Appendix B

This section indicates how a single cell population in $F^{(2)}$ can learn a spectial pattern of activity across $F^{(1)}$. Analogous arguments then show how many simultaneously active cell populations across $F^{(2)}$ can learn a spatial pattern across $F^{(1)}$, albeit not necessarily the same spatial pattern that would have excited a single cell in $F^{(2)}$ by interfield signaling from $F^{(1)}$ to $F^{(2)}$.

Associative Learning

Our laws for associative learning appeared in a monograph by Grossberg (1964), and were mathematically analyzed in a series of articles, leading to a universal theorem of associative learning in Grossberg (1969a, 1971b, 1972d). The universal theorem proves that if these associative learning laws were invented at a prescribed time during the evolutionary process, then they could be used to guarantee unbiased associative learning in essentially any later evolutionary specialization. That is, the laws are capable of learning arbitrary spatial patterns in arbitrarily many, simultaneously active sampling channels that are activated by arbitrary continuous data preprocessing in an essentially arbitrary anatomy. Learning of arbitrary space-time patterns is also guaranteed given modest requirements on the temporal regularity of stimulus sampling. (See Grossberg, 1974, for a review.) Herein I summarize the fact that the unit of LTM is a *spatial pattern*. This is done by considering the *minimal* anatomy that is capable of classical conditioning.

STM and LTM Laws That Factor Pattern from Activity

Let presentation of a CS create an input $I_0(t)$ that activates the cell population v_0. Let the UCS create an input pattern $(I_1(t), I_2(t), \ldots, I_n(t))$ that activates the cell populations $v_1, v_2, \ldots, v_n$, whose outputs elicit the UCR. Let the STM trace of v_i be $x_i(t)$, $j = 0, 1, \ldots, n$, and let the LTM trace of the axon pathway e_{0i} from v_0 to v_i be $z_{0i}(t)$, $i = 1, 2, \ldots, n$ (Figure A1). Suppose that the STM and LTM traces obey the laws

$$\frac{d}{dt}x_0 = -A_0x_0 + I_0(t) \qquad (A3)$$

$$\frac{d}{dt}x_i = -Ax_i + Bz_{0i} + I_i(t), \qquad (A4)$$

and

$$\frac{d}{dt}z_{0i} = -Cz_{0i} + Dx_i, \qquad (A5)$$

$i = 1, 2, \ldots, n$. The terms A_0 and A are STM decay rates. The term C is the LTM decay rate. The terms B and D are signals from v_0 along all the pathways e_{0i}, $i = 1, 2, \ldots, n$; for example, $B(t) = f(x_0(t - \tau))$, where $f(w)$ is a sigmoid function of w. The LTM trace z_{0i} is computed at the interface of the synaptic knob S_{0i} (at the end of e_{0i}) and the postsynaptic cell v_i—that is, at the synaptic knob and/or postsynaptic membrane—where it can gate the signals B on their way to v_i, as in term Bz_{0i} of Equation A4, and simultaneously time-average (term $-Cz_{0i}$) the product of signals D and postsynaptic STM trace x_i (term Dx_i), in A5. In particular, A2 is a special case of A5.

A *spatial pattern* is a UCS whose *relative* activities remain fixed, even though their absolute activities can fluctuate through time,

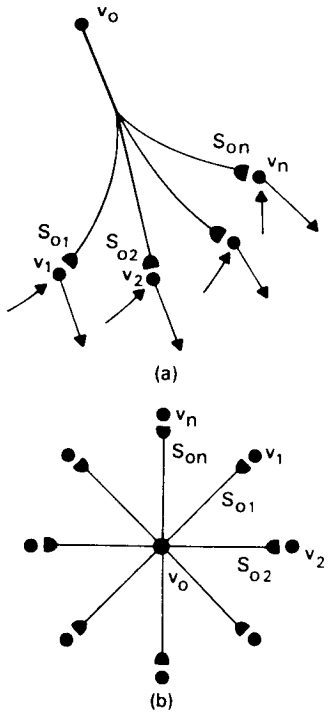

Figure A1. In (a) the conditioned stimulus (CS)-activated population v_0 samples the unconditioned stimulus (UCS)-activated populations $v_1, v_2, \ldots, v_n$; in (b) the *outstar* is the minimal network capable of classical conditioning.

namely, $I_i(t) = \theta_i I(t)$, $i = 1, 2, \ldots, n$, where θ_i is the fixed relative activity and $I(t)$ is the total UCS activity. The convention

$$\sum_{i=1}^{n} \theta_i = 1$$

guarantees the normalization

$$I(t) = \sum_{i=1}^{n} I_i(t).$$

The relative values $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ are like generalized "reflectances" that carry the information in the UCS pattern, whereas $I(t)$ provides the UCS activity that drives system changes in response to θ. It is shown below how this system, which I call an *outstar* (Figure A1), can factorize pattern information θ from information about total activity $I(t)$. This property has many important implications. For example, θ is a probability distribution, since each $\theta_i \geq 0$ and

$$\sum_{k=1}^{n} \theta_k = 1.$$

The system learns probabilities despite the fact that it can generate deterministic behavior. There exists a type of "wave-particle" dualism in these systems that helps to explain the partial successes of statistical learning models, and provides an interesting vantage point from which to think about the wave-particle dualism of quantum theory. Also, since there is no evolutionary advantage in perceptually discriminating data that cannot, in principle, be learned, we can expect the neural perceptual apparatus also to process spatial patterns. The brightness and hue constancies of vision illustrate this fact. These observations clarified how perceptual and learning mechanisms are matched to each other, and suggested study of the minimal neural networks that are capable of discriminating a spatial pattern θ; that is, reflectances. Some of these networks were constructed in Grossberg (1970, 1972a) and, not surprisingly, have an anatomy that is remarkably retinal.

System A3–A5 *factorizes* θ and $I(t)$ in the following sense. Equation A3 can be explicitly solved for $x_0(t)$ by integration, and the result used to solve for $B(t)$ and $D(t)$ as functions of time t. Then A4 and A5 can be rewritten in terms of the relative STM traces

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

and relative LTM traces

$$Z_i = z_{0i}\left(\sum_{k=1}^{n} z_{0k}\right)^{-1}$$

as follows:

$$\frac{d}{dt}X_i = E(Z_i - X_i) + F(\theta_i - X_i) \quad \text{(A6)}$$

and

$$\frac{d}{dt}Z_i = G(X_i - Z_i). \quad \text{(A7)}$$

The coefficients E, F, and G depend only on $I(t)$, on the total STM activity

$$x = \sum_{k=1}^{n} x_k,$$

and on the total LTM activity

$$z = \sum_{k=1}^{n} z_{0k}.$$

By A4 and A5,

$$\frac{d}{dt}x = -Ax + Bz + I \quad \text{(A8)}$$

and

$$\frac{d}{dt}z = -Cz + Dx. \quad \text{(A9)}$$

Equations A8 and A9 are independent of θ; they depend only on the total activity $I(t)$. These equations *decouple* total activity data (I, x, z) from pattern data (θ, X, Z), where $X = (X_1, X_2, \ldots, X_n)$ and $Z = (Z_1, Z_2, \ldots, Z_n)$ are also probability distributions. The total activity data influence the pattern data only via the coefficients E, F, and G, which are always nonnegative. No matter how wildly the CS input $I_0(t)$ and the UCS input $I(t)$ oscillate through time, these coefficients influence only the *rates* with which X and Z are influenced by θ, but not the *directions* in which X and Z can change in response to θ. It is this property that generalizes to yield the universal theorem cited above.

In particular, term $F(\theta_i - X_i)$ in A6 says that X_i approaches θ_i as learning proceeds (UCS read into STM). Term $E(Z_i - X_i)$ in A6 says X_i approaches Z_i (readout of LTM into STM). The net effect of these two terms shows how present demands of the UCS, expressed via θ, and past memories, expressed via Z, compete to change STM via X. Equation A7 shows that Z_i approaches X_i (transfer from STM to LTM). As X approaches θ, and Z approaches X, Z learns the spatial

pattern θ. On later performance trials, a CS input to v_0 activates x_0, which in turn activates the signal B. Signal B reads the pattern Z into STM via the terms Bz_{0i} in A4. Since $Z \cong \theta$, A4 shows that the x_is that are activated in this fashion are proportional to the θ_is, as desired.

Many aspects of associative learning can be understood using these STM and LTM laws in more complex anatomies. In particular, the Z_is are stimulus sampling probabilities whose properties explain in a neural setting the partial successes of statistical learning models. The distributions of STM and LTM traces also mimic and predict various data about serial learning, paired associate learning, and free recall experiments. See Grossberg (1974, 1978a, 1978e) for additional discussion.

Appendix C

This section summarizes how feedforward competitive interactions solve the saturation problem using automatic gain control by inhibitory signals, and how properties such as noise suppression, pattern matching, edge enhancement, and spatial frequency sensitivity follow as special cases.

Noise-Saturation Dilemma

All cellular systems face the following dilemma. If their inputs are too small, they can get lost in noise. If the inputs are too large, they can turn on all excitable sites, thereby saturating the system and rendering it insensitive to input differences across the cells. For example, suppose that the ith cell v_i receives an input I_i that can turn on some of its B excitable sites by mass action. Let $x_i(t)$ be the number of excited sites and $B - x_i(t)$ be the number of unexcited sites at time t. The simplest mass action law for turning on unexcited sites and letting excited sites spontaneously turn off is

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)I_i, \qquad \text{(A10)}$$

$i = 1, 2, \ldots, n$. Term $(B - x_i)I_i$ says that the input I_i turns on unexcited sites $B - x_i$ by mass action. Term $-Ax_i$ says that excited sites spontaneously becomes unexcited by mass action at rate A. Hence, when $I_i \equiv 0$, x_i can decay to the equilibrium point 0.

System A10 is inadequate for the following reason: Let the inputs form a spatial pattern $I_i = \theta_i I$. Given a fixed pattern $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, choose a background intensity I and let the system reach equilibrium. This equilibrium is found by setting $(d/dt)x_i = 0$ and solving for x_i:

$$x_i = \frac{B\theta_i I}{A + \theta_i I}. \qquad \text{(A11)}$$

Now keep θ fixed and increase I. That is, process the same pattern with different background activity. Then all x_i in A11 approach B even if the relative input intensity θ_i is small. This is saturation. How can the system preserve its sensitivity to θ even as I increases? In other words, how does the ith cell v_i compute its "reflectance" θ_i in response to a spatial pattern $I_i = \theta_i I$, $i = 1, 2, \ldots, n$, of inputs? Since

$$\theta_i = I_i I^{-1} = I_i\Big(\sum_{k=1}^{n} I_k\Big)^{-1},$$

cell v_i needs to know what all the inputs $I_1, I_2, \ldots, I_n$ are in order to compute θ_i. Since

$$\theta_i = I_i\Big(I_i + \sum_{k \neq i} I_k\Big)^{-1},$$

increasing the ith input I_i "excites" v_i (increases θ_i), whereas increasing any input I_k, $k \neq i$, "inhibits" v_i (decreases θ_i). When this intuition is most simply modeled by a cellular mass action network, we find the system

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)I_i - x_i\sum_{k \neq i} I_k, \qquad \text{(A12)}$$

$i = 1, 2, \ldots, n$. In Equation A12, I_i excites v_i via term $(B - x_i)I_i$, just as in A10. The new term

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

describes how the inputs I_k, $k \neq i$, inhibit (note the minus sign) the excited sites of v_i (which number x_i) by mass action. The *gain* of x_i is its decay rate. This is found by grouping together all the terms that multiply x_i. The sum of these terms is $A + I$, where

$$I = \sum_{k=1}^{n} I_k.$$