d\xi\right) dv\right].$$

Since

$$\exp\left(\int_0^t g\, dv\right) = v^{-1}(0)v(t)\exp(\mu t),$$

this equation is readily simplified, and when substituted into (28) yields the desired result.

Proof of Theorem 8. To show that the limiting distribution is uniform within a subfield, we consider three cases: (1) if $x_i(\infty) = 0 = x_j(\infty)$, we are done. (2) Suppose $x_i(\infty) \neq 0 \neq x_j(\infty)$. Then by (26), $g(x_i(\infty)) = g(x_j(\infty))$, and

T.B.

25

since $g(w)$ is strictly monotonic, $x_i(\infty) = x_j(\infty)$. (3) It cannot happen that $x_i(\infty) > 0 = x_j(\infty)$: if $B_i = B_j$, then

$$\frac{d}{dt} \log \left[\frac{x_j(\infty)}{x_i(\infty)}\right] = B_i[g(x_j(\infty)) - g(x_i(\infty))] > 0.$$

Thus $x_i(\infty) > 0$ implies $x_j(\infty) = \infty$ rather than 0.

It is obvious from (26) that $B_i \geq B_j$ implies $x_i(\infty) \geq x_j(\infty)$.

To show that $x_i(\infty) = 0$ if $B_i g(0) < A$, note by (10) that, for all $t \geq 0$,

$$\dot{x}_i \leq x_i(B_i g(x_i) - A)$$

$$\leq x_i(B_i g(0) - A),$$

which implies that x_i approaches 0 at an at least exponential rate.

To show that there is a unique limiting distribution corresponding to every set of positive $x_i(\infty)$ use (26) to find that $x_i(\infty)$ satisfies the equation

$$B_i g(w) = A + \sum_{k=1}^{n} g^{-1}\left(\frac{B_i}{B_k} g(w)\right) \frac{B_i}{B_k} g(w),$$

where $x_{l-1}(\infty) = 0 < x_l(\infty)$. The left-hand side of this equation is a strictly decreasing function of w. The right-hand side is a strictly increasing function of w, since $f(w) = wg(w)$ is such a function. This proves uniqueness. The condition for reverberation is proved as in Grossberg (1973).

Proof of Theorem 9. To prove that the limiting distribution is 0–1 or locally uniform relative to a subfield, define

$$X_j^{(i)} = x_j \left(\sum_{B_k = B_i} x_k\right)^{-1}$$

and find that

$$\dot{X}_j^{(i)} = B_i X_j^{(i)} \sum_{B_k = B_i} X_k^{(i)}(g_j - g_k). \tag{30}$$

This equation has the same form as in the faster-than-linear EW case treated by Grossberg (1973).

To prove the instability of all critical points with more than one $\xi_i > 0$, order the x_i's in such a fashion that $\xi_1, \ldots, \xi_m$ are positive, and $\xi_{m+1}, \ldots, \xi_n$ are zero. We will compute the Jacobian matrix of the system $\dot{x}_i = h_i(x_1, \ldots, x_n)$ defined by

$$h_i(x_1, \ldots x_n) = x_i\left[-A + (B_i - x_i)g_i - \sum_{k \neq i} x_k g_k\right].$$

Clearly,

$$\frac{\partial h_i}{\partial x_i} = -A + (B_i - x_i)g_i - \sum_{k \neq i} x_k g_k + x_i[(B_i - x_i)g_i' - g_i].$$

Suppose $0 \le i \le m$. Since

$$0 = h_i(\xi_1, \ldots, \xi_n) = \xi_i \left[-A + (B_i - \xi_i) g(\xi_i) - \sum_{k \ne i} \xi_k g(\xi_k) \right]$$

and $\xi_1 \ne 0$, we find

$$\frac{\partial h_i}{\partial x_i}(\xi) = \xi_i [(B_i - \xi_i) g'(\xi_i) - g(\xi_i)]. \tag{31}$$

Suppose $i > m$. Then $x_i = 0$ and (31) becomes

$$\frac{\partial h_i}{\partial x_i}(\xi) = -[A + F(\xi)] + B_i g(0). \tag{32}$$

For all $j \ne i$,

$$\frac{\partial h_i}{\partial x_j} = -x_i [x_j g'(x_j) + g(x_j)].$$

Thus for $0 \le i \le m$,

$$\frac{\partial h_i}{\partial x_j}(\xi) = -\xi_i [\xi_j g'(\xi_j) + g(\xi_j)] \tag{33}$$

and for $i > m$,

$$\frac{\partial h_i}{\partial x_j}(\xi) = 0. \tag{34}$$

Combining (31)–(34) yields a Jacobian matrix of the form

$$\mathbf{J} = \left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array} \right).$$

The $m \times m$ matrix $\mathbf{A} = \|a_{ij}\|$ has entries

$$a_{ij} = -\xi_1 [\xi_j g'(\xi_j) + g(\xi_j)] + \delta_{ij} B_i \xi_i g'(\xi_i),$$

where $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \ne j$. The matrix $\mathbf{B}$ is irrelevant. The $(n-m) \times (n-m)$ matrix $\mathbf{C} = \|c_{ij}\|$ has entries

$$C_{ij} = \delta_{ij} [-(A + F(\xi)) + B_i g(0)],$$

$m+1 \le i \le n$. The eigenvalues λ of $\mathbf{J}$ are solutions of the equation

$$0 = \det \left(\begin{array}{c|c} A - \lambda I & B \\ \hline 0 & C - \lambda I \end{array} \right) = \det (A - \lambda I) \det (C - \lambda I)$$

It therefore suffices to compute the eigenvalues of $\mathbf{A}$ and of $\mathbf{C}$ separately.

Matrix $\mathbf{A}$ can be written in the form $\mathbf{A} = \mathbf{D} - \mathbf{R}$, where $\mathbf{D}$ is a diagonal matrix whose ith diagonal entry is $d_{ii} = B_i \xi_i g'(\xi_i)$, and $\mathbf{R}$ is a matrix of rank 1 with entries $r_{ij} = \xi_i [\xi_j g'(\xi_j) + g(\xi_j)]$. Concerning such a matrix the following lemma holds.

Lemma 7. Let $\mathbf{E} = \|e_{ij}\|$ and $\mathbf{F} = \|f_{ij}\|$ be $m \times m$ matrices such that $\mathbf{E}$ is diagonal and $\mathbf{F}$ is of rank 1. Then

$$\det (\mathbf{E}+\mathbf{F}) = \prod_{k=1}^{m} e_{kk}\left(1+ \sum_{k=1}^{m} \frac{f_{kk}}{e_{kk}}\right). \tag{35}$$

Proof. Introduce the notation $e_k = e_{kk}$ and let $\det F_{ijk} \ldots$ be the minor of $\mathbf{F}$ with the $i, j, k, \ldots$th rows and columns removed. Expand $\det (\mathbf{E}+\mathbf{F})$ in terms of sums of products of the form

$$\det F_{w(1) \ldots w(j)} \prod_{k=1}^{j} e_{w(k)},$$

where $1 \leq w(1) \leq \ldots \leq w(j) \leq m$. The coefficient $\det F_{w(1) \cdots w(j)}$ vanishes either if the $w(i)$ are not all distinct or, because $\mathbf{F}$ is of rank 1, if $j < n-1$. The only nonvanishing coefficients correspond to distinct $w(i)$ with $j = n-1$ or n. If $j = n-1$ and the sequence $w(1), \ldots, w(n-1)$ is $1, \ldots, n$ with i removed, then

$$\prod_{k=1}^{j} e_{w(k)} = e_{ii}^{-1} \prod_{k=1}^{n} e_{kk}$$

and $\det F_{w(1)\ldots w(j)} = f_{ii}$. If $j = n$, then

$$\prod_{k=1}^{j} e_{w(k)} = \prod_{k=1}^{n} e_{kk}$$

and $\det F_{w(1)\ldots w(j)} = 1$. Adding these $m+1$ products yields (35).

Using Lemma 7, we can define a new matrix $\mathbf{A}' = \mathbf{D}-\mathbf{R}'$ with the same eigenvalues as $\mathbf{A} = \mathbf{D}-\mathbf{R}$, such that $\mathbf{R}'$, in addition to being of rank 1 and having positive entries, is also symmetric. By Lemma 7, the eigenvalues λ of $\mathbf{A}$ satisfy the equation

$$0 = \prod_{k=1}^{m} (d_{kk}-\lambda)\left(1- \sum_{k=1}^{m} \frac{r_{kk}}{d_{kk}-\lambda}\right).$$

Note that only the diagonal entries of $\mathbf{R}$ enter this equation. Define $\mathbf{R}' = \|r'_{ij}\|$ by $r'_{ij} = (r_{ii}r_{jj})^{\ddagger}$. Clearly $r'_{ii} = r_{ii}$ and $r'_{ij} = r'_{ji}$. If also $\mathbf{R}'$ is of rank 1, then our assertion is proved. This follows from the fact that $r'_{ij}(r'_{kj})^{-1} = r_{ii}^{\ddagger}r_{kk}^{-\ddagger}$ is independent of j. Using this fact, we can prove the following lemma.

Lemma 8. If $\mathbf{A}$ is an $m \times m$ matrix such that $\mathbf{A} = \mathbf{D}-\mathbf{R}$, where $\mathbf{D}$ is a diagonal matrix with positive diagonal entries, and $\mathbf{R}$ is a real matrix of rank 1 with positive entries, then $\mathbf{A}$ has at least $(m-1)$ eigenvalues of positive real part.

Proof. We consider $\mathbf{A}'$ instead of $\mathbf{A}$. Since $-\mathbf{R}'$ is of rank 1, $(m-1)$ of its eigenvalues are 0. We wish to show that, because $\mathbf{D}$ is a diagonal matrix

with positive diagonal entries, the eigenvalues of $\mathbf{A}' = \mathbf{D} - \mathbf{R}'$ corresponding to the 0 eigenvalues of $-\mathbf{R}'$ are positive. This can be seen in several ways. For example, by Courant & Hilbert (1953, pp. 31–34), if $\mathbf{K}$ is a real symmetric matrix, then its eigenvalues can be found by defining the quadratic form $\mathbf{K}(x) = x'\mathbf{K}x$, maximizing the value of $\mathbf{K}(x)$ subject to the constraints

$$\sum_{p=1}^{n} x_p^2 = 1$$

and

$$\sum_{p=1}^{m} a_{lp} x_p = 0,$$

$l = 1, 2, \ldots, r-1$, for some real values a_{lp}, and then minimizing this value over all sets of $\{a_{lp}\}$. The rth eigenvalue of k is the result of this minimax procedure.

Given Lemma 6, the instability result for the nonlinear system $\dot{x}_i = h_i$ follows from the fact that $\mathbf{J}$ has at least $m-1$ positive eigenvalues, using Theorem 4.1 in chap. 13 of Coddington & Levinson (1955).

The remaining assertions of the theorem are proved in Grossberg (1973).

Proof of Theorem 10. We will prove that inequality $x_i > M_{ij}x_j$ propagates; the inequality $x_i < m_{ij}x_j$ can be similarly treated. Let $x_i x_j^{-1} > M_{ij}$ at any time $t = \tau$. Then at $t = \tau$, (27) implies

$$\frac{\mathrm{d}}{\mathrm{d}t} \log\left(\frac{x_i}{x_j}\right) > B_i g(x_i) - B_j g(M_{ij}^{-1} x_i).$$

By definition of M_{ij}, $M_{ij} \geq x_i / g^{-1}(B_i B_j^{-1} g(x_i))$. Thus

$$\frac{\mathrm{d}}{\mathrm{d}t} \log\left(\frac{x_i}{x_j}\right) > 0, \tag{36}$$

and inequality $x_i x_j^{-1} > M_{ij}$ propagates.

If $B_i g(0) > B_j g(B_j)$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} \log\left(\frac{x_i}{x_j}\right) \geq K_{ij} > 0,$$

where $K_{ij} = B_i g(0) - B_j g(B_j)$. Integrating this inequality shows that

$$B_i \geq x_i \geq x_j \exp(K_{ij}t) \geq 0,$$

so that eventually $x_i > x_j$ and $x_j(\infty) = 0$.

Suppose $x_i(0) < m_i$ where $g(m_i) = AB_i^{-1}$. Since

$$\dot{x}_i \leq x_i[-A + B_i g(x_i)], \tag{37}$$

$\dot{x}_i(0) < 0$, which can only make the right-hand side of (37) more negative. Clearly x_i decreases to zero at an at least exponential rate.

Proof of Theorem 11. The proof follows that of Theorem 10, except for the statement about contrast enhancement relative to a subfield, which is proved just as in Theorem 4 of Grossberg (1973) for the EW case, using (30).

REFERENCES

ANDERSON, P., GROSS, G. N., LOMO, T. & SVEEN, O. (1969). In *The Interneuron* (M. Brazier, ed.) p. 415. Los Angeles: University of California Press.
BLAKEMORE, C. & COOPER, G. F. (1970). *Nature, Lond.* **228,** 477.
CODDINGTON, E. & LEVINSON, N. (1955). *Theory of Ordinary Differential Equations.* New York: McGraw-Hill.
COURANT, R. & HILBERT, D. (1953). *Methods of Mathematical Physics*, vol. 1. New York: Interscience.
ECCLES, J. C., ITO, M. & SZENTAGOTHAI, J. (1967). *The Cerebellum as a Neuronal Machine.* New York: Springer.
ELLIAS, S. A. & GROSSBERG, S. (1976). *Biol. Cybernetics* (in press).
ESTES, W. K. (1972). In *Coding Processes in Human Memory* (A. W. Melton & E. Martin, eds). Washington, D.C.: V. H. Winston and Sons.
GROSSBERG, S. (1970). *J. theor. Biol.* **27,** 291.
GROSSBERG, S. (1971). *J. theor. Biol.* **33,** 225.
GROSSBERG, S. (1972a). *Math. Biosci.* **15,** 39.
GROSSBERG, S. (1972b). *Math. Biosci.* **15,** 253.
GROSSBERG, S. (1972c). *Kybernetik* **10,** 49.
GROSSBERG, S. (1973). *Stud. Appl. Math.* **52,** 213.
GROSSBERG, S. (1975a). *Int. Rev. Neurobiol.* **18** (in press).
GROSSBERG, S. (1975b). *Biol. Cybernetics* (in press).
HEBB, D. O. (1949). *The Organization of Behavior.* New York: Wiley.
HIRSCH, H. V. B. & SPINELLI, D. N. (1970). *Science, N.Y.* **168,** 869.
HODGKIN, A. L. (1964). *The Conduction of the Nervous Impulse.* Springfield, Ill.: C. C. Thomas.
HONIG, W. K. (1970). In *Attention: Contemporary Theory and Analysis* (D. I. Mostofsky, ed.) p. 193. New York: Appleton-Century-Crofts.
KERNELL, D. (1965a). *Acta physiol. scand.* **65,** 65.
KERNELL, D. (1965b). *Acta physiol. scand.* **65,** 74.
LEVINE, D. S. & GROSSBERG, S. (1976). Submitted for publication.
RALL, W. (1955). *J. cell. comp. Physiol.* **46,** 413.
SPERLING, G. (1970). *Percept. Psychophys.* **8,** 143.
SPERLING, G. & SONDHI, M. M. (1968). *J. opt. Soc. Am.* **58,** 1133.
STEFANIS, C. (1969). In *The Interneuron* (M. Brazier, ed.) p. 497. Los Angeles: University of California Press.
WIESEL, T. N. & HUBEL, D. H. (1963). *J. Neurophysiol.* **26,** 1003.
WIESEL, T. N. & HUBEL, D. H. (1965). *J. Neurophysiol.* **28,** 1029.