THE ROYAL SOCIETY

Published online 4 November 2002

A new interpretation of thalamocortical circuitry

Paul Adams* **and Kingsley Cox**

Department of Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Almost all the information that is needed to specify thalamocortical and neocortical wiring derives from patterned electrical activity induced by the environment. Wiring accuracy must be limited by the anatomical specificity of the cascade of events triggered by neural activity and culminating in synaptogenesis. We present a simple model of learning in the presence of plasticity errors. One way to achieve learning specificity is to build better synapses. We discuss an alternative, circuit-based, approach that only allows plasticity at connections that support highly selective correlations. This circuit resembles some of the more puzzling aspects of thalamocorticothalamic circuitry.

Keywords: thalamus; neocortex; synaptic error; synapse splitting; layer 6

1. INTRODUCTION

Almost all the information that reaches the neocortex arrives via the thalamus, which, to a first approximation, acts as a simple relay. Since the simplest, most efficient relay would be an uninterrupted axon, why does the thalamus exist at all? The unknown functions of the thalamus must be rather general, because its basic anatomy and physiology are universal throughout the various nuclei. It seems unlikely that this universal circuitry is used differently in each nucleus to perform specialized processing appropriate to the particular type of information being relayed by that nucleus (visual, somatosensory, auditory, motor, hippocampal, etc.). This universality has become even clearer with the recent realization that much (in primates, most) of the relayed information arises from layer 5 of the neocortex itself, rather than from subcortical sources (Sherman & Guillery 1996).

The universal core circuits of thalamus include massive feedback from layer 6 of the cortical region to which a thalamic region projects, side projections to layer 6 of the main relay input to middle cortical layers, and relay cell feed-forward excitation to and feedback inhibition from reticular nucleus. All these circuits are roughly topographic, although in no case is the detailed pattern of the connections understood (Sherman & Guillery 2001).

Many aspects of thalamic physiology also seem universal. Some of these (and their anatomical correlates) reflect basic relay function. Thus, a particular incoming 'driver' axon (which can originate subcortically or from layer 5 of the neocortex) typically makes numerous powerful synapses on the proximal dendrites of a relay cell, such that an incoming spike is likely to trigger an outgoing spike with a very short, fixed delay. Each relay cell receives only one such major input, though it may also receive a few subsidiary synapses from other driver axons. However, other universal features seem inexplicable in a simple relay scenario. For example, all relay cells have two firing modes, 'tonic' and 'burst', both of which carry out effective (though slightly different) relay functions (Guido *et al.* 1995).

A popular view is that the universal function of the thalamus is 'gating' (Steriade & McCarley 1990). In its simplest form, this asserts that the thalamus selects which information is sent on to the cortex. In one extreme form, the gate would be closed throughout the thalamus during sleep, and the rhythmic bursting of relay cells would be a type of busy signal or screensaver. In the awake thalamus, bursting would also be a 'no signal' mode, but could be selective for individual nuclei, or even individual relay cells, and might correspond to an attentional 'spotlight' (Crick 1984). However, the discovery that, in the awake state, bursting is irregular and time-locked to driving input (Swadlow & Gusev 2001) is difficult to reconcile with the 'no-signal' hypothesis. Instead, it seems likely that both burst and tonic modes are relay modes, and that the firing mode instructs the cortex how to handle the incoming information, rather than fundamentally changing that information.

We present a speculative account of some of the universal features of thalamocortical circuitry and physiology, based on the idea that complex circuits should be built using accurate synaptic learning.

2. CONNECTIONISM

Clearly, whatever the thalamus does, it is intimately related to whatever the neocortex does. Although the neocortex carries out a vast range of different functions, a suitable starting point is that it probably uses rather basic 'connectionist' principles. These principles are: (i) cortical neurons integrate their synaptic inputs and provide outputs to other neurons (or more formally, neurons compute weighted sums of their inputs); and (ii) activity-dependent changes in synaptic strength programme these neuronal computations using purely local signals (for example, Hebbian rules in unsupervised learning), perhaps together

* Author for correspondence (padams@notes.sunysb.edu).

One contribution of 22 to a Discussion Meeting Issue 'The essential role of the thalamus in cortical functioning'.

DOI 10.1098/rstb.2002.1164

with global feedback. The ways in which these connectionist principles actually work out in terms of circuits, synaptic weights, coding strategies, etc., may vary greatly from region to region, but these important details are unlikely to explain the universal aspects of thalamocortical circuitry.

Connectionist principles have to be embodied in real neurons and synapses, which have limitations of accuracy and speed. A reasonable starting point might be that the thalamus exists to minimize the impact of these limitations, to which very elaborate networks such as the neocortex might be particularly sensitive. One obvious possibility is that real biological inputs and outputs, encoded as spike trains, contain noise caused by imprecise spike timing, ultimately traceable to the small sizes of neural components and finite ion-channel numbers. It has been proposed, for example, that intracortical recurrent circuitry can minimize such noise, by a sort of spatial averaging mechanism (Deneve *et al.* 1999). Some current ideas about the thalamus (Sillito *et al.* 1994; Dong & Atick 1995) might fit into this category.

Another aspect of 'biological connectionism' that has hitherto been largely ignored centres on the second half of the connectionist paradigm—local activity-dependent weight setting. Two obvious biological limitations that are usually ignored in connectionist modelling are: (i) biological networks are very sparsely connected (because neural numbers are huge and neural wires are expensive); and (ii) because synaptic learning is a physical event, using small numbers of molecules it cannot be anatomically completely precise.

The traditional view of the first problem is that appropriate sparse connectivity is 'precomputed' by Darwinian gene-based evolution, and hard-wired by suitable marker molecules (netrins, ephrins, etc.). Activity-dependent synaptic learning is then used to set appropriate strengths of existing connections. An example of this is the use of 'arbour functions' in models of visual cortex development (Miller 1990, 1994). Particularly striking evidence for such hardwiring comes from the olfactory system (Wang *et al.* 1998), where individual wiring is achieved using thousands of special-purpose markers. However, in a way this beautiful example actually shows the weakness of the hardwiring approach as a general strategy, because these markers monopolize a significant fraction of the entire genome. The strategy works here only because these markers, which are the odorant receptor molecules themselves, are available 'gratis' for the secondary task of hooking up olfactory neurons to the appropriate glomeruli in the olfactory bulb. It would be impossible to coarsely wire the neocortex using such a costly, precomputed marker approach.

There is a second, even more powerful, argument against extensive cortical hardwiring. Only those features of the environment that persist over thousands of generations can be exploited by gene-based evolution. Prewiring eliminates most of the advantages of flexible learning that are thought to be a neocortical hallmark. Finally, there is considerable experimental evidence against such hardwiring (Sur & Leamey 2001).

It is widely suspected that sprouting provides a bridge between genetically specified hardwiring and activity-dependent learning at fixed complete connections (Miller 1990). Thus, sprouting from existing connections could provide new trial connections, which are then tested by activity-dependent synapse adjustment (Willshaw & von der Malsburg 1979; Fraser & Perkel 1990). Unfortunately, this approach has not been very extensively tested. It is commonly assumed that sprouting provides a 'free lunch', in that it allows new configurations to be tested without seriously degrading the quality of the final set of connections and weights (which could be more speedily attained with non-biological complete connectivity). However, there is some evidence that this may not be so (Elliott *et al.* 1996).

The second biological limitation of synaptic learning is that it may be anatomically imprecise in that, not just the connection across which there is correlated activity may strengthen, but nearby inactive connections may also be affected. There is a good deal of evidence that this occurs in the hippocampus (Bonhoeffer *et al.* 1994; Schuman & Madison 1994; Engert & Bonhoeffer 1997) and in the neocortex (Kossel *et al.* 1990). It has been argued that spines exist to compartmentalize the calcium signals that are the immediate trigger for synapse strengthening (Koch & Zador 1993), which suggests that minimization of anatomical learning inaccuracy has been of enormous importance in the vertebrate nervous system, and it has been widely assumed that such compartmentalization is complete. However, this hope is unrealistic because there are concomitant opposing pressures to miniaturize synapses, which makes anatomical specificity harder to achieve.

A final limitation that biological realism imposes on connectionism concerns the dynamic range of synaptic learning. Typical models require a wide range of possible synaptic weights, which can be implemented biologically in two different ways: varying the strengths of individual synapses ('physiology') and varying synapse numbers ('anatomy'). There is, as yet, no good experimental evidence as to how long-term weight changes are distributed between these two routes, although there is some evidence that as time progresses physiological changes are converted to anatomical changes (Colicos *et al.* 2001). The most efficient arrangement would probably be to make the initial change at the level of existing synapses (increasing transmitter release, phosphorylating existing receptors, recruiting additional receptors), and then (perhaps if subsequent activity does not immediately countermand these temporary changes) convert these changes *pari passu* to a change in the number of synapses. *Pari passu* refers to the requirement that the switch from physiological substrate to anatomical substrate should preserve the 'strength' of the connection, a parity that raises some interesting cell biological questions.

If synaptic strengthening is anatomically imprecise, and involves the creation of new synapses, these new synapses may not always form at the connection across which triggering activity occurred, and may even involve the creation of new connections. This could be regarded as an activity-dependent 'sprouting' mechanism, and could be either presynaptic or postsynaptic, or both. Such 'accidental' new synapses could provide a solution to the problem of finding the best set of connections in a very sparsely connected network such as the neocortex. However, such anatomical errors could also severely degrade network

performance. We propose that the thalamus could set the balance between flexibility and accuracy during neocortical learning. In § 3, we explore these ideas more quantitatively.

3. A MODEL OF SYNAPTIC ERROR

As a first step, we constructed a very simple model of synaptic learning that allows a more quantitative discussion of these issues. We considered a single presynaptic neuron (such as a lateral geniculate relay cell) that can connect to a set of postsynaptic neurons (such as the set of layer-4 cells in striate cortex that are the potential targets of LGN cells). The aim is to accurately connect the relay cell to a subset of these cortical cells using activity-dependent mechanisms. In our model, the various layer-4 cells do not influence one another, so each set of connections can be considered independently. The postsynaptic neurons are 'linear', so their output V_i ('activity') is simply given by the weighted input activity $w_i\ V_{\mathrm{pre}}$, where w_i is the strength of the connection to the ith postsynaptic cell:

$$V_i = w_i V_{\mathrm{pre}}. \tag{3.1}$$

An important feature of the model is that when a connection strengthens it does so in a digital manner, by adding new synapses. Because synaptic strengthening occurs in an all-or-none manner (Petersen *et al.* 1998), the addition of synapses should occur probabilistically. Except when errors occur, the new synapses added as a result of coincident activity across a connection should have the same 'connectivity' as the synapses comprising the original connection, and the only reasonable way to accomplish this is for a new 'daughter' synapse to be closely associated with an existing 'parent' synapse, either because the original synapse 'divides' (Carlin & Siekevitz 1983; Toni *et al.* 1999; Luscher *et al.* 2000; but see Fiala *et al.* 2002) or because of *de novo* formation of a bouton/spine pair very close to the original synapse, the new bouton belonging to the same axon as the parent bouton, and the new spine belonging to the same dendrite as the parent bouton (Colicos *et al.* 2001). Either mechanism seems compatible with the expectation that the biochemical triggers for activity-dependent synaptogenesis act locally near, or at, the initiating synapses. In either case, the fundamental mechanism is one of synaptic 'replication', in which the connectivity of the new synapse is specified by, and triggered at, an existing synapse (Adams 1998).

In the model, we used a standard Hebb rule, according to which the change in strength of a connection due to coincident activity is proportional to the product of the presynaptic and postsynaptic firing rates (k_i is a, possibly connection-specific, learning rate):

$$\mathrm{d}w_i/\mathrm{d}t = k_i V_{\mathrm{pre}} V_i. \tag{3.2}$$

How is this rule to be interpreted at the level of single synapses? If the probability that each existing synapse gives rise to a new synapse depends simply on the coincident activity, then the effective gain of the Hebb rule over the whole connection would increase as more synapses are added, and equation (3.2) would not be obeyed (see figure 1*b*). To use a traditional Hebb rule in a framework of individual synapses (rather than the abstract 'synaptic weight' of traditional models) appears to imply that the probability that coincident activity causes synapse replication depends inversely on the number of synapses comprising a connection—a decidedly 'non-local' and rather implausible requirement. Instead, we suggest that when replication occurs, either the existing or the new synapse be 'implastic' (figure 1*a*). This would allow the Hebb rule to operate locally at individual synapses (which is where the machinery appears to be located), while preserving the Hebb rule quantitatively over the whole connection. Thus, coincident activity would promote 'replication' of the connectivity and strength of existing synapses, but not their plasticity. The rather anticlimactic result of this discussion is that we retain a conventional Hebbian learning rule, though expressed in terms of synaptic number rather than weight.

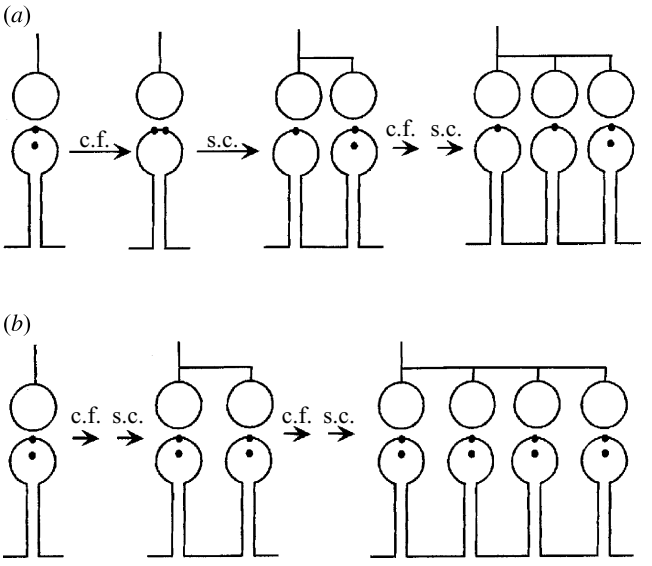

Figure 1. Asymmetrical and symmetrical synaptogenesis following correlated firing (c.f.) across a connection. In this figure, the new synapses preserve the connectivity of the original synapse, though this does not necessarily imply physical splitting. A group of AMPA receptors is shown as a black dot. AMPA receptors located in the membrane of the spine head are electrophysiologically functional and endow a synapse with its 'strength'. AMPA receptors located in the spine head interior are electrophysiologically silent, but constitute a reserve pool that endows a synapse with plasticity. In (*a*), correlated firing across the connection, initially comprising one synapse, leads to physiological strengthening (insertion of a group of AMPA receptors). The strengthened synapse is no longer plastic (the actual mechanism could be different from depletion of the reserve, which is used here as an iconic representation of plasticity). This is followed by structural changes (s.c.) that convert the temporary two-strength synapse to two one-strength synapses, only one of which, however, regains its plasticity. If the same amount of correlated firing occurs again, the same strengthening of the connection occurs (as in a conventional Hebb rule). In (*b*), both the orginal and the new, correlation-induced, synapses are plastic, so a second episode of correlated firing leads to a larger increase in synaptic strength (in contradiction of the usual quantitative formulation of Hebb's rule). Thus, in both parts 'strength replication' and 'connectivity replication' are symmetrical, but only in (*b*) is 'plasticity replication' symmetrical. If the new synapses do not have the connectivity of the original synapses, the result would be a 'synaptic mutation'.

Combining equations (3.1) and (3.2), we obtain

$$dw_i/dt = \phi_i w_i, \tag{3.3}$$

where $\phi_i = k_i V_{\mathrm{pre}}^2$. The parameter ϕ_i plays a rule rather similar to 'fitness' in evolution models. Straight Hebbian learning leads to unlimited synaptic growth, which is biologically unrealistic, especially when expressed in terms of synapse number. Most models introduce a 'normalization' process at this point, for example constraining the total number of synapses made by the presynaptic neuron to be constant. There are various ways to biologically implement such normalization, such as competition for growth factors, inclusion of a non-Hebbian forgetting term or use of a time-dependent learning rule, but in our simulations we used a brute force normalization, dividing the weights by a factor that kept their sum constant. Thus, activity merely triggered synapse rearrangement between the various target cells.

So far, this model exhibits very simple behaviour (von der Malsburg & Willshaw 1980). Connections with high ϕ_i values grow at the expense of low-ϕ_i connections, and eventually all but the 'fittest' connections disconnect. The Hebbian rule is able to detect and amplify small biases and generate a completely precise set of final connections. However, these final connections are irreversible, since there is no way to create new connections.

We now introduce synaptic error, by assuming there is some low probability, E, that a new synapse created by conjoint neural firing appears, not at the connection across which conjoint activity occurred, but at an adjacent connection (even if that 'connection' was nonexistent, with no synapses). The learning rule becomes

$$dw_i/dt = (1 - E)\phi_i w_i + E(\phi_{i-1} w_{i-1} + \phi_{i+1} w_{i+1})/2. \tag{3.4}$$

(This could also be regarded as a rule incorporating coactivity-dependent sprouting; the vital point is that it provides for the creation of new connections).

We simulated the behaviour of this model numerically, using a row of 13 target neurons and a probabilistic version of equation (3.4) for the formation of individual synapses. We assigned a high value, ϕ_{m}, for the neuron at the left end of the row, and a uniform low value, ϕ_{p}, over the remainder of the row. We found that eventually a steady-state distribution of synapses was attained, with (as expected) most of the synapses located near the preferred neuron, but a considerable tail of synapses straggling away from it (figure 2). This steady tail arises because synapses leak from the high-fitness neuron at a rate that exactly compensates for the excess production there. As expected, we also found that increasing the error rate or decreasing the fitness ratio $\phi_{\mathrm{m}}/\phi_{\mathrm{p}}$ led to a broader tail of synapses (figure 2). As expected, zero error leads to completely specific connections.

We also examined the transient behaviour of the model. Figure 3 shows the initial steady distribution of synapses generated when the leftmost neuron was fittest, and the subsequent migration of synapses when the rightmost neuron instead became the fittest. The initial trail of synapses acts as a seed for the growth of the rightmost connection. In figure 4, all the synapses were initially placed on the leftmost neuron, and then allowed to migrate to the fitter, rightmost neuron. The figure shows the number of synapses accumulating on this target neuron. There is an initial, rather variable delay before the first synapse reaches the target, which then rapidly flourishes. As expected, this delay grows smaller as the total number of synapses increases.

The model can be analysed straightforwardly in the continuum limit (large numbers of synapses and neurons). In the above equations w becomes a synapse density at the point x on a continuous neuronal line, and equation (3.4) becomes

$$\partial w/\partial t = (\phi - <\phi>)w + 0.5\phi E \partial^2 w/\partial x^2. \tag{3.5}$$

(The average fitness $<\phi>$ is introduced to enforce weight normalization; see von der Malsburg & Willshaw (1980)). Although equation (3.5) is nonlinear, in the steady state $<\phi>$ attains some constant value, leading to an ordinary second-order differential equation, whose solution in the particular case of a high-fitness mesa surrounded by a low-fitness plateau is, over the low-fitness plateau,

$$w_{\mathrm{p}} = C\exp(-x/\lambda_{\mathrm{p}}) + D\exp(x/\lambda_{\mathrm{p}}), \tag{3.6}$$

where $\lambda_{\mathrm{p}}^2 = E\phi_{\mathrm{p}}/2(<\phi> - \phi_{\mathrm{p}})$ and C and D are constants. For short mesas and long plateaux (the conditions of our simulations), the second term on the RHS can be neglected, and the decay outside the mesa is exponential, as observed. The space constant λ_{p} is given by

$$(2\lambda_{\mathrm{p}} + n)/n\lambda_{\mathrm{p}}^2 = 2(\phi_{\mathrm{m}}/\phi_{\mathrm{p}} - 1)/E, \tag{3.7}$$

where n is the number of neurons in the mesa. We found that the distributions in our simulations were close to exponential, with space constants described by equation (3.7) (figure 5).

This is a highly simplified model of synaptic rearrangement, probably the simplest that exhibits selective wiring, but it clearly shows an intuitively plausible phenomenon which is probably common to more elaborate models, that incorporation of anatomical error into a learning rule produces a blurring of connections away from the precise pattern attained in the absence of error. Furthermore, it illustrates quantitatively (equation (3.7)) the expectation that the extent of blurring is greater when error rates are high or when targeting signals are weakest. We regard the mutual interdependence of the factors that promote blurring (i.e. error) and those that enhance specificity (selective neural activity) of capital importance, because it means that a correlational mechanism (such as Hebb's rule) cannot build circuits of unlimited precision. Furthermore, even very low synaptic error rates could be of great significance in preventing the self-organization of large networks such as the neocortex because correlations in the real world are likely to be weak and near the limits of detectability (a tiger in the grass).

4. A SLIGHTLY MORE REALISTIC MODEL OF SYNAPTIC LEARNING

Although the above model shows the basic phenomena, it is rather unrealistic, because it neglects interactions between different inputs (other than the normalization process), and does not really specify the origins of the postulated fitness differences. To some extent, this can be remedied by flipping the model, so that a row of presynaptic cells projects onto a single, linear postsynaptic cell. We

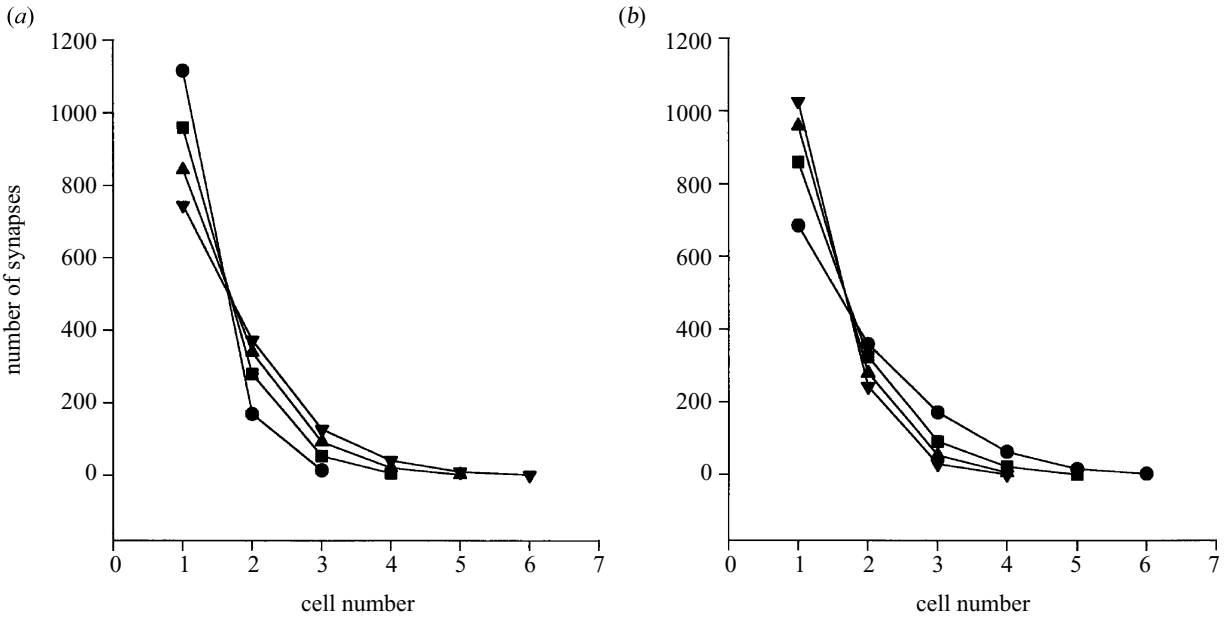

Figure 2. Simulations of the steady-state distribution of synapses on target neurons for various error rates (*a*) or fitnesses (*b*). In each case, the left connection was the fittest ($\phi = \phi_m$), and there were 13 neurons and 1300 synapses. Connections on neurons 2–13 had uniform low fitness ϕ_p. The average number of synapses on each neuron achieved at equilibrium is shown for various values of error rate E and fitness ratio ϕ_m/ϕ_p (no synapses were formed on neurons 7–13, which are not shown). In all cases, synapses are most numerous on the high-fitness neuron, but as error rates increase (*a*) or fitness ratios decrease (*b*) synapses become more spread out. Values of E used in (*a*) were 0.1 (circles), 0.2 (squares), 0.3 (triangles) and 0.4 (inverted triangles), with $\phi_m/\phi_p = 1.4$ throughout. Values of ϕ_m/ϕ_p used in (*b*) were 1.11 (circles), 1.25 (squares), 1.42 (triangles), 1.66 (inverted triangles), with $E = 0.2$ throughout.

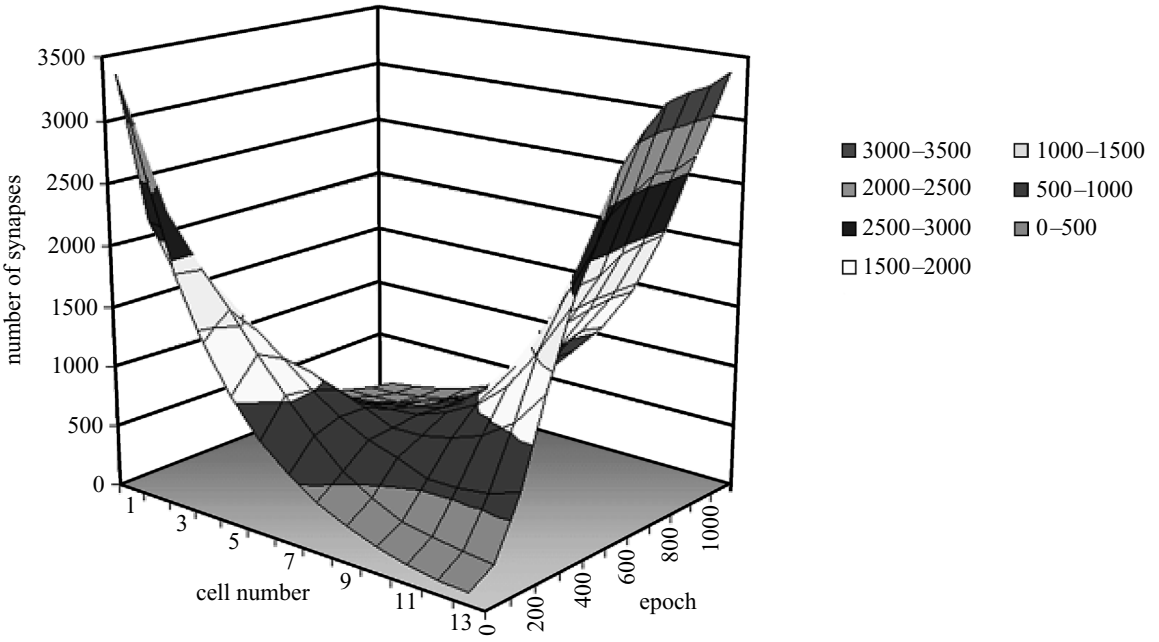

Figure 3. The migration of synapses following a mirror reversal of fitness. Initially synapses were equilibrated with the leftmost connection (on cell 1) being 5% fitter than the others. This resulted in the zero-epoch profile, with most of the synapses on cell 1, but with some spread across the entire set of target cells. The rightmost connection (on cell 13) was then made 5% fitter than the others, and the resulting profiles were plotted for successive epochs (after attainment of equilibrium and then binning results for 20 consecutive epochs to reduce noise). The total number of synapses was 13 000 and the error rate 0.25.

can now ask what stable set of weights emerges if the input neurons are subjected to a given series of patterns. This model (without synaptic error) is essentially the prototype of all connectionist unsupervised learning models (Diamantaras & Kung 1996). Because the Hebb rule detects correlations, in a fully connected error-free network the weight vector gradually aligns with the leading eigenvector of the covariance matrix of the set of patterns, also known as the principal component. Each pattern 'pulls' the weight vector towards itself, and an equilibrium

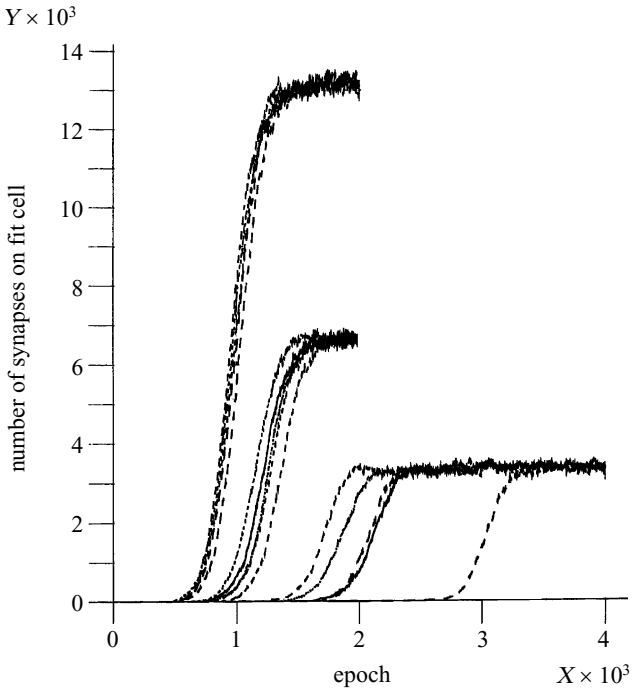

Figure 4. Kinetics of appearance of synapses at the fittest connection, on the rightmost neuron (cell 13). All synapses were initially placed on the leftmost neuron. The number of synapses on the right neuron was plotted at successive epochs. The total numbers of synapses M were 26 000 (left five runs), 13 000 (middle five runs) or 6500 (right five runs). Note that there is a variable delay before the formation of the first synapse on the fittest neuron, followed by rapid increase in the connection strength. However, because for the parameter values used ($E = 0.1$, $\phi_m/\phi_p = 1.05$) the length constant is quite high, the fittest neuron only gains about half of the total number of synapses.

is attained when all these little pulls balance out. If the set of patterns is visualized as a cluster of points in high-dimensional space, then the principal component is the least-squares line through these points, the direction along which the variability of the points is maximal. The postsynaptic neuron evolves to act as an ideal statistical filter, since its output in response to any particular pattern is the projection of that pattern on the principal component.

If an unvarnished Hebb rule is used, although the weight vector aligns with the principal component, the weights themselves grow without bound. This could be remedied by constraining the sum of the weights to be constant (as in the previous section). However, to allow for both positive and negative activities and weights, it is usual to normalize the sum of the squares of the weights. This can be done using a purely local rule, involving an additional 'forgetting' term given by the product of the square of the postsynaptic activity and the weight (Oja 1982).

We are currently investigating how introducing learning errors modifies the behaviour of this single neuron principal-component analyser. We find that the weight vector now stabilizes to a new direction that is intermediate between the principal component and the direction corresponding to a uniform weight distribution, so that the neuron no longer acts as a statistically optimal filter. The extent of degradation of the filter depends both on the error rate E and on the structure of the covariance matrix.

This behaviour can be seen particularly clearly in the limiting case where the patterns are uncorrelated, corresponding to a cloud of points whose main axis is aligned with one of the input coordinates. This coordinate represents the input neuron whose activity over the set of patterns has maximal variance. If the patterns are uncorrelated, then the optimum arrangement is for the postsynaptic neuron to be connected exclusively to the 'most interesting' input neuron, the one whose variance is maximal. In this particular case, since the patterns are uncorrelated, the evolution of any particular synaptic weight does not depend on the evolution of the other weights, but only on the variance of the relevant input neuron. For uncorrelated patterns, the model becomes identical with that in the previous section, with the variances replacing the fitnesses. In the simplest case, where the variance at one input neuron is high, and the variance for all the other input neurons is low, the main weight will be on the correct neuron, but it will not be exclusive, there being an exponential tail of weight distribution onto 'nearby' neurons. Thus, learning errors lead to a suboptimal filter.

Of course, it is unlikely that neurons in the visual system, or anywhere else, act as principal component filters, primarily because such a representation is suited only to the simplest type of Gaussian pattern statistics. Nevertheless, something like the core concept, statistically optimal representation, is probably at work in the neocortex, and it is likely that in all cases anatomical learning errors will impose limitations on the self-organization of neural networks in response to structured inputs. In § 5, we consider how the impact of such errors can be minimized.

5. LIVING WITH ERROR

One obvious way to minimize the impact of learning errors is to lower the error rate, by optimizing the machinery of synaptic plasticity. Perhaps the most obvious possible way for synaptic weight adjustment to be anatomically imprecise would be diffusion of second messengers involved in the plasticity cascade from active to inactive synapses. Of these messengers, calcium is the best understood and probably the most important, since calcium influx through NMDA receptors seems to be crucial for Hebbian learning. Consistent with the idea that activity-dependent learning should be synapse-specific, spines (which are found wherever vertebrate learning occurs) seem to exist to compartmentalize these calcium signals, by a combination of physical barriers and biochemical machinery (Koch & Zador 1993).

However, achieving the goal of complete synaptic independence is incompatible with the equally desirable goal of maximizing synapse numbers (since ultimately the amount of useful information that can be stored in a neural network depends both on the numbers of synapses and the precision with which their strengths can be independently set). For example, increasing the distances between synapses decreases both crosstalk and numbers. Over the open times of NMDA receptors (*ca.* 100 ms) even well-buffered calcium diffusion ($D \approx 0.01\ \mu m^2\ ms^{-1}$) can span typical intersynaptic distances. The problem is compounded by the requirement that synapses should not only

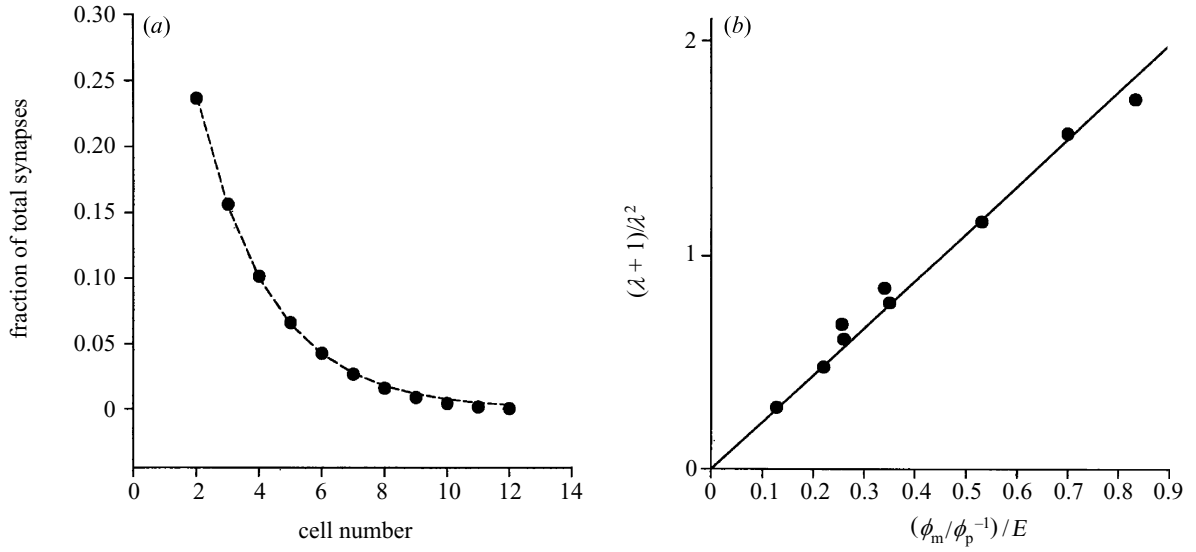

Figure 5. Comparison of simulation results (points) with theory (lines). (*a*) The fraction of the total number of synapses (13 000) that form at less fit connections (i.e. the plateau region) compared with an exponential curve of length constant $\lambda = 2.45$ neurons. $E = 0.2$; $\phi_m/\phi_p = 1.05$. Cell 13 received no synapses and was omitted. (*b*) The reciprocal effects of relative fitness and error rate. Each point was obtained for a different combination of ϕ_m/ϕ_p and E from data similar to those shown in (*a*). The linear regression through the origin has a slope of 2.2, compared with the predicted slope 2.0 (see equation (3.7)). Because in the simulation the fit neuron was at the end of the line, while in the analysis it is in the middle, $n = 2$.

be well separated but also small, since this means that rather small numbers of calcium ions must reliably trigger synapse modification.

It is therefore not surprising that recent experimental tests have shown that the synapse specificity of LTP can break down quite dramatically (Engert & Bonhoeffer 1997; Schuman & Madison 1994). In these experiments, one common feature was that the stimulation protocols used to generate LTP were rather drastic, typically involving hundreds of spikes or long-maintained depolarizations, precisely the circumstances in which calcium diffusion will be favoured. It seems probable that under more natural conditions, involving rather precise timing of pre- and postsynaptic spikes, and very brief and localized calcium signals, the specificity of synaptic strengthening or weakening will be much greater, but presumably not perfect.

So how can the brain live with ineluctable learning errors? The problem is that, because of the positive feedback inherent in the Hebb rule (which is what makes it so useful), errors can propagate and prevent useful learning, especially when the statistical regularities that guide learning are relatively weak. Ultimately the useful size of a biological neural network must be limited by the accuracy with which the individual 'bits' that constitute its 'programme' can be written, and this limitation is likely to be particularly severe in a vast network such as the neocortex which deals with relatively subtle statistical regularities (such as finding principles in masses of neuroscientific data).

More specifically, how can a single LGN relay axon be correctly wired to a handful of neurons amongst the hundred million or so in the primate striate cortex, using the statistical correlations that are generated either by intrinsic prenatal activity (such as retinal waves) or postnatal visual experience? Roughly speaking this wiring feat is accomplished because the firing of the incoming axon is more highly correlated, on average, with the firing of the 'correct' handful of cells than with the incorrect myriad of other potential targets, although it might also be aided by biochemical cues. However, we have seen in § 3 that if learning is anatomically imprecise, connections will also be formed onto 'neighbours' of the correct cells, and these incorrect connections will abound to an extent that depends inversely on how much more the axon's firing is correlated with 'correct cells' than with the 'incorrect' cells. If the selective correlations that guide wiring are strong, then wiring can be accurate even though the learning rule is anatomically imprecise (see equation (3.7)). If these correlations are internally generated (for example, by retinal waves) then they could, in principle, be sufficiently selective that precise wiring could still be achieved, but if they are generated by interaction with the real world, this cannot be guaranteed, and self-organization may fail.

It could be argued that although imprecise wiring is inevitable, its flexibility advantage (arising because currently incorrect connections may be useful in the future) always outweighs its drawbacks (poor performance). We suspect, however, that since errors are inevitable there will always be adequate flexibility, and that the main learning task is to ensure that residual errors do not prevent useful learning.

We therefore suggest the following formulation of the neocortical-learning problem. Given that a particular set of 'correct' connections has been established as a result of past selective correlations, how can we guarantee that these connections will remain correct if (i) they remain plastic and (ii) the correlations are not guaranteed to remain selective? The difficulty is, of course, that if the network continues to learn, less selective correlations can allow errors to persist, leading to incorrect wiring.

Formulating the problem in this way immediately suggests a solution. If the survival of errors depends on the selectivity of correlations across currently connected neurons compared with the correlations across incipient con-

nections, then a measurement of these correlations (across current and incipient connections) could be used to control learning at current connections. In particular, if the ratio of the correlations across current and incipient connections exceeds some critical value (which will depend on the error rate and the number of synapses comprising a connection), learning could be switched off entirely, guaranteeing that error will not occur. In this context, an 'incipient connection' is defined as one that could be immediately formed as a result of an anatomical error in strengthening of an existing connection and subsequent synaptogenesis. (Of course, further errors could result from strengthening of newly formed 'erroneous' connections; the gradual propagation of these errors away from the original 'correct' connections underlies the insidiousness of the error problem, but this is avoided if the initial error is prevented).

The circuitry that is required to implement this solution is shown in figure 6. Presynaptic neurons are labelled J, postsynaptic neurons are labelled I, and neurons which detect correlations are labelled K. Actually, two different though related circuits are needed, corresponding to two different types of synaptic learning error. If an anatomical error is made in creating a synapse at an existing connection, the erroneous synapse could be made either between the correct presynaptic neuron and an incorrect postsynaptic neuron (dubbed a 'presynaptic error' because the initial correct change is presumably initiated presynaptically, but triggers the selection of an incorrect postsynaptic target which happens to be available nearby), or between the correct postsynaptic neuron and an incorrect presynaptic neuron (a 'postsynaptic error'). (Error rates are assumed to be sufficiently low that the probability of a double error is negligible.) In figure 6 the set of incorrect 'neighbours' onto which errors can be made is depicted, for convenience, as the neurons whose cell bodies are 'neighbours' of the cell bodies of the correct neurons. However, the actual set of neurons is presumably determined instead by the anatomical disposition of the terminals of the presynaptic neurons and the dendrites of the postsynaptic neurons; difficulties with this assignment are discussed below.

In the case of 'presynaptic' errors, it would be necessary to measure the correlations between the currently connected neurons, and also the correlations between the currently connected presynaptic neuron and the postsynaptic neurons to which it could become connected if an error occurs (figure 6*a*). These correlations would be measured using a special type of coincidence-detecting 'K'-neuron, shown using a different symbol from the conventional pre- and postsynaptic J- and I-neurons. Excitation of K-neurons would be caused by a pair of spikes, one in a presynaptic neuron and one in a postsynaptic neuron, with a suitable timing delay corresponding to the paired spikes which cause the strengthening of the existing connection. K-excitation by a correlated spike pair would also reflect any time dependence of the strength change at the J–I connection triggered by that spike pair. One possible way to accomplish this would be to use similar NMDA receptors to trigger strength changes in J–I connections or excitation of K-neurons, but selective innervation of distal and proximal dendrites of K-cells (as sketched in figure 6) could also be employed.

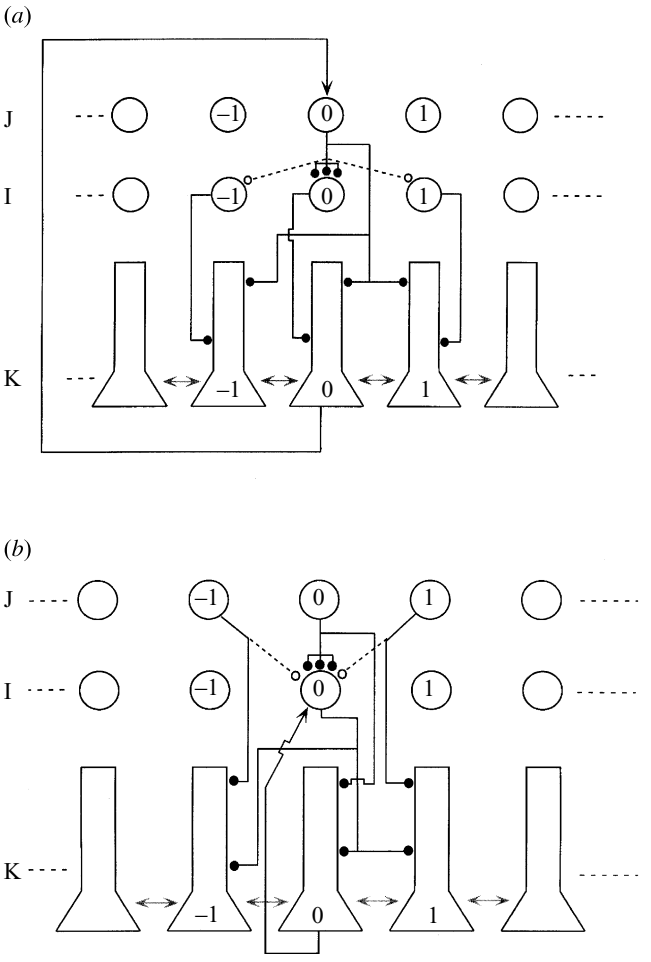

Figure 6. Circuits for error avoidance. In both (*a*) and (*b*) the middle neuron in the top row (J_0) has formed a selective connection (small solid dots) on the centre neuron of the middle row (I_0), as a result of selective correlations in the firing of these two neurons. If this connection undergoes further strengthening as a result of correlated firing, the added synapses could (with very low probability) be formed on the neighbours of I_0 (or, more exactly, on neurons whose dendrites are neighbours of the existing connection). These possible 'presynaptic error' synapses or 'presynaptic mutations', which are incipient connections, are shown as dotted lines and small open circles in (*a*). If these errors are to propagate they must be supported by adequate correlations in the firing of the J_0–I_{-1} or J_0–I_1 pairs, compared with J_0–I_0 correlations. These correlations are measured by K-neurons in the bottom layer. If propagation of errors is unlikely, K_0 fires, presynaptically enabling the plasticity of the current connection made by J_0. (*b*) An alternative 'postsynaptic' scenario, in which strengthening of the J_0–I_0 connection could cause incorrect synapses to form from neighbours of J_0 (or, more precisely, from axon branches which are neighbours of the existing connection). In this case, the correlations that must be measured by K-neurons are slightly different, and if their selectivity is favourable, plasticity of the current connection should be enabled postsynaptically. In real networks, each J-cell can make several connections on I-cells, and each I-cell can receive several inputs from J-cells. In these cases, the relevant correlations, measured by K-cells, are the average correlation across the current connections, compared with the average correlation across incipient connections (see figure 7).

The next step is to compare the excitation of K-cells corresponding to current connections with excitation of K-cells corresponding to incipient connections. If this relative excitation is strong enough, it must somehow enable the plasticity of the current connection, secure in the knowledge that if this connection strengthens, added synapses either will be correctly placed or will be inadequately supported by weak correlations. An obvious way to do this would be for the 'current' K-cell to be inhibited (via an interneuron) by the 'incipient' K-cells, illustrated in figure 6 by short horizontal arrows. Thus, firing of the current K-cell (denoted K_0 in figure 6) would signal that the present input is well correlated with the firing of the currently connected output, compared with the correlation with the firing of the 'neighbours' of the current output, and this firing should therefore enable the plasticity of the current connections (only one shown in figure 6). Although at first sight it might seem logical to lead the axon of this K-cell to the synapses comprising the current J–I connections, this would be quite complicated to wire up. It would be much simpler to lead the axon to the cell body of the appropriate J-neuron, and to multiplex onto the spike train emitted by that J-cell an additional command, which would automatically enable the plasticity of any connection that J-neuron makes onto I-neurons. If the average rate of the spikes of the J-neuron was used to convey the information about the input, then it would be natural to use a second-order statistical parameter such as spike clustering to convey the plasticity-control signal; in the simplest case, a 'tonic' firing mode could be used to enable plasticity and a 'burst' firing mode could be used to disable it. Thus, the J-cell associated with a current connection would continuously monitor the selectivity of the waxing and waning correlations across that connection (just as the strengthening of that connection would itself depend on the flux of the absolute correlations across it), only allowing learning under favourable conditions. (Of course, the I-cells would, under all circumstances, *process* the impinging J-spikes.)

Figure 6*a* illustrates the principle, but it is highly oversimplified. Not only might a given J-cell connect to several I-cells, but also each I-cell could obtain input from several J-cells. Indeed, part of the reason that the firing of the neighbouring I-cells I_{-1} and I_1 are to some extent correlated with the firing of J_0 is that they obtain input from J-cells whose firing is correlated with the firing of J_0 (for example, because of patterned visual input). These correlations have, in the past, been insufficient to cause direct wiring to J_0, even though the occurrence of errors tends to promote such wiring. All the other I-cells also have their plasticity controlled by suitable K-cells, which can be imagined as additional K-sublayers not shown in the figure. At first sight it might seem that if there are n I-neurons, and each I-neuron has m neighbours, a total of nm K-cells would be needed. However, insofar as each J-cell innervates several I-cells, all these I-cells constitute a 'current connection', and a single K-cell could compute the average correlation between that J-cell and its I-targets. Likewise, a single K-cell could also compute the average correlation between a J-cell and all the neighbours of its current I-targets. (In these average correlations, a K-cell would be excited by paired spikes originating in any of its J- or I-inputs).

What modifications to figure 6*a* should be made to allow for the fact that several J-cells can converge on a single I-cell? For example, in the cat striate cortex several LGN relay neurons converge onto single layer-4 cells, endowing these neurons with their characteristic receptive field properties (Reid & Alonso 1995; see figure 7). Presumably, this arrangement exists because in the past the activities of these particular relay neurons were highly correlated with the activities of this particular layer-4 target cell (Miller 1994). One way to maintain the specificity of these connections, even if present visual experience is less structured than in the past, would be to enforce a strong 'critical period' outside which plasticity is turned off, but this would make the system inflexible. Instead, the approach advocated here is that plasticity should merely be temporarily disabled, on a connection-by-connection basis, whenever the correlations that initially wired the connections wane. In this case, the output of the K-cell that corresponds to the I-cell receiving the convergent J-input should be led back to the all the appropriate J-cells, and enable their plasticity (figure 7). (Of course, this has the disadvantage that some of these J-cells also make synapses on other I-cells, and the plasticity of these connections will also be spuriously enabled. However, the K-partners of these I-cells will not be firing, limiting the enablement; plasticity would be controlled in a 'distributed' manner just like regular neural computation.)

Figure 6*b* shows the rather similar machinery that would be required to control 'postsynaptic errors'. In this case, a synaptic strengthening error would involve formation of a connection between a new spine originating at dendrites involved in the current connection, and a bouton forming on an axon terminal that is near to the axon making the original connection. These neighbouring axons would form synapses onto K-neurons that compute correlations across 'incipient' connections (K_{-1} and K_1 in figure 6*b*), which would, in turn, inhibit the 'current' K-cell K_0, which detects the correlation between the neurons contributing to the current connection. In this case, the output of K_0 would be used to control the plasticity of the current connection postsynaptically.

6. COMPARISON WITH THALAMOCORTICAL CIRCUITRY AND FUNCTION

Both figure 6*a* and 6*b* are reminiscent of some of the universal features of thalamocortical wiring. In particular, Callaway (1998) has pointed out that in primate striate cortex layer-6 neurons receive a copy of the input to layer-4 cells (via collaterals of relay cells) and of the outputs of these cells (via descending branches; see also Tarczy-Hornoch *et al.* 1999). In the present scheme, this would arise because layer-6 neurons are essentially evaluating the hypothesis that the spikes fired by layer-4 cells are 'caused by' spikes in the relay cells to which they are connected. They also evaluate the alternative hypothesis that some of these layer-4 cells spikes are 'caused' by relay cells to which they are not currently connected, but could easily become connected. The definition of 'causation' that layer-6 cells use is simply appropriate temporal contiguity. The relative strengths of these two hypotheses are then used to decide whether to allow the existing connections to be modified by the ongoing neural activity. The reason

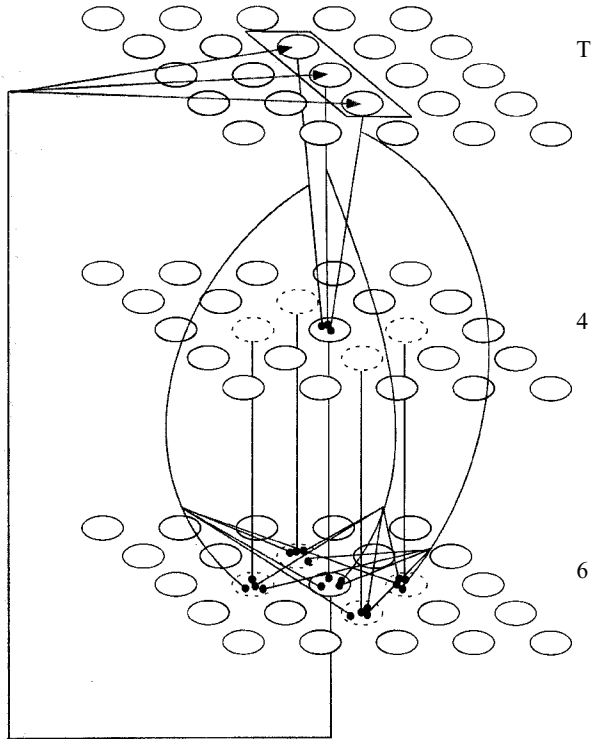

Figure 7. A more realistic version of figure 6*a*, based on the LGN (T) projection to striate cortex (layers 4, 6). In this figure the J-cells are shown explicitly as thalamic relay cells (T) , the I-cells as cortical layer-4 cells (4), and the K-cells as cortical layer 6. The corticothalamic feedback connections (arrows) terminate on the distal dendrites of relay cells, where they can activate metabotropic glutamate receptors, which depolarize relay cells and shift them from burst ('implastic') to tonic ('plastic') mode. Several relay cells converge on a given layer-4 cell, in this case three relay cells responding to a bar of light or darkness on the retina innervate a single 'simple' layer-4 cell. This simple cell in turn innervates, via a fixed connection, the soma of its corresponding layer-6 partner, which also receives, on its distal apical dendrite, input from branches of the relay cell axons that innervate the layer-4 simple cell. For simplicity, the extended electrotonic structure of layer-6 cells, sketched in figure 6, is not shown here. The central layer-6 cell would, by virtue of either of these two types of input, itself be simple. (Complex layer-6 cells also occur, but these would correspond to the K-cells shown in figure 6*b*.) Note that the activation of the layer-6 cells would depend, in this scheme, on the conjunction of action potentials in its input cells. The firing of the central layer-6 cell then depends on a comparison of its own activation with those of its neighbours (shown as dotted circles), which receive their proximal inputs from the *neighbours* of the layer-4 cell (also shown dotted). Note also that the central layer-6 cell feeds back to all the thalamic relay cells that innervate its layer-4 partner. Although these postulated connections and properties are consistent with the known anatomy and physiology of thalamic and cortical cells, they venture slightly beyond it. However, the circuitry shown here, and in figure 6, can easily be established by two types of offline calibration signals applied in alternation while either the T–6 or 6–T connections are selectively plastic.

why plasticity is rationed in this way is because it is a double-edged weapon—it allows refinement of the existing set of weights, but at the potential cost of forming inappropriate connections.

Is there any evidence that layer-6 cells do compute such 'correlation selectivity' signals? In cat striate cortex, layer-6 cells fall into two major physiological groups, 'simple' cells and 'complex' cells (e.g. Hirsch *et al.* 1998). The majority of those layer-6 cells that project back to the LGN appear to be simple (Grieve & Sillito 1995). These cells seem to have similar response properties to the 'simple' and 'complex' cells in the overlying cortical column. Simple cells in layers 4 and 6 could have identical receptive fields either because they receive identical afferents or because the layer-6 cell is driven by the layer-4 cell, but in either case the relevant connections would have to be quite strong. Since the connections are weak, it is more plausible that a layer-6 cell mimics a layer-4 cell because it receives both sets of input, which reinforce each other, as shown in figures 6 and 7. However, although neurons elsewhere in the brain act as coincidence detectors (Agmon-Snir *et al.* 1998), there is so far no direct evidence for this in layer 6.

The schemes shown in figures 6 and 7 suggest that a layer-6 cell's projection back to the thalamus should innervate the relay cells that drive the layer-4 cell that contributes to that layer-6 cell's receptive field properties. A rather similar wiring arrangement has been postulated on quite different grounds by Sillito *et al.* (1994), for which they have obtained some evidence. However, the detailed pattern of the connections to and from layer-6 cells is still unknown. There is no evidence that individual spikes or bursts of spikes in relay cells are more or less likely to generate LTP at thalamocortical synapses, as postulated above.

7. WIRING UP THE ERROR CONTAINMENT CIRCUITRY

Although the above account concentrates on the issue of maintaining appropriate connections in the presence of noisy input and synaptic learning errors, the postulated circuitry cannot prevent error entirely. Most of the errors that do still occur will impair the efficiency with which neural circuits process information, but occasionally a new connection will be useful, especially if environmental changes allow it to support a high level of correlated firing. In these circumstances, feed-forward circuits could actually rewire, requiring an adjustment of the error-prevention circuitry. These changes are illustrated in figure 8. Figure 8*a* shows the initial connections. If cell J_0, which (as a result of a previous high degree of correlation) is currently wired to cell I_0, accidentally makes a synapse on cell I_1, activity patterns might change such that the new connection replaces the old connection (as shown in figure 8*b*). The K-cell connections that 'guarded' the original J_0–I_0 connection are no longer appropriate for guarding the new connection. In particular, the connections shown as dotted in figure 8*a* must be broken, and the connections shown as dashed in figure 8*b* must be created. This could be done if the circuit is taken 'offline' so it is no longer exposed to the patterned activity that caused the initial J–I rewiring, and is instead subjected to certain internally generated 'calibration signals'. While offline, the J–I synapses would, by the action of a suitable neuromodulator, be rendered implastic, and unable to respond to the calibration signals. Conversely, the connections to and from

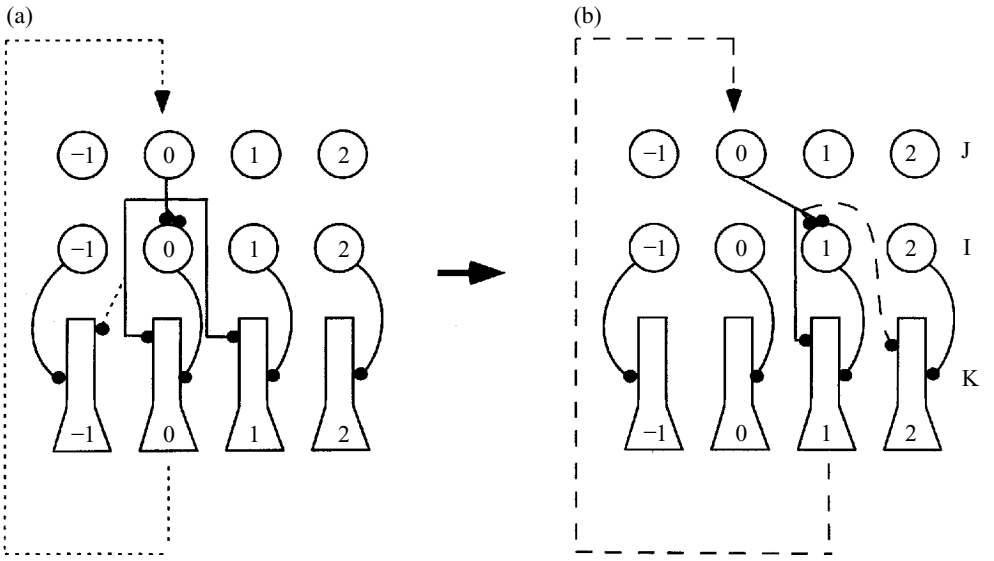

Figure 8. Generation of K-cell circuitry by offline recalibration. (*a*) Recapitulation of the circuitry sketched in figure 6*a*. If, despite the operation of this circuitry, an error does occur, with the formation of a 'mutant' synapse onto neuron I_1, and this new connection strengthens at the expense of the original connection (because the pattern of correlations that led to the original connection has changed), a different set of connections to and from layer K is required, if future errors in the strengthening of the new connection are to be prevented, as shown in (*b*). In order to reconfigure the K-layer connections from (*a*) to (*b*), the connections shown as dotted lines must be broken, and the connections shown as dashed lines must be created. This can be done in a two-stage process if the entire circuit is taken offline and suitable internally generated recalibration signals are played into the J-layer. First, individual J-cells (for example J_0) should be strongly activated while the J–I connections are all rendered implastic, and the J–K connections are rendered plastic (these global plasticity changes could be procured through release of suitable neuromodulators such as acetylcholine or norepinephrine). This strong J_0 activation should, in turn, cause all the I-cells that are its current targets (in this case, I_1) to fire. This could be achieved if a wave of 'burst' activity sweeps over the J-layer. This will strengthen the J_0–I_1 connection, which will create (via uncontrolled error) the desired new J_0–K_2 connection. The unsupported J_0–I_{-1} connection withers. In the second stage, the plasticity of K–J connections is selectively enabled, and J-cells are randomly activated (for example by random activation of cholinergic brainstem inputs). If the set of J-cells that innervates I_1 (in the case shown, just J_0) happens to fire, then K_1 will also fire (since it detects the relevant correlations). If (due to previous uncontrolled errors) K_1 makes a connection onto J_0, this connection will be appropriately strengthened, while the unsupported K_0–J_0 connection will wither.

the K-cells would be rendered plastic (they would be fixed when online). In particular, the required new J_0–K_2 connection could be created if cell J_0 fires selectively and powerfully so that it causes I_1 to fire. The conjoint firing of J_0 and I_1 would then cause K_1 to fire, strengthening the J_0–K_1 connection. Errors in the strengthening of this connection would then create the desired J_0–K_2 connection. The old, dotted J_0–K_{-1} connection is neither supported by correlated firing nor the beneficiary of errors, so it will disappear. The feedback connections from K to J would then be modified in a second phase of offline learning, via the correlated firing of J_0 and K_1 (Cox & Adams 2000). In both cases, the crucial neighbourhood relations underlying the circuitry are established by exploiting error, since a 'neighbour' is defined as a cell onto which, or from which, errant synapses form.

8. EVALUATION

The idea underlying these speculations is rather simple, though it has been little discussed. The creation of complex and precise neural networks requires that wiring errors be minimized, and that weight adjustments be anatomically specific. The size of any information-rich complex object, be it a genome, a neural network or a computer disk, is ultimately limited by the precision with which the information can be written. If new synapses created by strengthening existing connections do not inherit the original connectivity, they will impair network performance. It might be argued that if these new, erroneous, synapses are 'silent' (Isaac *et al.* 1995) they will not impair performance, while providing a reservoir of novel and potentially useful connections. However, since correlated activity across new silent synapses will cause unsilencing, network impairment cannot be avoided. In our terminology new connections are created as a result of errors in the strengthening of existing connections, but one could also regard them as coactivity-induced 'sprouts'. If, instead, sprouting occurred at some activity-independent basal rate S, then a similar equation to equation (3.5) would result, with the term ϕE replaced by S. Under these circumstances, the spread of synapses would depend on the difference, not the ratio, of the mesa and plateau fitnesses, each relative to S. Accurate connections could still be maintained by the circuitry of figure 6, except that K_0 would have to compute the differences in the absolute levels of correlations, rather than the ratio.

One way to ensure anatomically specific learning is to build better synapses, with improved insulation, greater separation and larger numbers of key molecules. However, some residual non-specificity is inevitable, and if neural correlations, deriving from subtle environmental regularities, are relatively weak, self-organization may fail entirely. Ultimately, the useful size of any complex system

is limited by the accuracy with which its components operate, and the neocortex, perhaps the most complex object known, is unlikely to be exempt. This suggests the possibility that the neocortex has developed some unusually effective way of minimizing the consequences of inevitable synaptic learning errors. It seems unlikely that neocortical synapses embody a new error-free design principle, and more probable that instead some of the unusual features of thalamocortical circuitry are involved.

Our viewpoint is orthogonal to traditional discussions of neocortical function, which naturally focus on the issue of how circuits can explain perception, memory, decision and behaviour, and how such circuits can be established. We propose that much circuitry is instead devoted to non-information processing tasks. In figures 6 and 7, 'information processing' is being done by one set of connections (from layer J to layer I), while the other three sets of connections are being used to prevent adverse consequences of synaptic learning errors. None of this extra circuitry would be needed if there were no errors in synaptic learning. It would not be needed either if mistakes could somehow be 'averaged out' over time; however, our analysis suggests that mistakes can only be averaged out if very strong correlations are present. The real world is sufficiently ambiguous, complex and noisy that efficient averaging is unlikely. Under favourable conditions (negligible error rates or strong correlation), our circuits default to a traditional 'wide-open' learning condition.

Our viewpoint is also orthogonal to the notion that somehow random wiring leads, magically, to efficient information processing (Braitenberg & Schuz 1991). Current evidence suggests, instead, that cortical wiring is very precise (Reid & Alonso 1995). In other situations, there are indications that wiring may be less precise, especially during development (Chen & Regehr 2000). For example, we have already alluded to the finding that while a typical LGN relay cell gets its major input from a single retinal ganglion cell, it also receives minor input from one or two other ganglion cells. It is not clear whether such convergence is 'deliberate' (so that LGN cells have non-relay functions) or 'accidental' (a result of imprecision in Hebbian wiring exacerbated by inevitable correlations in the firing of the 'major' and 'minor' inputs), though the finding that such overlap greatly decreases during development (Chen & Regehr 2000) favours the latter view.

The circuitry and physiology that we have proposed are detailed and complicated, and although they are largely consistent with the sketchy information that is available, it will be important to subject both to experimental tests (Elliott 2002). Is the logic itself faulty? We think there are two areas of potential weakness. First, we have not really justified the claim that learning errors can lead to complete failure of self-organization. If one could guarantee that such errors never 'run away', and at worse merely produce some controllable degradation of network performance, error avoidance would be less necessary. In particular, in the simple model presented in § 3, although errors render connections diffuse, the 'correct' neuron always receives the most synapses, and progressive addition of alternative target neurons never completely prevents selective wiring. Second, in our diagrams neurons are conveniently lined up in rows, so that neighbourhood relations are explicit. In reality, neighbourhood relations are less explicit, and presumably reflect the accidents of particular wiring histories. Nevertheless, some neurons are still more likely error targets than others, and in this sense, the neighbourhood concept is valid. The real problem is that the neighbourhood relations that are established for J–I connections must be echoed by J–K connections.

In summary, we propose that much thalamocortical circuitry exists to ensure accurate wiring in the face of anatomically imprecise learning rules combined with environmental complexity. Although our model is speculative and oversimplified, it focuses on issues that have been neglected, and which might be relevant to the assembly of extremely complex neural networks.

REFERENCES

Adams, P. R. 1998 Hebb and Darwin. *J. Theor. Biol.* **195**, 419–438.

Agmon-Snir, H., Carr, C. E. & Rinzel, J. 1998 The role of dendrites in auditory coincidence detection. *Nature* **393**, 268–272.

Bonhoeffer, T., Staiger, V. & Aertsen, A. 1994 Synaptic plasticity in rat hippocampal slice cultures: local Hebbian conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement. *Proc. Natl Acad. Sci. USA* **86**, 8113–8117.

Braitenberg, V. & Schuz, A. 1991 *Anatomy of the cortex*. Berlin: Springer.

Callaway, E. M. 1998 Local circuits in primary visual cortex of the macaque monkey. *A. Rev. Neurosci* **21**, 47–74.

Carlin, P. & Siekevitz, R. K. 1983 Plasticity in the central nervous system: do synapses divide? *Proc. Natl Acad. Sci. USA* **80**, 3517–3521.

Chen, C. & Regehr, W. G. 2000 Developmental remodeling of the retinogeniculate synapse. *Neuron* **28**, 955–966.

Colicos, M. A., Collins, B. E., Sailor, M. J. & Goda, Y. 2001 Remodeling of synaptic actin induced by photoconductive stimulation. *Cell* **107**, 605–616.

Cox, K. J. A. & Adams, P. R. 2000 Implications of synaptic digitisation and error for neocortical function. *Neurocomputing* **32–33**, 673–678.

Crick, F. H. C. 1984 Function of the thalamic reticular complex: the searchlight hypothesis. *Proc. Natl Acad. Sci. USA* **81**, 4586–4590.

Deneve, S., Latham, P. E. & Pouget, A. 1999 Reading population codes: a neural implementation of ideal observers. *Nature Neurosci.* **2**, 740–745.

Diamantaras, K. & Kung, S. Y. 1996 *Principal component neural networks: theory and applications (adaptive and learning systems for signal processing, communications, and control)*. New York: Wiley.

Dong, D. W. & Atick, J. J. 1995 Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network* **6**, 159–178.

Elliott, T. 2002 From synaptic errors to thalamocortical circuitry. *Trends Cogn. Sci.* **6**, 147–148.

Elliott, T., Howarth, C. I. & Shadbolt, N. R. 1996 Axonal processes and neural plasticity. 1. Ocular dominance columns. *Cerebral Cortex* **6**, 781–788.

Engert, F. & Bonhoeffer, T. 1997 Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388**, 279–284.

Fiala, J. C., Allwardt, B. & Harris, K. M. 2002 Dendritic spines do not split during hippocampal LTP or maturation. *Nature Neurosci.* **5**, 297–298.

Fraser, S. & Perkel, D. H. 1990 Competitive and positional cues in the patterning of nerve connections. *J. Neurobiol.* **21**, 51–72.

Grieve, K. L. & Sillito, A. M. 1995 Differential properties of cells in the feline primary visual cortex providing the corticofugal feedback to the lateral geniculate nucleus and visual claustrum. *J. Neurosci.* **15**, 4868–4874.

Guido, W., Lu, S.-M., Vaughan, D. W. & Sherman, S. M. 1995 Receiver operating characterstic (ROC) analysis of neurons in the cat's lateral geniculate nucleus during tonic and burst response mode. *Vis. Neurosci.* **12**, 723–741.

Hirsch, J. A., Gallagher, C. A., Alonso, J. M. & Martinez, L. M. 1998 Ascending projections of simple and complex cells in layer 6 of cat striate cortex. *J. Neurosc.* **18**, 8086–8094.

Isaac, J. T. R., Nicoll, R. A. & Malenka, R. 1995 Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434.

Koch, C. & Zador, A. 1993 The function of dendritic spines: devices subserving biochemical rather than electrical compartmentalization. *J. Neurosci.* **13**, 413–422.

Kossel, A., Bonhoeffer, T. & Bolz, J. 1990 Non-Hebbian synapses in rat visual cortex. *Neuroreport* **1**, 115–118.

Luscher, C., Nicoll, R. A., Malenka, R. C. & Muller, D. 2000 Synaptic plasticity and dynamic modulation of the postsynaptic membrane. *Nature Neurosci.* **3**, 545–550.

Miller, K. D. 1990 Correlation-based models of neural development. In *Neuroscience and connectionist theory* (ed. M. A. Gluck & D. E. Rumelhart), pp. 267–353. Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, K. D. 1994 A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity dependent competition between ON- and OFF-center inputs. *J. Neurosci.* **14**, 409–441.

Oja, E. 1982 A simplified neuron model as a principal component analyzer. *J. Math. Biol.* **15**, 267–273.

Petersen, C. C. H., Malenka, R. C., Nicoll, R. A. & Hopfield, J. J. 1998 All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl Acad. Sci. USA* **95**, 4732–4737.

Reid, R. C. & Alonso, J. M. 1995 Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* **380**, 281–284.

Schuman, E. M. & Madison, D. V. 1994 Locally distributed synaptic potentiation in the hippocampus. *Science* **263**, 532–536.

Sherman, S. M. & Guillery, R. W. 1996 The functional organisation of thalamocortical relays. *J. Neurophysiol.* **76**, 1295–1367.

Sherman, S. M. & Guillery, R. W. 2001 *Exploring the thalamus*. San Diego, CA: Academic.

Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. 1994 Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* **369**, 479–482.

Steriade, M. & McCarley, R. W. 1990 *Brainstem control of wakefulness and sleep*. New York: Plenum.

Sur, M. & Leamey, C. A. 2001 Development and plasticity of cortical areas and networks. *Nature Rev. Neurosci.* **2**, 251–262.

Swadlow, H. A. & Gusev, A. G. 2001 The impact of bursting thalamic impulses at a neocortical synapse. *Nature Neurosci.* **4**, 402–408.

Tarczy-Hornoch, K., Martin, K. A. C., Stratford, K. J. & Jack, J. J. B. 1999 Intracortical excitation of spiny neurons in layer 4 of cat striate cortex *in vitro*. *Cerebral Cortex* **9**, 833–843.

Toni, N., Buchs, P.-A., Nikomenko, I., Bron, C. R. & Muller, D. 1999 LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature* **402**, 421–425.

von der Malsburg, C. & Willshaw, D. J. 1980 Differential equations for the development of topological nerve fibre projections. *SIAM-AMS Proc.* **13**, 39–47.

Wang, F., Nemes, A., Mendelsohn, M. & Axel, R. 1998 Odorant receptors govern the formation of a precise topographic map. *Cell* **93**, 47–60.

Willshaw, D. J. & von der Malsburg, C. 1979 A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. *Phil. Trans. R. Soc. Lond.* B **287**, 203–243.

GLOSSARY

AMPA: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
LGN: lateral geniculate nucleus
LTP: long-term potentiation
NMDA: *N*-methyl-D-aspartate