Kybernetik 10, 49—57 (1972)

Neural Expectation: Cerebellar and Retinal Analogs of Cells Fired by Learnable or Unlearned Pattern Classes

STEPHEN GROSSBERG *
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A.

Received July 24, 1971

Abstract. Neural networks are introduced which can be taught by classical or instrumental conditioning to fire in response to arbitrary learned classes of patterns. The filters of output cells are biased by presetting cells whose activation prepares the output cell to "expect" prescribed patterns. For example, an animal that learns to expect food in response to a lever press becomes frustrated if food does not follow the lever press. It's expectations are thereby modified, since frustration is negatively reinforcing. A neural analog with aspects of cerebellar circuitry is noted, including diffuse mossy fiber inputs feeding parallel fibers that end in Purkinje cell dendrites, climbing fiber inputs ending in Purkinje cell dendrites and giving off collaterals to nuclear cells, and inhibitory Purkinje cell outputs to nuclear cells. The networks are motivated by studying mechanisms of pattern discrimination that require no learning. The latter often use two successive layers of inhibition, analogous to horizontal and amacrine cell layers in vertebrate retinas. Cells exhibiting hue (in)constancy, brightness (in)constancy, or movement detection properties are included. These results are relevant to Land's retinex theory and to the existence of opponent- and nonopponent-type cell responses in retinal cells. Some adaptation mechanisms, and arousal mechanisms for crispening the pattern weights that can fire a given cell, are noted.

1. Introduction

This paper describes neural networks containing cells U that fire in response to arbitrary learnable classes of input patterns. These input patterns need not be precoded in the anatomy or physiology of the networks at their birth, and can be changed at will throughout their life. The cells U can also learn to control arbitrary output patterns. U cells are thus "universal" cells whose input and output properties are entirely flexible and changeable by classical and instrumental conditioning.

The learned class of input patterns which can fire a cell U is controlled by presetting cells P. The cells P send axons to the filtering mechanism (e.g., inhibitory interneurons and dendrites) that processes inputs to U. Each P cell can learn a particular pattern that will bias U's filter when P is active. A given P cell can bias several U cells, each with a different learned—or unlearned—pattern. If more than one P cell is active, then U's filter is biased by a pattern which is a weighted average of the patterns controlled separately by each active P cell. The weighting pattern can be changed by altering the relative firing intensities of the active P cells.

The need for such presetting mechanisms is suggested by learning experiments in which an animal O's expectations of future events are learned. For example, let O learn to lever press for food. On a recall trial, O "expects" food when it lever presses in response to hunger. If food is not delivered, O can become "frustrated". Frustration is negatively reinforcing (Kimble, 1961; Wagner, 1969) and can modify O's future expectations. Frustration is biologically useful, since it permits O to eliminate learned, but later unsuccessful, instrumental behavior before irreversible damage is done by the absence of primary reinforcement. On recall trials in the lever press example, lever press cures are presumed to also preset con-

* Supported in part by the Alfred P. Sloan Foundation and the Office of Naval Research (N00014-67-A-0204-0051).
Address: Massachusetts Institute of Technology, 2-382, Cambridge, Massachusetts 02139.

summatory controls which can be released by expected sensory cues of the food reward. The presetting cells are analogous to the cells P and the consummatory control cells are analogous to the cells U. This paper approaches the study of neural expectation and its frustration by sketching some mathematical features of presetting mechanisms. Presetting mechanisms will later be merged with mechanisms of "negative incentive motivation", as developed in Grossberg (1972), to create a more global picture of the expectation-frustration mechanism.

Given a catalog of preset mechanisms, one naturally seeks analogies with known anatomies *in vivo*. This search is complicated by two factors. First, seemingly different anatomies can share common preset mechanisms; their differences might only be due to differences in arrangement ("symmetries") of P and U cells, and in the distribution of $P \to U$ axons. Second, the same preset mechanism can form part of different total processing schemes in different cell groups. A striking analogy with aspects of cerebellar anatomy is nonetheless discernible. This analogy includes the following cerebellar facts (Bell and Dow, 1967; Eccles, Ito, and Szentogothai, 1967): (1) existence of a diffuse excitatory mossy fiber input; (2) existence of a localized excitatory climbing fiber input that branches off from excitatory axons to cerebellar nuclear cells; (3) existence of an inhibitory Purkinje cell output to nuclear cells; (4) possibility of firing the Purkinje cell in response to either, or both, input sources; (5) possibility of plastic changes at parallel fiber—Purkinje spine contacts.

The network anatomies use inhibitory interneurons in several configurations. A particular configuration is partly determined by the physiology of its individual cells; for example, by whether the inhibition is subtractive or shunting (multiplicative) (Creutzfeldt, Sakmann, Scheich, and Korn, 1970; Sperling, 1970; Sperling and Sondhi, 1968). Each configuration determines a characteristic time response of the network to test inputs; for example, transient "on-off" responses, or saturated "on" responses, or graded responses to rectangular inputs can occur. Alternatives to inhibitory interneurons exist in some cases; for example, blockade of postsynaptic potential response at higher than prescribed spiking frequencies, or switch-over from net excitation to net inhibition of postsynaptic sites as presynaptic spiking frequency exceeds a critical value (Bennett, 1971; Blackenship, Wachtel, and Kandel, 1971; Wachtel and Kandel, 1971). The present mechanisms include as special cases some mechanisms of sensory adaptation, and the possibility of "crispening" the pattern weights that can elicit a response from U by altering the ambient arousal level.

U cell anatomies are related to the anatomies of cells R which can discriminate patterns without prior learning or presetting. R cell anatomies were introduced in Grossberg (1970a). They also have neural analogs, for example in vertebrate retinas. Such cells can, for example, exhibit essentially perfect hue constancy or brightness constancy (Cornsweet, 1970). Two successive stages of lateral inhibitory interactions, reminiscent of horizontal cell and amacrine cell interactions (Dowling and Werblin, 1969; Werblin and Dowling, 1969), prepare the inputs to R cells to achieve this constancy. R cell response can also exhibit shifts in perceived hues and brightnesses as background illumination increases. Perfect constancy cells need differ from imperfect constancy cells only in the spatial distribution of the inhibitory interneurons that prepare their outputs. Logarithmic transformation of peripheral inputs is unnecessary to achieve hue or brightness constancy of the R cell variety.

2. Theoretical Review

A previous paper (Grossberg, 1970a) introduced neural networks which can discriminate arbitrary input patterns and learn to release arbitrary output patterns in response to prescribed input patterns. These discrimination mechanisms are ritualistic in the sense that a given output cell responds to the same class of patterns at all times. The ritualistic discrimination and learning mechanisms can be modified to construct the cells U. They are therefore reviewed below.

Grossberg discusses the following situation. Let n cells v_i be given, $i = 1, 2, \ldots, n$. Let the input to v_i be $C_i(t)$. A *spatial pattern* is an excitatory input to $V = \{v_i : i = 1, 2, \ldots, n\}$ of the form $C_i(t) = \tilde{\theta}_i\, C(t)$, where $\tilde{\theta}_i$ is the fixed relative intensity of the input at v_i (hence $\tilde{\theta}_i \geqq 0$ and $\sum_{k=1}^{n} \tilde{\theta}_k = 1$), and $C(t)$ is the total input intensity. This concept of spatial pattern notes that recognition of a picture is invariant under considerable fluctuations in background illumination. Grossberg considers the following problem: what is the *minimal* anatomy and physiology of an input filtering device, fed by V, that fires an output cell R if and only if

$$\theta_i - \varepsilon < \tilde{\theta}_i < \theta_i + \varepsilon \tag{1}$$

for all $i = 1, 2, \ldots, n$; that is, if and only if the pattern $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_n)$ differs by less than ε from a prescribed pattern $\theta = (\theta_1, \ldots, \theta_n)$? It will be shown that such a cell R can exhibit perfect hue constancy or brightness constancy, within an error of ε.

The above definition of spatial pattern is compatible with mathematical properties of neural networks *(embedding fields)* derived from psychological postulates about classical conditioning (Grossberg, 1969a; Grossberg, 1971). In their simplest form, these networks are defined as follows.

$$\dot{x}_i(t) \quad \alpha_i\, x_i(t) + \sum_{k=1}^{N} [x_k(t \quad \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki}\, z_{ki}(t) \\ - \sum_{k=1}^{N} [x_k(t \quad \sigma_{ki}) \quad \Omega_{ki}]^+ \gamma_{ki} + C_i(t) \tag{2}$$

and

$$\dot{z}_{jk}(t) = -\delta_{jk}\, z_{jk}(t) + \varepsilon_{jk} [x_j(t - \tau_{jk}) - \Gamma_{jk}]^+ x_k(t), \tag{3}$$

where $i, j, k = 1, 2, \ldots, N$ and $[\xi]^+ = \max(\xi, 0)$ for any real number ξ. $x_i(t)$ denotes the stimulus trace (or average membrane potential) at time t of the cell body (or cell body cluster) v_i, and $z_{jk}(t)$ denotes the memory trace (or associational strength, or excitatory transmitter production activity) at time t of the synaptic knob (or knobs) N_{jk} found at the end of the axon(s) e_{jk} from v_j to v_k. The term $[x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki}$ in

(2) is proportional to the spiking frequency released into e_{ki} in the time interval $[t-\tau_{ki}, t-\tau_{ki}+dt]$. Γ_{ki} is the spiking threshold, β_{ki} is proportional to the anatomical connection strength from v_k to N_{ki}, and τ_{ki} is the time required for spikes to travel from v_k to N_{ki}. The term $\sum_{k=1}^{N} [x_k(t-\tau_{ki})-\Gamma_{ki}]^+ \beta_{ki} z_{ki}(t)$ in (2) is the total excitatory input from other cells to v_i at time t. At an excitatory synapse ($\beta_{ki}>0$), spiking frequency couples multiplicatively to transmitter to release transmitter that perturbs $x_i(t)$, and all such signals combine additively at v_i. The term

$$\sum_{k=1}^{N} [x_k(t-\sigma_{ki})-\Omega_{ki}]^+ \gamma_{ki}$$

is the total inhibitory input from other cells to v_i at time t. The term $C_i(t)$ is the experimental input (or stimulus) to v_i at time t. See Fig. 1.

In (3), the memory trace cross-correlates the presynaptic spiking frequency which reaches N_{jk} from v_j at time t with the value of average potential at v_k at this time.

To illustrate how the spatial pattern concept relates to learning in (2) and (3), consider the simplest embedding field that can learn by Pavlovian conditioning; namely, an *outstar* (Grossberg, 1969b, 1970b). Let one CS-activated cell v_{n+1} send equal signals to its synaptic knobs $N_{n+1,i}$, which about the UCS-activated cells $V=\{v_i: i=1, 2, \ldots, n\}$, see Fig. 2. v_1 can learn and perform at V a spatial pattern; that is, a UCS input to V of the form $C_i(t)=\tilde{\theta}_i C(t)$. The synaptic knobs $N_{n+1,i}$ encode the pattern, or "relative figure-to-ground", $\tilde{\theta}=(\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \theta_n)$ in their transmitters $z_{n+1,i}$ at a rate which depends on $C(t)$, among other factors. The total amount $\sum_{k=1}^{n} z_{n+1,k}$ of transmitter accumulation is not constant, however. It can be potentiated either by presynaptic spikes from v_{n+1} to the knobs $N_{n+1,i}$, or by a combination of pre- and postsynaptic activity. It can also spontaneously decay. These fluctuations can occur in the absence of learning.

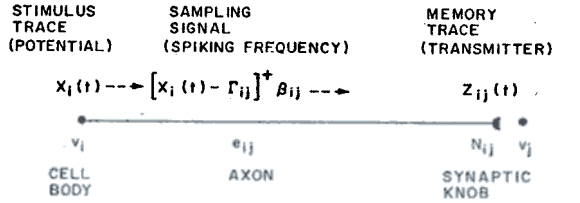

Fig. 1. Network interpretation

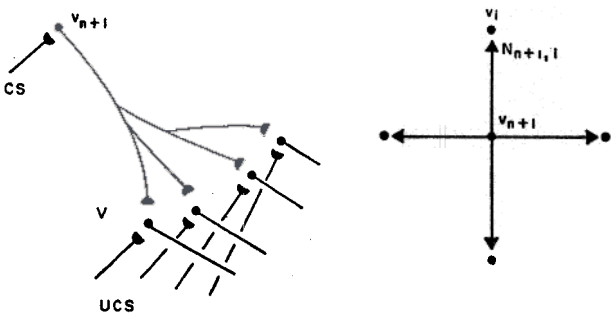

Fig. 2. An outstar

Learning alters the relative sizes of the $z_{n+1,i}$, which are attracted to the pattern weights $\tilde{\theta}_i$.

The pattern $\tilde{\theta}$ at V is learned only at times when the synaptic knobs $N_{n+1,i}$ receive CS-activated spikes from v_{n+1}. This is the property of "stimulus sampling". The relative transmitters $Z_{n+1,i}=z_{n+1,i}\left(\sum_{k=1}^{n} z_{n+1,k}\right)^{-1}$ are the "stimulus sampling probabilities" of an outstar (Grossberg, 1970b). Whenever v_{n+1} samples V, its synaptic knobs begin to learn the spatial pattern playing on V at this time. If a sequence of spatial patterns plays on V while v_{n+1} is sampling, then v_{n+1} learns a pattern which is a weighted average of all the patterns, rather than any single pattern in the sequence. Any nonnegative continuous input to V can be written as such a sequence of patterns.

During recall trials, an input to v_{n+1} reproduces at V the pattern which the outstar has learned. The total output to V can, however, vary wildly through time. For example, fluctuations in spiking frequency from v_{n+1} to the knobs $N_{n+1,i}$ can change the total output either directly, by altering the term

$$[x_{n+1}(t-\tau_{n+1})-\Gamma_{n+1}]^+$$

in the v_{n+1}-to-v_i signals

$$\beta[x_{n+1}(t-\tau_{n+1})-\Gamma_{n+1}]^+ z_{n+1,i}(t),$$

or indirectly, by changing the total transmitter level $\sum_{k=1}^{n} z_{n+1,k}(t)$. Such changes in total output do not destroy the encoded memory of pattern weights or the relative output to v_i (Grossberg, 1969b, 1970b). Nonetheless, v_i cannot discern what its pattern weight is if the total output fluctuates. It is therefore natural to seek anatomies, fed by outstar signals, which can discriminate the pattern that the outstar has learned in spite of fluctuations in total outstar output. This is the problem summarized by Eq. (1).

In the study of ritualistic pattern discrimination, no learning occurs. Hence $\delta_{jk}=\varepsilon_{jk}=0$ in (3) and $z_{ki}(t)$ is set equal to 1 in (2) for notational convenience. Two stages of inhibitory processing intermediate between V and a discriminative output cell R are needed to achieve the criterion in (1) (Grossberg, 1970a, Section 8):

I. Pattern Normalization and Low-Band Filtering

Pattern normalization guarantees that as the pattern intensity $C(t)$ becomes arbitrarily large, the potential $x_i(t)$ in the i-th filtering channel remains bounded without losing the pattern weight due to saturation at the maximal potential. In other words, the first inhibitory layer limits the range of dynamical response to inputs without saturating pattern weights. For example, normalization occurs if $x_i(t)=\tilde{\theta}_i \Gamma(t)$ where $0<\Gamma(t)\leqq\Gamma<\infty$.

Given a mechanism of pattern normalization, a suitably chosen positive spiking threshold can prevent an output from the i-th channel unless $\tilde{\theta}_i>\theta_i-\varepsilon$. For example, if $x_i(t)=\tilde{\theta}_i\Gamma(t)$, choose the spiking threshold of the i-th channel to equal

$$\Gamma_i=\Gamma(\theta_i-\varepsilon). \qquad (4)$$

Then $x_i(t) \leqq \Gamma_i$ unless $\tilde{\theta}_i > \theta_i - \varepsilon$. One half of the inequalities (1) is hereby achieved. By (4), given any pattern normalization device, it is necessary that

$$\max \sum_{k=1}^{n} x_k > \sum_{k=1}^{n} \Gamma_k,$$

in order that all low-band filtered channels be able to fire simultaneously.

II. High-Band Filtering

A second processing step is needed to guarantee the inequalities $\tilde{\theta}_i < \theta_i + \varepsilon$ in (1). This step inhibits the i-th low-band filtered signal when it exceeds a size corresponding to an input whose pattern weight $\tilde{\theta}_i$ exceeds $\theta_i + \varepsilon$. The doubly inhibited output from the i-th channel is therefore large only if (1) holds.

All n of these doubly inhibited signals summate at R. R's threshold is chosen so high that R can emit a signal only if it receives large signals almost simultaneously from all n channels; that is, only if (1) holds for all $i = 1, 2, \ldots, n$.

The pattern normalization and high-band filtering steps do not determine which pattern will pass through the filter. The choice of spiking threshold as in (4) does this. A U cell filter differs from an R cell filter only because active P cells can change the "threshold pattern" which must be overcome to fire U, and can learn arbitrary threshold patterns. The network that accomplishes this does not, however, use variable spiking thresholds. Rather it uses a physiologically more plausible mixture of excitatory and inhibitory signals. This mixing process will be motivated by a review of some concrete ways to realize steps (I) and (II) of the ritualistic filtering process.

Pattern normalization and low-band filtering can be achieved in at least two ways.

(i) Subtractive Nonspecific Nonrecurrent Interneuron (see Fig. 3). Let the net signal in the i-th channel after operation of the inhibitory interneuron be

$$J_i(t) = [C_i(t) - \Gamma_i]^+ - \zeta \left[\sum_{k=1}^{n} C_k(t) - \Gamma \right]^+, \qquad (5)$$

with $\Gamma > \sum_{k=1}^{n} \Gamma_k$ and $\zeta \geqq 1$. Since $\Gamma > \sum_{k=1}^{n} \Gamma_k$, Γ_i can be written in the form $\Gamma_i = \Gamma(\theta_i - \varepsilon)$, as in (4). Also write $C_i = \tilde{\theta}_i C$ where $C = \sum_{k=1}^{n} C_k$. By (5), $J_i(t) \leqq 0$ for all $t \geqq 0$ unless $\tilde{\theta}_i > \theta_i - \varepsilon$. Thus no output occurs unless $\tilde{\theta}_i > \theta_i - \varepsilon$. This is because $\tilde{\theta}_i \leqq \theta_i - \varepsilon$ implies that C exceeds Γ before C_i exceeds Γ_i, and that $dJ_i(t)/dt \leqq 0$ whenever $C > \Gamma$. In the case $\tilde{\theta}_i > \theta_i - \varepsilon$, Grossberg (1970a, Section 10) studies the function J_i as C oscillates, and extends the results to cases in which time lags and exponential averaging rates in the excitatory and inhibitory terms of (5) differ. For example, unless $\tilde{\theta}_i = \zeta = 1$ in (5), $J_i(t)$ responds to growth of C from zero to large values with a transient "on" response, and to decay of C from large values to zero with a transient "off" response. If $\tilde{\theta}_i = \zeta = 1$, then $J_i(t)$ saturates at a constant value when C takes on large values.

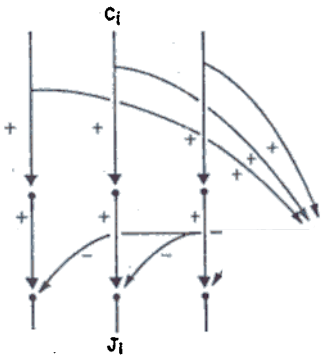

Fig. 3. Subtractive nonspecific nonrecurrent interneuron

(ii) Multiplicative On-Off Field. The subtractive form of the interaction in (5) depends on inputs and thresholds which are small compared to the saturation level of cell body potentials. Given large inputs that drive the potentials towards saturation, a different inhibitory anatomy can preserve input pattern weights in the response of cell body potentials. Consider Fig. 4.

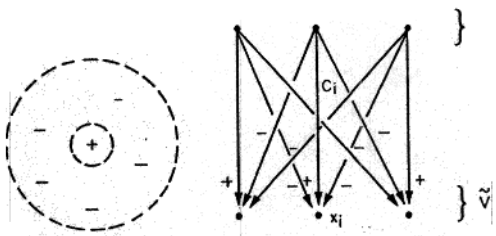

Fig. 4. Shunting on-center off-surround field

The i-th cell in V excites the i-th cell in $\tilde{V}$ and equally inhibits all other cells in $\tilde{V}$ by interacting multiplicatively with ("shunting") cell body potential. The potential x_i of the i-th $\tilde{V}$ cell thus obeys the equation

$$\dot{x}_i = (M - x_i)\, C_i - \alpha x_i - x_i \sum_{k \neq i} C_k. \qquad (6)$$

This is a passive membrane equation with equilibrium scaled to zero for convenience, and inputs C_i representing depolarizing or hyperpolarizing conductance changes (Sperling, 1970; Sperling and Sondhi, 1968). Grossberg (1970a, Section 14A) proves that each x_i is bounded and is attracted to $\tilde{\theta}_i$. If, in fact, C varies slowly relative to x_i, then

$$x_i \cong \tilde{\theta}_i \frac{MC}{\alpha + C}, \qquad (7)$$

so that x_i is monotone increasing as a function of C. Contrast mechanism (i). Eq. (7) accomplishes pattern normalization without low-band filtering. The next positive spiking threshold in the i-th channel accomplishes low-band filtering.

High-band filtering can be achieved by a specific inhibitory interneuron, as in Fig. 5a (Grossberg, 1970a, Section 11). For example, let the input in the i-th low-band filtered channel have the form

$$K_i(t) = [O_i(t) - \Omega_i]^+ - \zeta [O_i(t) - \Gamma_i]^+, \qquad (8)$$

where $O_i(t)$ is the i-th output after low-band filtering, and the parameters satisfy $\Omega_i < \Gamma_i$ and $\zeta > 1$. K_i is monotone increasing as a function of O_i when $\Omega_i < O_i < \Gamma_i$, but decreases in O_i as soon as O_i exceeds Γ_i.

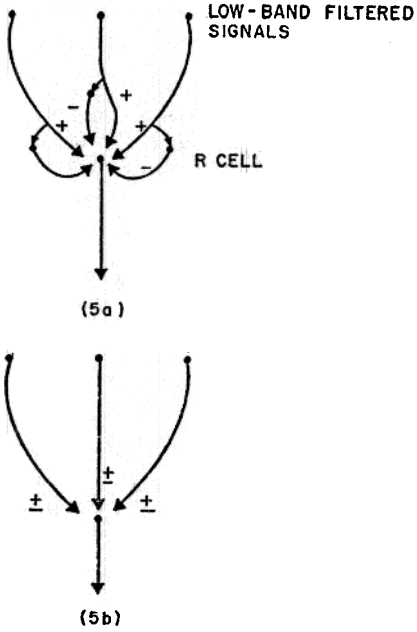

Fig. 5. Some high-band filters

In the present case, letting $\Gamma_i = \Omega_i + \eta$ suffices, where $O_i = \Gamma_i$ before $\tilde{\theta}_i = \theta_i + \varepsilon$. Other mechanisms accomplish the same functional goal; for example, blockade of postsynaptic potential response if presynaptic spiking frequency exceeds the level corresponding to $\theta_i = \theta_i + \varepsilon$, or switch-over as spiking frequency reaches the $\tilde{\theta}_i = \theta_i + \varepsilon$ level from net excitation to net inhibition of postsynaptic response (Bennett, 1971; Blackenship, Wachtel, and Kandel, 1971; Wachtel and Kandel, 1971), as in Fig. 5b.

3. Retinal Analog of *R* Cells

Cells of type R can exhibit hue constancy or brightness constancy because of the following considerations. Suppose, for example, that an R cell receives peripheral inputs from three classes of cones, v_1, v_2, and v_3, each class containing a pigment with different but overlapping absorption spectra (Cornsweet, 1970, p. 200). Suppose momentarily that a given hue is characterized by a fixed relative excitation of each cone class; that is, by a fixed spatial pattern $\theta = (\theta_1, \theta_2, \theta_3)$. An increase in background illumination by white light does not change the pattern of inputs to the three cone classes (Cornsweet, 1970, pp. 243f.). Thus the cell R responds to the hue characterized by θ in spite of fluctuations in the background intensity of white light. In this example, output from a given cone class is inhibited by output from one or more other cone classes at the pattern normalization step. This fact is compatible with the existence of spectrally opponent cells at the ganglion cell layer of the macaque retina (Abramov, 1968).

A similar minimal construction exists for cells R exhibiting brightness constancy. It suffices to change the source cells in the above example. Thus let the pattern $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ describe the relative intensities of illumination at center cells v_1 and surround cells $v_2, \ldots, v_n$, such that the source cells are either all rods, or are all members of a single cone class. The R cells responding to these source cells will yield a brightness constancy, respectively, to either white light or to a fixed wavelength to which the cone class responds. The brightness constancy dissolves as soon as the surround is no longer excited. The existence of inhibitory interactions between cells fed by cones and between cells fed by rods has been reported in the goldfish retina. Stell (1967) showed that the external horizontal layer is fed by cones and the intermediate horizontal layer is fed by rods. Kaneko (1970) recorded luminosity- and chromaticity-type S-potentials from the external and internal horizontal layers of the goldfish retina. Naka and Rushton (1966) reported that S-potentials from color units in the fish retina are influenced by at least one stage of shunting inhibition; cf. Eq. (7). A shunt acts also at the level of the cones in the turtle retina (Baylor and Fuortes, 1970). In all these cases, the absolute range of dynamical response in retinal cells can be small, because relative responses at different spatial loci are the crucial quantities.

More sophisticated forms of color constancy are known, and some can be approached by modifying cell connections. The above simple form of color constancy is achieved by pattern normalization among cone classes, followed by low- and high-band filtering. Land (1964), in his retinex theory, notes data suggesting that each cone class establishes its own "lightness" scale before the cones interact. See also Land and McCann (1971). This idea is related to the following modified anatomy, which is included more as a directive for further studies than as a purported explanation of all lightness phenomena. First let each cone class interact within itself to yield pattern normalization. Choose a multiplicative on-off field as in Section 2(ii) for definiteness. Let the off-surround be of much broader spatial extent than the on-center. Each on-off field can, for example, feed the bipolar cell layer of the network. The normalization step reduces the variability of response due to variations in background illumination that excite each cone class, as Eq. (7) illustrates. It can be followed by low- and high-band filtering to yield R cells exhibiting brightness constancy within each cone class. Even if it isn't, however, some degree of constancy is assured. The intracone class normalization step (perhaps followed by low- and high-band filtering) provides a "retinex" for our network. Each cone class is given its own retinex, as Land suggested. Then let the different retinexes interact "classically"—namely, as in the original color constancy construction—to yield spectrally-opponent cell types in the ganglion cell layer. This construction, were it to exist in vivo, could employ bipolar cells which respond to inputs derived from only one cone class in an on-center off-surround configuration. A more local color theory, without a "lightness" scale, could require less bipolar cell selectivity; namely, bipolars could respond to more than one cone class in an on-center off-surround configuration.

The mudpuppy retina exhibits analogs with the minimal mechanisms of Section 2. Since our constructions say little about the *global* properties of neuronal fields of discriminative cells, no more than qualitative similarities will be mentioned. An analog exists between mechanisms of pattern normalization followed by low-band filtering and processing in the horizontal cell layer of the mudpuppy retina (Dowling

and Werblin, 1969). Consider the multiplicative on-off field in Section 2 (ii) for definiteness. An on-center and off-surround response can be measured intracellularly from bipolar cells, at which mudpuppy receptor and horizontal cell inputs converge (Werblin and Dowling, 1969). The horizontal cells hyperpolarize bipolar cells and integrate receptor signals from a broad retinal area, as is also true of a pattern normalizer. Werblin and Dowling (p. 347–348) stress the importance of the ratio of center-to-surround illumination in determining the bipolar cell response; cf., Eq. (7). They note that horizontal cell inhibition counteracts the depolarizing effects of receptor input on bipolar cell response without hyperpolarizing the bipolar cell (p. 347), that the horizontal cell apparently operates in a nonrecurrent manner relative to its receptor input sources (p. 347), and that the bipolar potential can be maintained at graded sizes in response to graded inputs (p. 347–348). These facts also hold in Eq. (6). In principle, however, Eq. (6) can also yield hyperpolarization if a tonic input term is added. Bipolar responses resemble color-coded *C*-type *S*-potentials (p. 347). This is compatible with the use of pattern normalization and low-band filtering to prepare inputs to cells R whose responses exhibit color constancy.

High-band filtering mechanisms can be compared with amacrine cell responses in the mudpuppy retina. Werblin (1970, p. 348) notes that an amacrine cell response can be elicited from a limited number of bipolar cells, much as a specific mixed excitatory-inhibitory connection forms a high-band filter. Werblin and Dowling (1969, p. 351) strengthen this interpretation by noting that the amacrine cell is activated by the initial part of the transient in the bipolar cell response. They hypothesize that a return synapse from amacrine cell to bipolar cell is inhibitory, and that the amacrine cell inhibits the excitation which it receives from the bipolar cell. Such a mechanism could suffice to high-band filter the amacrine cell input. Given this horizontal cell and amacrine cell analogy of R-cell filters, ganglion cells, receiving amacrine-bipolar output, would be a natural class of cells in which R cells, if they exist, might be sought in the mudpuppy retina. Grossberg (1970a) shows that other types of discriminative cells than hue or brightness constancy detectors can be constructed using similar mechanisms; e.g., movement detectors (Werblin, 1970).

The above constancies can be weakened by changing the relative strengths of inhibitory signals. For example, replacing (6) by

$$\dot{x}_i = (M - x_i)\, C_i \gamma_{ii} - \alpha\, x_i - x_i \sum_{k \neq i} C_k \gamma_{ki}, \qquad (9)$$

where the γ_{ki} determine the strengths of signals (excitatory if $k = i$, inhibitory if $k \neq i$) from the k-th to the i-th cone class, yields an asymptotic response to the pattern $C_i(t) = \tilde{\theta}_i\, C(t)$ of the form

$$x_i \cong \frac{\tilde{\theta}_i \gamma_{ii} MC}{\alpha + \left(\sum_k \tilde{\theta}_k \gamma_{ki}\right) C}. \qquad (10)$$

Unless the sums $\sum_k \tilde{\theta}_k \gamma_{ki}$ are independent of i, the pattern recorded in the relative potentials $x_i \left(\sum_{k=1}^{n} x_k\right)^{-1}$ shifts as a function of total intensity C. This shift in pattern excites those R cells which are sensitive to the new pattern, and yields an analogous shift in perceived hue. Transformations of the form $A(C) = K_1\, C(K_2 + C)^{-1}$ have been used to explain properties of hue and brightness inconstancies as a function of background illumination (Cornsweet, 1970, p. 252). The choices of coefficients in these examples differ from that in (10), which illustrates only one possible cause of hue shifts. The problem of choosing coefficients in $A(C)$ is complicated by the appearance of shunting terms at retinal layers beyond the cone and horizontal layers. Creutzfeldt, Sakmann, Scheich, and Korn (1970) use such a transformation to describe ganglion cell responses in the cat retina. They also reference earlier efforts.

4. A Learnable Preset Mechanism: Subtractive Case

To motivate the construction of U cells, consider, as in Section 1, the problem of learning to lever press for food. Activity of P cells is associated with the lever press response. Activity of U cells releases consummatory motor activity. During training trials, P cells learn the patterns which unconditionally activate the U cells. During recall trials, P cells prevent the U cell filter from firing U unless the learned input patterns reoccur. Thus we must answer three questions: (A) How does P learn input patterns at U? (B) Before learning, how do the input patterns at U unconditionally drive U cell firing? (C) After learning, how does P activity bias the U cell filter to prevent all but the learned patterns from activating U? We seek the *minimal* anatomies that answer these questions. Only the U cell analog of Eq. (4) need be studied; all other properties of R-filters carry over to the U-filter case without change.

Consider question (A). Denote the collection of P cells generically by $V_1 = \{v_j : j = 1, 2, \ldots, m\}$. Let each P cell be the source cell of an outstar whose axon collaterals terminate at the cells $V_2 = \{v_i : i = m+1, \ldots, m+n\}$. Let the cells V_2 receive the input patterns which unconditionally activate U from cells

$$V_3 = \{v_{i+n} : i = m+1, \ldots, m+n\}.$$

The $V_1 \to V_2$ synaptic knobs can learn these input patterns by Pavlovian conditioning.

Consider question (C). Denote by

$$V_4 = \{v_{i+2n} : i = m+1, \ldots, m+n\}$$

the cells at which the input pattern is low-band filtered by the output from P cells. As in (5), the difference between the input pattern and the threshold pattern must be computed. Hence excitatory $v_{i+n} \to v_{i+2n}$ axons deliver the input pattern from V_3 to V_4. Since the threshold pattern is reproduced at V_2 by active V_1 cells, inhibitory $v_i \to v_{i+2n}$ axons carry the threshold pattern from V_2 to V_4. Note that the input pattern is multiply represented at V_2 and V_4, and that two excitatory input sources converge at V_2 cells.

Consider question (B). In the absence of P cell activity, input patterns can unconditionally activate U. The i-th input channel activates a net excitatory signal $v_{i+n} \to v_{i+2n}$ and a net inhibitory signal $v_{i+n} \to v_i \to v_{i+2n}$ from V_3 to V_4. Thus the absolute value of the excitatory signal exceeds the absolute value of the inhibitory signal. Moreover the spiking thresholds of

$v_i \to v_{i+2n}$ axons are set equal to zero to avoid distortion of threshold pattern weights.

Output from V_4 cells is high-band filtered on the way to the output cell $V_5 = v_{m+3n+1}$, which is of type U. The spiking threshold of $V_4 \to V_5$ axons are set equal to zero to avoid distortion of low-band filtered signals. V_5 can be the source of an outstar which (say) samples motor control patterns at cells V_6, see Fig. 6. In Fig. 6, synaptic knobs at which classical conditioning occur

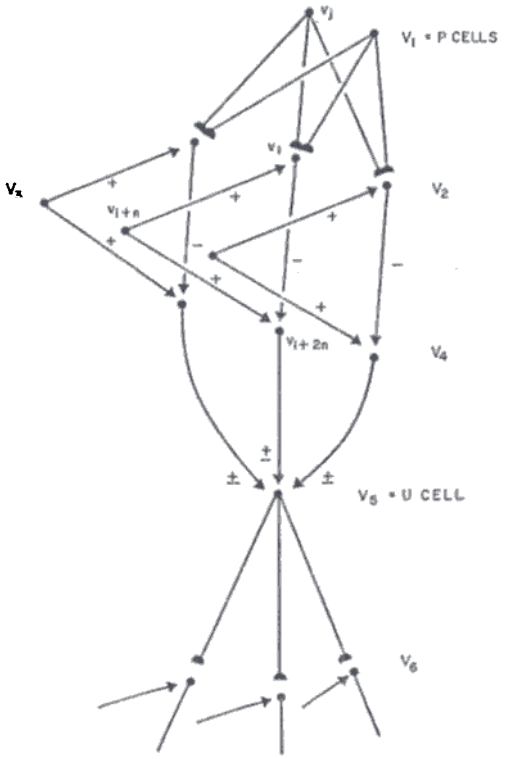

Fig. 6. A learnable preset mechanism: subtractive case

are denoted by semicircles. All other knobs are denoted by arrows. A brief summary of network dynamics follows.

V_3 can create unconditional outputs from V_5 in the absence of V_1 activity. To see this, let the pattern weight $\tilde{\theta}_i$ be emitted from v_{i+n} to v_i and v_{i+2n}. Let the strength of the excitatory $v_{i+n} \to v_{i+2n}$ signal be $\tilde{\theta}_i I$. Then the net strength of the inhibitory $v_{i+n} \to v_i \to v_{i+2n}$ signal can be written in the form $\tilde{\theta}_i \eta I$, where $0 < \eta < 1$. By additivity of inputs, the net signal to v_{i+2n} created by V_3 is $\tilde{\theta}_i(1-\eta)I$, which is positive. An output from v_{i+2n} occurs because the v_{i+2n} spiking threshold is zero. The constraint on $v_{i+n} \to v_i \to v_{i+2n}$ signal size does not constrain the size of the $v_{i+n} \to v_i$ signal, nor consequently the rate of pattern learning in $V_1 \to V_2$ synaptic knobs.

To see how the presetting mechanism works, suppose that $v_j \to V_2$ synaptic knobs are active while the pattern $\theta = (\theta_1, \ldots, \theta_n)$ is emitted by V_3. Then the j-th outstar learns this pattern by Pavlovian conditioning. On recall trials, let v_j be the only active cell in V_1. It reproduces the pattern θ at V_2. V_2 communicates this pattern as inhibitory signals to V_4. Let the net $v_j \to v_i \to v_{i+2n}$ inhibitory signal be $-\theta_i K$. Now let V_3 emit the test pattern $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_n)$. The $V_1 \to V_2 \to V_4$, $V_3 \to V_2 \to V_4$, and $V_2 \to V_4$ signals combine additively at V_4. The net signal to v_{i+2n} is

$$J_i = \tilde{\theta}_i(1-\eta)I - \theta_i K. \tag{11}$$

Since the spiking threshold of v_{i+2n} is zero, (11) implies that v_{i+2n} will fire only if

$$\tilde{\theta}_i > \theta_i K(1-n)^{-1} I^{-1}. \tag{12}$$

Eq. (12) can hold for all i only if $(1-\eta)I > K$; that is, only if the V_3 channel is stronger than the V_1 channel. This achieves low-band filtering by the conditioned pattern θ. High-band filtering of *any* low-band filtered input is then automatically achieved by $V_4 \to V_5$ signals. Hence, when V_1 presets the U-filter with pattern θ, the U cell only fires if the test pattern emitted by V_3 is θ, within an error of ε. Note that increasing the total $v_j \to V_2$ input increases K, and thus the minimal pattern weights that can fire V_4. By contrast, increasing the total V_3 output increases I, and thereby decreases the minimal pattern weights that can fire V_4. This "crispening" effect can thus be controlled by varying the arousal, or adaptation, levels of V_1 and V_3, respectively.

Suppose that more than one cell in V_1 fires to V_2 at a given time. Let the $v_j \to v_i$ synaptic knob encode the pattern weight θ_{ji}. Then the net signal from v_j to v_{i+2n} has the form $-\theta_{ji} K_j$, where K_j depends on the spiking frequency in $v_j \to V_2$ axons. The total input to v_{i+2n} in response to all active V_1 cells and to V_3 output is

$$J_i = \tilde{\theta}_i(1-\eta) I - \sum_{j=1}^{m} \theta_{ji} K_j$$

Thus v_{i+2n} fires only if

$$\tilde{\theta}_i > \sum_{j=1}^{m} \theta_{ji} K_j (1-\eta)^{-1} I^{-1}.$$

The right-hand side of (13) is a weighted average of all pattern weigths in knobs abutting v_i. The weights are determined by the intensity of signals from each cell v_j. The above conclusions can be mathematically extended to consider the influence of time lags and exponential averaging rates using the analysis of the analogous ritualistic case in Grossberg (1970a).

5. Cerebellar Analog of U Cells

By redrawing the network in Fig. 6, a striking analogy with aspects of cerebellar anatomy emerges (Bell and Dow, 1967; Eccles, Ito, and Szentagothai, 1967). This analogy includes the following identifications: V_1 = mossy fiber glomeruli; V_2 = Purkinje cells; V_3 = inferior olive cells; V_4 = cerebellar nuclear cells; $V_1 \to V_2$ axons = excitatory parallel fibers; $V_2 \to V_4$ axons = inhibitory Purkinje cell axons; $V_3 \to V_2$ axons = excitatory climbing fibers; $V_3 \to V_4$ axons = excitatory collaterals of climbing fibers.

This analogy becomes more evident when Fig. 6 is redrawn as follows. The $V_1 \to V_2$ outstar axons can be drawn as in Fig. 7a, rather than as in Fig. 7b. The mossy fiber ends in a glomerulus (rosette) that feeds the dendrites of a band of contiguous granule cells. The granule cell axons are parallel fibers, which

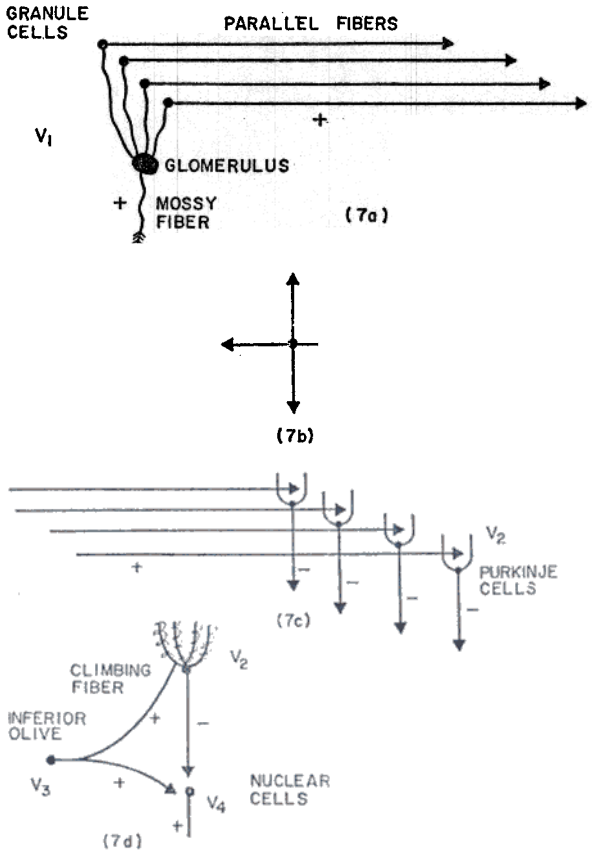

Fig. 7. Cerebellar analog of preset mechanism

activate Purkinje cell dendrites. The abstract outstar anatomy of Fig. 7b is functionally identical with Fig. 7a if the cluster of parallel fibers in Fig. 7a is driven in phase by its glomerulus. In Fig. 7c, the overlapping dendritic trees of Purkinje cells receive input from contiguous bands of parallel fibers. In Fig. 7d, each Purkinje cell receives a climbing fiber input from V_3. V_3 also sends an input to a cluster of nuclear cells. Also in Fig. 7d, Purkinje cells send inhibitory signals to those nuclear cells which mutually share the same V_3 sources. Suppose that this analogy with cerebellar anatomy also extends to cerebellar physiology, beyond a mere labelling of axons as excitatory and inhibitory. Then a possible functional reason for the convergence of excitatory mossy fiber and climbing fiber inputs on inhibitory Purkinje cells, and for diverging excitatory signals to Purkinje cells and nuclear cells, is the following: the mossy fiber input biases the nuclear cells to fire in response to prescribed patterns in the climbing fiber channel. If this interpretation is correct, then both mossy fiber and climbing fiber inputs can separately fire Purkinje cells, and simultaneous inputs from different channels can summate. Freeman (1970) reports analogous data. This interpretation is compatible with the suggestion (Grossberg, 1969c; Miller and Oscarsson, 1970) that cross-correlation of mossy fiber and climbing fiber inputs occurs at the Purkinje cells. Moreover, (12) requires that the climbing fiber channel exert a more profound influence on Purkinje firing than the mossy fiber channel. On the other hand, Bloedel and Roberts (1971) emphasize the possible functional importance of the refractory period in Purkinje cell spiking that follows a climbing fiber input. Possibly this refractory period helps to break up the temporal processing of cerebellar inputs into sequences of spatial patterns (Grossberg, 1969b, Section 12). Quantization of temporal processing seems to occur in some sensory systems. For example, exploratory sniffing and tactile input from facial vibrissae seem to be synchronized with the theta rhythm and heart beat in rats (Komisaruk, 1970). After the rat's head is fixed in position, the vibrissae twitch forward and a brief inhalation sniff occurs. The vibrissae are then retracted and the head moves to a nearby fixation point. Then the cycle of coordinated vibrissae motion and inhalation repeats itself. This mechanism seems to break up the sensory input into sequences of spatial patterns. Different sensory channels admit their next spatial pattern in phase with each other. Thus patterns in different modalities can be filtered simultaneously and correlated with each other. This interpretation of Purkinje refractoriness is at best speculative at the present time. Nonetheless, the anatomical analogy in Fig. 7 clearly shows that the abstract minimal anatomy in Fig. 6 is constructed using plausible anatomical principles.

6. A Learnable Preset Mechanism: Multiplicative Case

A mechanism for low-band filtering and pattern normalization using shunting inhibition, such as that in Section 2 (ii), will now be sketched. Consider Fig. 8. V_1 consists of outstar sources. V_2 receives outstar signals. V_3 sends test inputs to V_2 and V_4. V_5 (V_6) receives signals from V_2 (V_4) that have been preprocessed by a multiplicative on-off field. V_6 sends excitatory signals to V_7, whereas V_5 sends inhibitory signals to V_7. The low-band comparison between the input patterns controlled by

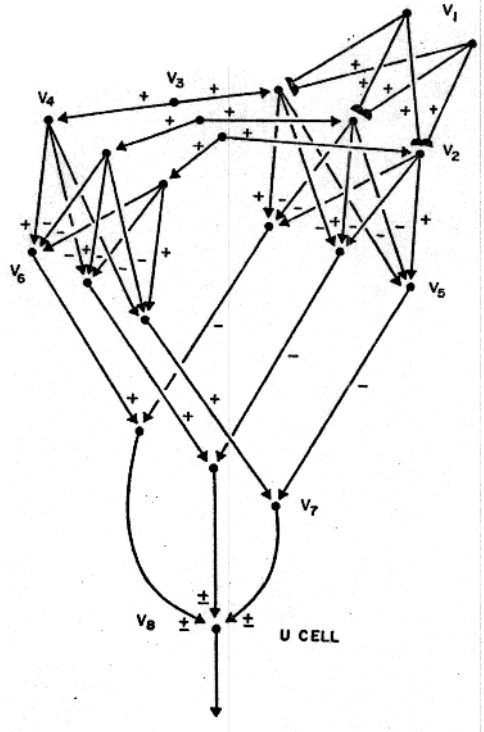

Fig. 8. A learnable preset mechanism: shunting case

V_3 and the threshold patterns controlled by V_1 thus occurs at V_7. Signals from V_7 to V_8 are high-band filtered. V_8 is a cell of type U. The inhibitory signal $V_3 \rightarrow V_2 \rightarrow V_5 \rightarrow V_7$ must be weaker than the excitatory signal $V_3 \rightarrow V_4 \rightarrow V_6 \rightarrow V_7$. This can be accomplished in several ways; for example, let the saturation level of V_5 potentials be smaller than the saturation level of V_6 potentials. The spiking thresholds of $V_2 \rightarrow V_5$, $V_4 \rightarrow V_6$, $V_5 \rightarrow V_7$, $V_6 \rightarrow V_7$, and $V_7 \rightarrow V_8$ axons are set equal to zero to avoid biasing filtered pattern weights. This completes the construction. Note that in all low-band filters of this paper, the statistical dispersion of signals in the excitatory channels is the same as the statistical dispersion of signals in the parallel inhibitory channels. This yields decision rules for cellular firing that retune themselves as the statistics of the input change; cf., Sperling (1970).

More elaborate variations on these themes readily suggest themselves. Given the existence of such striking retinal and cerebellar analogs to the minimal anatomies, it is to be hoped that some of these variations will have more quantitative neural analogs, whose functional meaning will be evident from an inspection of their psychologically derived counterparts. At the very least, the minimal anatomies show how different anatomies can be, even if they carry out similar discrimination tasks, one ritualistically and one with a learnable, or unlearned, preset mechanism.

References

Abramov, I.: Further analysis of the responses of LGN cells. J. Opt. Soc. Amer. 58, 574 (1968).

Baylor, D.A., Fuortes, M.G.F.: Electrical responses of single cones in the retina of the turtle. J. Physiol. (Lond.) 207, 77 (1970).

Bell, C.C., Dow, R.S.: Cerebellar circuitry. In: Neurosciences Research Symposium Summaries, vol. 2 (Schmitt, F.O., Melnechuk, T., Quarton, G.C., and Adelman, G., eds.). Cambridge, Mass.: M.I.T. Press 1967.

Bennett, M.V.L.: Analysis of parallel excitatory and inhibitory synaptic channels. J. Neurophysiol. 34, 69 (1971).

Blackenship, J.E., Wachtel, H., Kandel, E.R.: Ionic mechamisms of excitatory, inhibitory, and dual synaptic actions mediated by an identified interneuron in abdominal ganglion of Aplysia. J. Neurophysiol. 34, 76 (1971).

Bloedel, J.R., Roberts, W.J.: Action of climbing fibers in cerebellar cortex of the cat. J. Neurophysiol. 34, 32 (1971).

Cornsweet, T.N.: Visual perception. New York: Academic Press 1970.

Creutzfeldt, O.D., Sakmann, B., Scheich, H., Korn, A.: Sensitivity distribution and spatial summation within receptive-field center of retinal on-center ganglion cells and transfer function of the retina. J. Neurophysiol. 33, 654 (1970).

Dowling, J.E., Werblin, F.S.: Organization of retina of the mudpuppy Necturus maculosus. I. Synaptic structure. J. Neurophysiol. 32, 315 (1969).

Eccles, J.C., Ito, M., Szentagothai, J.: The cerebellum as a neuronal machine. Berlin-Heidelberg-New York: Springer 1967.

Freeman, J.A.: Responses of cat cerebellar Purkinje cells to convergent inputs from cerebral cortex and peripheral sensory systems. J. Neurophysiol. 33, 697 (1970).

Grossberg, S.: Embedding fields: A theory of learning with physiological implications. J. Mathematical Psychology 6, 209 (1969a).

— Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, I. J. Math. Mech. 19, 53 (1969b).

— On learning of spatiotemporal patterns by networks with ordered sensory and motor components, I. Excitatory components of the cerebellum. Studies in Applied Mathematics 48, 105 (1969c)

— Neural pattern discrimination. J. theor. Biol. 27, 291 (1970a).

— Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II. Studies in Applied Mathematics 49, 135 (1970b).

— Pavlovian pattern learning in nonlinear neural networks. Proc. nat. Acad. Sci. (Wash.) 68, 828 (1971).

— A neural theory of punishment and avoidance. Mathematical Biosciences 1972, submitted for publication.

Kaneko, A.: Physiological and morphological identification of horizontal, bipolar and amacrine cells in goldfish retina. J. Physiol. (Lond.) 207, 623 (1970).

Kimble, G.A.: Conditioning and learning. New York: Appleton-Century-Crofts 1961.

Komisaruk, B.R.: Synchrony between limbic system theta activity and rhythmical behavior in rats. J. comp. physiol. Psychol. 70, 482 (1970).

Land, E.H.: The retinex. Amer. Scientist 52, 247 (1969).

— McCann, J.J.: Lightness theory. J. Opt. Soc. Amer. 61, 1 (1971).

Miller, S., Oscarsson, O.: Termination and functional organization of spino-olivocerebellar paths. In: The cerebellum in health and disease (Fields, W.S., and Willis, W.D., eds.). St. Louis: W.H. Green 1970.

Naka, K.I., Rushton, W.A.H.: *S*-potentials from color units in the retina of fish (*Cyprinidae*). J. Physiol. (Lond.) 185, 536 (1966).

Sperling, G.: Model of visual adaptation and contrast detection. Perception and Psychophysics 8, 143 (1970).

— Sondhi, M.M.: Model for visual luminance discrimination and flicker detection. J. Opt. Soc. Amer. 58, 1133 (1968).

Stell, W.K.: The structure and relationship of horizontal cells and photo-receptor-bipolar synaptic complexes in goldfish retina. Amer. J. Anat. 121, 401 (1967).

Wachtel, H., Kandel, E.R.: Conversion of synaptic excitation to inhibition at a dual chemical synapse. J. Neurophysiol. 34, 56 (1971).

Wagner, A.R.: Frustrative nonreward: A variety of punishment. In: Punishment and aversive behavior (Campbell, B.A., and Church, R.M., eds.). New York: Appleton-Century-Crofts 1969.

Werblin, F.S.: Response of retinal cells to moving spots: Intracellular recording in Necturus maculosus. J. Neurophysiol. 33, 342 (1970).

— Dowling, J.E.: Organization of the retina of the mudpuppy, Necturus maculosus. II. Intracellular recording. J. Neurophysiol. 32, 339 (1969).

Prof. S. Grossberg
Dept. of Mathematics
Mass. Inst. of Technology
Cambridge, Mass. 02139, U.S.A.