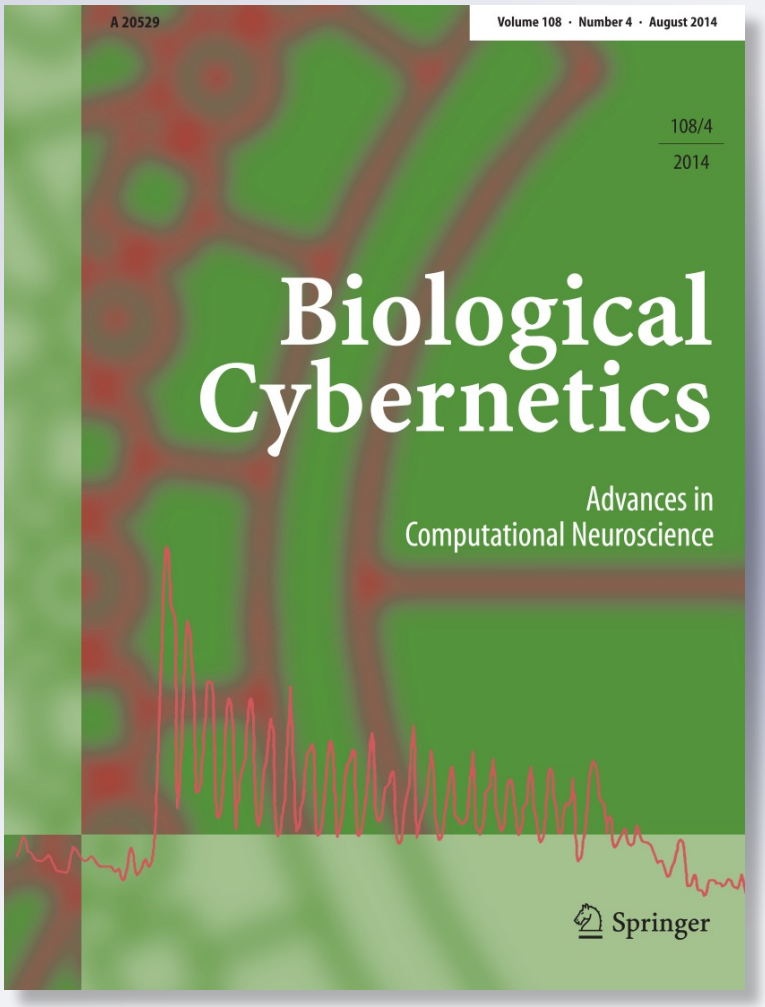
Hebbian learning from higher-order correlations requires crosstalk minimization

K. J. A. Cox & P. R. Adams

Biological Cybernetics
Advances in Computational Neuroscience

ISSN 0340-1200
Volume 108
Number 4

Biol Cybern (2014) 108:405-422
DOI 10.1007/s00422-014-0608-4



Springer

Biol Cybern (2014) 108:405–422
DOI 10.1007/s00422-014-0608-4

ORIGINAL PAPER

Biological Cybernetics

Hebbian learning from higher-order correlations requires crosstalk minimization

K. J. A. Cox · P. R. Adams

Received: 30 September 2012 / Accepted: 6 May 2014 / Published online: 27 May 2014

Abstract Activity-dependent synaptic plasticity should be extremely connection specific, though experiments have shown it is not, and biophysics suggests it cannot be. Extreme specificity (near-zero "crosstalk") might be essential for unsupervised learning from higher-order correlations, especially when a neuron has many inputs. It is well known that a normalized nonlinear Hebbian rule can learn "unmixing" weights from inputs generated by linearly combining independently fluctuating nonGaussian sources using an orthogonal mixing matrix. We previously reported that even if the matrix is only approximately orthogonal, a nonlinear-specific Hebbian rule can usually learn almost correct unmixing weights (Cox and Adams in Front Comput Neurosci 3: doi:10.3389/neuro.10.011.2009 2009). We also reported simulations that showed that as crosstalk increases from zero, the learned weight vector first moves slightly away from the crosstalk-free direction and then, at a sharp threshold level of inspecificity, jumps to a completely incorrect direction. Here, we report further numerical experiments that show that above this threshold, residual learning is driven instead almost entirely by second-order input correlations, as occurs using purely Gaussian sources or a linear rule, and any amount of crosstalk. Thus, in this "ICA" model learning from higher-order correlations, required for unmixing, requires high specificity. We compare our results with a recent mathematical analysis of the effect of crosstalk for exactly orthogonal mixing, which revealed that a second, even lower, threshold, exists below which successful learning is impossible unless weights happen to start close to the correct direction. Our simulations show that this also holds when the mixing is not exactly orthogonal. These results suggest that if the brain uses simple Hebbian learning, it must operate with extraordinarily accurate synaptic plasticity to ensure powerful high-dimensional learning. Synaptic crowding would preclude this when inputs are numerous, and we propose that the neocortex might be distinguished by special circuitry that promotes extreme specificity for high-dimensional nonlinear learning.

Keywords Crosstalk · Neocortex · Hebbian · Bifurcation · Learning

K. J. A. Cox (✉)· P. R. Adams
Department of Neurobiology and Behavior, Stony Brook University, Stony Brook, NY 11794–5230, USA
e-mail: kcox@syndar.org

1 Introduction

It is widely believed that activity-dependent adjustments of the strengths, or "weights", of synaptic connections contribute significantly to learning underlying cognition and behavior. Such changes may reflect a combination of new synapse creation or elimination (in extreme cases, forming new connections or removing old ones), unsilencing or resilencing of existing silent synapses, and graded or discrete changes in individual synapse strengths. Such changes would occur partly in response to local signals, such as the conjoint activity of pre- and/or postsynaptic neurons, in a Hebbian manner. Hebbian learning is driven by input–output firing correlations and therefore by input statistical regularities, in a connection-specific manner. However, the chemicals mediating synaptic plasticity inevitably diffuse, which, combined with the high synapse density needed when neurons receive many inputs, makes it difficult or impossible that the adjustments could be completely connection specific. For example, in the case of NMDAR-mediated LTP, it is thought that the narrow spine neck, endowed with calcium pumps (Feng et al. 2007), reduces calcium escape to the shaft and thence to neighboring synapses (Koch and Zador 1993; Sabatini et

al. 2002; Wickens 1988; Yuste and Denk 1995), minimizing inspecificity or "crosstalk." However, the spine neck must also be wide enough to ensure that synaptic currents reach the dendritic shaft. Related conflicts may limit effective use of synapse-like nanoscale memristors (Kim et al. 2011; Vontobel et al. 2009).

It has been suggested that adequate electrical coupling and chemical isolation can both be achieved (Koch and Zador 1993) but data suggest that neither is perfect (Araya et al. 2006; Matsuzaki et al. 2004; Noguchi et al. 2005; Palmer and Stuart 2009), and that compromise is inevitable. Indeed, classical LTP is not completely synapse specific (Bi 2002; Bonhoeffer et al. 1989; Engert and Bonhoeffer 1997; Harvey and Svoboda 2007), and it seems unlikely that other forms of Hebbian learning, such as LTD, could be either (Reynolds and Hartell 2000). It is possible that under such circumstances at least certain types of learning, involving repeated gradual exposure to patterned stimuli might completely fail as a result of the accumulation of small errors, even though they would succeed if adjustments were more accurate, in a manner analogous to the way natural selection fails if polynucleotide copying is insufficiently accurate (Eigen 1971b). We are interested in the possibility that certain hitherto poorly understood neocortical microcircuits might be involved in preventing such an "error catastrophe" (Adams and Cox 2002a,b; 2006), thereby facilitating sophisticated learning underlying complex cognition and behavior. The ultimate potential of "machine learning" to augment or emulate human intelligence might also depend on the development of massively parallel hardware that can accurately read and adjust connection weights (Likharev 2008); issues of weight readout or adjustment independence may also become paramount here. Learning often involves combinatorial explosions which might only be tamable if connections can be read and updated in massively parallel and specific fashion. A fundamental problem for effective high-dimensional learning might simply be ensuring that weight adjustments do not directly and adversely affect each other despite inevitable close-packing.

Perhaps very small inaccuracies could be neglected, but even tiny errors might gradually compound (Eigen 1971a). We recently studied (Cox and Adams 2009; Radulescu et al. 2009) the behavior of some simple classical "connectionist" Hebbian rules, modified to reflect possible connectional inspecificity in weight updating, "crosstalk." We made the simplest possible assumption: Weight update vectors are modified by multiplication by an "error matrix" $\mathbf{E}$, which reflects the way that the update initiated by local activity is redistributed, on average, over all the available synaptic connections. Standard neural network learning would correspond to the special case $\mathbf{E} = \mathbf{I}$ (the identity matrix). We have focused on unsupervised learning, because it seems likely that this is important in the cortex (Hinton and Sejnowski 1999), and involves repeated slow adjustment driven by input statistical regularities. We neglect the role of spike timing and use positive or negative, continuous, activities and weights. The situation in the brain would be more complicated, but simple connectionist models do seem to capture basic aspects of neural learning, such as representation of the underlying hidden causes of the input data stream.

Naïve correlation-driven Hebbian learning is unstable, but can be stabilized by preventing unlimited weight increases or decreases. We have considered 2 versions of stabilized Hebbian learning, either linear or nonlinear in the postsynaptic activity (Adams and Cox 2002a; Cox and Adams 2009; Radulescu et al. 2009). Of course, the overall rule is always nonlinear, because of assumed stabilizing normalization, but throughout this paper we use "linear" to refer to the case where the Hebbian part is linear in the postsynaptic activity. In the linear case, specific learning typically converges to the dominant eigenvector of the input covariance matrix $\mathbf{C}$ (Oja 1982). With crosstalk, learning converges to the dominant eigenvector of $\mathbf{EC}$ (Botelho and Jamison 2004; Radulescu et al. 2009); this can deviate significantly from that of $\mathbf{C}$, especially at high error or when the eigenvalues of $\mathbf{C}$ are quite close to each other (weakly patterned correlation). However, except in the special "unbiassed" case, when inputs do not privilege particular connections, a situation corresponding to development rather than learning, there are typically no bifurcations in the dynamics at critical error levels, and the outcome is qualitatively unaffected by crosstalk. The unbiassed case leads to either equal weights or broken-symmetric "segregation"; the weight-equalizing effect of crosstalk tends to favor the former (Radulescu and Adams 2013); see also (Elliott 2012).

However, sophisticated cognition and learning seem to require higher-order statistics (Field 1994), which give clues to the underlying causes of observations. A very simple, transparent and popular assumption has been the ICA model, where the inputs are generated by linearly mixing, via a square matrix $\mathbf{M}$, independently fluctuating "sources" at least one of which has a nonGaussian distribution (Amari et al. 1996, 1997; Amari 1998; Bell and Sejnowski 1997; Hoyer and Hyvarinen 2000; Hyvarinen and Hoyer 2000; Hyvarinen et al. 2001). In particular, a neuron using an accurate, normalized, correctly signed nonlinear Hebbian rule can always successfully learn a row of the inverse of an orthogonal $\mathbf{M}$, $\mathbf{M}^{-1}$, so its output "tracks" the fluctuations of a corresponding non-Gaussian source (Hyvarinen et al. 2001; Hyvärinen and Oja 1998). Orthogonal mixing of independent sources maintains the absence of pairwise input correlations; pairwise decorrelation could plausibly be achieved by suitable preprocessing (Atick and Redlich 1990; Srinivasan et al. 1982; Kuang et al. 2012). When the neuron's weight vector lies in the direction of an unmixing row, it lies parallel to the column of $\mathbf{M}$ that matches the corresponding source, so that average changes

in the weight vector must also lie in that direction and are unaffected by the behavior of the other sources. Imposing a constraint on the length of the weight vector and choosing an appropriate sign for the learning rule (Hebbian or antiHebbian, depending on the source distribution and the nonlinearity) stabilizes this equilibrium. Thus, a completely accurate nonlinear learning rule allows the output to exactly recover an underlying "cause" of the inputs, a fluctuating source. Our recent simulations suggest that this is no longer exactly true if the rule is inaccurate, and above a sharp error threshold, no longer true at all (Cox and Adams 2009). Furthermore, we found that even if **M** is not perfectly orthogonal, while an accurate rule still almost always converges close to a column of **M**, it does not do so at all above a threshold crosstalk level. However, above this threshold, the inspecific rule still appeared to converge to definite weights. Here, we clarify the nature of this failure. An illuminating mathematical analysis of an important special case of the ICA-with-crosstalk model recently appeared (Elliott 2012). This analysis assumed that the inputs were perfectly white (i.e., orthogonal mixing). We use this analysis as a guide in presenting and discussing our results.

2 Methods

The basic approach was to simulate nonlinear Hebbian learning by a model neuron driven by input patterns derived by mixing independently fluctuating sources one of which has a nonGaussian distribution (which induces the informative input higher-order correlations (HoCs) see also (Rattray 2002)). We studied only learning by single neurons (and therefore, only "postsynaptic" crosstalk, occurring between different connections on the same neuron, caused for example by dendritic intracellular diffusion of calcium or downstream signals). Presumably similar issues would also affect "presynaptic" crosstalk (Cox and Adams 2009; Schuman and Madison 1994) caused by intra-axonal diffusion between connections made by the *same* axon on *different* postsynaptic neurons. Because such learning is distorted or prevented by noninformative second-order correlations (SoCs) (see Sect. 2.4), ideally the mixing should be "orthogonal", so that inputs SoCs are eliminated ("whitening"). However, since it is unlikely that perfect whitening could be achieved biologically, we somewhat relaxed the strict orthogonality assumption, as well as relaxing the assumption that the Hebb rule is completely synapse specific. The remainder of the Methods details the procedures and necessary background (see also Cox and Adams 2009). It should be noted that related recent mathematical analyses (summarized in Sect. 2.4) by Elliott (2012, and personal communications) provide much needed additional rigor and insight. The published detailed analytic results use more restrictive assumptions, and, for analytic tractability, only a cubic nonlinearity. We consider some aspects of the relation between the analysis and our numerical results in the Sect. 4.

2.1 Mixing matrix, source and input vectors

We use the well-known ICA model (square linear mixing of independent univariate "sources" by an invertible matrix **M**, or $\mathbf{M_o}$ derived therefrom; see below) because it exhibits, in the simplest possible form, the core feature of learning from higher-order correlations HoCs by a nonlinear Hebbian rule. The sources s_i take on successive random real values distributed symmetrically around zero according to a defined distribution, which we take to be either Gaussian or Laplacian (i.e., superGaussian). Note that while it is necessary that all but one source be nonGaussian to recover all the rows of $\mathbf{M^{-1}}$, to learn only one row and thus recover the corresponding source, only that source must be nonGaussian (Rattray 2002). Throughout this paper, we use this one nonGauss source condition.

The input vector **x** was calculated using Eq. (3) below. In this paper, untransposed vectors are column vectors. We use the standard one-unit ICA rule (Hyvärinen and Oja 1998) which requires an orthogonal mixing matrix to guarantee convergence.

It is usually assumed that a suitable preprocessing step (for example, using PCA, or "ZCA", which produces center-surround receptive fields for natural images (Bell and Sejnowski 1997), "whitens" the inputs (i.e., decorrelates and equalizes their variances, so $\mathbf{C} = \mathbf{I}$), such that if they were generated by linear mixing, the effective overall mixing matrix is orthogonal (Hyvarinen et al. 2009; Hyvarinen et al. 2001). However, it seems unlikely that perfect decorrelation could be achieved biologically, partly because of sampling or finite learning rate problems, and partly because crosstalk would also distort any PCA/ZCA-like learning required for preprocessing (Radulescu et al. 2009). Fortunately, we found that in almost all cases (i.e., starting with various randomly generated **M** and weights), perfect whitening is not necessary for good ICA learning, even though it is necessary to guarantee successful learning for arbitrary **M**. We therefore typically relaxed the requirement that the effective **M** be exactly orthogonal, in the following manner (see also Cox and Adams 2009). A small batch (typically 1,000–10,000) of N_B input vectors $\mathbf{x}_B$, generated using a starting mixing matrix **M** whose elements were chosen randomly, with uniform distribution between 0 and 1 [Eq. (1) below], was used to calculate a small-sample covariance matrix $\mathbf{C_B}$. This imperfectly, but unbiassedly, estimates the true **C** (which is defined for an unlimited sample). By varying the batch size N_B, we could vary how well $\mathbf{C_B}$ estimated **C** and thus how efficient the "offwhitening" of the input vectors **x** used for learning would be. New examples of source vectors were

then mixed using $\mathbf{M_o}$ [an approximately orthogonal mixing matrix; Eq. (2) below] where $\mathbf{C_B}^{-1/2}$ is a decorrelating matrix (which is equivalent to ZCA whitening, Bell and Sejnowski 1997) to generate a new, larger batch of N_L offwhite inputs vectors, which was used to drive learning. This procedure mimics what might happen biologically: Early stages of the brain (e.g., in the retina) find approximately decorrelating weights, and then, later stages (e.g., cortex) find weights that reduce higher-order correlations of nearly white inputs. The procedure can be summarized:

$$\mathbf{x_B} = \mathbf{Ms} \tag{1}$$

$$\mathbf{M_o} = \mathbf{C_B}^{-1/2}\mathbf{M} \tag{2}$$

$$\mathbf{x} = \mathbf{M_o s} \tag{3}$$

We did not investigate other whitening matrices (e.g., a PCA-based, rather than ZCA-based, matrix), which would generate slightly different nearly orthogonal mixing matrices $\mathbf{M_o}$ from the same starting $\mathbf{M}$.

We found that under these circumstances, when one starts with random weights, learning would almost always converge (within the limit set by the finite learning rate), in the absence of crosstalk, quite close to a row of $\mathbf{M_o^{-1}}$, provided the batch number N_B was reasonably large ($\geq$1,000), so the input vectors driving learning were well whitened [see also Eq. (9) Sect. 2.4]. But since $\mathbf{M_o}$ is not exactly orthogonal, the appropriate row of $\mathbf{M_o^{-1}}$, the "IC", was not exactly equal to the corresponding column of $\mathbf{M_o}$. We found that the specific rule converged to a direction which was even closer to the appropriate column of $\mathbf{M_o}$ rather than the row of $\mathbf{M_o^{-1}}$. For example with a learning rate $k = 10^{-4}$, for 15 randomly chosen mixing matrices, the average difference of the cosine of the column weight vector angle from exactly 1 (i.e., perfect alignment) was 0.00882, while the row weight vector cosine difference from 1 was 0.01269. When the learning rate was decreased by a factor of 200, both cosines decreased even further, as expected, but the decrease in the column cosine was much greater (fivefold) than in the row cosine, suggesting that the rule converges exactly to the appropriate column of $\mathbf{M_o}$ [see Sect. 2.4 Eq. (9)]. However, our plots show the weight vector direction relative to the true IC, the appropriate row of $\mathbf{M_o^{-1}}$.

Throughout the paper, we used the estimated covariance matrix $\mathbf{C_L}$ based on the actual vectors $\mathbf{x}$ used in the run, rather than the theoretical covariance matrix approached in the large number of vectors limit (which is equal to $\mathbf{M_o M_o^T}$) in order to calculate the eigenvectors of **EC**. For all specific examples illustrated, we give the seed number used to create the relevant initiating random $\mathbf{M}$. If the batch number $\mathbf{N_B}$ was too small, or sometimes if the weights were started exactly at the appropriate eigenvector of $\mathbf{C}$, the rule would converge to or stay at this principal component ("PC") eigenvector (see Sects. 2.4 and 3). We often slightly misuse this abbreviation and for brevity refer to an eigenvector of **EC** or even **E** as a "PC" (see also Elliott 2012); whether this is an eigenvector of **EC**, **E** (perfect whitening) or **C** (perfect specificity) should be clear from the context. An alternative, and apparently less arbitrary, procedure would have been to start with an exactly orthogonal $\mathbf{M}$ and introduce small perturbations (Cox and Adams 2009). However, we wanted to study various random mixing matrices, which would have entailed constructing a column at random and then making the other columns mutually orthogonal. This again involves a rather arbitrary choice of which column should correspond to the nonGaussian source.

We measured the orthogonality of $\mathbf{M_o}$ by defining a quantity O_F equal to the Frobenius norm (square root of the sum of the squares of all the elements) of $(\mathbf{I} - \mathbf{M_o M_o^T})$ where $\mathbf{I}$ is the identity matrix. For the case shown in Fig. 2b, the partial whitening made $\mathbf{M_o}$ twenty times "more orthogonal" than $\mathbf{M}$, since the O_F for $\mathbf{M_o}$ was 0.0807 and for $\mathbf{M}$ it was 1.609. When $\mathbf{M_o}$ was not sufficiently orthogonal, the error-free 1-unit rule instead converged close to the appropriate eigenvector of $\mathbf{C}$, even with nonGaussian inputs.

2.2 Learning rule

We used a standard online negentropy maximizing 1-unit rule (Hyvarinen et al. 2001; Hyvärinen and Oja 1998), with output $y = \mathbf{w^T x}$, where $\mathbf{w}$ is the neuron's weight vector, and $\mathbf{x}$ its input. We therefore ignore the important problem of coordinating learning in different neurons so that different rows of $\mathbf{M_o^{-1}}$ can be learned. The necessary conditions for the standard 1-unit rule to converge to a row of $\mathbf{M_o^{-1}}$ in the low learning rate (= k) limit are that $\mathbf{M_o}$ is orthogonal (i.e., the inputs to the neuron are white and have equal variance, so that the input covariance matrix $\mathbf{C} = \mathbf{I}$) and the weight vector is normalized to unit length after each update. Thus, convergence to a row of $\mathbf{M_o^{-1}}$ implies convergence to a column of $\mathbf{M_o}$. As described above, in this paper, $\mathbf{M_o}$ is almost orthogonal which in practice still usually allows convergence. Also, the sign of the nonlinear Hebbian term must be chosen to "match" the input statistics (sub or superGaussian) for any given nonlinearity; the exact form of the nonlinearity f(y) is otherwise unimportant. We used either $f(y) = y^3$, which requires a positive Hebbian term, or $\tan h(y)$, which requires a negative, "antiHebbian," term for superGauss sources (Hyvarinen et al. 2001). Thus, the basic rule used in this paper is:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + (\text{sgn})\ \text{kf}(y)\,\mathbf{x} \quad \text{(first, Hebbian part)} \tag{4}$$

followed by

$$\mathbf{w}(t+1) = \mathbf{w}(t+1) / ||\mathbf{w}(t+1)|| \quad \text{(second, normalizing part)} \tag{5}$$

where k is the learning rate and (sgn) the appropriate sign.

The first part of this rule is local and biologically plausible. It could be approximated by spike coincidence detection. The second, normalizing, step is apparently nonlocal, but could perhaps be achieved biologically by some combination of rescaling, homeostasis, an Oja-type local process (e.g., subtraction of $\mathbf{w}y^2$) or STDP. We do not want to make any specific assumptions about how this term, which guarantees that the dynamics has stable fixed points, is actually implemented biologically. The main focus here is on the nature of these stable fixed points. Therefore, we simply divided the weight vector by its current norm after each update. Note that 1-unit ICA can also be done using other nonlinear rules, which might show quite different crosstalk effects. For example, a kurtosis maximization rule has an additional negative weight-dependent term (Hyvarinen et al. 2001); this rule is resistant to the effect of crosstalk [Elliott, personal communication; see also Eq. (8) below]; however, this rule might be difficult or impossible to implement biologically.

Our key assumption is that neither of the 2 parts of the learning rule (i.e., Hebb and normalization) can be accurately implemented: The calculated weight changes are redistributed using an error matrix **E** or **F**, reflecting crosstalk in whatever processes implement the Hebbian part (**E**) or the normalizing part (**F**). The biological pattern of crosstalk is, on a spike-to-spike basis, likely to be stochastic and to reflect the current detailed connectivity. However, recent data (Wilbrecht et al. 2010; Xu et al. 2007) show that this connectivity varies throughout the learning process. In the simplest case, neither the crosstalk stochasticity nor the fluctuating connectivity would be correlated with the current input pattern, so we approximate the exact unknown and fluctuating pattern of crosstalk with a single average pattern, represented by **E** (or **F**). This might be invalid if spine necks dilate/lengthen in response to strengthening, or if weight changes are mainly achieved by anatomical changes. For simplicity, we usually assumed that crosstalk does not itself introduce a particular bias into learning, with offdiagonal elements of **E** (or **F**) equal to $E/(n+1)$ (E is a "total error" parameter, and n is the dimensionality) and the diagonal elements equal to $1-E$. E reflects the combined effect of a connection-intrinsic parameter b and the number of inputs n (Cox and Adams 2009; Radulescu et al. 2009), with $b = E/(n(1-E)) \approx E/n$. In all the figures, we use b as a crosstalk parameter, rather than E. The "isotropicity" implied by equal offdiagonal elements is likely to emerge on average because individual connection weights are composed of multiple, anatomically plastic, synapses scattered over the dendritic tree (Jia et al. 2010; Radulescu et al. 2009). However, similar results are obtained with an error matrix ($n \geq 3$) that is not exactly uniform [Results; (Cox and Adams 2009)]. The modified learning rule therefore becomes (t is the iteration step number):

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \mathbf{E}[(\mathrm{sgn})\ \mathrm{kf}(y)\ \mathbf{x}] \quad \text{modified Hebbian part} \tag{6}$$

or

$$\mathbf{w}(t+1) = \mathbf{w}(t+1) - \mathbf{F}\left[\mathbf{w}(t+1) - (\mathbf{w}(t+1)/||\mathbf{w}(t+1)||)\right] \quad \text{modified normalization part} \tag{7}$$

where $y = \mathbf{w}^{\mathbf{T}}\mathbf{x}$ and $\mathbf{x} = \mathbf{M}_o\mathbf{s}$. We always set either **F** or, for the last section of the Results, **E** as the identity matrix, to investigate the 2 forms of crosstalk separately (see Sect. 4). All sources had equal variance.

2.3 Foldiak bars

N^2-dimensional binary input vectors were created by concatenating the N rows of "images" of combinations of individual N-dimensional "bars" (rows or columns of 1's against a background of −1's or 0's). The 2N possible individual bars were combined to form images and hence input vectors in one of 2 ways. In a "standard" 0,1 protocol (Foldiak 1990; Triesch 2007), the number of input "bars" was random (with probability p usually set at $1/N$, with $N = 6$ or 10). Where bars coincided, pixels were set to 1 not to 2, making the combinations nonlinear. Input vectors are then divided by their norm, centered by removing the mean and used as inputs to an ICA neuron with a fixed, cubic, normalized learning rule as described above. Triesch (Triesch 2007) notes that a single unit with a fixed nonlinearity can learn a bar. Alternatively, in a "2-bar" protocol, 6 by 6 binary (−1,1) uncentered image vectors were generated that always consisted of just 2 bars, chosen randomly from the set of 12 possible bars, with overlaps set to 1. The learning rate in both models was constant but varied between 0.0001 and 0.000001 between different runs. Neither protocol used whitening.

2.4 Elliott's analysis

Here, we summarize Elliott's recent analysis of the important special case where the mixing matrix is orthogonal and thus the inputs are "white," maintaining the sources' lack of second-order correlation. He notes that because of the source independence, one can average the learning equations to yield, in the low rate limit, with the cubic nonlinearity and only one nonGaussian source (kurtosis k) the deterministic equation

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} = \mathbf{PE}[k\,(\mathbf{m}^{\mathbf{T}}\mathbf{w})^3\,\mathbf{m} + 3\mathbf{w}] \tag{8}$$

$\mathbf{P} = (\mathbf{I} - \mathbf{w}\mathbf{w}^{\mathbf{T}})$ is a projection matrix that corresponds to the normalization step in the stochastic rule, and **m** is the appropriate column of the mixing matrix. The fixed point behavior

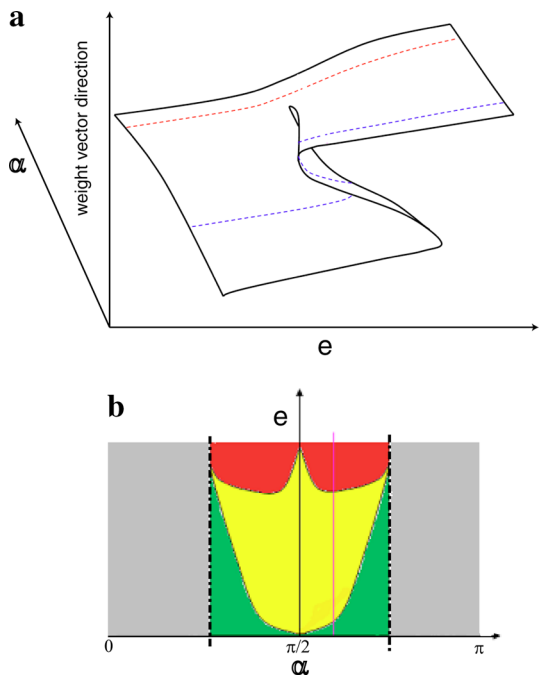


Fig. 1 Sketches of the qualitative behavior of the white 1-nonGauss model according to the analysis in (Elliott 2012). **a** shows the fixed points as a function of crosstalk and mixing angle α. **b** shows how critical crosstalk values depend on α. See text for further explanation

of this equation depends on the angle α between the principal eigenvector of **E** (the "PC") and **m**. When α is less than a critical value α_c defined by Elliott's equation 4.37, there is only one fixed point, which is stable and moves smoothly from the **m** direction to the PC of **E** direction, as crosstalk increases from zero to the maximum value where all specificity is lost. Above α_c, 2 dynamical bifurcations arise. The first occurs at a lower crosstalk threshold e_{c1}, with the creation of a stable approximate PC of **E** fixed point, and the second at a higher threshold e_{c2}, with the annihilation of the approximate **m** fixed point by collision with an unstable fixed point created at the first bifurcation. Between e_{c1} and e_{c2} a bistable regime exists, where the starting weights determine whether the stable approximate **m** or PC of **E** fixed point is selected. Elliott graphs (his figure 16) the FP direction(s) as a function of both α and crosstalk. For the reader's convenience, we present a sketch of this figure in Fig. 1a. The ordinate represents the angle between fixed point directions and the column of **M** corresponding to the supragaussian source, and the abscissa the crosstalk parameter e, for all possible combinations of α and e.

The FPs lie on a curving sheet in this three-dimensional space, which develops a fold at the critical α value (a "cusp catastrophe"; (Strogatz 1994)), with the 2 positive sloped surfaces corresponding to the stable fixed points and the negative sloped surface to the unstable fixed point. Thus, as crosstalk decreases from initial complete inspecificity ($e = (n-1)/n$), the steady state weights track the approximate PC solution until they reach the edge of the upper fold, at which point they "fall" to the lower surface, the approximate IC solution that was created at the upper critical (mixing angle dependent) error; conversely, if one starts at the IC at complete specificity, then with increasing error the weights track the approximate IC solution until they reach the right-hand fold, at which point they "jump" up to the now exclusively stable approximate PC solution on the upper surface—a solution that was created at the lower, mixing angle dependent, threshold.

The dotted lines correspond to 2 constant-α cuts through this 3D surface. The red line cut corresponds to a constant subcritical angle; the single, stable, equilibrium weight vector direction changes smoothly with error. The blue cut corresponds to a fixed supracritical angle; the cut forms a sigmoid relation, between equilibrium directions and error. At very low error, there is only one, stable, equilibrium; at a low critical error, 2 new equilibria are born in a saddle-node bifurcation; one of these is unstable, and the other is the stable approximate PC solution. At a second higher critical error, the approximate IC solution disappears in a second saddle-node bifurcation, by collision with the unstable solution, leaving only one equilibrium, the stable approximate PC.

Elliott also plots the 2 critical crosstalk values as a function of α_c (his figure 15b dotted line), which form a "badge" perimeter, the upper line of which corresponds to e_{c2} and the lower line to e_{c1}. We present a sketch of this figure in Fig. 1b, with the 4 regions of the model's parameter space defined by the badge colored differently. The vertical lines through the lateral "cusp" points of the badge at α_c define the transition from the smoothly deforming or "sliding" regime (gray; one fixed point) to the bifurcating regime. The badge perimeter encloses a yellow zone, where both approximate **m** and PC of **E** solutions are stable. Above the badge lies a red zone where the only fixed point is the approximate PC of **E**, and below the badge a green zone where only the approximate **m** fixed point is stable. Only in the complete absence of crosstalk (the standard ICA model; Hyvarinen et al. 2001; Rattray 2002) is the exact **m** the only stable fixed point. However, at sufficiently low crosstalk, the only stable fixed point is a good approximation to **m**; above α_c the crosstalk level must be very low to ensure good learning irrespective of the starting weights (i.e., the green zone). In the yellow zone, good learning can be achieved only with low crosstalk and if the starting weights happen to lie close to **m**.

The key feature of Eq. (8) generating these interesting dynamics is the competition between the 2 parenthetical right-hand terms. The first term drives the weight vector toward **m**, and the second term drives it in its current direction. However, in the absence of crosstalk, normalization cancels growth in the **w** direction, so the weights stabilize at **m**.

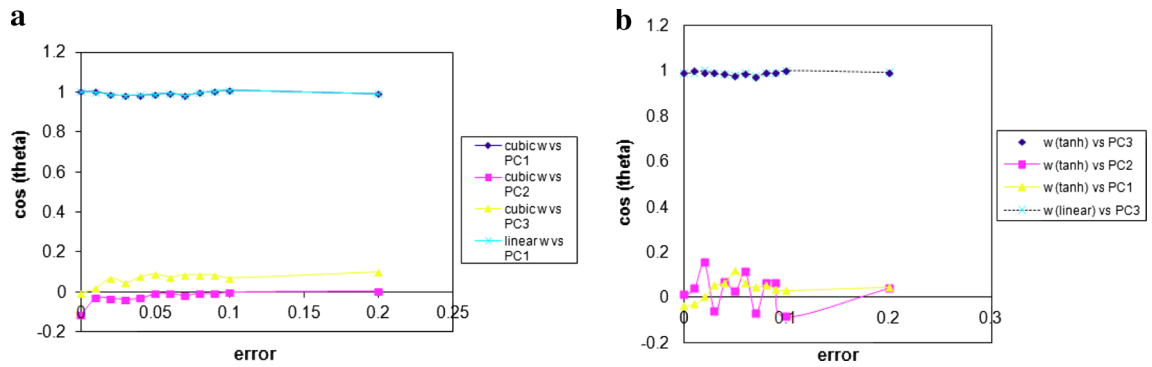


Fig. 2 Learning with all Gauss sources. The plots show the cosines of the angles (theta) between the averaged weight vector and the 3 eigenvectors of $\mathbf{EC_L}$ (PC 1,2,3 with decreasing eigenvalues) at various crosstalk levels ("error," measured using the single connection parameter b defined in the Sect. 2). Standard errors are less than the symbol sizes. With a linear rule, the results were indistinguishable from those for the cubic rule (panel **a**). Seed 140, $O_F = 0.1931$, $k = 0.00005$, batch size $N_B = 1{,}000$. Panel **b** shows similar results using a tan*h* nonlinearity. Seed 23, batch 1,000, $k = 0.002$, $O_F = 0.0807$

Crosstalk modifies both competing terms, and normalization no longer exactly cancels growth due to the second term. The outcome of the competition can either be a smooth interpolation between **m** and the PC (gray zone), or near complete victory by one or the other (green or red zones), or by both (depending on the starting conditions).

Our numerical experiments used nonorthogonal mixing matrices and were mostly obtained prior to the availability of Elliott's analysis. In this case, one can derive a similar averaged deterministic "colored" equation:

$$\frac{\mathrm{d}\mathbf{w}}{\mathrm{d}t} = \mathbf{PE}[k\,(\mathbf{m^T w})^3\,\mathbf{m} + 3\mathbf{w^T C w C w}] \tag{9}$$

Note that in this case setting, $k = 0$ (i.e., using all Gaussian sources) reduces the equation to the standard PCA form (Radulescu et al. 2009; Oja 1982). Elliott (Elliott 2012) does not analyze the dynamics of Eq. (9), but the close resemblance to Eq. (8) suggests that one might expect qualitatively similar behavior, now with competition between the **m** term and a PC of **EC** term, as reported below.

For other odd power function nonlinearities, or for functions that can be expressed as sums of odd power, the basic form of the equations remain the same, though the scalars associated with the **m** and **w** variables differ (Elliott, personal communication).

3 Results

3.1 All Gauss sources

A linear normalized Hebb rule responds only to SoCs and in the presence of crosstalk generically converges to the maximal eigenvector of **EC** (Radulescu et al. 2009), or "PC." Here, we study the ICA learning, model, with inputs generated by linear square mixing of independent sources only one of which has a nonGaussian distribution but first we present, as background, some results using all Gauss sources, for which ICA learning is not possible.

Jointly Gaussian signals have higher-order moments, but no higher-order cumulants (or crosscumulants), and one might expect that the nonlinear Hebbian rule driven by jointly Gaussian, dependent, inputs would therefore also converge to the appropriate eigenvector of **C**, or, in the presence of crosstalk, of **EC** [see Methods Sect. 2.4, Eq. (9)]. In simulations with finite learning rates, the relevant **C** would presumably be that calculated for the sample of vectors that is actually used for learning, $\mathbf{C_L}$, rather than the long-term expectation, $\mathbf{M_o M_o^T}$. Figure 2a shows that indeed a cubic Hebbian (positive sign) rule converges to the explicitly calculated dominant eigenvector of $\mathbf{EC_L}$, which was typically very close to the eigenvector of $\mathbf{M_o M_o^T}$. Using a linear, SoC-driven, Hebb rule under the same conditions also gave this result. Even when $\mathbf{C_L}$ was extremely close to **I**, as in Fig. 2a, the cubic Hebb rule would converge to the dominant eigenvector even at zero error.

Single-unit ICA is often done using the statistically more robust tan*h* nonlinearity, which requires an antiHebb rule for superGaussian sources (Hyvarinen et al. 2001; Hyvärinen and Oja 1998). Figure 2b shows that a linear antiHebb rule converges to the least eigenvector of **EC**, as one might expect. Exactly the same behavior was seen using the tan*h* antiHebb rule and all Gauss inputs (Fig. 2b). Again, there was a smooth movement, as crosstalk increased, from the exact least eigenvector of **C** at zero crosstalk to an exact least eigenvector of **EC** at high crosstalk. However, presumably because the difference between the two minor eigenvalues of **C** was very small, a very low learning rate was required for convergence without crosstalk. These results with antiHebb rules

are somewhat counterintuitive: making weight adjustment less specific favors outcomes with more unequal weights. Also, while the use of an antiHebb linear rule is expected to make the attracting fixed point of the linear Hebb rule unstable, it is not entirely clear theoretically that it would always selectively stabilize the least PC, rather than for example the second PC ($n \geq 3$), and a fortiori in the case of a nonlinear rule. Nevertheless, we typically found empirically that the same PC was selected using either a nonlinear or a linear rule.

3.2 Nonlinear rules with one nonGauss source and no crosstalk

Probably, no biological process could perfectly whiten inputs, and therefore, the usual assumption made in 1-unit ICA, and in Elliot's analysis, that the mixing matrix is exactly orthogonal might be inappropriate. However, biology could achieve good decorrelation, and therefore, we studied a slightly modified ICA model, using approximately orthogonal mixing matrices, which we designate $\mathbf{M_o}$, constructed using limited samples (typically $N_B = 1{,}000$) to estimate $\mathbf{C_B}$, the whitening covariance matrix (see Sect. 2). One might anticipate that since the correctly signed 1-unit rule always converges to the appropriate row of $\mathbf{M_o^{-1}}$ (which equals the column of $\mathbf{M_o}$ corresponding to the 1-nonG source) when $\mathbf{M_o}$ is orthogonal, it would almost always converge very close to that row (i.e., the IC) if $\mathbf{M_o}$ is almost orthogonal, though not exactly. We found that indeed this almost always occurred, if the batch number N_B used to construct $\mathbf{M_o}$ was quite large ($\geq$1,000) and the weights were randomly initialized, although (see Sect. 2) convergence was even closer to the corresponding column of $\mathbf{M_o}$, as expected from Eq. (9). Many examples can be seen throughout this paper. However, as expected for small batches ($\leq$100) and inadequate whitening, learning often (depending on the initial weights) failed to converge to the appropriate column of $\mathbf{M_o}$, and instead converged to the expected eigenvector of $\mathbf{C}$ (or, in the presence of crosstalk, of $\mathbf{EC}$). Interestingly, even though for large estimating batches weights almost always converged to the expected column with random weight initialization, we could find occasional examples of $\mathbf{M_o}$ where if weights were started at the appropriate eigenvector of $\mathbf{C}$, they would remain there indefinitely.

While we did not investigate these behaviors systematically, they are compatible with the notion that there are generally 2 basins of attraction of the nonlinear error-free unwhite dynamics, an "IC basin" that centers close to the row of $\mathbf{M_o^{-1}}$ and a "PC basin" that centers close to the appropriate eigenvector of $\mathbf{C}$, as in Elliott's analysis of the white crosstalk model [see Eqs (8) and (9)]. The only case in which the PC basin always vanishes is of course when $\mathbf{C} = \mathbf{I}$ (all eigenvalues degenerate), as usually assumed for the 1-unit rule. This would be why convergence is only guaranteed in the (practically unobtainable) perfectly white case, but fortunately almost always occurs as long as the whitening is good.

3.3 Cubic rule and one nonGauss source with crosstalk

The 1-unit ICA learning rules for perfectly white inputs allow almost any smooth nonlinearity to be used, which makes them biologically plausible, but do require that the sign of the Hebbian term be appropriately matched to the source statistics and the nonlinearity (Hyvärinen and Oja 1998). If the sign is mismatched to the single nonGauss source, we observed instead convergence to the expected eigenvector of $\mathbf{EC}$, as described above for all Gauss sources. In this section, we describe results using a positive Hebbian rule appropriate to a Laplacian (superGauss) source using a cubic nonlinearity.

In many cases, particularly with only 2 input neurons, we found that the stable learned weight vector rotated in a smooth fashion as crosstalk increased (Fig. 4c), moving gradually away from the row of $\mathbf{M^{-1}}$ toward the leading eigenvector of $\mathbf{EC}$. We call this type of behavior "sliding," and the $\mathbf{M_o}$ that produced it a "slider." This behavior is very similar to that seen in Elliott's analysis of the orthogonal mixing case for $\alpha < \alpha_c$ (Methods Sect. 2.4). Sliding behavior was also sometimes seen using a tan$h$, antiHebb rule (not shown). However, particularly for $n \geq 3$, we often observed, for many randomly generated mixing matrices, especially when the IC direction differed greatly from the equal weight direction (as expected from Elliott's analysis, see Sect. 2.4), that as crosstalk was gradually increased, the equilibrium weight vector would suddenly change direction at a threshold value, as described in (Cox and Adams 2009) for the tan$h$ case using 2 Laplacian sources, see below. Example runs for one case are shown in Fig. 3, and 4a shows compiled results using many runs, at various crosstalk levels, using another $\mathbf{M_o}$ (the same $\mathbf{M_o}$ as in Fig. 2a). In the previous paper (Cox and Adams 2009), we did not attempt to determine the meaning of this new direction, which continues to change slightly, relative to the fixed IC, as crosstalk increases further beyond this threshold. Given the results in Fig. 2, and Elliott's analysis of the orthogonal case, it seemed plausible that this new direction would correspond to that of the appropriate eigenvector of $\mathbf{EC}$. Thus, the nonlinear rule would become almost blind to the HoCs present in the input above a sharp threshold and would respond chiefly to SoCs, as though the input distribution had switched to being equivariant jointly Gaussian. Figure 4a shows that this surmise appears correct. At the threshold, the weights and the direction of the weight vector suddenly shift to nearly match the calculated PC of $\mathbf{EC}$, and then track that direction with further increases in crosstalk. Furthermore, we found that if we began with all Gauss sources at an error rate above the threshold and

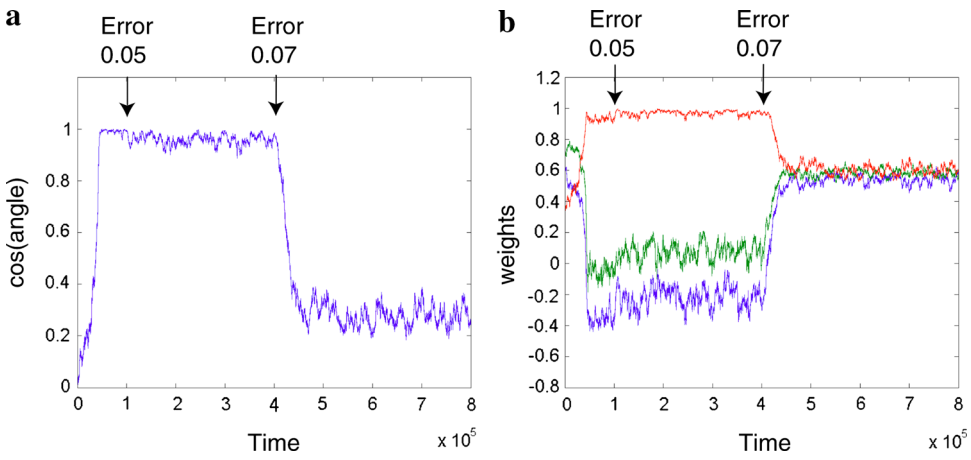


Fig. 3 Effect of crosstalk on ICA learning using a cubic rule and 1 Laplace and 2 Gauss sources. **a** shows the cosine of the angle between the IC and the current weight vector, starting at zero error, when the initially random weights (shown in the *right plot*) rapidly move very close to the IC. Error was then increased to 0.05, resulting in a small adjustment, and then to 0.07, producing a large shift (and weight equalization) at the indicated times. Seed 156, NB = 1,000, LR = 0.00005, $O_F = 0.1058$

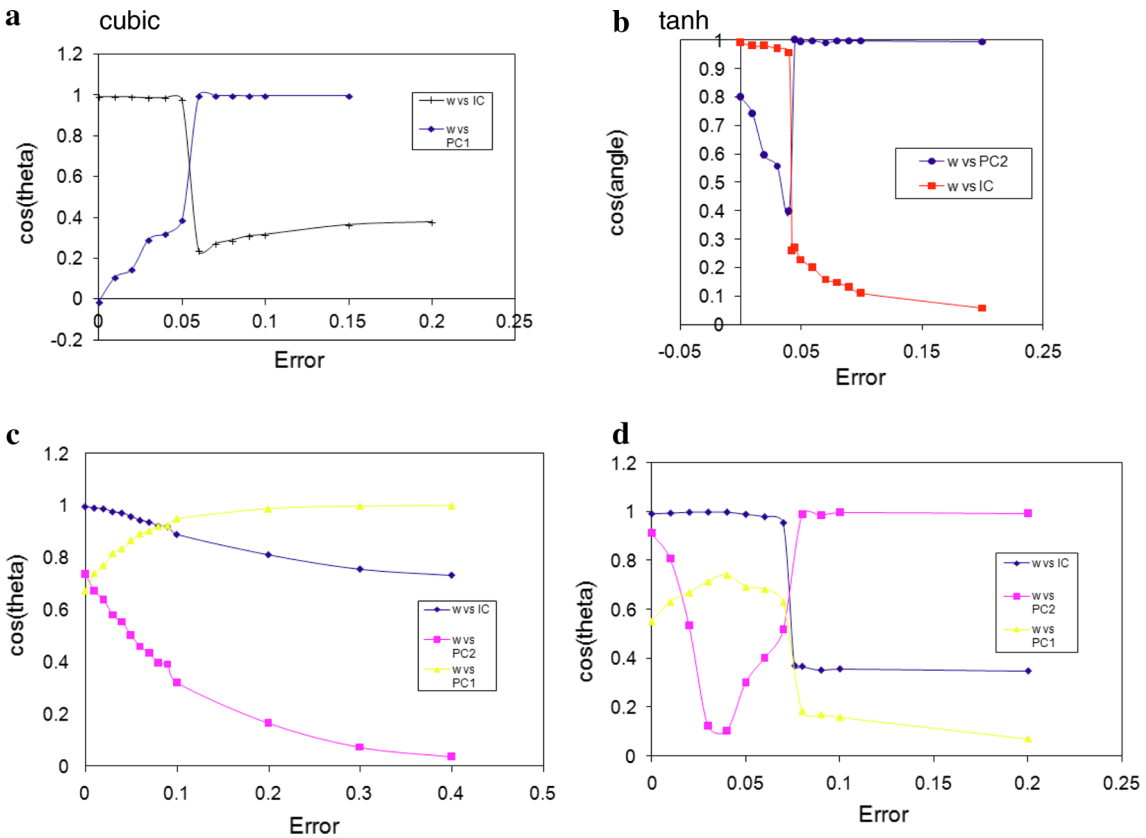


Fig. 4 Effect of crosstalk on learning with one Laplacian source. In all the panels *a–d*, the *y*-axis shows the cosine of the angle theta between the weight vector (**w**) and either the IC (appropriate row of $\mathbf{M_O^{-1}}$), or an eigenvector of **EC**, at equilibrium in the presence of various degrees of crosstalk. The abscissa shows the error (crosstalk) parameter b. Panel **a**: The angle between **w** and the IC suddenly swings by almost 80° at a threshold error near 0.055 (*blackline*). The blue line is the angle between **w** and the calculated first PC of **EC**. $N = 3$, $O_F = 0.1931$, $k = 0.00005$, seed 140, batch 1,000, cubic nonlinearity. (Compare with Fig. 2 A which is uses the same conditions except all GGG Gauss sources). Panel **b**: The red line shows how the cosine of the angle between the weight vector and the first row of $\mathbf{M_O^{-1}}$ (corresponding to the Laplacian source) changes with error, with a sharp change at $b = 0.0425$). The *blue line* is for **w** against the least PC of **EC**; $k = 0.002$. Similar results were obtained with $k = 0.0002$, $N = 3$, $O_F = 0.0807$, seed 23, $N_B = 1{,}000$, $k = 0.002$, tan*h* nonlinearity. (Compare against Fig. 2b which is the same except all Gauss sources). Panel **c**: Sliding behavior. The weight vector smoothly shifts toward the first PC of **EC**, and away from the IC, as crosstalk increases. $N = 2$, $O_F = 0.0825$, seed 2, $k = 0.00005$, batch = 1,000, cubic nonlinearity. Panel **d**: A near-orthogonal $\mathbf{M_O}$ was constructed from an initial **M** (seed 5) by partial whitening ($N_B = 1{,}000$). The cosine of the angle between the IC found at zero crosstalk, and that found at equilibrium in the presence of various degrees of crosstalk is plotted (*blue line*). This angle suddenly swings away from the IC by almost 60° at a threshold error of 0.076. The weight vector then aligns close to the direction of the minor PC of **EC** (*pink line*) rather than to that of the dominant PC (*yellow*). $N = 2$, $O_F = 0.0895$, seed 5, $k = 0.002$, Batch = 1,000, tan*h* nonlinearity (color figure online)

then switched to one Laplacian source halfway through the run, the weight vector hardly changed. For instance, for seed 140 ($N = 3$, $k = 0.00005$, $O_F = 0.1931$) at crosstalk $b = 0.08$ with all Gauss sources, the weight vector converged to [0.48,0.71,0.50] and when one source was made Laplacian the weight vector changed to [0.50,0.70,0.50]. Under identical conditions but with crosstalk $b = 0.03$ (well below the upper threshold, see below) with initially all Gauss sources the equilibrium weight vector [0.40,0.81,0.41] changed minimally, to [0.41,0.81,0.40], when one source was switched to a Laplacian distribution. We also found that when **E** had unequal off-diagonal elements, but remained stochastic with the same diagonals (an "anisotropic" case), the erroneous cubic rule still converged to the (new) principal eigenvector of **EC**. For instance, for seed 140 after the error threshold (0.08 in this case), **E** was changed from

$$\begin{bmatrix} 0.8475 & 0.0763 & 0.0763 \\ 0.0763 & 0.8475 & 0.0763 \\ 0.0763 & 0.1763 & 0.8475 \end{bmatrix} \text{ to } \begin{bmatrix} 0.8475 & 0.0163 & 0.1363 \\ 0.0163 & 0.8475 & 0.1363 \\ 0.0163 & 0.1363 & 0.8475 \end{bmatrix}$$

The weight vector converged to [0.67,0.57,0.47], and the theoretical PC of **EC** was [0.67,0.57,0.46]. Other seeds showed similar behavior.

3.4 tan*h* rule with one nonGauss source and crosstalk

Although the cubic rule is mathematically simpler, and more straightforward because it uses a positive Hebb rule, for which the outcome is unambiguous when using all Gauss inputs, it carries the practical disadvantage that it can be sensitive to outliers (Hyvarinen et al. 2001). A more robust rule uses the tan*h* nonlinearity, which requires an antiHebb rule with a Laplacian source. Here, as noted above, the stable PC for all Gauss inputs corresponds to a minor (typically the least) PC. This is the rule we used in our initial report (Cox and Adams 2009), using 2 Laplacian sources. Figure 4d shows an example using one Laplacian and one Gaussian source, where there is a sudden swing to the (unambiguous) minor PC direction at a critical crosstalk value b around 0.076.

A more completely studied case, with 3 sources, one Laplacian, is shown in Fig. 4b. Here, there are 2 possible minor PCs, but the PC selected at suprathreshold crosstalk corresponds to the least PC. Once again, after the weights move very close to the IC at zero crosstalk, gradually increasing crosstalk produces further slight shifts away from the IC. Then, at a sharp threshold, there is a dramatic swing in the direction of **w** (mostly arising from a large change in one of the weights), to align very close to the calculated least PC of **EC**. With further increases in crosstalk, there is a second drift in direction, closely tracking the shift in the PC of **EC** direction.

3.5 Bistability

In order to see a clear jump over a small range of crosstalk values in numerical experiments, it is obviously necessary that the direction of the IC and PC be substantially different. We surveyed about 15 examples of random matrices $\mathbf{M}_o$ derived from initial random matrices **M** ($n = 3$) with a good separation between the IC and the direction seen at with crosstalk = 0.1, and of these around 1/2 showed a jump at a threshold value of crosstalk. As already noted (Cox and Adams 2009), the actual threshold appeared to depend on the particular starting $\mathbf{M}_o$ used (see Elliott 2012). It also depended to some slight degree on the learning rate (see Cox and Adams 2009); typically, a large decrease in the learning rate (twofold to tenfold) produced a small (<20 %) increase in the error threshold estimate.

To estimate the threshold more exactly, the learning rate should be very low, as in the example shown in Fig. 5, where the learning rate it was set at 10^{-5}. After initially allowing the weights to stabilize very close to the IC at zero error, crosstalk was introduced ($b = 0.066$). This produced a small shift to an approximate IC. But then, after a highly variable delay, without any further increase in crosstalk, the weights moved much further, over a period of around 0.5 million updates, to the PC direction (Fig. 5). Note that instead, at $b = 0.065$, the weights, started at the IC, moved to and remained indefinitely, at the approximate IC (at least for 5 million updates), whereas at $e = 0.067$ they shifted promptly (within 0.5 million updates) to the PC (not shown).

This behavior strongly suggests that at certain crosstalk levels there are 2 stable states of the weights, corresponding to an approximate IC, and an approximate PC, as seen in Elliott's analysis of the orthogonal case. At slightly higher crosstalk, only the PC is stable, and at slightly lower crosstalk, the IC is indefinitely stable in the face of very small fluctuations (and remains there if it starts there). We conclude the true threshold lies very close to $b = 0.066$ in this case.

Just below this threshold, the direction of the weight vector shows large slow fluctuations, but just above it this slow noise almost disappears. This presumably corresponds to the marginal stability of the IC (very small basin of attraction) and the almost global stability of the PC. Eventually, this noise would always push the weights to the stable PC direction, while at a slightly lower error (0.065) the use of an exceedingly low learning rate would indefinitely protect the IC.

We also examined the tan*h* rule behavior very close to the threshold at very low learning rates, as shown for the cubic case in Fig. 5. Exactly the same result was found: A very narrow range of crosstalk existed for which the approximate IC, having been learned initially at zero error, lingered noisily for long periods, but then eventually abruptly gave way to the almost noiseless approximate PC, which then remained

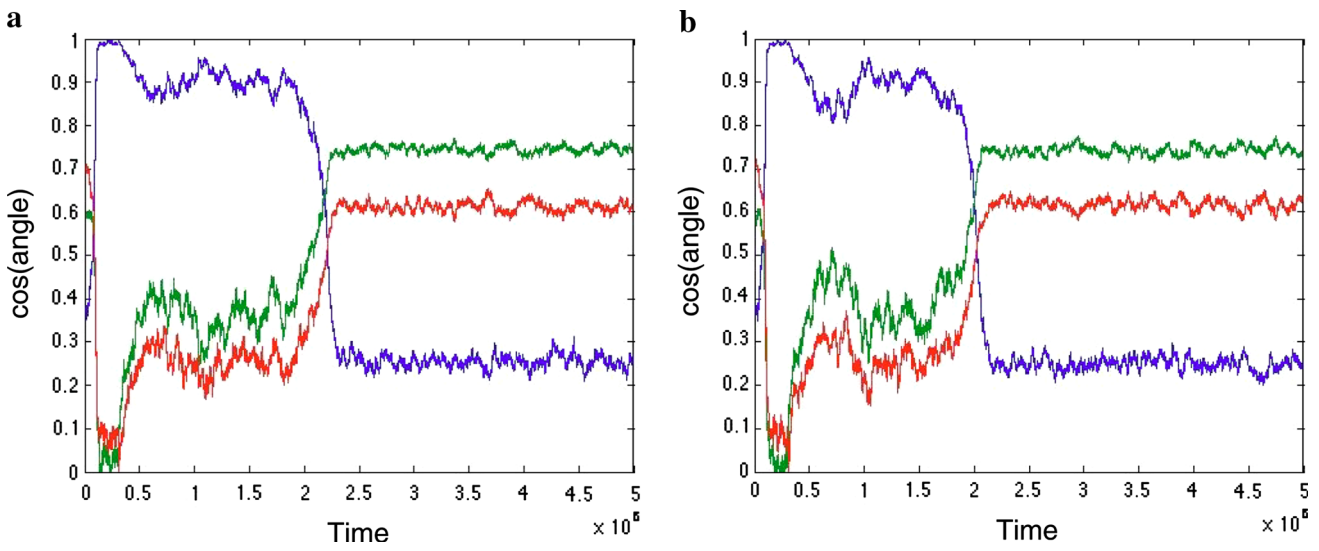


Fig. 5 Drop from just stable IC to fully stable PC at a single error rate. Seed 140, cubic rule and a lower k (10^{-5}) than in Fig. 3a. The 2 panels are in identical conditions, but different runs, and show the cosines of the angle between the weight vector and the IC (*blue*) and the other rows of $\mathbf{M}^{-1}$. (*green*, *red*). Zero error initially brought the weights rapidly to the IC. At 300 K epochs crosstalk was increased from 0 to $b = 0.066$ and maintained; the angle then dropped, to a steady (but very noisy) level corresponding to an approximate IC

indefinitely stable. Very slightly below this threshold, the approximate IC was indefinitely stable, while very slightly above it the approximate PC was reached immediately, without any lingering at the approximate IC.

In all the numerical experiments described so far, the weights were first brought very close to the IC at zero crosstalk, since we wanted to see whether successful learning from HoCs could survive imposition of update inspecificity. Elliott (2012), while confirming and greatly clarifying our main conclusion, that IC stability can be lost at a critical crosstalk level, also shows that there is not simply a stability swap at a unique threshold, but a fully bistable regime at intermediate crosstalk values (i.e., between "lower" and "upper" thresholds). Comparison of Eqs. (8) and (9) (Sect. 2) indeed suggests that the colored and white cases should behave similarly.

Bistability can easily be seen at fairly high learning rates, and crosstalk values below but fairly close to the upper threshold, where the fluctuating inputs cause the weights and their direction to switch between discrete stable states (Fig. 6 b,c).

Similarly, at very low learning rates, one could set the initial weights at the calculated PC and then gradually increase crosstalk until the weights suddenly shift close to the IC (i.e., in the opposite manner to that seen in Fig. 5). This protocol defined a second, threshold, at lower crosstalk than the one previously described, below which the PC lost stability, and the IC became uniquely stable, similar to the orthogonal case (Elliott 2012 and Fig. 1b). An example is shown in Fig. 6a.

As previously noted, as the upper threshold is approached from below, slow fluctuations in the weights appear, a typical sign of the approach to criticality (Scheffer et al. 2009). This "slow noise" level dramatically decreased when the weights underwent the shift from PC to IC (Fig. 6a). Interestingly, there were rare cases (including the **M** used in Fig. 5), where the PC remained indefinitely stable at very low learning rates even when crosstalk was reduced to zero. Of course, if the weights were started at random values, completely specific learning almost always converged on the IC. However, as already noted, when the whitening was very poor (very low batch numbers), only the PC was stable even at zero crosstalk. Clearly, the model shows hysteresis: It "remembers" whether the starting weights are close to the IC or to the PC.

A particularly striking case arises when the IC is exactly orthogonal to the PC (Elliott 2012), as illustrated in Fig. 7. An orthogonal **M** ($n = 2$) with one column having equal opposite-signed entries and the other equal entries was used. The weights were started at the PC (i.e., equal and same-signed); in the presence of very low crosstalk (panels **a** and **b**, angle and weights respectively), they remained there indefinitely, but if crosstalk was completely removed, after a delay they snapped permanently to the IC (panels **c** and **d**). This behavior corresponds to the central axis of the "badge" in Elliott (2012, Fig. 15B; see Methods Sect. 2.4) and illustrates that extremely low crosstalk levels can completely prevent effective learning.

3.6 The effect of normalization crosstalk

In the numerical experiments described so far, normalization crosstalk was absent ($\mathbf{F} = \mathbf{I}$). We briefly describe results obtained by instead setting the Hebbian crosstalk matrix **E** to the identity, and varying the (equal) offdiagonal elements

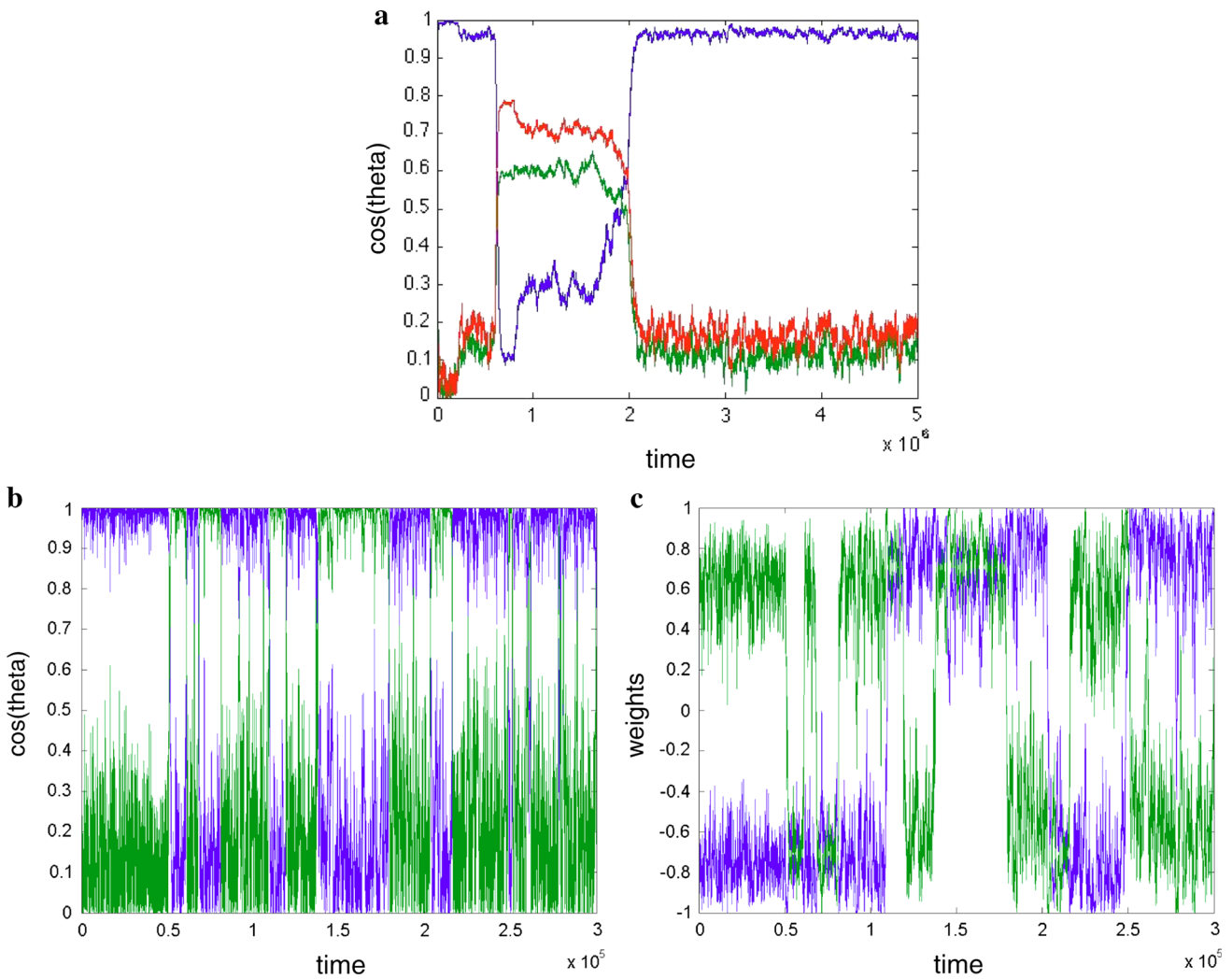


Fig. 6 Bistability. *Panel* **a**: k = 0.00001, seed 140 (as in Fig. 5), batch = 1,000, cubic nonlinearity. The initial part shows convergence to the IC, and then at 200 K epochs $b = 0.05$ was applied, so the weight vector moves to the approximate IC (*blue plot*; the *red* and *green plots* show the angle cosines relative to the other rows of $\mathbf{M_0}^{-1}$). Then, at 600 K epochs, crosstalk b was increased to 0.2, driving the vector to the PC of **EC**. Then, at 800 K, crosstalk was set back to 0.05. However, initially the vector did not move to the IC but to the appropriate approximate PC and stayed there (with slow large fluctuations) almost a million epochs; then, at about 1.7 M epochs, it spontaneously switched back to the appropriate approx IC, at a much lower noise level. Panels **b** and **c**: The angle is in the *left panel* (IC blue line and PC green line) and weights in the *right*. Initially, **w** was set to the IC with 0 error. Error of $b = 0.03$ was introduced at 50,000 epochs, at the arrow. The learning rate was 0.001 throughout. Over the remaining epochs **w** moves erratically between the IC and PC. The cubic rule was used with the following orthogonal mixing matrix: $M = \begin{bmatrix} -0.817 & 0.573 \\ 0.573 & 0.817 \end{bmatrix}$

in **F**. An example (chosen from 4 studied) is shown in Fig. 8. Here, we compare the weights obtained with all Gauss inputs with those obtained with 2 Gaussian inputs and one Laplacian, using the tan*h* nonlinearity at various crosstalk levels. At a critical crosstalk level, the tan*h* weights suddenly shift to match those obtained in the all Gauss case. With purely Gaussian inputs, the weights change smoothly as crosstalk increases. Note that in this case the PC corresponds quite closely to weight equality, i.e., the principal eigenvector of **E**, despite the use of an antiHebb rule. Of course, applying error to the Hebbian term in this situation leads to a minor PC. We also did normalization crosstalk experiments using a cubic nonlinearity and a Hebbian rule; in this case, above the error threshold rather than moving to the expected least PC of **EC**, the weights moved to the second eigenvector. However, because the data are well whitened, the 2 lesser eigenvalues of **EC** are quite close.

3.7 Foldiak bars

The only situation in which Hebbian adjustment of a single layer of feedforward connections can always find weights that directly invert a real-world generative model is when that model itself is linear mixing. In order to extend our results

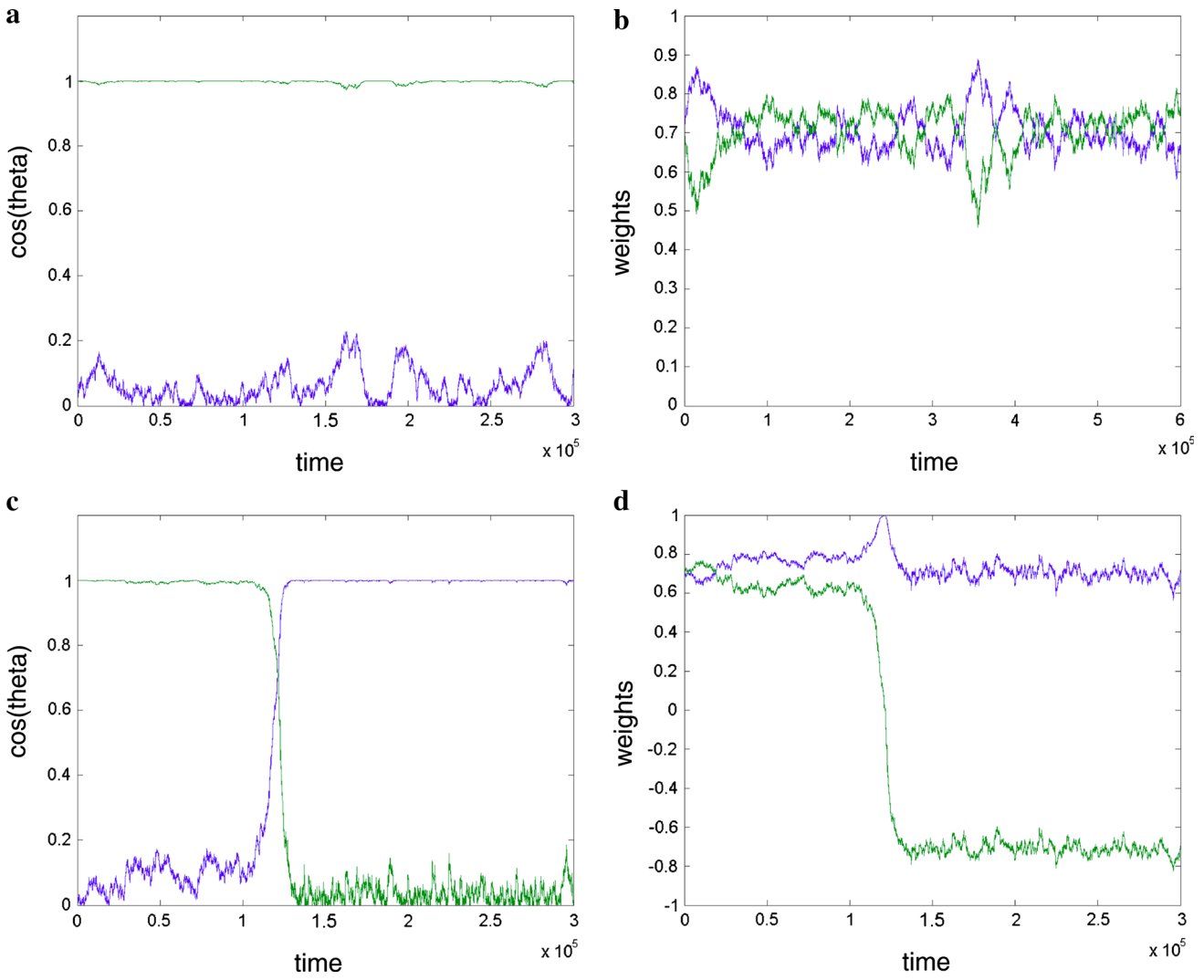


Fig. 7 Results with an exactly orthogonal **M** ($n = 2$), showing that the PC can be stable even at very low error. Here, the IC is orthogonal to the PC of **E** (i.e., equal weights). This corresponds to the central axis of the "badge" shown in Fig. 1. The angle cosines between the weight vector and either the PC of E (*green*) or the IC (*blue*) are plotted in the *left panels* (**a**, **c**) and the weights in the *right* (*panels* **b**, **d**). In panels **a** and **b**, b was 0.02 throughout, and the weights stayed at the PC. In *panels* **c** and **d**, b was zero throughout, and the weights shift to the IC (color figure online)

beyond the standard ICA framework, we also studied the effect of crosstalk in a popular nonlinear ICA model, Foldiak bars (Foldiak 1990; Triesch 2007). In this model, input vectors (dimensionality N^2) are generated as concatenations of combinations of variable numbers of vertical and/or horizontal "bars" at N possible vertical and/or random locations. When a bar is present, a pixel (input element) is set at 1, and the background is set at 0 or -1. Bar intersections are also set at 1. We studied 2 versions of this model: the standard case, where the number of bars is random (bar probability $= p$), and a simpler "2-bar" case, where 2 bars are always present (see Sect. 2). The input vectors derived from such "bars" images can be preprocessed in a variety of ways (centering, length normalization and whitening). For simplicity, and for comparison with our ICA experiments, we used a fixed cubic nonlinearity. The dynamical behavior of Hebbian learning in such protocols can be quite varied and complicated and depends on both SoCs and HoCs (Elliott, personal communication). We studied 2 protocols which, in the absence of crosstalk, reliably converged to a set of weights that represent a single bar (an "independent component" of the input images); the actual bar found depended on the randomly chosen starting weights.

In Fig. 9, we used a standard, 0,1 centered, normalized, but unwhite protocol ($N = 10,100$ inputs, $p = 0.1$). After a bar was quickly and reliably learned, crosstalk was gradually increased. The bar persisted until crosstalk reached a level of $b = 0.06$, when the bar first became rather noisy and

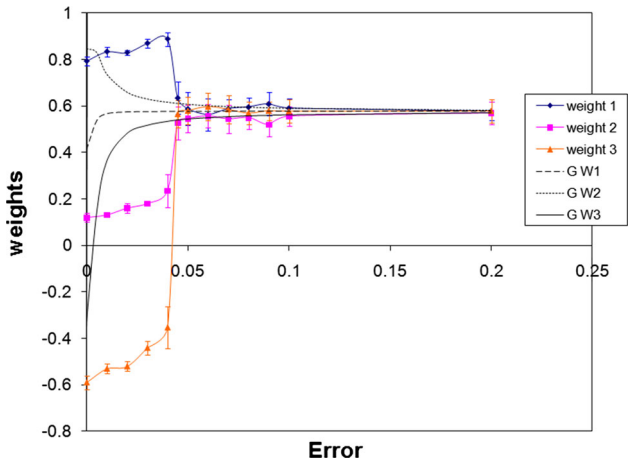


Fig. 8 The *symbols* (with standard deviation *bars*) and the associated *colored lines* show data obtained with a tan*h* nonlinearity and an antiHebb rule, at various normalization crosstalk levels, using one Laplacian source. The *lines without symbols* show data for the same $\mathbf{M_0}$ (seed 30) but all Gauss sources. $N = 3$, $k = 0.002$, $O_F = 0.0231$, batch = 10,000. Weights show a sudden shift to equality in the LGG case, but move gradually to equality in the GGG case (color figure online)

then abruptly disappeared as the weights equalized. It should be noted that if the weights were started exactly equal, they remained so even in the absence of crosstalk. Therefore, we could not test whether there was a second, "lower" threshold which below which a bar cannot be learned when starting from equal weights. However, this equal weight fixed point seems to have a very small basin of attraction. Similar results were obtained with $N = 6$ inputs. Increasing the bar probability produced a decrease in the crosstalk threshold.

A similar result, but with an even lower threshold ($b = 0.02$) was obtained using a 2-bar, uncentered unwhitened protocol ($n = 6$).

4 Discussion

4.1 Summary of new results

Activity-dependent synaptic plasticity plays a central role in theoretical neuroscience and connectionist models of learning. Recent evidence suggests that such adjustments are not completely connection specific (Engert and Bonhoeffer 1997; Harvey and Svoboda 2007). In previous work, we incorporated such inspecificity, quantified by a crosstalk matrix **E**, into simple standard models of Hebbian learning (Adams and Cox 2002a; Radulescu et al. 2009; Cox and Adams 2009). We reported that while minor crosstalk typically only modestly affects linear Hebbian learning, it can have catastrophic effects on nonlinear learning. In particular, we reported that in a 1-neuron ICA model using only one non-Gaussian source low levels of crosstalk could completely prevent useful learning (Cox and Adams 2009). A recent analysis (Elliott 2012) of this model for the special case of orthogonal mixing confirmed and expanded this finding. Here, we present numerical experiments which further extend and clarify these previous results. Most importantly, we show that even when mixing is only approximately orthogonal, ICA learning is usually possible, unless crosstalk reaches a critical value, and that when it fails, the weight vector converges instead very close to the direction expected if the sources are all Gaussian (Figs. 2, 4), and therefore, the inputs possess no higher-order cumulants. Thus, crosstalk can make nonlinear Hebbian learning almost blind to HoCs. In particular, we show that above a critical crosstalk threshold, the weights align with an eigenvector of **EC** (Fig. 4). Furthermore, we find this outcome not only with the cubic nonlinearity analyzed by Elliott, but also with a tan*h* nonlinearity. In this case, competition between HoC and SoC terms also controls the dynamics [see Eq. (8)], though the associated scalars are more complicated. We also show that the bistability revealed by Elliott's analysis in the orthogonal case also occurs for the nonorthogonal case. Our new numerical results accord qualitatively with the similarity between the averaged learning equations for the orthogonal case [Eq. (8), analyzed by Elliott] and the nonorthogonal case [Eq. (9)], although no analysis of the latter is currently available. We also show that normalization crosstalk is equivalent to Hebbian crosstalk (Fig. 8) and that crosstalk causes catastrophic failure in a nonlinear ICA problem (Fig. 9).

4.2 Comparison with Elliott

Our data suggest a close similarity between the effects of crosstalk in the specific orthogonal mixing case (Elliott 2012) and the more general nonorthogonal case, as expected from a comparison of Eqs. (8) and (9). In particular, even in the absence of crosstalk, spherical normalization no longer cancels growth in the **w** direction when there are input SoCs, which if strong enough completely dominate learning. However, with adequate though incomplete whitening HoCs dominate learning unless crosstalk exceeds critical values, just as seen in the orthogonal case. Our data show that there are 2 critical values, as in the orthogonal case. Above a critical value, e_{c2}, IC learning is impossible for all initial weights. Below that value, but above a second, lower critical value e_{c1}, IC learning is only possible if the weights happen to start close to the IC. Only below e_{c1} is good IC learning almost always possible. Elliott showed that in the orthogonal case, e_{c1} approaches zero when the relevant column of **M** is orthogonal to the PC of **E**. Our "colored" results suggest that e_{c1} can be very low, but we have not explored the conditions under which it might approach zero. Since a random direction in n-dimensional space lies on average increasingly orthogonal to a reference direction (such as the PC of **E** or of **EC**)

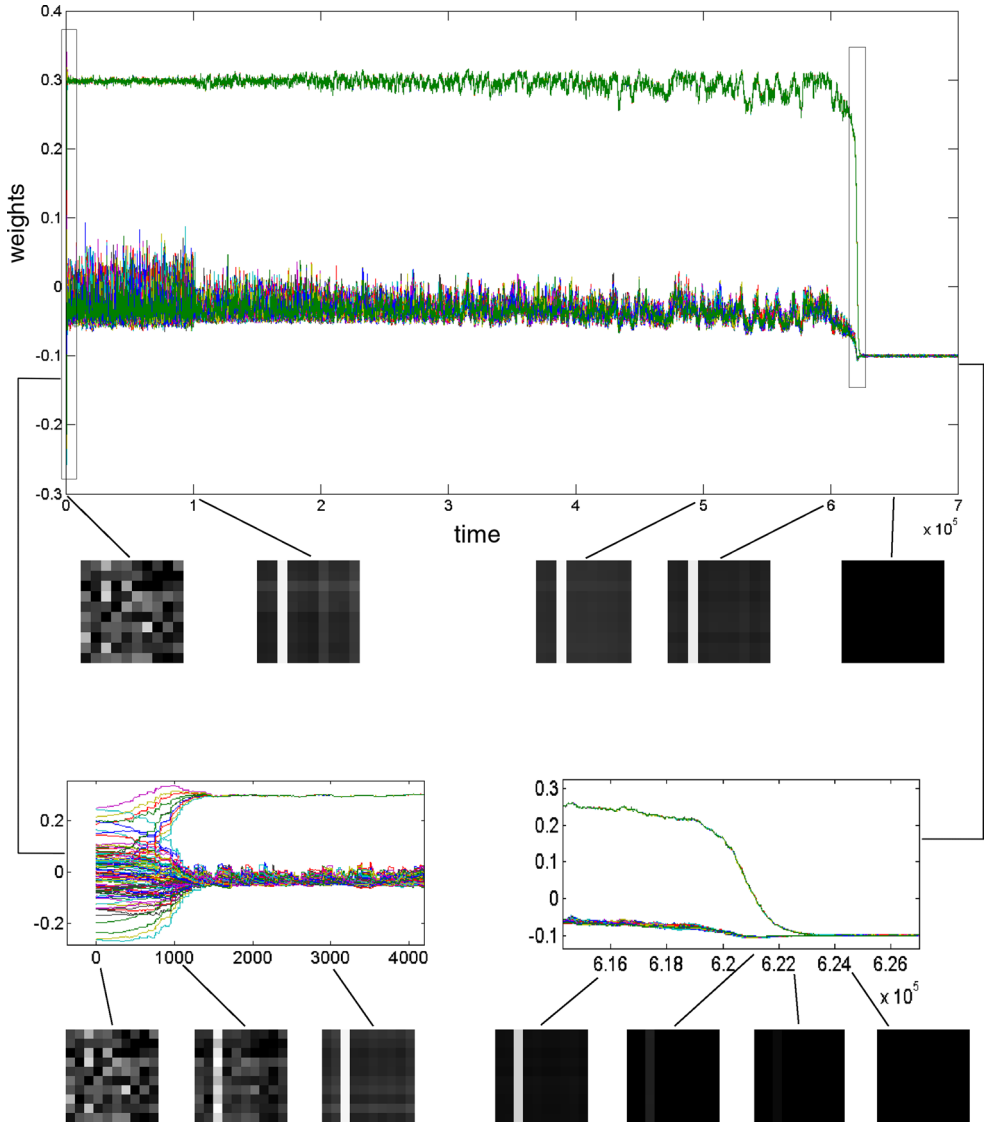


Fig. 9 *Foldiak Bars*. Standard protocol inputs were used with a cubic, normalized learning rule, learning rate of 0.00005. The top plot shows the complete run of 800,000 epochs with **w** starting from random weights. After converging to a bar at zero error, error of 0.01 was applied at 100,000 epochs, 0.02 at 200,000 epochs and so on until 600,000 epochs where at an error value of 0.06 **w** collapses down to the all equal weight vector. Snapshots of the weight vector converted to a gray-scale image ($-0.1 = black$, $0.31 white$) are shown at the end of each block of 100,000 epochs. The *inset* on the *bottom left* is a close-up of the initial convergence of **w** to a *bar* at zero crosstalk, with snapshots of **w** as indicated. On the bottom right is a close-up of the collapse to equal weights with $b = 0.06$, again with snapshots of **w** at various points

as n grows, these results strongly suggest that the effect of crosstalk on learning from HoCs would become increasingly toxic as the number of inputs to a neuron increases. Since neurons can receive thousands of inputs, these results raise the possibility that even extremely low levels of crosstalk could prevent systematic learning from HoCs in the brain.

4.3 Significance

We have only explored the effect of crosstalk in particularly simple models. Indeed, the 1-unit ICA model with all but one sources set to Gaussian distributions is probably the simplest nontrivial Hebbian learning model. Simplicity confers robustness. For example, columns of orthogonal Ms corresponding to nonGaussian sources are always stable fixed point of an appropriately signed completely accurate normalized nonlinear Hebbian rule, regardless of the form of the nonlinearity (Hyvarinen et al. 2001). But our results show that low levels of crosstalk often prevent such learning, and in particular, Elliott shows that any degree of crosstalk, however, small, can prevent learning when the column of **M** corresponding to a single nonGauss is orthogonal

to the PC of **E**. For the large number of inputs typical of real neurons, even near-zero crosstalk could usually be fatal to HoC-driven learning . It has been argued (Sabatini et al. 2002) that only 1–10 % of the NMDAR-dependent calcium increase that underlies classical forms of Hebbian plasticity escapes from a synaptic spine, and that this would be "negligible." However, at least in the ICA model, any nonzero escape could be catastrophic. This model shows an error threshold reminiscent of that seen in the Eigen molecular evolution model (Eigen 1971a), for much the same reason: The elementary machinery underlying self-organization fails catastrophically when subject to noise, and the critical noise level approaches zero as the dimensionality increases.

Could these robust simple models be misleading? One obvious possibility is that there exist simple crosstalk-resistant Hebbian ICA rules, for which averaging over inputs removes the contaminating "SoC" term in Eq. (9). For example, the rules we used work by maximizing approximations to negentropy, but a kurtosis maximization rule contains an additional negative weight-dependent term (e.g., Hyvarinen et al. 2001, Eq 8.13), which on averaging over inputs exactly cancels the weight dependent "SoC" component inside the brackets in Eq. (8) or Eq. (9). However, this rule is more complicated, and it could be difficult to biologically implement the required readout of individual weights to sufficient accuracy. Implementing crosstalk-resistant rules might always require extremely accurate neural machinery (e.g., specific nonlinearities) to overcome Hebbian inaccuracies.

Real inputs are unlikely to be generated by linear square mixing, but having neurons that use nonlinear Hebbian rules to provide sensitivity to HoCs to learn weights that confer desirable statistical properties, such as output nonGaussianity or independence, could be a useful component of a general strategy. Indeed, it has been suggested (Friedman 1987; Chen and Gopinath 2000; Shan et al. 2007; Hyvärinen 2013) that repeated ICA, coupled with suitable pre- and postprocessing so the overall transform is nonlinear, could be such a strategy. Examples of appropriate processing, which could be done in separate cortical layers from the ICA-like transform, would be whitening, pooling and divisive normalization. Our results suggest that the nonlinear Hebbian learning required for the ICA-like step should be particularly accurate. This raises the possibility that certain brain structures, for example the neocortex, might be specialized for learning from HoCs using nonlinear Hebbian rules because they possess dedicated crosstalk reduction machinery. One simple way to reduce crosstalk would be to lengthen or constrict spine necks, but recent evidence suggests that they are already at the limit required for good electrical coupling (Palmer and Stuart 2009). Another direct way would be to separate synapses, and it is interesting that the thalamocortical synapses responsible for the main receptive field properties of cortical neurons are rather sparse (Banitt et al. 2007; Da Costa and Martin 2011). This sparsity entails that thalamic input must be amplified, and from this point of view, recurrent amplification (Somers et al. 1995) could be viewed as a crosstalk reduction strategy.

We have proposed (Adams and Cox 2002a, 2006; Cox and Adams 2012) that additionally much cortical circuitry, involving layer 6 corticothalamic (CT) neurons, might perform a type of "Hebbian proofreading" operation. In this picture, input–output spike pairing would be detected both by a thalamocortical (TC) layer 4 synapse and by a corresponding layer 6 CT cell, which would get input from branches of the thalamic and layer 4 cell axons and act as a coincidence detector. The TC-4 synapse would store the detected coincidence as a "draft" trace which would only be finalized if the corresponding CT cell provides confirmation that a coincidence occurred at the synapse. The confirmation would be delivered specifically to the relevant synapses via axon branches to both the relevant thalamic relay (which would enter burst mode) and the appropriate layer 4 cell. While the required circuits and physiology have all been observed, this proposal must be regarded as speculation inspired by analogy with the mechanism underlying accurate DNA replication and evolutionary learning. Proofreading circuitry might allow the neocortex to systematically learn complex models of the world.

Acknowledgments We are deeply grateful to Terry Elliott for extensive advice, helpful criticism and for providing advance copies, and patient explanations, of his analysis of crosstalk models. We also thank him for detailed, pointed and constructive criticism of earlier drafts of this paper, which he would have written quite differently, and for extensive discussions and dissections of Foldiak bars. We also thank Giancarlo La Camera and Luca Mazzucato for helpful reading of a draft of our paper.

References

Adams P, Cox K (2002) A new interpretation of thalamocortical circuitry. Philos Trans R Soc Lond Ser B Biol Sci 357(1428):1767–1779. doi:10.1098/rstb.2002.1164

Adams PR, Cox KJA (2002b) Synaptic Darwinism and neocortical function. Neurocomputing 42(1–4):197–214. doi:10.1016/s0925-2312(01)00591-4

Adams PR, Cox KJA (2006) A neurobiological perspective on building intelligent devices. Neuromorphic Eng 3(1):2–8

Amari S-I (1998) Natural gradient works efficiently in learning. Neural Comput 10(2):251–276. doi:10.1162/089976698300017746

Amari S-I, Chen T-P, Cichocki A (1997) Stability analysis of learning algorithms for blind source separation. Neural Netw 10(8):1345–1351. doi:10.1016/s0893-6080(97)00039-7

Amari S, Cichocki A, Yang HH (1996) A new learning algorithm for blind signal separation. In: Touretzky D, Mozer M, Hasselmo M (eds) Advances in neural information processing systems. MIT Press, Cambridge, pp 757–763. doi:10.1016/s0893-6080(97)

Araya R, Jiang J, Eisenthal KB, Yuste R (2006) The spine neck filters membrane potentials. Proc Natl Acad Sci USA 103(47):17961–17966. doi:10.1073/pnas.0608755103

Atick JJ, Redlich AN (1990) Towards a theory of early visual processing. Neural Comput 2(3):308–320. doi:10.1162/neco.1990.2.3.308

Banitt Y, Martin K, Segev I (2007) A biologically realistic model of contrast invariant orientation tuning by thalamocortical synaptic depression. J Neurosci 27(38):10230–10239. doi:10.1523/jneurosci.1640-7

Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. Vis Res 37(23):3327–3338

Bi GQ (2002) Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. Biol Cybern 87(5–6):319–332. doi:10.1007/s00422-002-0349-7

Bonhoeffer T, Staiger V, Aertsen A (1989) Synaptic plasticity in rat hippocampal slice cultures: local "Hebbian" conjunction of pre- and postsynaptic stimulation leads to distributed synaptic enhancement. Proc Natl Acad Sci U S A 86(20):8113–8117

Botelho F, Jamison JE (2004) Qualitative behavior of differential equations associated with artificial neural networks. J. Dyn. Diff. Equ. 16(1):179–204. doi:10.1023/B:JODY.0000041285.36221.bf

Chen S, Gopinath R (2000) Gaussianization. In: Advances in Neural Information Processing Systems, pp 423–429

Cox KJA, Adams P (2009) Hebbian crosstalk prevents nonlinear unsupervised learning. Front. Comput. Neurosci. **3**: doi:10.3389/neuro.10.011.2009

Cox KJA, Adams PR (2012) From life to mind: 2 prosaic miracles? In: Simeonov PL, Smith LS, Ehresmann AC (eds) Integral biomathics: tracing the road to reality. Proceedings of iBioMath 2011, Paris and ACIB'11. Springer, Stirling

Da Costa N, Martin K (2011) How thalamus connects to spiny stellate cells in the cat's visual cortex. J Neurosci 31(8):2925–2937

Eigen M (1971a) Molecular self-organization and the early stages of evolution. Experientia 27(11):149–212

Eigen M (1971b) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58(10):465–523

Elliott T (2012) Cross-talk induces bifurcations in nonlinear models of synaptic plasticity. Neural Comput. 24:1–68

Engert F, Bonhoeffer T (1997) Synapse specificity of long-term potentiation breaks down at short distances. Nature 388(6639):279–284. doi:10.1038/40870

Feng D, Marshburn D, Jen D, Weinberg RJ, Taylor RM 2nd, Burette A (2007) Stepping into the third dimension. J Neurosci 27(47):12757–12760. doi:10.1523/JNEUROSCI.2846-07.2007

Field DJ (1994) What is the goal of sensory coding? Neural Comput 6(4):559–601. doi:10.1162/neco.1994.6.4.559

Foldiak P (1990) Forming sparse representations by local anti-Hebbian learning. Biol Cybern 64(2):165–170

Friedman JH (1987) Exploratory projection pursuit. J Am Stat Assoc 82(397):249–266

Harvey CD, Svoboda K (2007) Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature 450(7173):1195–1200. doi:10.1038/nature06416

Hinton GE, Sejnowski TJ (1999) Unsupervised learning: foundations of neural computation. MIT Press, Cambridge

Hoyer PO, Hyvarinen A (2000) Independent component analysis applied to feature extraction from colour and stereo images. Network 11(3):191–210

Hyvärinen A (2013) Independent component analysis: recent advances. Philos Trans R Soc A Math Phys Eng Sci **371**(1984). doi:10.1098/rsta.2011.0534

Hyvarinen A, Hoyer P (2000) Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. Neural Comput. 12(7):1705–1720

Hyvarinen A, Hurri J, Hoyer P (2009) Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. Springer Publishing Company, Incorporated

Hyvarinen A, Karhunen J, Oja E (2001) Independent component analysis. Wiley. New York. Available via http://worldcat.org. http://www.myilibrary.com?id=26480

Hyvärinen A, Oja E (1998) Independent component analysis by general nonlinear Hebbian-like learning rules. Signal Process 64(3):301–313. doi:10.1016/s0165-1684(97)00197-7

Jia H, Rochefort NL, Chen X, Konnerth A (2010) Dendritic organization of sensory input to cortical neurons in vivo. Nature 464(7293):1307–1312. doi:10.1038/nature08947

Kim K-H, Gaba S, Wheeler D, Cruz-Albrecht JM, Hussain T, Srinivasa N, Lu W (2011) A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. Nano Lett 12(1):389–395. doi:10.1021/nl203687n

Koch C, Zador A (1993) The function of dendritic spines: devices subserving biochemical rather than electrical compartmentalization. J Neurosci 13(2):413–422

Kuang X, Poletti M, Victor JD, Rucci M (2012) Temporal encoding of spatial information during active visual fixation. Curr Biol CB 22(6):510–514. doi:10.1016/j.cub.2012.01.050

Likharev KK (2008) Defect-Tolerant Hybrid CMOS/Nanoelectronic Circuits. Paper presented at the Proceedings of the 2008 IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems

Matsuzaki M, Honkura N, Ellis-Davies GC, Kasai H (2004) Structural basis of long-term potentiation in single dendritic spines. Nature 429(6993):761–766. doi:10.1038/nature02617

Noguchi J, Matsuzaki M, Ellis-Davies GC, Kasai H (2005) Spine-neck geometry determines NMDA receptor-dependent Ca2+ signaling in dendrites. Neuron 46(4):609–622. doi:10.1016/j.neuron.2005.03.015

Oja E (1982) A simplified neuron model as a principal component analyzer. J Math Biol 15(3):267–273

Palmer LM, Stuart GJ (2009) Membrane potential changes in dendritic spines during action potentials and synaptic input. J Neurosci 29(21):6897–6903. doi:10.1523/JNEUROSCI.5847-08.2009

Radulescu A, Cox K, Adams P (2009) Hebbian errors in learning: an analysis using the Oja model. J Theor Biol 258(4):489–501. doi:10.1016/j.jtbi.2009.01.036

Radulescu A, Adams P (2013) Hebbian crosstalk and input segregation. J Theor Biol (337)133–149. doi:10.1016/j.jtbi.2013.08.004

Rattray M (2002) Stochastic trapping in a solvable model of on-line independent component analysis. Neural Comput 14(2):421–435. doi:10.1162/08997660252741185

Reynolds T, Hartell NA (2000) An evaluation of the synapse specificity of long-term depression induced in rat cerebellar slices. J Physiol 527(Pt 3):563–577

Sabatini BL, Oertner TG, Svoboda K (2002) The life cycle of Ca(2+) ions in dendritic spines. Neuron 33(3):439–452

Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, Sugihara G (2009) Early-warning signals for critical transitions. Nature 461(7260):53–59. doi:10.1038/nature08227

Schuman EM, Madison DV (1994) Locally distributed synaptic potentiation in the hippocampus. Science 263(5146):532–536

Shan H, Zhang L, Cottrell GW (2007) Recursive ICA. Adv Neural Inf Process Syst 19:1273–1280

Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J Neurosci 15(8):5448–5465

Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. Proc R Soc Lond B Biol Sci 216(1205):427–459

Strogatz SH (1994) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Westview Press, Boulder

Triesch J (2007) Synergies between intrinsic and synaptic plasticity mechanisms. Neural Comput 19(4):885–909. doi:10.1162/neco.2007.19.4.885

Vontobel PO, Robinett W, Kuekes PJ, Stewart DR, Straznicky J, Stanley Williams R (2009) Writing to and reading from a nano-scale cross-

bar memory based on memristors. Nanotechnology 20(42):425204. doi:10.1088/0957-4484/20/42/425204

Wickens J (1988) Electrically coupled but chemically isolated synapses: dendritic spines and calcium in a rule for synaptic modification. Prog Neurobiol 31(6):507–528

Wilbrecht L, Holtmaat A, Wright N, Fox K, Svoboda K (2010) Structural plasticity underlies experience-dependent functional plasticity of cortical circuits. J Neurosci 30(14):4927–4932. doi:10.1523/JNEUROSCI.6403-09.2010

Xu H-T, Pan F, Yang G, Gan W-B (2007) Choice of cranial window type for in vivo imaging affects dendritic spine turnover in the cortex. Nat Neurosci **10**(5):549–551. http://www.nature.com/neuro/journal/v10/n5/suppinfo/nn1883_S1.html

Yuste R, Denk W (1995) Dendritic spines as basic functional units of neuronal integration. Nature 375(6533):682–684. doi:10.1038/375682a0