COMMENTARY

THE THALAMOCORTICAL ALGORITHM

Paul R. Adams

Department of Neurobiology and Behavior
SUNY Stony Brook NY 11794

Phone: 516 632 6938
Email: padams@notes.cc.sunysb.edu

Key Words: Synaptic Error, Thalamus, Neocortex, Sleep

Abbreviations: LGN: lateral geniculate nucleus; LTP: Long Term Potentiation; BCM: Bienenstock-Monro-Cooper; ART: adaptive resonance theory; NMDA: N-methyl D-aspartate; REM: rapid eye movement; AMPA: alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate; R: receptor; Glu: glutamate. CA1,3 : Cornu Ammonis 1,3.

ABSTRACT: A hypothesis for the function of some of the most characteristic, and puzzling, circuitry of neocortex is presented. It is proposed that this circuitry is concerned, not with conventional information processing, but with assuring cortical wiring accuracy, and neural encoding efficiency. If specific wiring is generated by correlation-based mechanisms, then the accuracy of that wiring is set by (1) error rates for synapse placement, and (2) the sharpness of the correlations. There are thus 2 ways to improve efficiency: lower error rates or sharpen correlations. Error rates can be reduced by improving the biophysical machinery of synapses, up to a ceiling set by miniaturisation requirements. Correlations cannot be sharpened, since they depend on the structure of the real world and of previous stages of processing, but they can be measured, by a special type of cell. If these cells can suppress the plasticity of connections across which correlations are not sharp, great wiring accuracy can be achieved even in the presence of significant synapse placement errors. It is suggested that layer 6 corticothalamic feedback neurons measure correlation sharpnesses and control the plasticity of feedforward connections made by or received from individual relay or cortical cells. If feedforward cortical rewiring does occur as a result of daytime, online, learning, the corresponding layer 6 circuitry must be appropriately updated offline by a process resembling sleep.

CONTENTS

1. INTRODUCTION

In this Commentary I outline an interpretation of some of the most characteristic, and puzzling, machinery of the neocortex. I suggest that neurons in layer 6 measure correlations between neurons located in other, paired, successive, layers. These correlation signals are then translated into a format that regulates the plasticity of connections between the neurons of the appropriate pair of layers. This arrangement would provide local control of the learning rate of connections formed by, or onto, individual neurons.

A requirement to control the learning rate of individual sets of connections is developed within the framework of a simple general model of the process of creating new connections. It is proposed that new connections are formed by the anatomically erroneous strengthening of old connections. However, individual control of plasticity in continuously learning networks is likely to be important for other reasons too, and it is likely that the correlation signals that may be computed by layer 6 neurons could also be useful in a variety of other contexts.

In the first section I sketch a caricature of cortical anatomy and physiology, as background for the subsequent discussion. I then outline a model of the formation of new connections, and point out its limitations. The core proposal, that errors in synaptic strengthening can be minimised by using correlation signals to regulate plasticity, is then introduced, and shown to correspond to many previously puzzling features of neocortex. In the fourth section, a surprising consequence of this idea, a requirement for sleep, is discussed. Section 5 relates this account of sleep to the idea of hippocampal replay. A sixth section then applies these ideas to a few other aspects of neocortical function. The Commentary finishes with a list of difficulties and experimental tests.

2 A SKETCH OF THALAMUS AND NEOCORTEX

Neocortex is dauntingly complex and any attempt to summarise its main features is doomed to inadequacy, but despite the complexity it remains possible that the underlying principles are relatively straightforward, and that much of the rococo detail is merely elaboration on a basic plan. If there is such a neocortical plan, it has been very successful and has allowed tremendous expansion in the mammalian radiation. My sketch, summarised in Fig 1, is largely based on the cat striate cortex [30,38,42,54,55,100,112] but many of the features I describe are shared by other, possibly all, neocortical areas [92,131].

Neocortex is composed of 6 layers, classically summarised as (1) a mostly cell-free layer of horizontal axons (2) a layer of small pyramidal cells (3) a layer of medium pyramids (4) a layer rich in granule cells (5) a layer of large pyramids and (6) a layer of multiform cells. In addition 2 other structures, the dorsal thalamus and the claustrum, are so intimately related with all parts of neocortex that they could be considered as additional layers, layers 0 and layer 7. Almost all information reaching the neocortex from the rest of the brain arrives via the axons of thalamic "relay" cells. These thalamocortical inputs arrive in both middle layers of cortex, and in layers 1 and 6. To a first approximation the major relay population of sensory thalamic nuclei such as lateral geniculate nucleus (LGN) sends its main input to layer 4, with sidebranches to layer 6, while a minor relay cell population synapses in layer 1.

The classical concept of the "relay" function of thalamus, though retained in the naming of the principle, cortically-projecting, cells, has been recently modified in 2 important ways. First, there is some evidence [34,114] that relay cells do not merely faithfully transmit the information they receive, whether from subcortical sites such as retina (in the case of "first order" nuclei such as LGN) or from neocortex itself (in the case of "higher-order" nuclei[110] such as pulvinar). However, such adjustment of the spatiotemporal input-output relations of relay cells tends to be minor, and the anatomical evidence, that each relay cell receives a single main "driver" input [110], squares best with the traditional "relay" picture. Although these adjustments may be important, particularly since they are may involve neocortical feedback, it is likely that their nature and extent is specific to the particular function being computed by the cortical area to which the relay cell projects, and therefore not susceptible to a unifying general explanation.

The second modification of the traditional "relay" picture stems from the seminal discovery that relay cells exhibit 2 distinct firing modes[58]. In the "burst" mode, which occurs when brief excitation is imposed on a relatively hyperpolarised neuron, a low-threshold calcium spike occurs which then triggers a short burst of sodium action potentials. The channels underlying the low threshold spike inactivate during the spike, and require around 100 milliseconds to recover following the return to resting potential [50,134]. In the "tonic" mode the neuron starts out more depolarised, so the low threshold calcium channels are already inactivated, and brief excitation directly triggers individual sodium action potentials. However, in both modes only the sodium spikes, and the information contained in their temporal pattern, reaches neocortex. In both modes the brief, triggering, excitation is normally due to action potentials arriving over the principle "driving" input axons[113]. Thus information relayed to cortex arrives in 1 of 2 slightly different packages, labelled "burst" or "tonic". Inevitably, because relay cells in burst mode cannot follow rapidly varying inputs very precisely (since recovery from low threshold inactivation is rather slow), there is some degradation of the temporal fidelity of information transfer, but it seems unlikely that burst mode acts merely as a low pass filter. Although burst mode tends to filter out rapidly varying information, it can also relay information more accurately in the presence of noise[43], and the overall rates of information transfer in burst and tonic mode are rather similar[101]. The arrangement rather resembles the contrast control on a photocopier. The "tonic" mode corresponds to a low contrast setting which can preserve the full gray-scale content of the image. The "burst" mode is a high contrast, black-or-white setting in which each pixel is either "on" or "off". Intuitively one might suppose that the grey mode, in which each pixel is represented at several bits of resolution, will always be superior to the black/white mode, with one bit. However, if the resulting image is to be transmitted along a noisy channel, the black/white setting may be best, since the resulting washed-out output image can then be restored to the original simply by setting all low pixels to black and all high pixels to white. The threshold for the black/white encoding has to be correctly set. Murray Sherman and his colleagues have proposed that the burst mode is best suited to stimulus detection and the tonic mode to signal analysis [43,111,110], a concept closely related to this distinction between black/white and gray modes.

The problem the brain faces is that it does not know, *a priori*, whether a complex, and confusing, stream of spikes arriving in cortex represents a familiar object degraded by noise, or an unfamiliar scene, and therefore does not know which setting, grey or black/white, to use. Is there a simple strategy, based on the immediate computation performed by the area of cortex to which a relay cell projects, for appropriate adjustments to the burst/tonic setting? This point will be taken up in Section 6.

In the mean time it can be noted that the cortical zone to which a relay cell projects can indeed regulate the burst/tonic setting of that relay cell, because layer 6 neurons in that zone typically feed back to the distal dendrites of that relay cell, where they release glutamate, which via a metabotropic receptor lowers the resting potassium conductance of the target relay cell, depolarising it and shifting it to tonic mode [90,111]. The driver inputs are instead located on the proximal dendrites of the relay cell, where they are best suited to supplying rapid, strong, ionotropic, excitation[110]. The layer 6 feedback inputs can also activate ionotropic receptors[105], but it is unclear whether, or when, such input can drive the relay cell to fire.

Whatever the detailed interpretation of the significance of the burst/tonic transition, it seems plausible that it is involved more in "packaging" or "formatting" information than in actually changing that information. In the simplest view, it is only the mean rate of a neuron's sodium spikes that conveys information about the driving input to that neuron[70,108,109]. Higher-order temporal statistics of spike sequences, such as "burstiness", could then be used to multiplex an additional signal, not derived from the driver input, which instructs the recipient of the spike train how to deal with the arriving information. It will be argued below that the "recipient" here is not the cell postsynaptic to the relay cell neocortical terminals (which cannot easily read out the burst label, especially since it is also receiving inputs from numerous other relay cells), but the terminals themselves. One obvious possibility is that the terminals respond to the bursts more reliably[75] than to single spikes. Reliability will affect both the effectiveness of the synaptic action (more vesicles released) and the noisiness of the action (less variance in release). However, it is also possible that other, more subtle and vital aspects of synaptic action are also affected, such as the longer term effectiveness of that terminal (its synaptic strength), or the ability

of that terminal to modify its effectiveness in response to contingencies like correlated firing (its Hebbian[49] plasticity).

Returning now to the main sequence of information processing in cortex, and hewing to a traditional simple relay view of the thalamus, it appears that the convergence of a small group of relay cell axons on specific layer 4, granule or spiny stellate cells, is the crucial first step, as originally surmised by Hubel and Wiesel[54]. These "simple" cortical cells thus do not relay information – they compute. From an abstract point of view such a cell is a generic "connectionist" neuron – it computes the weighted sum of its synaptic inputs[4]. This sum is the "dot product" of the input vector (the pattern of light on the retina) and the weight vector (the pattern of synaptic strengths between the retina – via the thalamus – and the cortex). If the inputs and outputs are appropriately normalised by divisive synaptic inhibition, this sum is the cosine of the angle between these vectors, which is a natural measure of their similarity. More concretely, the simple cell acts as a feature detector, responding most vigorously if the input pattern contains the oriented short bar to which it is tuned. The simple cell of striate cortex is tuned because it receives input from a set of LGN cells which are in turn wired to a set of colinear ganglion cells. In this case the feedforward weights are either large or absent.

The Hubel-Wiesel view has been controversial[38], mainly because it does not account for the fact that the great majority of synapses on spiny stellate cells are not of geniculate origin[28]. It has therefore been tempting to suggest that the intracortically-originating synapses are vital for orientation tuning [29,96]. However, at least qualitatively, there is now overwhelming evidence for the Hubel-Wiesel arrangement [22,25,37,38,99,100].

The main intracortical sources of excitatory synapses onto spiny stellate cells are (1) recurrent connections onto spines from other spiny stellate cells and (2) collaterals of layer 6 pyramidal cells, which ascend into layer 4 and contact dendritic shafts[28]. Several authors have suggested that the recurrent excitatory connections sharpen the orientation tuning by an "attractor" mechanism [9,46,115]. Imagine a ring of simple cells around which "bare" orientation preference, generated by Hubel-Wiesel type feedforward connections, systematically rotates. If neighboring cells on the ring, with similar orientation preference, mutually excite each other, they will tend to amplify any initial excitation bias (representing the effect of a weakly oriented stimulus). Such mutual excitation is

unstable, and must be countered by more broadly spreading inhibition. The tendency of mutually coupled cells to reinforce each others' firing underlies the "cell assembly" concept of Hebb[49], and also the attractor network proposed by Hopfield as an associative memory mechanism[53]. In the latter case the mutually reinforcing groups become coupled together as a result of exposure to input patterns, which strengthen synapses by a Hebbian mechanism. After learning has created the groups, degraded versions of the memorised patterns can be selectively "amplified" until they match the stored prototypes. In the case of the "line attractor" mechanism for orientation sharpening, the "prototypes" are essentially the set of possible oriented bars. The input is treated by the network as a degraded, noisy version of the prototype that it most resembles. Initially therefore a broad set of neurons on the ring fires, but gradually the activity narrows to that set of neurons whose preference best matches the stimulus. The term "line attractor" arises because the stable states of mutually-reinforcing activity correspond to a set of connected points on the ring.

Clearly the line attractor mechanism works, could be useful, and does correspond roughly to the observed anatomy. However, it is a double edged weapon[17]. It can rapidly produce good estimates of orientation (or of any other parameter to which the network is systematically sensitive), but it can generate false estimates to ambiguous stimuli, or even hallucinations (since it is built to interpret any input as conforming to its own simplistic view of the world). It is likely therefore that it is used in a controlled way by the cortex, somewhat in the way that the different advantages of burst or tonic encoding by thalamus are balanced. It is therefore rather intriguing that the same layer 6 signals that determine the burst/tonic setting are also sent, via collaterals, to layer 4 cells [28,51,133]. The terminals formed by these layer 6 inputs in both cases are of an unusual "drumstick" type[15].

There is an obvious circumstance under which the line attractor mechanism would be particularly useful in forming rapid preliminary estimates of local orientation (or other parameters). If cortical neurons use a rate code, and if (at least as viewed by the decoding neuron) arriving individual spikes have arbitrary timing, then chance fluctuations in spike arrival will corrupt estimates of position, orientation etc. Thus if the receiving neuron only has access to a short sequence of Poisson spikes it can only form a noisy estimate of the underlying rate. This mean that a simple cell that responds to the onset of an oriented

bar will initially form a noisy estimate of that bar's degree of match to its preferred orientation. The line attractor mechanism can rapidly clean up that initial estimate, essentially by pooling information from other simple cells, whose relative lack of firing is informative. Combatting this noisiness of a rate code may be the main use of the line attractor mechanism.

If the bar persists, then the simple cell can improve its estimate of the bar's orientation by temporal averaging over the bare, feedforward connections. Thus the line attractor mechanism could be disengaged (for example by lowering the gain of the recurrent synapses). It would be now be less susceptible to the adverse effects of the line attractor process. However, this disconnection strategy suffers from the drawback already noted for cortical control of the burst/tonic transition to optimise information transfer- the cortex has to know what it is seeing in order to intelligently set the gain of the line attractor mechanism. If it is actually witnessing an unpredictably rotating or moving bar, and not a fixed bar corrupted by noise, it should not disengage the line attractor process. Once again the cortex seems to need a simple, immediately calculable, local heuristic to regulate its own internal information-processing strategies. In both cases the cortex must develop a rough on-line measure of how well it is doing – how much information an output layer is preserving about its input. Crudely speaking, it needs to know how well correlated the output is with the input, using some simple pairwise measure of that correlation. This point will be further developed in Section 5. For the moment, we will just note that these 2 on-line decisions (setting the burst/tonic transition and the gain of the recurrent synapses) seem to require rather similar signals, and that layer 6 neurons seem to have the appropriate wiring.

Once local orientation has been computed by simple cells, primarily spiny stellate cells in layer 4 of cat striate cortex, the classic Hubel-Wiesel view is that it is generalised to a slightly less local orientation estimate in the complex cells of layer 2 and 3. Their suggestion[54] that this is once again achieved by feedforward convergence of several simple cells responding to the same orientation in a small patch of visual space has received some recent support [2]. Hubel and Wiesel imagined that the basic operation of a complex cell was a logical "OR", as opposed to the logical "AND" of the simple cells, achieved by linearly summing and then thresholding the convergent inputs. The main

problem with simple summing and thresholding is ambiguity between cases where a few simple cells are firing strongly and many are firing weakly, rather like the problem with the contrast invariance of orientation tuning itself. At bottom this is again the problem of vector normalisation faced by the generic connectionist neuron, and one possible solution is divisive inhibition, possibly in a push-pull arrangement[37]. Koch [68] has pointed out that simple shunting inhibition is not divisive in spiking neurons, but recurrent networks can implement such division, and have been proposed as substrates of the "OR" operation. More abstractly, Riesenhuber and Poggio[102] suggest that instead of a thresholded sum, the complex cells compute a "MAX" operation – they locate their simple input which is firing fastest, and fire at that rate.

Their paper is also of interest because it explicitly advances a viewpoint that was implicit from the earliest discoveries of Hubel and Wiesel – that the simple/complex strategy of striate cortex is the basic plan of all neocortex. Thus each area (or subarea) would compute a simple feature of its input, arrived at by systematically combining groups of incoming axon signals, and then generalise that over "space". Thus striate cortex would combine position information to extract orientation, and then combine local orientation signals to generalise over position, and generic cortex would combine X information to extract a parameter Y, and then generalise over X.

The layer 2/3 complex cells then pass their estimate of local orientation to 2 main targets : middle layer cells in "higher" cortical areas, and layer 5 pyramidal cells. The layer 5 cells are also complex, and it is not clear what is computed in this step. Layer 5 cells are traditionally regarded as the output cells of cortex. Thus in striate cortex many layer 5 cells project to the superior colliculus, a structure that is particularly involved in controlling eye movements (and which also receives direct retinal signals). A useful, though simplistic, view is provided by theoretical "backpropagation" neural networks, in which a hidden layer is followed by an output layer [4,48,104]. The synaptic weights of such a network are trained by comparing the actual outputs to the desired outputs. The resulting local error signals are sent to the hidden layer after weighting by "backpropagation" through the hidden-to-output synapses. If such supervised learning occurs in all cortical networks, then each cortical area should have the possibility of undergoing separate training, since otherwise the "credit-assignment" problem posed by

backpropagation through numerous cascaded cortical areas becomes overwhelming. Naively, striate cortex would undergo supervised learning for a simple task like appropriate eye movement to structured parts of the visual scene. Such programming would have to occur before the brain had any representation corresponding to "objects" or more complex ideas. The representations thus developed in the "hidden layer" neurons (actually , layers 4, 2 and 3), as a by-product, would be useful in further cortical operations, such as object recognition (for example, following the Riesenhuber-Poggio scheme).

The axons of layer 5 neurons, on their way to subcortical sites, send branches to "higher order" thalamic nuclei, such as pulvinar[110]. These layer 5-originating terminals are quite different from the "drumstick" terminals originating from layer 6 [45]. They closely resemble driver afferent terminals, such as those to LGN from retina, or to the ventrolateral nucleus from cerebellum. The puzzle set by these arrangements is not so much that an area of cortex seems to need to know what other areas of cortex are doing, but that information from layer 5 has to be sent via thalamus, while information from layer 2/3 can be sent directly. If the thalamus' main job is to apply the burst/tonic label to information reaching cortex, why does the layer 2/3 information not need the label?

One additional general puzzle concerns the anatomy of the excitatory cells in these various layers. Conventional pyramidal cells, by definition, carry a stout apical dendrite. Conventionally the apical dendrites of layer 2, 3 and 5 cells end in an apical tuft in layer 1, where they receive inputs from the horizontally running terminals of axons of (1) certain relay cells (2) deep layer cells in higher cortex. Most of the pyramidal cells of layer 6 do not form apical tufts in layer 1 (those that do send their axons to the claustrum, not the thalamus[64]). Obviously the spiny stellate cells have neither an apical dendrite nor a layer 1 tuft. Although these arrangements are far from universal, they are characteristic. Perhaps synaptic inputs onto apical dendrites, which are subject to the modulatory infuence of layer 1, differ in some general way from inputs onto basal dendrites. One appealing possibility, partly supported by detailed reconstructions[128], is that thalamocortical inputs to middle cortex, when they do not impinge directly onto spiny stellate cells, are made on basal dendrites of pyramidal cells, while feedforward intracortical or corticocortical (e.g. layer 2/3 to layer 5 or to higher cortex) are made on

apical dendrites. According to this view spiny stellate cells would simply not need apical dendrites, because their primary driving input would be from thalamus. Also, the "packaging" provided by the burst/tonic transition of thalamic relay cells would instead be provided by the layer 1 inputs, in the case of intra- or cortico-cortical feedforward connections.

The above sketch omits the various classes of inhibitory neurons. Perhaps these can be viewed as performing various vital housekeeping roles, along the lines of "vector normalisation", rather than carrying the main burden of detailed computation. From the perspective developed below, the smoothness of inhibitory neurons, and their mostly shaft or somatic targets, would be accompanied by some lack of precision in their wiring which would be compatible with housekeeping functions.

3 WIRING THE CORTEX

The above sketch of cortex, itself already a caricature, boils down to layer-to-layer feedforward processing, supplemented by a variety of circuits of plausible (e.g. recurrent excitation; inhibition) or obscure (corticothalamic feedback, layer 1, segregation between apical/basal dendrites, burst/tonic transition) function. All these circuits have to be wired up with considerable accuracy for the cortex to work. Possibly wiring must be done with neuron-to-neuron precision, especially in the case of the fundamental feedforward operation. Since neural tissue and function is extremely costly, it is unlikely that the wiring is sloppy, since it would be possible to replace a large, sloppily-wired, network by a much smaller exactly wired network, even supposing the former could work at all.
The core thesis of this Commentary is that the precision of neural wiring is of at least as much importance as the precision of other neural operations, such as synaptic transmission, spike generation, and encoding and computation, and that many features of the brain, and especially of cortex, stem from the need for exact wiring.
In order to discuss this issue, it is essential to construct a model which allows quantitative analysis of wiring precision. The model I present below is probably the simplest that allows such a discussion. It is basically a minor extension of the conventional Hebb

synapse, in which connection strengths are set by the accumulated past history of correlated firing across that connection[49]. I will not attempt to explain why useful networks can be wired by correlated neural activity, nor even to define such correlations. This well-worked issue is central to both neurophysiology and neural theory [4,48,104]. I will simply assume that when 2 connected neurons undergo correlated firing, and if the connection is "plastic", there is the possibility of an increase in strength not only between the connected neurons, but also between one of the connected cells and neighbors of the other connected cell (Fig 2). This closely resembles "volume learning" observed between CA3 and CA1 hippocampal neurons [13,33,94,107], and could arise in densely-packed neuropil by spread of neurotransmitters or second messengers. However, the erroneous strengthening of nearby connections could occur at either the functional or structural stages, with the latter being most likely, since it might involve withdrawal of glia which normally enhance compartmentalisation.

This assumption can be spelled out more formally as follows. There are 2 possibilities – that a presynaptic neuron can make errors in forming synapses onto a set of postsynaptic cells, or that a postsynaptic cell can make errors in receiving synapses from a set of presynaptic cells. The first situation might arise if the primary signals initiating synaptogenesis are presynaptic, and the second if they are postsynaptic. Let us call the first case presynaptic plasticity and the second, postsynaptic plasticity. (Real connections may evince both types). The following discussion is based on presynaptic plasticity, but identical arguments apply to postsynaptic plasticity [1]. First, let us distinguish between a connection between 2 specific cells, comprised of y synapses, and the total set of connections made by a presynaptic cell to a homologous set of postsynaptic cells, comprised of all the synapses that presynaptic cell makes, its "power'. Suppose a presynaptic cell is firing at a rate V_j and a postsynaptic cell at a rate V_i. Assume the growth of the connection's strength y is governed by a Hebb-type [7,49,67] rule:

$$dy/dt = k\, V_j V_i \qquad (1)$$

where k is a "learning constant" that is specific to that connection. In the absence of any other input to the cell we assume that $V_i = gyV_j$, where g is a connection-specific synapse effectiveness. Writing $w = kgV_j^2$ we have

$$dy/dt = w\,y. \qquad (2)$$

This equation implies that the connection increases exponentially in strength, because as it grows the firing rate of the postsynaptic cell induced by the steady activity of the presynaptic cell increases, in turn increasing the connection strength, in an autocatalytic manner. We could derive exactly the same equation by assuming that the connection is comprised of y synapses, each of which can functionally double with probability w. If "doubling" is caused by the correlated firing of the pre- and post-synaptic cells, then w can be viewed as the "correlation" between the cells. Petersen et al recently studied quantitatively the strengthening that occurs between CA3 and CA1 cell pairs linked by single synapses[97]. They found that paired activity does indeed lead to "doubling", with a probability that depends on the degree of correlation. However, Eq 2 implies that after initial doubling identical pairing would again produce further doubling, with similar probability. However, the authors found instead that no further change in strength occurred. This is not surprising, since it is well known that after "early" LTP, synapses become refractory to further pairing. Only after several hours have passed does the connection become again susceptible to LTP, and during this period protein synthesis must occur[39]. The existence of refractoriness has made it impossible to test quantitatively rules governing Hebb synapses. Eq 2 predicts that after complete recovery from refractoriness a further pairing episode would add 2 more units of synaptic strength, but it is quite possible that only one unit of strength would be added, so that the rule followed would be

$$dy/dt = w. \qquad (3)$$

If w is again interpreted as the degree of correlation across the connection, then Eq (3) would be a more conventional Hebb rule. However, because as the connection

strengthens the pair would become more correlated, the outcome would again be autocatalytic, and similar to Eq 2.

If correlated firing leads to synaptic doubling, as reported by Petersen et al, one can view synapse strengthening as a replication process, especially since the added functional synapse usually connects the same cells as the initial synapse. However, the initial doubling is merely functional. True replication would require the appearance of a complete new synapse, either by splitting of the original synapse (refs 12,18,40,124 but see 116) , or sprouting[24] of filopodia [32,80], which would become spines, and elaboration of a new active zone. It is important to emphasise that the present model assumes that cortical learning is a slow accumulative process in which episodes of correlation and functional doubling are well spaced to allow refractoriness to completely wane.

If neural correlations produce doubling, and doubling then increases these correlations, then the combined effect will be hyperbolic, not exponential growth. As discussed by von der Malsburg and Willshaw[83], in this regime connections tend to lock in at values dictated by random initial fluctuations, and be irresponsive to the statistical patterning of their inputs. This "double whammy" drawback of combining both correlation-induced replication and firing-induced correlation could be avoided if after each bout of functional doubling, only a conditional form of structural doubling were allowed: structurally doubled synapse would be automatically briefly halved in strength whenever that connection was plastic. Conceivably one of the purposes of tonic firing is to lower the strength of plastic synapses.

Popular models of neocortical wiring always contain the element of autocatalysis, captured by Eq 2. These models are of course concerned with particular, detailed information-processing networks, such as ocular dominance, topographical maps, or orientation selectivity [81,82,91,31,129,130]. However, here we are only concerned with connection growth, not with its usefulness, and Eq 2 forms a useful starting point. Because w is a replication rate, it is analogous to "fitness" in population dynamics[47]. However, it is also related to "degree of correlation", since synaptic doubling is produced by the near coincident firing of connected cells.

Eq 2 predicts that a connection will grow without bound. This undesirable feature can be removed by assuming that as the connection grows it becomes more difficult to add

further synapses, until growth stops when some maximum number of synapses y_t is reached:-

$$dy_i/dt = w\, y_i(1-y_i/y_t) \qquad (4)$$

We can generalise this to the case where the presynaptic neuron can make synapses on any of a set of postsynaptic neurons, indexed i. y_t is then the maximum power of the presynaptic cell. If we assume that the presynaptic cell has already reached its maximum power, then the equation becomes[83]

$$dy_i/dt = (w_i - \langle w \rangle)\, y_i \qquad (5)$$

where $\langle w \rangle$ is the average fitness of a connection defined by

$$\langle w \rangle = \sum w.y_i \; / \, y_t \qquad (6)$$

Note that now the fitness of a connection is defined relative to its brethrens' – there is competition. This equation, which describes the correlation-induced rearrangement of a fixed number of synapses, has the interesting consequence that all the synapses must end up on the fittest neuron – survival of the fittest. Even if a postsynaptic neuron is only slightly more correlated with the presynaptic neuron than are the others, it will eventually win all the synapses. This is a strong amplification mechanism, which is able to detect small peaks in correlation, just as the line attractor mechanism can pick out the neuron receiving a slight hump of excitation. Even if all the neurons have equal fitness, the neuron that starts out with the most synapses will end up with them all. In any practical situation even a completely flat distribution of synapses and fitnesses will undergo slight fluctuations, like a pencil balanced on its point, and one neuron will win. The cellular mechanism of the competition need not be defined. One possibility is that synapses simply compete for space or a growth factor. Another is that after a period of growth the total synaptic power of a neuron is measured offline (for example by playing in a calibration signal and measuring the combined output of all the postsynaptic cells it

connects to). A third is that a special-purpose correlation-measurement neuron computes the total synaptic power online, and continuously adjusts the firing levels of plastic neurons so that connections are effectively competing directly for correlation. A BCM connection, in which a feedback mechanism adjust a threshold at which learning flips from hebbian to antihebbian, could also work [7,10,127]. Possibly all of these mechanisms operate together, on different time scales and with different degrees of precision.

The amplifying but competitive nature of Hebb synapses lies at the heart of theories of neural self-organisation[82]. Typically such simulations start with roughly uniform synaptic weights. In the absence of imposed correlations, small initial differences will be amplified and locked-in. However, if the imposed neural activity is strong enough, it will dominate the selection process, and the final pattern of connections will reflect the statistical properties of the inputs, subject to additional cooperative constraints produced by lateral connections or local growth factor uptake.

We now introduce the possibility of heterosynaptic error. The simplest way to represent this is to imagine that when a new synapse is created as a result of correlated firing, it is placed at the correlated connection with a probability $(1\text{-}E)$ which is slightly less than one, and on the neurons neighboring the correlated connection with probability E. E is the (heterosynaptic) error rate for synapse placement. To make this more concrete, imagine that the postsynaptic neurons are laid out in a row, so each neuron has 2 neighbors, either of which will receive the extra synapse with probability $E/2$ (Fig 1). We can represent this as

$$dy_i/dt = (1\text{-}E)\,(w_i - \langle w \rangle)\,y_i + (E/2)\,(w_{i-1} - \langle w \rangle)\,y_{i-1} + (E/2)\,(w_{i+1} - \langle w \rangle)\,y_{i+1} \qquad (7)$$

To procede, it is convenient to move to a continuous model, in which x represents position along the line, and $y(x)$ is the density of synapses at the point x. The dependence of $y(x)$ and $w(x)$ on x will not be explicitly written.

$$dy/dt = (w - \langle w \rangle)\,y + w\,[E/2]\,d^2y/dx^2 \qquad (8)$$

where now $\langle w \rangle$ is defined by

$$<w> = \int wy\, dx / \int y\, dx \qquad (9)$$

The integrals run over the whole row of cells. The meanings of the 2 terms on the right hand side of Eq 8 are as follows. The first term expresses the growth (or decay) of connections which are stronger (or weaker) than average. By itself, it generates the winner-take-all behavior noted above. The second term represents the erroneous left or right placement of new synapses as a diffusion process[68]. By itself it gradually smears out (at a rate that depends on w and E) existing profiles of synapses, without altering their number. The equation describes the redistribution of synapses under the combination of correlation-induced growth, error and competition.

The pair of equations (8) and (9) are nonlinear and difficult to solve. However, in the steady state there is some fixed profile of synapses which means that $<w>$ is a constant, and Eq 8 reduces to

$$d^2y/dx^2 = 2(<w> - w)\, y/\, w\, E \quad . \qquad (10)$$

In order to solve this equation the boundary conditions must be specified. A particularly simple and useful case is where there is a central mesa of high fitness connections ($w = w_m$) surrounded on both sides by a plateau of lower fitness neurons ($w = w_p$), as depicted in Fig 2. Under these conditions inside the mesa w_m must be greater than $<w>$ so

$$y_m = A \sin x/\lambda + B \cos x/\lambda \qquad (11)$$

while outside the mesa w_p is less than $<w>$ and thus

$$y_p = C \exp x/\lambda + D \exp - x/\lambda \qquad (12)$$

This equations satisfy Eq 10 inside and outside the mesa respectively. In both cases λ is a length constant given by

$$\lambda^2 = (+/-)\, wE/2\,(w - \langle w \rangle) \qquad (13)$$

where $w = w_m$ (+) or w_p (-) as appropriate.

Let us now assume that both the length of the mesa and λ are very short compared to the whole line. Since there can be no synapses at the ends of the line, C must be zero.

Thus

$$y_p = D \exp - x/\lambda \quad (14)$$

It is now straightforward to evaluate λ, using (14), (13) and (9) and performing the appropriate integrations.

The result is (writing n for the number of neurons in the mesa)

$$N\lambda^2/(2\lambda + n) = E/2(w_m/w_p - 1) \quad (15)$$

The meaning of this math is as follows. Inside the mesa synapses are being formed (because all these connections are fitter than average). However, because of heterosynaptic error there is a flux of synapses across the mesa/plateau borders. In the plateau regions synapses are being destroyed (because all these connections are below average fitness). In the steady state the flux of synapses across the border exactly balances the excess production within the mesa, and the underproduction in the plateaus. Furthermore, if the plateaus are sufficiently long, and the mesa sufficiently short, the synapse distribution is flat within the mesa, with exponential tails (length constant λ) extending into the plateaus.

The significance of this result is that it explicitly relates the accuracy of neuronal wiring to the sharpness of neuronal correlations (w_m/w_p) and to heterosynaptic error rates. λ measures to within how many neurons a network is correctly wired (assuming that perfect wiring corresponds to all the synapses being located at the peak of the neural correlations). The factor 2 arises because in the line model each neuron has only 2 neighbors.

So far we have implied that the spread of synapses beyond the mesa is deleterious. Indeed, it will inevitably degrade the precision of maps of ocular dominance, position, orientation etc established by Hebbian learning, perhaps to unacceptable levels. However, the cloud of synapses that extends beyond the high correlation region is also useful in updating the maps, for example to allow for neuron death, is necessary. In the absence of error, once a connection has been eliminated, there is no way to restore it. The error cloud is a reserve pool of weak connections that is available in case the statistical structure of the outside world, or of preceding layers of neural analysis, change. Furthermore, if the network continues to learn, new erroneous synapses, extending beyond the error fringe, can be generated, and these may flourish in a new statistical input regime.

Such flexibility is only achieved in conventional Hebbian models by assuming permanent, fixed and complete anatomical connectivity (perhaps restricted to an arbor range) even though individual connections may have functionally zero strength. In the extreme case this would imply that, if every cell in the brain were connected to every other cell by an axon 1 cm long and 1 um wide, the brain would be 100,000 km wide and spikes would have to travel at the speed of light. In the more realistic case of 10^6 primate lgn cells projecting to 0.5×10^9 granule cells [11,85], the volume of wires would exceed that of the entire body. Of course one can always sidestep the problem by postulating that genetic mechanisms prewire the brain at low resolution to within an arbor radius, and that error-free Hebbian mechanisms then refine the coarse maps. Indeed, this is the assumption usually made, with some acknowledgement that supplementary sprouting mechanisms may also occur[83]. However, this simply pushes the problem of wiring the brain into the lap of gene-based Darwinian evolution, and the whole point of having a brain is to be able to adapt more swiftly to the environment than genes allow. Creation of connections by erroneous placement of new synapses provides a sort of virtual full connectivity, without an intolerable burden of silent wires, but at a cost of lowered performance and slower learning (these 2 factors being reciprocally related). It is likely that all parts of neocortex show some degree of adult learning [5,62,61], and that its unusual flexibility and computational power stems from an ability to form new connections.
It is of course possible that the overall formulation adopted above, that new connections are created by errors in the strengthening of existing connections, is wrong. For example,

new connections could be formed by continuous background sprouting. But if these connections are not pruned back by weaker than average correlations, they will simply act as a spreading fungus that gradually degrades performance. If they are pruned back by antiHebbian learning, the final result will be as described above.

Even though the error fringe can be useful in new contexts, it degrades current performance and must be kept under control. It is possible that the accuracy of neural wiring is an important factor in setting the size of neural networks, just as the accuracy with which silicon chips can be fabricated sets the power of computers. This would provide an explanation for the existence of spines, which are designed to compartmentalise chemical signals, not just neurotransmitters, but also second messengers like calcium that are crucial for synaptic strengthening[69]. The argument that spines confine calcium to the synapse where the Hebbian signal originates is really just an argument that error rates must be low. One reason why the neocortex might be particularly susceptible to the adverse consequences of erroneous wiring is that it seems to require numerous successive layer to layer operations. Consider the problem of preserving a topographic representation (identity mapping) through n successive layers. The dispersion in the final layer will be n times the initial dispersion. In a ruminating brain, where the signals may make thousands of passages, they will soon be hopelessly scrambled.

The above model (e.g. Eq 10) is linear and synaptic smearing just gets gradually worse as errors increase or correlations weaken (Eq 15). In a nonlinear model this degradation will not be so graceful, and it is likely that if connections smear beyond some critical length constant λ_c networks will fail to self organise at all. Any network that performs a useful function must compute, and have nonlinearities.

For all these reasons I assume that controlling heterosynaptic strengthening errors is vital for the correct function of large networks such as neocortex. I further assume that connections are allowed to smear as much as is compatible with adequate network performance (so as to achieve maximum flexibility) but no further. Neocortical networks should operate close to λ_c.

4 ERROR CONTROL BY THALAMOCORTICAL CIRCUITRY

If it is true that containment of heterosynaptic error has been paramount in brain evolution, then it appears that neocortex has been able to grow massively, not by making more efficient synapses (since neocortical synapses are very similar to those in hippocampus, olfactory cortex etc) but by inventing some other trick. There must be limits to the accuracy of synapses, set by their molecular composition and design, their size and their density. Neither the size nor spacing of synapses can be greatly increased without impoverishment of computational power. Spine heads are already so small, containing typically just one free calcium ion[69], that accidental strengthenings, triggered by Poissonian fluctuations in resting calcium levels, must be quite frequent. These provide an irreducible background fitness on which neural correlations are superimposed. Furthermore, spurious correlations are inevitable in neurons using rate codes, because of Poissonian spike fluctuations. One reason why hippocampal expansion has not kept pace with that of neocortex[59] may be that its size is limited by the error rates of spiny synapses. (Cerebellar expansion has not been so limited, but here the anatomy of the granule cell/ parallel fiber system provides a natural "arbor function" restricting the number of possible postsynaptic cells).

At first glance Eq 15 appears to imply that the only other way to improve the accuracy of connections is to increase w_m/w_p – the sharpness of correlations. However, these correlations are a given, reflecting the statistics of the external world and of previous layers of processing. While they may slowly become sharper as a result of weight adjustment and rewiring in previous layers, this does not solve the immediate problem. Although neocortex cannot, in the short term, control correlation sharpness, it can measure it, since it has in principle access to the required signals. Then if E exceeds $2(w_m/w_p-1)$ by a factor that would produce an unacceptably large spread of synapses ($\lambda > \lambda_c$, where λ_c is a critical spread value), the neuron making or receiving the relevent connections can be made "implastic" – unable to respond to any degree of correlation across those connections with strengthening or weakening. At first glance this requires measuring all possible pairwise correlations, of V_j with V_i. This is even less feasible than wiring all pre- and postsynaptic neurons together, since it requires an immense number

(N^2) not of axon segments but of whole neurons, in a supplementary correlation-measurement layer. However, because individual synapses are not infinitely weak, and the fringes of erroneous synapses in the plateau region do not extend very far, the problem is rendered tractable – only correlations between currently connected cells, and their immediate neighbors, matter.

The circuitry required to accomplish this is shown in Fig 3, for the specific case of presynaptic plasticity. The layer of presynaptic cells is labelled J, the postsynaptic layer I, and the correlation-measurement layer K. Consider the connections to I made by the central presynaptic cell J_0. Let us suppose that as a result of previous sharply focussed correlations J_0 has made a strong connection to the central postsynaptic cell I_0 (solid line and synapses). If it undergoes further strengthening (if it is again plastic) it can form connections to the 2 neighbors of I_0, I_{-1} and I_1. These incipient synapses are shown dotted and open. The central correlation-detecting cell K_0 has to compute $w_m/w_p - 1$. It should therefore receive an excitation equal to the correlation between J_0 and I_0 (i.e. w_m), while its 2 neighbors K_{-1} and K_1 receive excitations equal to w_p. K_0 should then receive lateral, divisive inhibition from its neighbors, such that it fires only if $(w_m/w_p - 1) > E(2\lambda_c - 1)/2\lambda_c^2$. If K_0 fires, it should render all the connections in layer I made by J_0 plastic – able to strengthen in response to correlated pre- and postsynaptic spikes. The most efficient way to accomplish this would be to feed the output of K_0 back to the originating cell J_0 in such a way that it applies a label to the stream of spikes emitted by J_0 so that connections receiving those labelled spikes are plastic. On the other hand, if $(w_m/w_p - 1) < E(2\lambda_c - 1)/2\lambda_c^2$, K_0 should remain silent, so J_0 stays in its default, implastic, condition. The result will be that the J_0 to I_0 connection is only allowed to be plastic if it is highly likely that if when it strengthens it accidentally makes erroneous synapses onto the neighbors of I_0, those errant synapse will be eliminated by competition with the parent synapses, since they are supported by relatively weaker correlations.

This arrangement limits the spread of synapses in the same way that a real reduction of error rates does. It amounts to a virtual lowering of the error rate, without incurring the costs of larger or less densely packed synapses. Of course neither this virtual decrease in error nor real decreases in error are guaranteed to eliminate error – they merely lower its

probability to acceptable levels. The strategy sketched in Fig 3 will of course lower learning rates – there are no free lunches.

This strategy boils down to : do not attempt to learn from confusing inputs. It is related to a proposal of S. Grossberg – "adaptive resonance theory" (ART)[19]. I quote :- "Thus the system allows one of its prior learned codes to be altered only if an input pattern is sufficiently similar to what it already knows to risk a further refinement of its knowledge". However, the background to ART, and the circuits it prescribes, are rather different from those developed here.

The circuitry depicted in Fig 3 resembles several of the most puzzling aspects of the neocortex, if layer J is the thalamus, layer I layer 4 and layer K layer 6. Layer 6 cells typically receive rather weak but direct input from relay cells. They also receive weak input from layer 4 cells [14,15]. Because a K-cell is a correlation-detector, it should respond to the conjunction of 2 weak inputs, but not to the weak inputs separately. Many layer 6 cells then project back to thalamus, where as discussed above, they can flip relay cells from burst to tonic mode. This in turn provides a label superimposed on the spike trains arriving at the relay cell connections in layer 4. There is however no evidence that the burst/tonic transition regulates the plasticity of thalamocortical synapses, though this is not inherently implausible. If indeed moment-to-moment cell-by-cell regulation of the presynaptic plasticity of a set of connections is required it is difficult to envisage any mechanism other than a burst/tonic transition.

A couple of other details were glossed over above. First, the precise biophysical mechanism of the evaluation of $(w_m/w_p\text{-}1)$ was not spelled out. First the correlations w_m and w_p have to measured. One possibility is that shown in Fig 3 – if the J cells synapse on the distal dendrite of K cells, and the I cells on proximal dendrites, then coincident pre- and postsynaptic spikes will produce closely overlapping epsps which temporally summate[98]. Another possibility is that I to K synapses use weakly voltage-dependent NMDA receptors [35] which are unblocked by concurrent J excitation. Various dendritic nonlinearities could also play a role. All that the model requires is that whatever the basis of correlation-detection in J-I synapses (leading to strengthening), a quantitatively similar evaluation should affect K excitation. It is not necessary to enter into discussion of what exactly constitutes "correlation", "coincidence", "conjoint firing" or "synchronous

activity", as long as K cells use a similar definition to J-I synapses. It is also irrelevent what is being computed in layer I (weighted sums, MAX functions, orientation, movement, color etc). It is also not necessary to worry about what is the definition of neighborhood (depicted as just 2 cells in Fig 3), as long as whatever rule defines the neighborhood of a layer I cell is used in wiring up the appropriate connections to layer K. These are all important issues but they are not relevent to the logic being pursued in this Commentary.

One issue that *is* potentially of concern is that the arrangement shown in Fig 3 needs to be duplicated for each and every layer J cell. However, each J cell does not need its own complete sublayer of K cells. There need be only as many sublayers in K as there are neighbors of an I cell – in the extreme case shown in Fig 3, only 3. This is a far less severe problem than that of computing a complete N^2 correlation matrix. It can be further alleviated if correlation levels are usually low, so that layer 6 can use population coding. The remainder of the computation of w_m/w_p-1, involving a ratio, is probably easily done using shunting inhibition in a recurrent excitatory network.

How long following the firing of a layer 6 cell should a relay cell's connections remain plastic? In other words, how long should recovery from low threshold inactivation take? Presumably this should match roughly the window of time during which a postsynaptic I spike is "coincident" with a presynaptic J spike – the duration of the NMDAR component of the epsp, around 100 msec. This is roughly true[35,50]. Over a longer period of time, the degree of plasticity would be set be the spike rate of the layer 6 cells. In this model the timing of layer 6 action potentials need not be very precise, consistent with their relatively slow conduction velocity.

The arrangement shown in Fig 3 matches nicely the observed relay projection to layer 4 (see Fig 1). If neocortex is to learn continuously similar strategies should also be used for intracortical feedforward projections, for example from layer 4 to layer 2/3 and from 2/3 to layer 5. However, in this case it does not seem possible to control plasticity presynaptically, since layer 4,2 and 3 neurons do not have well developed, sharp, burst-tonic transitions. Also, using a burst/tonic label to regulate presynaptic plasticity would be much more difficult in a computing cortical cell than in a relay thalamic cell. Therefore now plasticity should be initiated, and controlled, postsynaptically, as shown in

Fig 4. Here again there is an initial strong connection from J_0 to I_0. However, if this connection erroneously strengthens the errant synapses appear between I_0 and the neighbors of J_0. In the presynaptic plasticity case, it was supposed that the molecular cascade that initiates plasticity (or that converts functionally new to structurally new synapses) is presynaptically controlled. In the postsynaptic case, it is presumed to be postsynaptically located and controlled. There has been considerable discussion as to whether plasticity is initiated pre- or postsynaptically [71,72,95]. Both may occur, conceivably at the same synapses. All the other arrangements shown in Fig 4 are exactly analogous to those shown in Fig 3.

How could layer K control the postsynaptic plasticity of I cells? There is one attractive possibility, suggested by the observation that the intracortical feedforward connections terminate on pyramidal cells. If these synapses form on the apical dendrite (or its side branches), then their plasticity could be regulated by influencing dendritic spike backpropagation[120,121,78]. (It is likely, notwithstanding claims by Markram et al [84], that somatic spikes can spread passively back along basal dendrites; individual plasticity of synapses on basal dendrites would therefore have to be regulated presynaptically). In the simplest case, the spike would either actively backpropagate, or not, switching the neuron from plastic to implastic. There is now good evidence that neuromodulators such as acetylcholine can regulate both K-currents and Na-inactivation, and influence backpropagation[125]. However, this action would be too diffuse for the precise cell-by-cell control required. It is more likely that neuromodulators provide a general plasticity-enabling signal that is multiplied by a cell-specific signal. The cell-specific signal could be the state of the apical tuft. In the simplest case, layer 1 input would generate a local tuft Na/Ca spike [106,73], which would enable backpropagation by counteracting the unfavorable geometrical assymmetry of a branching cable. The duration of this local tuft spike should match the NMDAR open time, as observed [106,73,35]

Fig 1 shows how these arrangements for controlling pre- or post-synaptic plasticity would map onto actual cortical circuitry. 4 classes of layer 6 pyramidal cell are depicted. The 6/C cells, which maintain apical tufts in layer 1, project to claustrum[64]. The claustrum could act as an associative memory for correlation signals, but is not further discussed here. Recent data suggest that in cat striate cortex there are 2 main classes of

layer 6 cell[51]. The "simple" type have simple receptive field properties and send axon collaterals to layer 4, where they synapse on the dendritic shafts of spiny stellate, simple, cells. These are marked as "6/4" cells in the figure. The "complex" type have complex receptive field properties, and collateralise in either layer 2/3 (marked here "6/3") or in layer 5 (marked "3/5"). Probably the simple type (i.e. 6/4 cells) are simple because they receive simple input. Examining Fig 5 it can be seen that a 6/3 cell could be simple both because it receives a pattern of convergent input from relay cells that generates a "simple" type field, and because it receives excitation from a simple cell. This arrangement could be called "double simplicity", and is implicit in the notion that layer 6 cells act as correlation detectors. However, the simplicity of 6/3 cells could instead arise singly, either from their pattern of relay cell input, or because they get direct input from layer 4 simple cells. There is no strong evidence as to whether the simplicity of 6/3 cells is double or simple, but double simplicity is required for the heterosynaptic error control model. The collaterals of layer 6 cells in upper layers are shown synapsing on red arrows in Fig 1. These red arrows represent within layer recurrent excitatory collaterals of the primary computing cells of that layer (shown in red in layer 4, in blue in layer 2/3 and in orange in layer 5). This depiction is symbolic (since these collaterals actually terminate on the dendritic shafts of these primary computing cells), and as discussed further below represents their postulated role in modulating the dynamics of the recurrent "line attractor" process. The collaterals' significance here is that they confirm the existence of 3 separate populations of layer 6 cell, as required.

Examining Fig 1 again, it will be seen that pyramidal cells in layers 2,3 and 5 are depicted with apical tufts in layer 1, where they receive input from horizontally running axons. The simplest arrangement would be for 6/3 and 6/5 cells to control the postsynaptic plasticity of layer 2/3 and 5 cells respectively by directly forming synapses on the appropriate apical tufts, but this is not seen. Fig 1 instead shows the complex-type layer 6 cells as feeding back to thalamus, where they would control the firing of "matrix" type relay cells that synapse heavily in layer 1[60]. In the case of the lgn this would be a subpopulation of W cells (in cats) or koniocellular relay neurons (in primates)[27]. However there is no evidence for this specific arrangement, which is postulated because it allows updating of layer 6 connections during sleep (see below). Fig 1 also shows that

relay cells can receive input either from subcortical drivers (black circles) or from lower level cortical areas (orange; e.g. striate to pulvinar). These are shown as converging on relay cells, but in reality they probably target separate cells.

Fig 1 also shows schematically the feedforward connections that underlie the main cortical computations (relay to 4; 4 to 2/3; 2/3 to 5). However, the patterning of these connections, which determines the type of computation performed (e.g. simple or complex), is omitted. The figure also shows the feedforward projection of layer 2/3 cells to higher cortical areas. Because their plasticity cannot be controlled presynaptically, they should impinge on the apical dendrites of pyramidal cells in higher cortex.

5 SLEEP

The circuitry depicted in Figs 3 and 4 will limit spread of erroneous synapses but cannot eliminate it. If an erroneous synapse should form and persist, it may flourish as a result of changes in the real world or in previous layers of processing. In the extreme case, the errant connection may outcompete the parent connection. Such an allegiance transfer is depicted in Fig 6 for the case of presynaptic plasticity. Originally cell J_0 connected to I_0 (Fig 6a), but as a result of a shift in the statistics of the input ensemble to the J layer, it has become connected instead to the adjacent cell I_1 (Fig 6b). Unfortunately the layer 6 error-containment circuitry, copied in Fig 6a from Fig 3, is no longer appropriate to this new condition, and if it is not updated turnover of the new connection will generate errant synapses which will degrade its precision, and perhaps eventually lead to loss of function. 2 main types of updating are required. The existing layer 6 connections which must be eliminated are shown as dotted lines in Fig 5, and those that must be created are shown as dashed lines. It is assumed that the I-to-K connections are permanent. (They are presumably laid down in fetal life as a cortical backbone, possibly as a result of the spontaneous activity of Cajal-Retzius cells in layer 1, in which the layer 6 cells initially do maintain apical tufts). These permanent vertical connections would define the cortical column. The remaining connections can be updated offline if they are rendered plastic, while the main feedforward J to I connections should be rendered implastic. The connections must be updated offline because they require the application of appropriate

"calibration" signals that are quite different from those caused by the real world, and which are incompatible with normal function.

First, consider the rewiring of the J to K (relay to layer 6) connections . Suppose a strong calibration signal is played into cell J_0. Following the allegiance transfer, this will cause cell I_1 to fire. (As discussed below, because I_1 firing must here be driven by spikes in just one layer J cell, this J cell's spikes should be in a powerful, reliable, burst mode). Since both J_0 and I_1 are firing , cell K_1 will fire (since K_1 has inherited an input from J_0 from the pre-allegiance transfer condition). Neither K_0 nor K_2 will fire, because they do not receive conjoint J and I input. Thus the J_0 - K_1 connection will selectively strengthen. However, this will generate some erroneous J_0-K_0 and J_0-K_2 synapses, especially since there is no error-control circuitry for the layer 6 connections. These are precisely the connections that need to be created or maintained. The crucial point here is the definition of "neighborhood". When discussing the spread of feedforward connections from layer J to layer I, it was assumed that there was a set of I cells onto which erroneous synapses could form in a single generation, by imprecise replication. In the figures this is represented as the 2 left or right neighbors in the row, but in the real brain this neighborhood will be more complicated. First, the layer occupies at least 2 dimensions. Second, the dendrites of each I cell spread widely, and interdigitate with a large number of neighbors. But we do not need to worry about what constitutes the neighborhood of a J cell. All we require is that whatever the definition of the neighborhood of a J cell, K cells use the same definition. Thus rewiring will work provided that J to K synapses have similar error rates to J to I synapses, and provided that there is no error containment circuitry associated with layer 6. Thus the problem of what contains errors in wiring the error-containment circuitry – "quis custodiet custodies?" – not only vanishes, but transmutes to a solution. Note that the surplus J_0 to K_{-1} connection will likely be eliminated by competition, since it is supplied neither by conjoint firing, nor by direct errors.

In summary then, the proposal is that relay to layer 6 connections have to be updated offline to keep pace with possible online rewiring of the relay to layer 4 connections. This offline rewiring should be accomplished by playing narrowly focussed, strong, calibration signals into individual relay cells. The simplest way to accomplish this would

be for the relay cells to successively burst, in a wave-like pattern. Such a burst-wave has been described in thalamic slices [6,65]. It appears to arise from the circuit properties of the reticular nucleus [65,89,119] and has been proposed as a model of slow wave sleep. However, in naturally sleeping cats this burst-wave is somewhat modified [117,118]. This proposed mechanism resembles the way that retinal waves wire up the geniculate [36,131].

Rewiring the corticothalamic feedback connections would be more difficult. This difficulty is illustrated in Fig 5, which sketches the Hubel-Wiesel wiring required for orientation selectivity in cat striate cortex (yellow relay connections to purple spiny stellate cells). An I cell does not receive inputs from a single J cell, as assumed in the diagrams of Figs 3 and 6, but from a set of J cells. This is because the cortex does not simply form a topographic map of its input (as implied in Fig 3), but *computes*. However, the problem of heterosynaptic error is the same in setting up complex circuits that compute, or simple circuits that merely preserve information. Both require point-to-point wiring for maximum efficiency. We may therefore extend the principles embedded in Fig 3 to argue that a K cell should project back to that set of J cells that comprise the "receptive field" of its layer I permanent partner, and control their plasticity. This arrangement is sketched in Fig 5. Evidence that this proposed wiring actually exists is discussed below, but for the moment let us consider how it could be generated by a suitable calibration signal played into layer J.

The nub of the difficulty is that a K cell must determine the set of J cells currently comprising the receptive field of its I-partner. The obvious way to do this would be for the brain to perform reverse correlation analysis, in much the way that a modern neurophysiologist would determine the receptive field of a simple cell [99,103]. This is possible if K cells act as correlation detectors. Thus rather slowly changing random patterns (filtered white noise) should be played into the J layer, while J to I or K (thalamocortical) connections are implastic, and the K to J (corticothalamic) connections are plastic. From time to time, by chance, the pattern on the J layer will activate the set of cells that comprise the receptive field of I_1. (Note that this requires that slow wave sleep should always occur prior to paradoxical sleep, as observed). This set of cells should be firing in tonic mode, so that firing of I_1 is only triggered by their concurrent activation (in contrast to the updating in slow wave sleep; the resulting plasticity of tonic feedforward

connections must be over-ridden by a global neuromodulatory signal). Now cell K_1 receives concurrent inputs both from that set of J cells and from the I_1 cell, and so will discharge. If the synapses that it makes on the distal dendrites of relay cells are Hebbian, and plastic in REM sleep, then connections to the correct set of J cells will be reinforced (since these will still be firing if the white noise changes relatively slowly). If this set of cells differs only slightly from the set of cells to which it was previously strongly coupled (i.e. in Fig 5 cell J_1 is near cell J_0), it is likely that these connections already exist as fringe connections by virtue of errors in previous rounds of offline updating.

These requirements correspond quite closely to the observed properties of dream (REM) sleep. In REM sleep brainstem cholinergic neurons in nuclei in the parabrachial area (eg pedunculopontine tegmental) fire irregularly [26,52]. These neurons form synapses on relay cells, where they activate both nicotinic and muscarinic receptors [112]. Nicotinic receptors produce brief epsps which may act as "pseudodriving" inputs, capable of driving relay cells to threshold, and generating irregular sequences of relay sodium spikes. Muscarinic receptor activation inhibits a resting potassium conductance, depolarising the relay cell, and switching it to tonic mode[110]. In addition, acetylcholine is released abundantly into cortex itself[118], where it could affect which types of connection are plastic. Now the cortex does not need to know that the irregular relay spikes that arrive in REM sleep are fundamentally different in origin and nature from the irregular spikes that are driven by the environment in the awake state, and will interpret them (perhaps assisted by recurrent attractor dynamics) as conscious experience, i.e. dreams [76,77]. The rapidity with which dreams fade is consistent with the requirement that learning via feedforward connections does not occur during the tonic activity of REM sleep; the persistence of the tail of the dream presumably reflects activity reverberation. The requirement for Hebbian learning by the corticothalamic feedback projection is consistent with the finding that these synapses activate fast epsps mediated by AMPA- and NMDA- GluRs[105].

Is there any evidence for the specific pattern of corticothalamic feedback shown in Fig 5? This is much more specific than the well-characterised coarsely retinotopic arrangement[110]. One hint comes from the finding that a corticothalamic drumstick axon synapses in both left and right eye geniculate layers[93]. If such an axon synapses exclusively on the relay cells that synapse on a binocularly driven layer 4 cell which in

turn supplies the layer 6 cell that provides that feedback axon, this would nicely match with the principle embodied in Fig. 5. It is also consistent with the otherwise mysterious matching of topographic mappings in adjacent laminae of lgn. A further hint comes from experiments in which spike synchrony between pairs of relay cells was examined before and after cortical cooling [114]. It was suggested that the cortex-dependent synchrony reflected common excitation by a layer 6 feedback cell. This common excitation was caused by a moving bar that was aligned with the axis of the receptive fields of the relay cell pair, as predicted by Fig 7. A third hint comes from the morphology of the thalamic endings of the layer 6 cells [45,93]. Their overall shape is quite different from that of the driving terminals, which form the roughly spherical cloud expected if they were to synapse on the proximal dendrites of a single relay cell. The layer 6 endings, from which drumsticks emerge, penetrate the entire thickness of a geniculate lamina (indeed, as noted above, beyond it). The drumsticks seem to occur in clusters[93], as though they were selectively innervating a few cells along the axis of the terminal. Obviously this does not tell one which cells are being targetted, but it is consistent with the idea that these targetted cells comprise the receptive (or projective) field of a cortical cell. A further indication of considerable convergence is the fact that the cortical feedback axons are 10 – 100 times more numerous than the relay cells themselves[112].

Rather similar requirements and principles also apply to updating the postulated layer 6 circuitry controlling the postsynaptic plasticity of cortical pyramidal cells receiving feedforward excitation on their apical dendritic spines (see Fig 4). In this case the permanent connections are from J_0 to K_0 and so forth, and the feedforward I to K connections and the feedback K to I connections both require updating following an allegiance transfer, for example if I_0 initially gets input from J_0, and then instead receives input from J_1. Initially, I_0 was connected to K_{-1}, K_0 and K_1. It must now disconnect from K_{-1} and connect to K_2. Also, the feedback connection from K_0 to I_0 must be broken and replaced by a new connection from K_1 to I_0. More generally K_1 should supply the set of I cells that currently receives input from J_1, i.e. its "projective field". All these changes can be made by selectively activating J_1 while J to I synapses are implastic, and I to K and then K to I synapses are plastic.

How can J_1 be selectively activated? In the case where J is a spiny stellate, simple cell, this is automatically achieved during REM sleep, whenever the white noise relay input happens to match its receptive field. However, if J is a layer 2/3 pyramidal cell, this idea may not work. Instead it is possible that the individual activations of 2/3 cells are achieved by "matrix" relay cell drive to apical tufts in slow wave sleep. This would entail enabling a mechanism for orthodromic dendritic spike propagation, along the lines seen in mitral cells[23]. It would also explain why layer 6 control of postsynaptic plasticity is exerted indirectly, via thalamic "matrix" relay cells. If this is so it would make sense for the matrix-layer 1 connections to be fixed, and for the feedback connections to be updated in thalamus (just as for the simple layer 6 cell feedback). This would work quite well if these matrix cells can individually fire cortical cells via their apical tufts in an orthodromic dendritic propagation mode.

In summary the connections of layer 6 cells can be appropriately updated during sleep. The simple layer 6 cells' input from thalamus can be updated in slow wave sleep, and their feedback to thalamus in REM sleep. The complex 6/3 cells' connections can be updated in REM sleep, and those of complex 6/5 cells during slow wave sleep. Such updating is inevitable if (1) layer 6 cells compute correlation ratios (2) there is daytime rewiring of feedforward connections which exploits heterosynaptic error.

6 A COMPLEMENTARY VIEW OF SLEEP

McClelland and his colleagues[88] have developed a superficially quite different view of neocortical learning and sleep. They argue that backpropagation neural networks can only develop structured knowledge from complex data sets if they learn slowly, and such slow, interleaved, learning can only be accomplished if the raw data is put in a temporary store which is then repeatedly replayed. They propose that neocortex corresponds to the backpropagation network and the hippocampus to the temporary store, and that replay occurs offline during sleep. However, there are several interesting relationships between this view and the thesis developed in this Commentary. In both cases the central argument is the necessity for neocortical slow learning. Artificial neural networks suffer information catastrophes (memory blackouts or failures to detect structure) if they are

made to learn too rapidly [3,48]. A key to the success of the backpropagation neural network is the nonlinearity of the neurons' response function (usually the steepness of a logistic function). However, this nonlinearity is itself intimately related to the learning rate[123]. The steeper the response function the faster the learning. If the neuron is nearly linear, learning is very slow, but it does not fail catastrophically. The neurons in the treatment in Section 3 above are linear, and learning degrades gracefully with increases in error rates or decreases in correlation. However, a nonlinear network would probably fail catastrophically with weak correlations or large error rates. Thus both the fully-connected backpropagation paradigm and the sparsely connected heterosynaptic error paradigm fail when complex inputs are combined with nonlinearities, and in both this can be avoided by slow learning.

In the treatment by McClelland et al the neocortex is assumed to learn slowly at some overall rate, with raw data filed for future processing. In the view developed in this Commentary, individual neocortical neurons always learn at the maximum possible rate (such that λ is below some critical value). Nevertheless under most conditions (in a complex and noisy real world) these individual learning rates may be quite slow, and a hippocampal backup file system is still required. Furthermore, the idea of hippocampal replay in sleep would still hold. This replay would presumably occur via the mammillary nuclei and the thalamic anterior nuclei to cingulate cortex, and be subject to layer 6 correlation sharpness evaluation and plasticity control.

7 OTHER USES OF CORRELATION RATIOS AND PLASTICITY CONTROL

Although the above proposal, that layer 6 neurons measure correlation ratios and control pre- or post-synaptic plasticity, is based on consideration of errors in the placement of new synapses, correlation ratio measurements and local plasticity control signals could be used in a variety of other cerebral operations. The following outline sketches some of the possibilities.

In section 2 two critical decisions that cortex must make were identified. First, a cortical area (or individual column or neuron) has to tell a thalamic nucleus innervating that area (or column or neuron) what response mode (burst or tonic) to adopt. The black/white

burst mode is better suited to detection of noisy signals, while the gray, tonic, mode is better suited to signal analysis. It was argued that a major source of noise is the use of a rate code. Thus a novel, unexpected, stimulus (which by definition cannot be cleaned up by temporal averaging) is best transmitted in burst mode. Recordings from relay cells of awake cats support this inference[44]. If the instruction to thalamus is conveyed by layer 6 axons, what criterion can layer 6 cells use to decide whether a stimulus is novel or unexpected? This criterion must be developed locally and on-line. The firing rate of a simple cell is not sufficient. One needs to know to what extent the spikes of the simple cell allow one to determine the orientation of the input in the appropriate patch of visual space – how well correlated the simple output is with its thalamic input. Thus layer 6 cells may be well suited, as correlation-detectors, to this mode-switching role.

A second rather similar decision must also be reached about the setting of the recurrent excitation within a cortical layer that underlies the line attractor dynamics. Theoretical analysis suggests that this recurrent mechanism operates in one of three regimes or phases, depending on the relative strengths of short-range excitation and longer-range inhibition[46]. In the homogeneous phase the neuron responses are dictated by the feedforward connections. In the marginal phase the line attractor dynamics dominate, amplifying weak orientation signals. In the unstable phase positive feedback dominates and the input is disregarded. If there is much input noise (for example if the stimulus is novel), the marginal phase should be selected. Under these circumstances the output of a simple cell will be poorly correlated with the thalamic input, and its corresponding layer 6 cell's collaterals in layer 4 will be silent. It is suggested that this silence allows the layer 4 recurrent feedback synapses to operate. If the simple cell's spikes are tightly correlated with the incoming relay spikes, then the collaterals will release glutamate, which should, perhaps via a metabotropic receptor, lower the gain of the recurrent excitatory synapses. This control mechanism is depicted in Fig 1 as an intracortical feedback green arrow terminating on a red arrow representing within-layer recurrent excitation. Some evidence that there is an initial synchronisation (presumably reflecting recurrent feedback) which collapses as stimulus analysis procedes has recently appeared[20].

How can the appropriate pattern of recurrent connections be set up? In the line attractor model of orientation selectivity cells with similar orientation preference tend to excite

each other, and usually these are also assumed to be physically close to each other. However, because striate cortex forms a 2-dimensional map of both orientation and visual position, there must be map discontinuities[122], and thus function and physical proximity are not always compatible. Thus the line attractor lateral connections should be set up on the basis of *functional* similarity. These connections must be updated by Hebbian learning, to match any recent change in the feedforward connections. In this case learning must be online, because it requires structured input. For example an oriented bar will strongly excite an appropriate simple cell (via the feedforward connections) and, somewhat less strongly, other simple cells tuned to similar orientations. If the lateral connections between these cells are strengthened by this synchronous activity, simple cells are automatically wired into the correct recurrent pools. However during this process it is essential that the recurrent excitation itself be turned off, pushing the lateral network into the homogeneous regime. This could be achieved if the AMPAR component of the recurrent epsps were selectively suppressed during the layer 6 collateral firing[66]. Thus the recurrent excitation gain change postulated in the previous paragraph should be selective, suppressing the AMPAR but not the NMDAR component. Because the recurrent connections merely supplement the basic computation performed by the feedforward connections, they need not be as precisely wired, or as temporally accurate, and do not require the extra, individual-plasticity control machinery postulated above. Conversely, the feedforward J to I connections should *not* undergo online updating if the line attractor mechanism is operating, since here correlations will be at least partly due to recurrent dynamics and not solely dictated by the real world, previous layers of processing, and the feedforward connections themselves. Of course this is automatically achieved since the layer 6 signals control plasticity and line attractor dynamics in tandem. However, this could be supplemented by neuromodulatory inactivation of feedforward plasticity produced by layer 6 collateral firing.

8 REWARD-BASED LEARNING

So far all cortical learning was assumed to be unsupervised, with the goal of exploiting redundancies in complex inputs. However, much cortical learning must be guided by

error signals that the brain derives from its owner's actions. In neural networks the most popular paradigm for such supervised learning is the "delta rule", in which local error signals (e.g. the difference between the current output of I_1 and its desired output) replaces the postsynaptic cell's activity in the Hebb rule. If the connections to be modified are not on the output cells, but onto earlier "hidden" cells, then the error signal is "backpropagated" – sent back to the hidden connections after weighting by the feedforward, hidden-to-output, connection strengths. An approximation to this scheme can be achieved biologically using the following modification of the Hebb rule, called the Associative Reward penalty (ARP)[87] rule:-

$$dy/dt = ar(f_i - p_i)f_j + b\,(1-r)\,(1-f_i-p_i)f_j$$

where a and b are learning rate constants, r is a "reward", $(1-r)$ is a "penalty", f_i is a binary stochastic variable describing the firing of the postsynaptic neuron (1 = firing, 0 = silent), f_j is a similar variable for the presynaptic neuron, and p_i is the usual dot-product weighted sum of inputs. f_i is a logistic function of p_i, such that if the neuron is strongly depolarised the neuron is more likely to fire. The reward is some suitable scalar measure of the overall error (such as the root mean square error), presumed to be computed by other brain regions. Note that if $r = 1$ and $p_i = 0$ this reduces to the standard Hebb rule. In words the rule says that if a neuron accidentally fires when it "shouldn't" (because it is only weakly depolarised), then the connection from any input cell which also fired should be strengthened by an amount proportional to the reward (since those connections appear to be useful), and *vice versa*. Essentially a neuron jitters its responses, and explores its weight space, adjusting its synapses appropriately based on the outcome, somewhat in the way that the heterosynaptic errors discussed in section 3 allow a sparesely wired network to explore alternative connections. The biophysical assumptions underlying this rule are (1) there is a noise source between overall postsynaptic depolarisation (p_i) and firing (f_i), (2) the signals f_i, p_i and f_j are all available at a connection, the first due to spike backpropagation, the second due to summed postsynaptic depolarisation, and the third via local glutamate release (3) rewards and penalties affect the degree of spike backpropagation (or some other local control of plasticity).

There are several problems with this biologically plausible approach to supervised learning. First, it seems likely that most neural noise is not in the spike firing mechanism [16,79], but in the transmitter release process [8,63]. Quantal fluctuations will contribute to both f_i and f_j (as perceived by the postsynaptic neuron), and cancel out. Second, because there is not nearly as much information in a scalar reward signal as in a complete error vector, learning is inefficient. Small "good" fluctuations can be masked by larger "bad" fluctuations. Third, the stochastic firing will degrade the precision of the output. The last problem can be solved by gradually annealing the noise as performance improves, but at the expense of future flexibility. (The third problem and its solution are closely related to the problem of heterosynaptic error).

A mechanism for controlling the plasticity of individual neurons, like that proposed for layer 6, should greatly help with the last 2 problems. This can be most clearly seen in the limiting case, where only one neuron participates in sufficiently sharp correlations that it is allowed to be plastic. Only fluctuations across its plastic connections are considered in the learning process, greatly reducing or eliminating the masking problem. In addition, it should be possible to eliminate the fluctuations in implastic neurons, especially if they are synaptic in origin. These fluctuations merely generate output noise without contributing to learning. For example, if a relay to layer 4 connection is allowed to be plastic. transmitter release could be allowed to fluctuate ("low p" synapse[21]). If the connection is implastic transmitter release could be made reliable ("high p" synapse). This may be what happens – the burst mode produces reliable transmitter release[75].

9 DIFFICULTIES AND TESTS

The aim in this Commentary is not to provide a comprehensive theory of neocortical function but to sketch a new approach to the problem of flexible learning by networks of bulky, and balky, components. Nevertheless, many concrete interpretations of thalamocortical anatomy and physiology have been proposed, and if most of these prove wrong then the general approach is probably misguided. In this section some of the logical and experimental weaknesses of the main argument are spotlighted, and then some key predictions listed.

The approach builds on a very simple concept of weight adjustment, in which synapses replicate or die in response to nearly synchronous pre- and post-synaptic activity. However the precise rules governing long term potentiation and depression, especially to repeated activity pairings, are not known. It does appear that synapses increase in strength in digital fashion, and of course this is consistent with the quantal hypothesis [8,63]. The key assumption made here is that the added synapses do not always form between the correct neurons (the ones whose pairing triggered strengthening). Support for this hypothesis comes from experiments on "volume learning". Further support comes from the impossibility that any replication process can be exact. Even if synaptic strength is analog, not digital, one could still postulate that there is some spillover of added strength (as indeed the volume learning experiments indicate). However, if synapses are truly analog, then networks must be fully connected. This means that (1) there is no need to form new connections (2) the set of possible connections, and correlations, becomes astronomical.

If specific connections are to be made between neurons, there must be directing signals. These signals can be either genetic or activity-dependent. Neural correlations are likely to be crucial for activity-dependent wiring, but such correlations can never be completely specific. If connections grow linearly at rates that depend on correlations (Eq 3), then the effect of competition is that connection strengths evolve to reflect the strengths of those correlations. If there are a large number of possible target neurons, then the great majority of synapses will be inappropriate, even if they are all "correctly" (i.e. in proportion to correlations) placed. If connections grow exponentially with rate constants that depend on correlations (Eq 2), then under competition and no placement errors, then all synapses will be appropriate. However, if synapses are sometimes placed erroneously, such perfection is not achievable (unless correlations are completely specific. If connections grow hyperbolically (i.e. the exponent of y in Eq 2 is greater than unity), they will eventually "lock-in" and will not adjust to changes in correlations.

Although the proposed circuitry can be viewed as just an indirect way to improve the accuracy of cortical wiring, it becomes particularly interesting if there is an error catastrophe – that is, if a network fails completely to wire if it is too large, if imposed patterns are too weak, or error rates are too great. Somewhat related models, in which

new connections are formed by a random local process and selected by a Hebbian mechanism, have been successfully applied to the organisation of topographic, eye-dominance and orientation maps [31,130]. In the latter cases it was explicitly shown that maps failed to organise if a "temperature" parameter was set too high. "Temperature" appears to be analogous to error rate. If there is a cortical wiring catastrophe at excessive error rates, because of the huge number of possible wires and the weakness of the patterns in the real world, the proposed algorithm could be used to keep the network just on the right side of catastrophe, so that ordered connections can be achieved at the maximum possible rate. Complex, possibly random, input patterns that cannot be learned by cortex (because the current cortical networks are not yet sufficiently structured) would be filed in hippocampus, just as in the McClelland account.

Moving to more concrete difficulties, the accounts of the layer 6 lateral interactions underlying correlation ratio determination, and of allegiance transfer, are oversimplified, in related ways. The heart of this difficulty is the difference between "connected" and "unconnected". This difference is quite clear in the extreme case where a neuron makes or receives only one connection, as assumed in Figs 3 and 4. Indeed, in the absence of error and at equilibrium the algorithm of Eq 8 does lead to such exclusive connections, since it is winner-take-all. However, *during* the allegiance transfer there will be distributed connections, and even at equilibrium there will be an error fringe. Thus Figs 3, 4 and 5 have to be extended to account for the possibility that there are several connected neurons in the high fitness region. This means that the pool of layer 6 neurons corresponding to this connected region should mutually excite each other, so that their excitation (derived both from the correlation-sensitive feedforward connections and the lateral interactions) corresponds to the *average* correlation in the high fitness zone, $\langle w_m \rangle$. Inhibition by the neighboring layer 6 cells that correspond to the unconnected neighbors in the upper layers would allow the computation of $\langle w_m \rangle / w_p$. This is turn means that offline updating of the lateral layer 6 connections must also take place. For example in the case of presynaptic plasticity (Fig 3), a focussed calibrating burst signal should allow correct formation of a layer 6 high fitness pool if the layer 6 recurrent excitatory synapses are Hebbian and plastic in slow wave sleep.

An even more specific problem is that there is little evidence for thalamocortical plasticity, especially in the adult[24]. Slice experiments suggest that thalamocortical synapses are only plastic during an critical period [56], and this plasticity seems to be *postsynaptic*. Nevertheless there is some evidence for anatomical plasticity of geniculocortical connections[61], and the somatosensory thalamocortical synapses may be rather special.

Related to this difficulty is the lack of direct evidence for the postulated mechanisms of plasticity control. In the case of postsynaptic plasticity, where it was suggested that apical tufts control backpropagation which in turn is essential for plasticity, the difficulty is not severe, since it is known that apical tuft excitation interacts with backpropagation 73 and that backpropagation controls plasticity[78]. However, in this case there is no evidence that layer 6 neurons preferentially influence this process. For presynaptic plasticity the difficulty is exactly reversed. There is ample evidence that at least some layer 6 neurons do control the burst/tonic status of relay neurons, and it is likely that this control is central to the mystery of thalamic function. However, there is no evidence that the burst/tonic transition controls the presynaptic plasticity of thalamocortical afferents.

Several of these difficulties, and many other aspects of the theory, are susceptible to direct test, though in most cases the experiments are likely to be technically very demanding. Here are 4 obvious possibilities. First, in a hippocampal slice it should be possible to look for digital strengthening errors, by recording simultaneously from adjacent pairs of CA1 neurons, and looking at LTP induced by minimal stimulation [107,97]. Of course if pairing of 1 of these CA1 neurons with a single CA3 axon leads, with some low probability, to the appearance of a new functional synapse at the other previously "absent" connection, this could be interpreted as unmasking of a silent synapse at an existing connection lacking AMPARs [57,74].

Second, one could see whether either of the proposed mechanisms for plasticity control (presynaptically via the burst/tonic transition, or postsynaptically by layer 1 control of backpropagation) actually occur. The former would have to be done in a young somatosensory thalamic slice preparation, by testing the effectiveness of different patterns of spikes evoked in thalamus in producing thalamocortical LTP. The latter could be done by dual recording in neocortical slices, by testing whether local layer 1 excitation

does indeed affect backpropagation – essentially the reverse of the Larkum et al [73] experiment.

Third, it would be important to check that layer 6 pyramidal cells actually do operate as coincidence detectors. One possibility would again be a somatososensory thalamocortical slice, with a stimulating electrode in thalamus, and microelectrodes in layer 4 and layer 6 cells. The difficulty would be in finding a monosynaptically coupled cell pair. The prediction is that a thalamocortical volley that evokes a monosynaptic potential in the layer 6 cell would only fire a spike in that cell if coupled with a suitabled timed spike in a layer 4 cell that responded monosynaptically to that volley. Another approach would be to simulate such cellular coincidence detection, incorporating suitable dendritic synapse placement, NMDA receptors etc into a model layer 6 pyramidal cell.

Fourthly, the pattern of the layer 6 feedback connections could be studied. Ideally one would want to record *in vivo* from a layer 6 cell and show that it innervates a set of geniculate relay cells whose retinal inputs are collinear with the orientation preference of that layer 6 cell. This could be done by filling the layer 6 cell with HRP after recording its receptive field, and then "burning" with a high intensity light a line of cells on retina corresponding to that receptive field. The prediction is that the HRP-filled corticothalamic boutons would terminate on relay cells that receive degenerating retinal terminals.

All of these tests are difficult, and all have the drawback that even if they agree with predictions, they do not prove the model. Most of the tests would be interesting even in the absence of the model.

10 SUMMARY

The central problem faced by the brain is generating appropriate but complex behavior in a regular but complex environment, using biological machinery that is susceptible to error and genetically-programmed algorithms that may be suboptimal. To a first approximation brain operations fall into 2 categories, with quite different time scales. First, there is fast processing of electrical signals, whose essence is captured by the standard connectionist neuron. Second, there is slow adjustment of synaptic weights,

using locally available information such as firing rates and reward signals. The first process influences the second, and the second determines the outcome of the first. With luck and good management the coupled processes converge to generate behavior which increases the animal's probability of reproduction. Both processes are subject to errors, which are of 2 general types. First, there is noise (the wrong signal going to the right place). Second, there is crosstalk (the right signal going to the wrong place). In silicon-based computing these arise largely from thermal noise and parasitic capacitances respectively. How the brain deals with these 4 types of error (fast and slow, noise or crosstalk) depends critically on the statistical structure of the input. The problem of input or output complexity and of internal noise are intimately linked. (Of course even a noise-free neural algorithm can fail to generate appropriate behavior; however this failure can ultimately be traced to noise in the genetic evolution algorithm itself, since this determines the efficiency of the neural algorithm).

The main focus here has been on crosstalk in the learning process. Such crosstalk errors arise from the compactness of synapses and neuropil, much as in electronic computing. The core thesis of the Commentary is that because internal errors and external complexity are linked, the naïve and bounded strategy of lowering error rates through hardware improvements can be supplemented by a "software" or virtual strategy of matching errors rates to measured input (or output) complexity. This sounds grandiose and vague, but in the simple case of a single presynaptic neuron projecting to a set of linear postsynaptic neurons via feedforward Hebbian connections it proved possible to describe explicitly the relation between error rate, connection smearing and "input complexity", using a crude measure of "correlation". The measure used here, "fitness", reflects the degree of synchrony between pre- and postsynaptic spikes produced by existing networks in response to ensembles of input (just as genetic fitness reflects the degree of match between phenotype and environment). However, the exact nature of this relationship need not be spelled out – the analysis holds true provided that layer 6 neurons use the same measure of "correlation" as the feedforward connections. Even though synapses only form crude judgements about "correlation", "regularity", "coincidence" etc, rather sophisticated processing can be achieved given their immense numbers. The rule that is proposed is "measure the way the correlations between a neuron and its current or

possible immediate targets varies across those targets; if these correlations are much the same everyhere, keep the current strengths (and connections). If the correlations peak sharply at the currently connected neurons, allow those current connections to be plastic". I call this rule the "thalamocortical algorithm" because it appears to be implemented by some of the most characteristic neural machinery of cortex and thalamus. Less formally the algorithm says that plasticity is a precious resource that can only be allocated to connections across which there are strong correlations.

The layer 6 correlation sharpness signals represent a simple local measure of how well the current pattern of feedforward synaptic connections capture lower order statistics of the input ensemble. These correlation signals can also be used to control aspects of rapid timescale processing, such as burst/tonic packaging, line attractor dynamics, and neural noise. The addition of a third correlation-sensitive layer to a pair of input-output layers is only feasible when feedforward connectivity is sparse. It represents a simple extension of the general principle that neocortex combines numerous parallel and hierarchical analyses that are individually trivial but collectively stupendous. From this viewpoint the thalamus retains its traditional relay role, but with a new twist that makes it the key to accurately wiring the universe's most complex structure, the neocortex.

Acknowledgements. I thank Kingsley Cox, Murray Sherman and the late David Fox for help,discussion and advice.

REFERENCES

1. Adams, P.R. (1998) Hebb and Darwin. J. Theoretic. Biol.195, 419-438

2. Alonso, J.-M. and Martinez, L.M. (1998). Functional connectivity between simple cells and complex cells in cat striate cortex. Nature Neuroscience, 1, 395-403

3. Amit, D.(1989) Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press.

4. Anderson. J.A. (1995) An Introduction to Neural Networks. MIT Press, Cambridge

5. Antonini, A & Stryker, M.P. (1993). Rapid remodeling of axonal arbors in the visual cortex. Science, 260, 1819-1821.

6. Bal, T., Krosigk, M. von and McCormick, D.A. (1995). Role of the ferret perigeniculate nucleus in the generation of synchronised oscillations in vitro. J. Physiol. 483, 665-685.

7. Bear, M. (1996). A synaptic basis for memory storage in the cerebral cortex. Proc. Nat. Acad. Sci. 93, 13453-13459

8. Bekkers, J.M. (1994). Quantal analysis of synaptic transmission in the central nervous system. Current Opinion in Biology, 4, 360 - 365.

9. Ben-Yishai, R. , Bar-Or, R. and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Nat. Acad. Sci. 92, 3844-3848.

10. Bienenstock, E.L., Cooper, L.N. and Munro, P.W. (1982). Theory for the development of neuron and binocular interactions in visual cortex. Neurosci. 2, 32-48.

11. Bishop, G (1959) The relation between nerve fiber size and sensory modality: phylogentic implications of the afferent innervation of cortex. J. Nerv. Ment. Disease. 128, 89-114.

12. Bolshakov VY., Golan H., Kandel ER. & Siegelbaum SA. Recruitment of new sites of synaptic transmission during the cAMP-dependent late phase of LTP at CA3-CA1 synapses in the hippocampus. Neuron. 19, 635-651 (1997)

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

14. Buhl,E.H., Tamas,G., Szilagyi, T., Stricker, C , Paulsen,O and Somogyi, P. (1997). Effect, number and location of synapses made by single pyramidal cells onto aspiny interneurones of cat visual cortex. J. Physiol. 500, 689-713

15. Callaway EM. (1998) Local circuits in primary visual cortex of the macaque monkey. Annual Review of Neuroscience. 21,47-74.

16. Calvin, W.H. and Stevens, C.F. (1968). Synaptic noise and other sources of randomness in motoneuron interspike intervals. J. Neurophysiol. 31, 574-578

17. Carandini,M. and Ringach, D.L. (1997). Predictions of a recurrent model of orientation selectivity. Vision Res. 37, 3061-3071

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

19. Carpenter, G.A. & Grossberg, S. (1990) Self-organising neural network architectures for real-time adaptive pattern recognition. In : An Introduction to Neural and Electronic Networks. Ed by Zornetzer, S.F., Davis, J. L. and Lau, C. Academic press. San Diego

20. Cardoso de Oleivra, S., Thiele, A. and Hoffmann, K.-P. (1997) Synchronisation of neuronal activity during stimulus expectation in a direction discrimination task. J. Neurosci. 17, 9248-9260

21. Castro-Alamancos, M. A. and Connors, B.W. (1997) Thalamocortical synapses. Prog. Neurobiol. 51, 581 -606

22. Chapman,B, Zahs, K.R. and Stryker, M.P. (1991). Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborise within a single orientation column in ferret visual cortex. J. Neurosci. 11, 1347-1358

23. Chen WR. Midtgaard J. and ;Shepherd GM. (1997) Forward and backward propagation of dendritic impulses and their synaptic control in mitral cells. Science. 278(5337):463-7

24. Darian-Smith, C. & Gilbert, C.D. (1994) Axonal sprouting accompanies functional reorganisation in adult cat striate cortex. Nature, 368, 737-740.

25. Das, A. (1996) Orientation in visual cortex: a simple mechanism emerges. Neuron 16, 477-480

26. Datta,S. and Siwek, D.F. (1997) Excitation of the brain stem pedunculopontine tegmentum cholinergic cells induces wakefulness and REM sleep. J. Neurophysiol. 77, 2975-2988

27. Ding, Y. and Casagrande, V.A. (1997) The distribution and morphology of LGN K pathway axons within layer and CO blobs of owl monkey V1. Vis. Neurosci. 14, 691-704

28. Douglas, R.J.,Mahowald, M, Martin, K.A.C. and Stratford, K.J. (1996) The role of synapses in cortical computation. J. Neurocytol. 25, 893-911

29. Douglas, R. and Martin,K. (1991) A functional microcircuit for cat visual cortex. J. Physiol. 440, 335 - 337

30. Douglas, R.J. & Martin, K.A.C. (1998). Neocortex . pp 459-510 In : The Synaptic Organisation of the Brain. Ed. G.M. Shepherd. OUP, New York

31. Elliott,T., Howarth, C.I. and Shadbolt, C. I. (1996). Axonal processes and neural plasticity. 1: Ocular dominance columns. Cerebral Cortex 6, 781-788.

32. Engert F. Bonhoeffer T. (1999) Dendritic spine changes associated with hippocampal long-term synaptic plasticity. Nature. 399, 66-70

33. Engert, F. and Bonhoeffer, T. (1997). Synapse specificity of long-term potentiation breaks down at short distances. Nature 388, 279-284.

34. Fanselow EE. Nicolelis MA. (1999) Behavioral modulation of tactile responses in the rat somatosensory system. J. Neuroscience. 19, 7603-7616

35. Feldmeyer D. Cull-Candy S. (1996) Functional consequences of changes in NMDA receptor subunit expression during development. Journal of Neurocytology. 25(12):857-867

36. Feller, M.B. (1999) Spontaneous correlated activity in developing neural circuits. Neuron 22, 653-656

37. Ferster,D., Chung, S. and Wheat,H. (1996) Orientation selectivity of thalamic input to simple cells of cat visual cortex. Nature 380,249-252

38. Ferster,D and Miller, K.D. (2000). Neural mechanism of orientation selectivity in the visual cortex. Ann. Revs. Neurosci. 23,

39. Frey,U. (1997) Cellular mechanisms of long-term potentiation: late maintenance. In Neural-Network Models of Cognition. Ed. by Donahue,J.W., Dorsel, V.P.North-Holland (Elsevier)

40. Geneisman, Y., de Toledo-Morell, L, Morell, F., Heller, R.E., Rossi, M & Parshall, R.F. (1993). Structural synaptic correlate of long-term potentiation: formation of axospinous synapses with multiple, completely partitioned transmission zones. Hippocampus. 3, 435-446.

41. Gil,Z., Connors, B.W. and Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity. Neuron. 19, 679 - 686

42. Gilbert, C.D. (1977). Laminar differences in receptive field properties of cells in cat primary visual cortex. J. Physiol. 268, 391-421

43. Guido, W., Lu,S.M., Vaughan, J.W. Godwin, D.W. and Sherman, S.M. (1995) Receiver operating characteristic (ROC) analysis of neurons in the cat's lateral geniculate nucleus during tonic and burst response mode. Visual Neuroscience 12, 723-741

44. Guido, W. and Weyand,T. (1995). Burst responses in thalamic relay cells of the awake behaving cat. J. Neurophysiol. 74, 1782-1786

45. Guillery,R.W. (1995). Anatomical evidence concerning the role of the thalamus in corticocortical communication: a brief review. J. Anat. 187, 583-592

46. Hansel,D. and Sompolinsky,H. (1998). Modeling feature selectivity in local cortical circuits. In "Methods in Neuronal Modeling: From Ions to Networks". Ed. Koch, C & Segev, I. MIT Press Cambridge.

47. Hastings,A. (1997) Population Biology: Concepts and Models. Springer New York

48. Haykin, S. (1994). Neural Networks. A Comprehensive Foundation. Macmillan, New York

49. Hebb, D.O. (1949). The Organization of Behavior. Wiley, New York.

50. Hernandez-Cruz,A. and Pape, H.C. (1989). Identification of two calcium currents in acutely dissociated neurons from the rat lateral geniculate nucleus. J. Neurophysiol. 61, 1270-1283

51. Hirsch JA. Gallagher CA. Alonso JM. Martinez LM. (1998)Ascending projections of simple and complex cells in layer 6 of the cat striate cortex. Journal of Neuroscience. 18, 8086-8094

52. Hobson, J.A. and Stickgold, R. (1997) The conscious state paradigm: a neurocognitive approach to waking, sleeping and dreaming. 1373-1389 In: The Cognitive neurosciences, Ed Gazzaniga,M.S.

53. Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Aced. Sci. 79, 2554.

54. Hubel,D.H. and Wiesel,T. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106-154

55. Hubel,D.H.(1995) Eye, Brain and Vision. Scientific American Library, New York

56. Isaac, J.T.R. , Crair, M.C., Nicoll, R.A. and Malenka, R.C. (1997). Silent synapses during development of thalamocortical inputs. Neuron, 18, 269-280

57. Isaac, J.T.R., Nicoll, R.A. and Malenka, R. (1995) Evidence for silent synapses: implications for the expression of LTP. Neuron 15, 427-434

58. Jahnsen, H. & Llinas, R. (1984). Electrophysiological properties of guinea-pig thalamic neurones: an in vitro study. J. Physiol. 389, 187-203

59. Jones,E.G. & Peters, A. Cerebral Cortex: Further Aspects of Cortical Function, Including Hippoampus. Plenum.

60. Jones,E.G. (1998) Viewpoint: the core and matrix of thalamic organisation. Neuroscience 85, 331-345

61. Kaas,J.H. , Florence, S.L. and Jain, N. (1999) Subcortical contributions to massive cortical reorganisations. Neuron 22, 657-660

62. Karni, A, Tanne, D, Rubenstein, B.S. , Askenasy, J.J.M. & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. Science, 265, 679-682.

63. Katz, B. (1971). Quantal mechanism of neural transmitter release. Science 173, 123-126

64. Katz,L.C. (1987). Local circuitry of identified projection neurons in cat visual cortex brain slices. J. Neurosci. 7,1223-1249

65. Kim, U., Bal, T., & McCormick, D.A. (1995). Spindle waves are propagating synchronised oscillations in the ferret LGNd in vitro. J. Neurophysiology, 74, 1301-1323.

66. Kinney, G.A. & Slater, N.T. (1993). Potentiation of NMDA receptor-mediated transmission in turtle cerebellar granule cells by activation of metabotropic glutamate receptors. J. Neurophysiol. 69, 585-59

67. Kirkwood, A., Rioult.M.G. and Bear, M.F. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. Nature 381, 526-528

68. Koch,C. (1999). Biophysics of Computation. Oxford University Press, New York

69. Koch, C. and Zador, A. (1993). The function of dendritic spines devices subserving biochemical rather than electrical compartmentalisation. J. Neurosci. 13,413-422

70. Konig, P, Engel, A.K. and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. Trends. Neurosci. 19, 130-137

71. Kullmann, D.M. & Sigelbaum, S.A. (1995) The site of expression of NMDA receptor-dependent LTP: new fuel for an old fire. Neuron, 15, 997-1002

72. Kullman, D.M., Erdemli, G. and Asztely,F. (1997) LTP of AMPA and NMDA receptor-mediated signals: evidence for presynaptic expression and extrasynaptic glutamate spillover. Neuron, 17, 461-474

73. Larkum,M.E. , Zhu, J.J. and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. Nature. 398, 338-341

74. Liao, D., Hessler, N.A. & Malinow, R. (1995). Activation of postsynaptically silent synapses during pairing-induced LTP in CAl region of hippocampal slice. Nature, 375, 400 – 404

75. Lisman,J.E. (1997) Bursts as a unit of neural information: making unreliable synapses reliable. Trends Neurosci. 20, 38-43

76. Llinas,R.R. and Pare,D. (1991) Of dreaming and wakefulness. Neuroscience.44, 521-535

77. Llinas,R.R., Ribary, U., Contreras, DE. and Pedroarena, C. (1998). The neuronal basis for consciousness. Proc. Roy. Soc. B, 353, 1841-1849

78. Magee, J.C. and Johnston, D. (1997) A synaptically controlled, associative signals for hebbian plasticity in hippocampal neurons. Science, 275, 209-213

79. Mainen,Z.F. and Sejnowski, T.J. (1995) Reliability of spike timing in neocortical neurons. Science, 268, 1503-1505

80. Maletic-Savatic M. Malinow R. Svoboda K. (1999) Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. Science. 283(5409):1923-7

81. Malsburg, C. von der (1972) Self-organisation of orientation sensitive cells in the striate cortex. Kybernetik. 14 85-100

82. Malsburg, C. von der (1998) Self-organisation and the brain. In : The Handbook of Brain Theory and Neural Networks. Ed. Arbib, M.A. pp 840-843. MIT Press, Cambridge.

83. Malsburg, C. von der and Willshaw, D.J. (1980). Differential equations for the development of topological nerve fibre projections. SIAM-AMS Proceedings 13, 39-47

84. Markram, H., Lubke, J, Frotscher, M. and Sakmann, B. (1997) Regulation of synaptic efficacy by coincidence of postsynaptic Aps and EPSPs. Science 275, 213-215

85. Matthews G. (1996) Further observations on transneuronal degeneration in the lateral geniculate nucelus of the macaque monkey. J. Anat. Lond., 98, 255-263

86. Markram,H., Lubke, J., Frotscher,M. Roth, A. and Sakmann,B. (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J. Physiol. 500, 409 -440

87. Mazzoni, P, Anderson, R.A. & Jordan, M.J. (1991). A more biologically plausible learning rule for neural networks. Proc. Nat. Acad. Sci. 88, 4433-4437.

88. McClelland, J. L., McNaughton, B.L. and O'Reilly, R. C.. (1995). Why are there complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Review 102, 419-457.

89. McCormick, D.A. and Huguenard, J.R. (1992). A model of the electrophysiological properties of thalamocortical relay neurons J. Neurophysiol. 68, 1384-1400.

90. McCormick, D.A. & Krosigk, M. von. (1992) Corticothalamic activation modulates thalamic firing through glutamate "metabotropic" receptors. Proc Nat. Acad. Sci. 89, 2774 - 2778.

91. 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

92. Mountcastle,V.B. (1998) The Cerebral Cortex. Harvard University Press, Cambridge, MA.

93. Murphy, P.C. and Sillito, A.M. (1996). Functional morphology of the feedback pathway from area 17 of the cat visual cortex to the lateral geniculate nucleus. J. Neurosci. 16, 1180 –1192

94. Murthy,V.N. (1997). Synaptic plasticity: neighborhood influences. Current Biology, 7, 512-515

95. Nicoll, R.A. and Malenka, R.C. Contrating properties of two forms of long-term potentiation in the hippocampus. Nature 377, 115-118

96. Pei, X., Vidyasagar, T.R., Volgushev, M. and Creutzfeldt, O.D. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. J. Neurosci. 14, 7130-7140

97. Petersen, C.C.H., Malenka, R.C., Nicoll, R.A. and Hopfield, J.J. (1998). All-or-none potentiation at CA3-CA1 synapses. Proc. Natl. Acad. Sci. 95, 4732-4737

98. Rall, W. (1989) Cable Theory for Dendritic Neurons. In "Methods in Neuronal Modeling", ed. Koch, C and Segev, I. MIT Press, Cambridge

99. Reid,R.C. and Alonso, J.M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. Nature, 380, 281-284

100. Reid, R.C. and Alonso, S.-M. (1996). The processing and encoding of information in the visual cortex. Current Opinion. Neurobiol. 6, 475-480

101. Reinagel,P.M., Guido,W., Koch,C. and Sherman, M.S. (1999) Encoding of visual information by LGN bursts. J. Neurophysiology. 81,2558-2569

102. Riesenhuber,M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neurosci. 2, 1019-1025

103. Ringach, D.L., Hawken, MJ. and Shapley, R. (1997) Dynamics of orientation tuning in macaque primary visual cortex. Nature, 387, 281-2

104. Rojas, R. (1996) Neural Networks: A Systematic Introduction . Springer Berlin

105. Scharfman, H.E., Lu, S.M., Guido, W., Adams, P.R. & Sherman, S.M. (1990). N-methyl-D-aspartate (NMDA) receptors contribute to excitatory postsynaptic

potentials of cat lateral geniculate neurons recorded in thalamic slices. Proc. NatI. Acad. Sci. USA, 87: 4548-4552.

106. Schiller, J., Schiller, Y. , Stuart, G. and Sakmann,B. (1997). Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. J. Physiol. 505, 605-616

107. Schuman E.M. & Madison DV. (1994) Locally distributed synaptic potentiation in the hippocampus. Science. 263, 532-536

108. Shadlen, M.N. & Newsome, W.T. (1994) Noise, neural codes and cortical organisation. Current Opinion in Neurobiology. 4, 569 - 579.

109. Shadlen, M.N. and Newsome, W.T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and information coding. J. Neurosci. 18, 3870-3896

110. Sherman, S.M. & Guillery, S.M. (1996). Functional organisation of thalamocortical relays. J Neurophysiol. 76, 1367-1395

111. Sherman, S.M. (1996) Dual response modes in lateral geniculate neurons: mechanism and fuctions. Visual. Neurosci. 13, 205-213

112. Sherman, S.M. & Koch, C. (1998) Thalamus. In: The Synaptic Organisation of the Brain. edited by G.M. Shepherd. Oxford.

113. Sherman,S.M. and Guillery, R. (1998) On the actions that one nerve cell can have on another. Proc. Natl. Acad. Sci. 95, 7121-7126.

114. Sillito AM. Jones HE. Gerstein GL. and West DC.(1994) Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex Nature. 369,479-482

115. Somers, D., Nelson, S. and Sur., M. (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neuroscience 15, 5448-5465

116. Sorra, K.E., Fiala,J.C. & Harris, K.M. Critical assessment of the involvement of perforations, spinules, and spine branching in hippocampal synapse formation. J. Comp. Neurol. 398, 225-240. (1998)

117. Steriade,M. and Contreras,D. (1996) Relations between cortical and thalamic cellular events during transition from sleep patterns to paroxysmal activity. J. Neurosci. 15, 623-642

118. Steriade, M., Jones, E.G. and Llinas, R. (1990). Thalamic Oscillations and Signalling. Wiley and Sons, New York.

119. Steriade, M, McCormick, D.A. & Sejnowski, T. J. (1993) Thalamocortical oscillations in the sleeping and aroused brain. Science, 262, 679-685.

120. Stuart, G. . & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature, 367, 69-72.

121. Stuart,G., Schiller, J and Sakmann, B. (1997). Action potential initiation and propagation in rat neocortical pyramidal neurons. J. Physiol. 505, 617-632

122. Swindale,N. (1994). Looking into a Klein bottle. Current Biology.

123. Thimm,G., Moerland,P. and Fiesler,E. (1996). The interchangeability of learning rate and gain in backpropagation neural networks. Neural Computation 8, 451-460

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

125. Tsubokawa,H. and Ross, W.D. (1997). Muscarinic modulation of spike backpropagation in the apical dendrites of hippocampal CA1 pyramidal neurons. J. Neurosci. 17, 5782-5791

126. Tsumoto, T., Creutzfeldt, O.D. and Legendy, C.R. (1978). Functional organisation of the corticofugal system from visual cortex to lateral geniculate nucleus in the cat. Exp. Brain. Res. 32, 345-364

127. Turrigiano,G.G., Leslie,K.R., Desai,,N.S., Rutherford, L.C. and Nelson, S.B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 391, 892-896

128. White,E.L. (1989). Cortical Circuits: Synaptic Organisation of the Cerebral Cortex: Structure, Function and Theory. Birkhauser.

129. Willshaw, D.J. and Malsburg, C. von der (1976). How patterned neural connections can be set up by self-organisation. Proc. R. Soc. B. 194, 431-445.

130. Willshaw,D.J. & von der Malsburg, C. A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retino-tectal problem. Phil.Trans. R. Soc. Lond B 287, 203-243 (1979).

131. Wong, R.O.L. (1993). The role of spatio-temporal firing patterns in neuronal development of sensory systems. Curr. Opin. Neurobiol. 3, 595-601.

132. Zeki, S. (1993). A Vision of the Brain. Blackwell, Oxford

133. Zhang, Z.-W. and Deschenes,M. (1997). Intracortical axonal projections of lamina VI cells of the primary somatosensory cortex in the rat: a single-cell labelling study. J. Neurosci. 17, 6365-6379

134. Zhou, Q., Godwin,D., O'Malley, D.M. and Adams,P.R. (1997) Visualisation of calcium influx through channels that shape the burst and tonic firing modes of thalamic relay cells. J.Neurophyiol. 77, 2816-2825

135. Ziv, N.E. & Smith, S.J. (1996). Evidence for a role of dendritic filopodia in synaptogenesis and spine formation. Neuron, 17, 91-102.

Figure Legends

Fig 1. A sketch of neocortex, based on cat striate cortex, but with borrowings from other areas and species. Relay cells in the lgn are shown in yellow at the top (marked T). These either receive driver inputs from subcortical sources (shown black) or from layer 5 of lower order cortex (shown orange). Two classes of relay cell are shown, using a distinction made in monkey cortex[60]. The "core" class, (shown on the left) innervates mainly middle layers of cortex, in this case spiny stellate cells (purple) in layer 4, and also layer 6 pyramidal cells (green). The "matrix" class (shown at center and right in layer T) preferentially innervate layer 1, where they are shown synapsing on apical tufts of certain pyramidal cells (those in layers 2,3 and 5, and the claustrally-projecting layer 6 cells marked 6/C). (However, there may also be matrix-like cells that do not preferentially innervate layer 1). Note that it is likely that relay cells receive either subcortical or layer 5 driving input, not both as shown. Layer 4 spiny stellates provide feedforward input to layer 2/3 pyramidal cells (blue), which provide input to layer 5 cells (orange) and to higher cortex (broken blue line and arrow). Some layer 5 cells provide output to subcortical sites, including relay cells in higher order thalamic nuclei. Spiny stellate cells, and layer 2, 3 and 5 pyramidal cells provide input to layer 6 pyramidal cells.

Non-claustral layer 6 cells collateralise extensively in upper layers (green arrows). There appear to be 3 classes of non-claustral layer 6 cell. The simple class (marked 6/4) collateralise in layer 4. The complex classes (6/3 and 6/5) collateralise in layers 2/3 or 5. All types of cell are shown with within-layer recurrent excitation (curved red arrows). This is not autaptic. Relay cells receive feedback from non-claustral layer 6 cells (green arrows in layer T). Certain details of the arrangement in the diagram are speculative. First, it is not certain that 6/4 cells receive input from that set of relay cells that provide input to layer 4 simple cells (though that would explain the simplicity of 6/4 cells), nor is it certain that only 6/4 cells get input from spiny stellates (though again this would explain their simplicity). Second, it is not certain that 6/3 cells get selective input from spiny stellates and from layer 2/3 cells (though that would explain their complexity), and correspondingly for 6/5 cells. Third, the proximal and distal dendritic placement of the feedforward input to layer 6 cells shown is symbolic only: it represents the postulated correlation sensitivity. Fourth, the upper layer collaterals of layer 6 cells are shown influencing the recurrent collaterals (red arrows) of upper layer cells, while anatomically they form drumstick synapses on dendritic shafts of upper layer spiny cells. Fifth, it is not known whether the core class of relay cell receives input preferentially from layer 6 simple cells, and the matrix class from complex layer 6 cells, as shown. Note that the exact pattern of inputs that underlies information processing is not shown (see Fig 5), and inhibitory cells are omitted.

Figure 2. A model for the formation of new synaptic connections. A layer of presynaptic cells can connect to a layer of postsynaptic cells. In the top part, an existing connection, composed of 3 synapses (solid) , is shown. As a result of correlated or synchronous firing, an additional synapse is created. This new synapse could appear either on the postsynaptic component of the connection (show bottom left) or on a neighbor of that postsynaptic cell (an example is shown bottom right as an open circle). Both the creation of new synapses and their misplacement occur stochastically, the former with probability w, and the latter with probability E. The former but not the latter varies across the array (according to the strengths of correlations across the connections). The text discusses the simple case where fitness of one possible

connection (w_m) is higher than that of surrounding possible connections (w_p; see fitness profile sketched at the top of the figure).

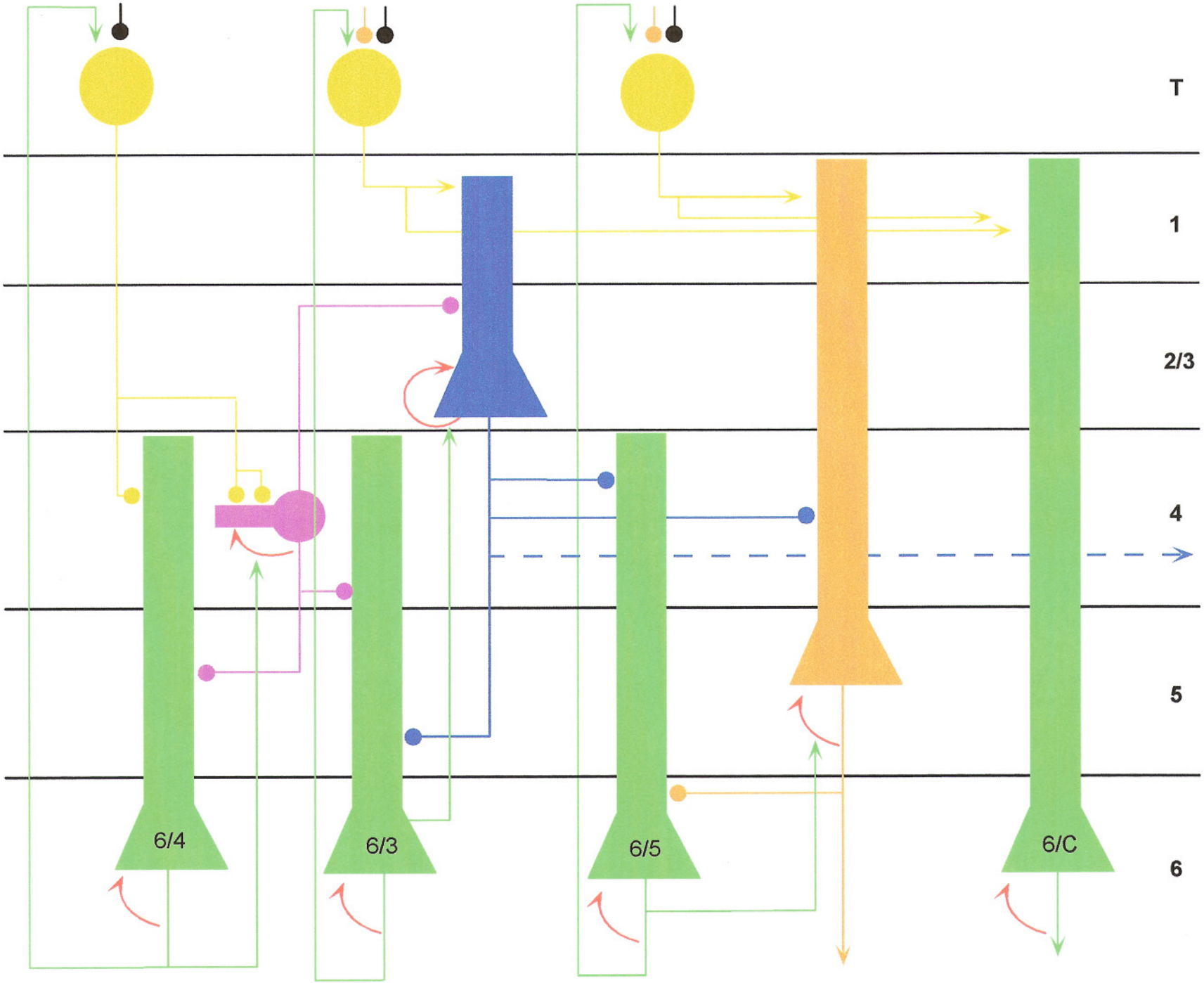
Figure 3. Postulated circuit allowing containment of heterosynaptic, presynaptic, errors. The central top layer presynaptic cell, labelled J_0, has, as a result of previous patterned activity in the J and I layers, formed a strong connection on the central postsynaptic cell I_0 (solid dots). If this connection should again become plastic, and if there is again correlated firing of J_0 and I_0, this connection could strengthen (or at least turn over), and heterosynaptic replication errors could generate errant new synapses on I_{-1} or I_1 (shown dotted with small open circles). I_{-1} and I_1 represent schematically the neighbors of I_0. A third layer of cells, marked K, act as correlation detectors. Thus K_0 receives input from both J_0 and I_0, and is excited whenever these 2 input cells fire nearly together, but not otherwise. Its excitation thus represents w_m. Similarly, K_{-1} computes the correlation of J_0 with I_{-1}, and K_1 that of J_0 with I_1, and thus w_p. The proximal and distal placement of synapses on the layer K cells is mostly symbolic – it represents correlation detection. The horizontal gray arrows in the K layer represent lateral interactions which allow K_0 to compute w_m/w_p ("correlation sharpness"). This cell then fires if w_m/w_p exceeds a critical value, enabling the presynaptic plasticity of the connections made by J_0 (black arrow in layer J). This arrangement minimises the probability that errant synapses (shown as small circles) will survive, and ensures precise connections. In the text layer J is interpreted as thalamic relay cells, layer I as spiny stellate cells in layer 4, and layer K as simple layer 6 cells.

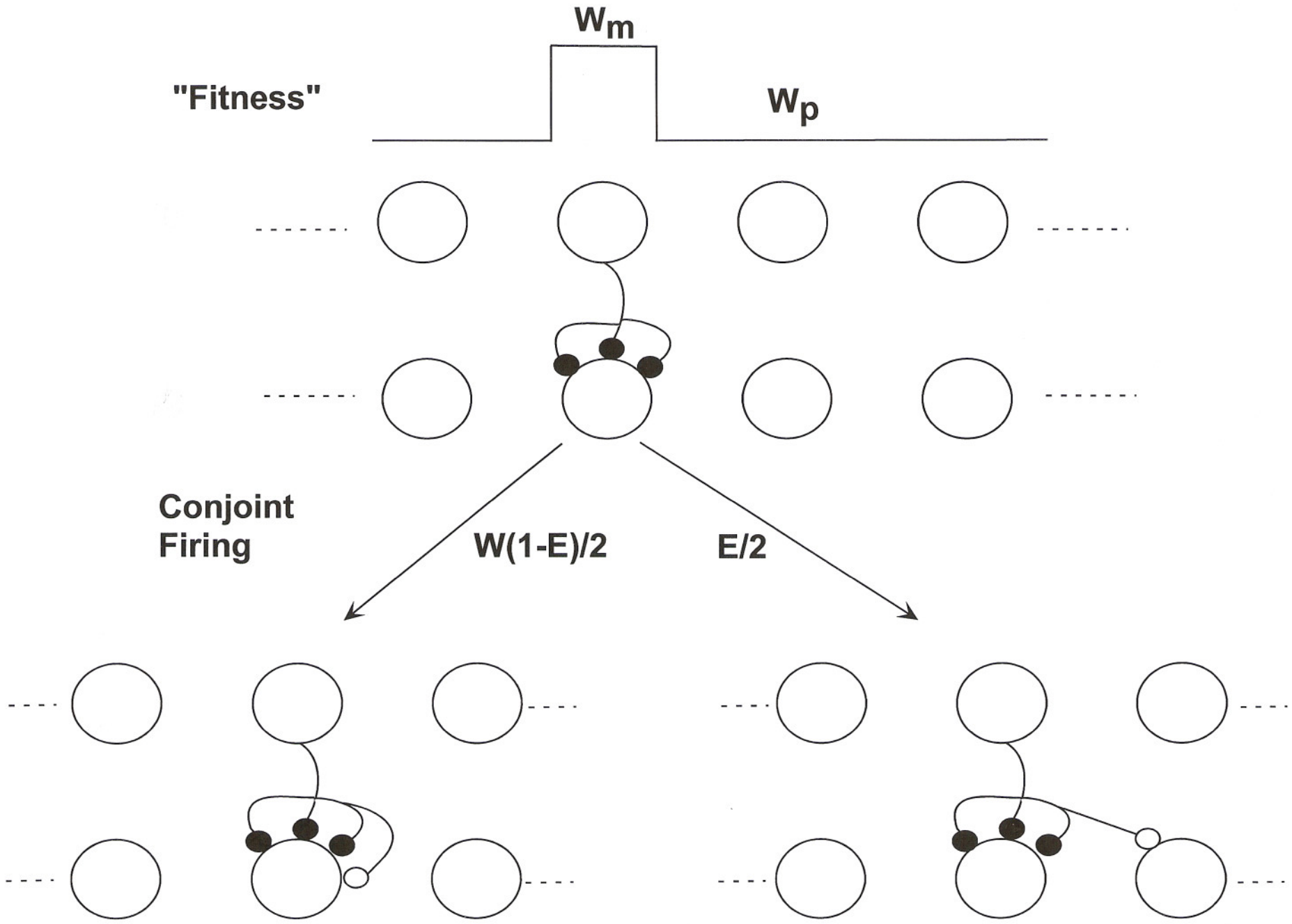
Fig 4. Postulated circuit to contain postsynaptic errors. In this case the presynaptic cell J_0 has again formed a strong connection on the postsynaptic cell I_0. However, now plasticity is postsynaptic, so that heterosynaptic strengthening errors of the existing connection could result in the formation of synapses with either J_{-1} or J_1 (dotted line and small open circles). K_0 computes w_m, and K_{-1} and K_1 compute w_p, as before, and K_0 fires when w_m/w_p exceeds a threshold. However, now the firing of K_0 enables the postsynaptic plasticity of I_0 (solid arrow). Layer J corresponds to layer 4 (or 2/3), layer I to layer 2/3 (or 5) and layer K to layer 6 complex cells.

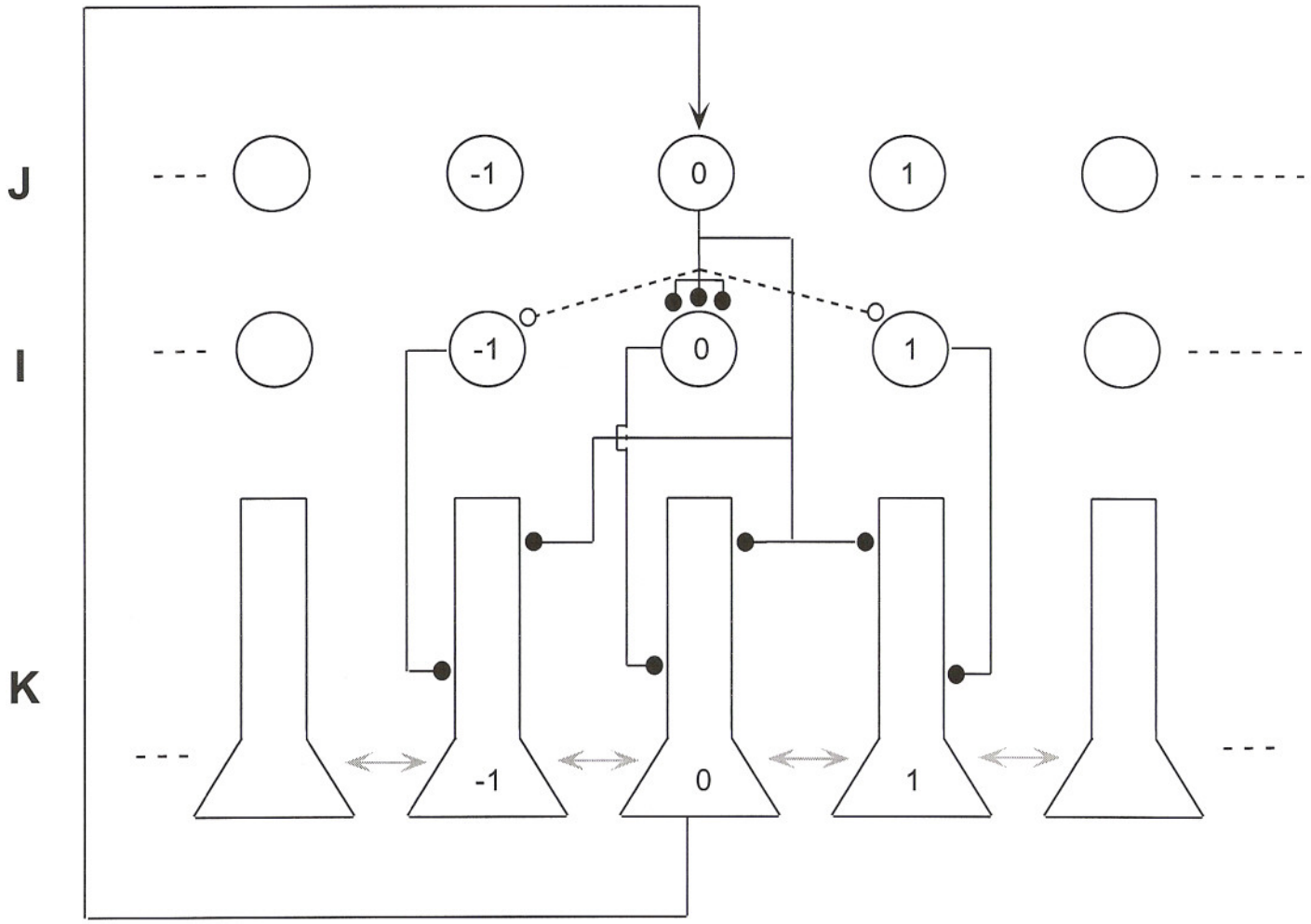
Fig 5. A more realistic view of the organisation of layer 6 circuitry controlling presynaptic plasticity. In this case J cells are interpreted as LGN relay cells (top row, yellow, marked T), I cells are interpreted as layer 4 spiny stellate simple cells (purple), and K cells as layer 6 simple cells (green). The diagram represents the feedforward connections (yellow T cells, equivalent to J_0 in Fig 3) underlying the orientation sensitivity of a layer 4 simple cell (show as solid purple, equivalent to I_0). The nearest neighbors of this layer 4 simple cell (equivalent to I_{-1} and I_1) are shown as dashed purple circles – these cells could receive synapses if erroneous strengthening of any of the current T to 4 connections strengthen. (The 4 selected are symbolic only; the neighborhood may encompass other cells). The vertical 4 to 6 connections (purple) are permanent - they connect topographically corresponding cells, and define the cortical column. These vertical connections identify a set of cells in layer 6 which are the neighbors of the layer 6 cell (solid green, equivalent to K_0) which receives a vertical connection from the simple cell which currently receives input from the set of boxed relay cells. These "neighboring" layer 6 cells (equivalent to K_{-1} and K_1) are shown as dashed green circles. Both solid and dashed green circles receive thalamocortical input (curved yellow lines) from the set of relay cells (boxed) that comprise the receptive field of the layer 4 simple cell. The solid layer 6 cell then feeds back to the boxed relay cells, and controls the presynaptic plasticity of all the layer 4 connections that these cells make (green arrows). This figure should be compared to Fig. 3. Note that a similar logic would underlie the actual connections of layer 6 complex cells (see Fig 4 and text).

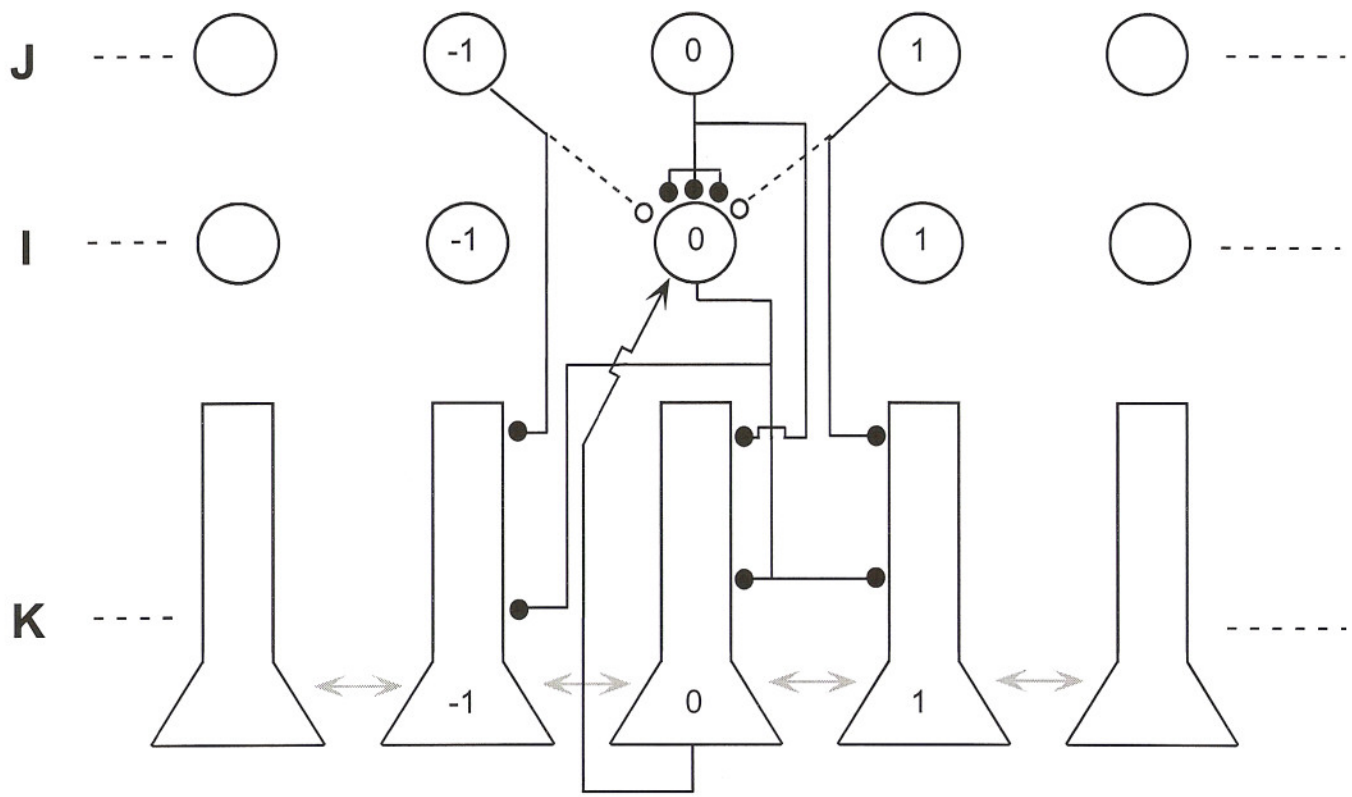
Fig 6. Offline updating of the connections of the correlation layer (K, here corresponding to layer 6 simple cells) following an allegiance transfer. Part A (left) shows the situation before the allegiance transfer, and corresponds to Fig 3. The connections shown dotted have to be broken after the allegiance transfer. Part B shows the new J to I connection after the allegiance transfer, which occurred because online, daytime learning expoited a heterosynaptic error. The dashed connections must be created. After the allegiance transfer, and updating of the K layer connections, the situation is

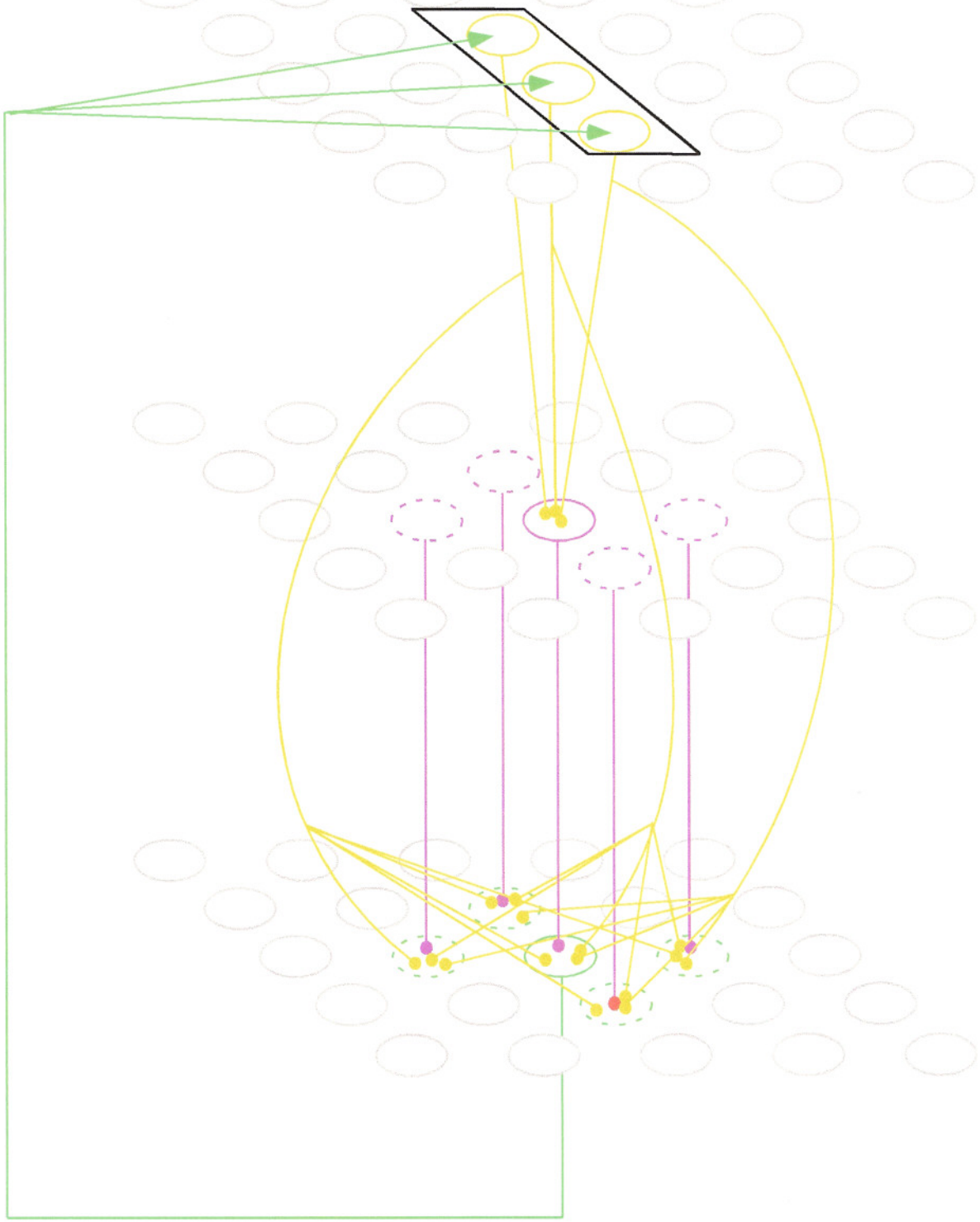
once again equivalent to Fig 3, so that future synaptic errors can be contained. The figure does not show intermediate stages in the allegiance transfer, where J_0-I_0 and J_0-I_1 synapses coexist. During this stage lateral connections within layer 6 (not shown) have to be transiently reorganised, so that both K_0 and K_1 contribute to the measurement of w_m, and K_{-1} and K_{+1} to w_p. This reorganisation can be accomplished offline if the lateral connections are Hebbian.

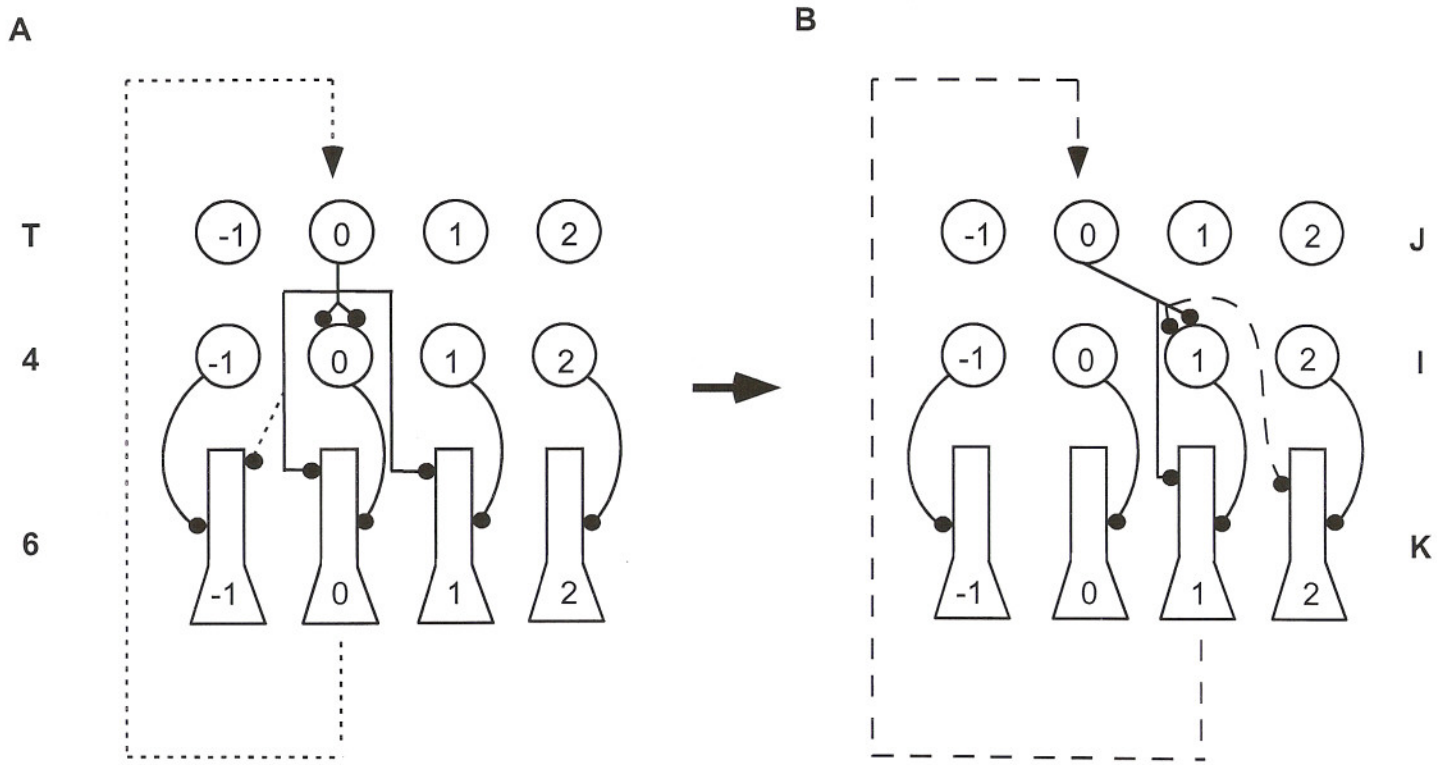

T
1
2/3
4
5
6
6/4
6/3
6/5
6/C


W_m
"Fitness"
W_p
Conjoint
Firing
W(1-E)/2
E/2


J
I
K
-1
0
1
-1
0
1
-1
0
1


J
I
K
-1
0
1
-1
0
1
-1
0
1


T
4
6


A
B
T
4
6
J
I
K
-1
0
1
2