From Life to Mind: 2 Prosaic Miracles?

Paul Adams and Kingsley Cox

Stony Brook University, Stony Brook, NY 11794, USA
padams@notes.cc.sunysb.edu, kcox@syndar.org

Abstract. The origin of life from matter and the subsequent emergence of mind were fundamental events. Our work is based on the idea that the chemical/genetic/mathematical framework developed over the last 150 years to explain the first is conceptually similar to the neural/psychological/mathematical framework needed to understand the second. First we outline the first, seemingly adequate, framework and then we explain some related, unusual and controversial, ideas that offer a "translation" into neural terms. The core idea is that the extraordinary, mysterious and qualitatively unique features of "life" and "mind" arise because of extraordinary (though completely explicable) levels of accuracy of the relevant elementary processes (base-copying and synaptic strengthening). The living and the mental might hinge on prosaic, though accurate, lower-level machinery.

Keywords: Hebbian Proofreading, Crosstalk, Neocortex, Mind, Neural Sex.

1 Chemical Machinery of Darwinian Evolution

The key transitions[1] that led to complex life were (1) Onset of Darwinian evolution in the RNA world; (2) emergence of the dna/protein world and prokaryotic life (3) sexual, eukaryotic, evolution.

(1) Spontaneous formation of an RNA sequence that could act as a high-fidelity selfreplicase. The length of this sequence must have been under the per-base copying error rate (Eigen threshold), allowing onset of Darwinian evolution, in a phase transition. But search was restricted to compact sequence spaces.
(2) Searchable sequence space vastly enlarged ($> 4^8$ fold) as a result of replicase fidelity improvements, notably proofreading. But the Eigen threshold prevented more complex forms of organization than prokaryotes. The problem is that near-neutral mutations cannot accumulate in a finite population for long enough to combine with other individually near-neutral mutations with which they are synergistic, because the mutation rate must be below the Eigen threshold. Instead, selection in slowly changing environments favors low mutation[2].
(3) Advent of eukaryotes and sex allowed the threshold to be surpassed[2].

The crucial factor for life is proofreading, which lowers the copying error rate by $\sim 10^4$, though other smaller factors also play roles. Proofreading copies bases twice, and only if the 2 attempts agree is replication allowed.

2 Neural Machinery of Learning for Understanding

In our view causal learning (a neural equivalent of Darwinian adaptation) is the key to intelligence and mind. We learn to (partly) understand the world, and infer underlying causes (objects, ideas etc) from sensations, by adjusting vast networks of synaptic connections in response to local spiking traffic across those connections, as well as more global signals. Networks learn to track possible hidden causes given the current inputs, based on past statistics, and gradually narrow the range of likely causes. Repeated past temporal pairing of input and output spikes at specific connections leads, slowly, to more frequent future pairing and ultimately to improved inference and understanding. However, different from most approaches, we focus on crucial details of the relevant synaptic hardware. We believe that the accuracy of synaptic detection of such spike-pairing plays a fundamental role in the sophisticated learning underlying cognition in much the same way that accurate base-pairing drives Darwinian evolution. In this view, the essential problem confronting the brain is to ensure that pairing-based adjustment is connection-specific, despite extremely high synapse density. Mind could only emerge, in a type of phase-transition, if synapse adjustment were extraordinarily specific, and such specificity would be attainable only using specialized neural circuitry found throughout the neocortex and associated thalamus.

We studied this novel thesis in the simplest possible general model of the synaptic learning of weights that allow underlying causes to be extracted from neural inputs (**x**). We assume, for simplicity, that causes (the independently fluctuating components of **s**) are veiled by linear mixing: $\mathbf{x} = \mathbf{Ms}$, where **M** is an n by n matrix. To extract a cause, one must learn a row of $\mathbf{M}^{-1}$. Fortunately this can be easily done using (completely-accurate) nonlinear Hebbian, spike-pairing based, learning, which is driven by the higher-order correlations between inputs generated by the mixing of causes, which must have nonGauss distributions. Hebbian learning is driven by recently described synaptic processes, such as localized calcium entry through spike-pair activated NMDA receptors. Such machinery has 2 conflicting requirements: a synapse must *transmit* current to the spike-trigger region of the neuron, but calcium etc. must be *confined* to the synapse. This conflict implies that the Hebbian spike-pair detection cannot be completely synapse-specific. A similar "read/write dilemma" arises in DNA replication: Crick-Watson basepairing must be strong (to give accuracy) but weak (to allow replica separation). We[3] therefore modified the standard Hebbian rule to incorporate inevitable inaccuracy (via a matrix **E** which specifies how different connections slightly affect each other). In the simplest most plausible case, this matrix has equal small offdiagonal elements $e/n << 1$ that reflect inaccuracy.

The key result of this bifurcation analysis (to which T. Elliott has crucially contributed[4]) is that there is a maximal value of e, e_c, allowing reliable learning of causes; e_c approaches zero as n increases. This result is similar to that underlying the Eigen error catastrophe, and implies that sophisticated learning (i.e. of causes, driven by higher-order correlations between numerous inputs) is only possible given extraordinary Hebbian accuracy. Above this crosstalk threshold, correct learning (which corresponds to "understanding") can only be achieved if one starts very close to the correct solution (e.g. via luck, genetics or supervision); if weights start equal or random,

learning is driven only by the combined influence of **E** and (causally-uninformative) pairwise correlations. To reliably learn from higher order correlations, and gain individual insight into novel problems, crosstalk must be very low, and in some cases (especially for large n) even negligible.

While this analysis is rooted in machine learning (the problem of assigning meaning to observations), it is also rooted in recent ideas about the underlying detailed neural mechanisms. Progress in understanding learning, the neocortex and mind has been retarded by the almost complete isolation of 3 relevant fields: machine learning; synapse biophysics; circuit/system neuroscience. Dramatic progress in biology ensued by bringing together molecular, genetic and ecological levels of description. Understanding the world requires analyzing the structure of the higher-than-pairwise correlations that it generates: these are the clues that can reveal underlying causal structure. Structure in higher-order correlations can be revealed by nonlinear Hebbian learning (or variants thereof), but only when it is extremely accurate. In this view the simplest observation about brains is the most relevant: they have a lot of synapses! Highly specific synapse adjustment would allow circuits to develop powerful representations capturing underlying realities hidden by the apparently random flux of experience: truth from trash; meaning from observation. These trillions of synapses must each be regulated by the tiny aspect of the world they see: the impulse traffic across them. Extracting meaning from data thus resembles efficiently evolving DNA sequences, bit by bit.

3 Cerebral Proofreading

The general view that high accuracy is needed for the sorts of elementary "local" processes underlying neural network learning is not revolutionary; most theorists assume that synapses can reliably do this. Experimenters know that they cannot, but they assume instead that the theories have adequate slack. Darwin knew that organisms reproduce, but he did not know how; what it essentially requires is copying the entire genome, with a per base error rate approaching 10^{-10}. The "miracle" of life lies in that extraordinary number, achieved by a combination of processes, of which molecular proofreading is the most important. We propose that the miracle of mind is similarly, and rather prosaically, achieved, by a "neural proofreading" operation that is unique to the neocortex, the brain structure that first appears in mammals, and reaches its acme in humans.

Hebbian learning boils down to detecting paired, pre- and post-synaptic, spikes, manifesting as a *local* (e.g. calcium) signal. The reason why this crucial synaptic process (very rarely) makes mistakes is that signals can diffuse from neighboring synapses that belong to different connections experiencing different impulse traffic. This problem is *biophysically* inevitable but it can be greatly alleviated by a proofreading operation: one needs a second independent, extrasynaptic, measure of the relevant spike-pairing, which has to "approve" the first, synaptic measure. Because the 2 measures are independent, their error rates multiply. This principle drives accurate basepairing, and ultimately, life.

The problem of implementing this necessary "Hebbian proofreading" operation may have been solved by the special characteristic circuitry and physiology of the neocortex (and the associated thalamus; see Figure). We believe[5] that each thalamocortical connection, primarily responsible for the tuned responses of cortical neurons, is equipped with a "proofreading neuron", which gets copies of input and output spikes arriving at that connection. This proofreader would be a corticothalamic neuron in layer 6. If it also detects a "coincidence" (a spike-pair) it swiftly sends signals to both the input and output side of the relevant synapses comprising the connection. This double-signal then confirms that the synaptically-detected coincidence was valid, in a procedure that is closely analogous to that operating during DNA proofreading. This analogy arises because proofreading is the only effective strategy for overcoming physical limitations.

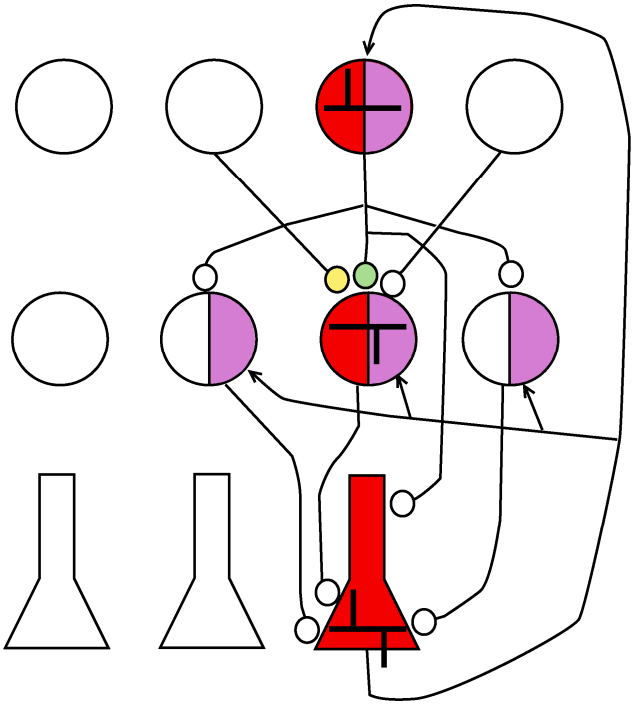

Figure. Cortical proofreading circuit for superaccurate learning (postsynaptic error version). A circuit that would allow a single cortical layer 6 cell (bottom row, red) to proofread many connections, all formed by the same presynaptic thalamic "relay" cell (top row, colored). However, the connections formed onto a particular layer 4 cell (middle row) by different relay cells each get there own layer 6 proofreader, only one of which is shown in detail here. One of a set of relay cells fires (denoted by the left red semicircle), as does one of set of layer 4 target cells (red, middle row). The timing of the relevant paired spikes is shown by the vertical lines within the circles; presynaptic spike up, and postsynaptic spike down. In this case, the pre-spike is closely followed by a post-spike (a "pairing" or "coincidence"), which triggers the generation of a second messenger within the relevant postsynaptic spine. The spine itself is not shown, and

the small circles show synapses, without specific reference to boutons or spines. The coincidence occurring at one of the connections is marked by green, and this produces crosstalk ("false pairing"), because of postsynaptic messenger spread to another synapse, made by a different relay cell on the same target cell, shown in yellow. The neurons shown in the bottom row are coincidence-detecting "proofreading" neurons in layer 6; the relevant proofreading neuron (colored), which detects coincidences between a specific partner relay neuron and any of the thalamorecipient neurons on which it currently synapses, fires in response to this coincidence (the firing is shown as red color, and the coincidence detecting function is shown schematically within the cell body). Such pre-post coincidence-detection can be implemented if the relay cell makes weak distal synapses on the proofreader, and the target cells makes proximal synapses, as shown. Both types of inputs must fire, in sequence, to trigger proofreader firing, which then feeds back both to the whole set of neurons targeted by the relay being proofread by the given layer 6 cell, and to its "partner" relay cell; this feedback is modulatory (arrows). This modulatory feedback briefly (~100 msec) "half-enables" (purple semicircles) the expression of the coincidence-induced plasticity change (held in "draft" or temporary form) both presynaptically and postsynaptically. However, although the relevant output cell is half-enabled, the relevant relay cell (that contributing the synapse receiving the crosstalk) is not, and therefore the erroneous "false pairing" induced by spillover from the activated synapse is not expressed as a strength change. Note that the colored proofreader shown here can perform a similar operation at any of the connections (only 3 are shown) made by its thalamic partner (also colored). For example, if paired spikes occurred in this thalamic cell and its rightmost layer 4 target, the proofreader would enable that connection (but not false pairings erroneously induced at other connections on that rightmost cell). But if a spurious coincidence occurs at that same connection shortly afterwards, it would be falsely approved, because of inevitable proofreading delays and persistences. This "distributed crosstalk" makes proofreading imperfect, especially with large numbers of inputs. If most connections are merely potential, such errors are reduced, at the expense of slower learning. These circuits must be continuously updated by separate sleep-like offline learning to track ongoing online rewiring (e.g. conversion of potential to actual connections). A different but closely related circuit, using anticoincidence, would be needed to handle presynaptic errors, and we think these are the dominant type, and that this second form of proofreading is the one that is actually used. Since presynaptic errors are probably associated with anticoincidence detection, the connections onto layer 6 proofreading neurons must be reversed (input from 4 is distal, and input from relays is proximal, as observed).

4 Proofreading Machinery

Because there are far more thalamocortical connections than layer 6 corticothalamic cells, proofreading must be done in a distributed fashion: each proofreader services all the connections made by a given thalamic (or thalamorecipient) cell (see Figure). This can work well because the close spiking-pairings that drive learning are quite rare, and become rarer as learning proceeds and weak connections are eliminated. Merely potential connections, prior to dendritic spine insertion at close axodendritic approaches, do not require proofreading. There are 2 interesting consequences. First, sophisticated learning will be very slow (since potential connections cannot immediately learn). Second, proofreading neurons must be continuously rewired to match current connectivity created by recent learning. Both input and output connections

must be rewired; this may be the purpose of the alternating slow-wave and paradoxical phases of sleep.

This view shifts the balance in the study of mind from machine learning or psychological principles to the associated neural hardware, which is where neuroscience makes the most distinctive contribution. We focus on the tremendously difficult problem of implementing basic learning rules at quadrillion-element scales, and less on clever "AI" algorithms built around assumed perfect rules. Rather than complex rules that work despite hardware imperfections, nature uses simple rules but complex hardware. The figure diagrams the proposed neocortical "proofreading" hardware that would allow extremely accurate adjustment of a particular thalamocortical (top 2 layers) synapse (marked in green) in response to pre-post spike-pairing (red colors and vertical black lines) despite inevitable postsynaptic chemical spread to an inappropriate synapse (yellow). A layer 6 neuron (bottom layer) detects the coincident pairing (pre-post spikes and red color) and fully enables potentiating plasticity only at the appropriate synapse. Note that although approval is also delivered to other synapses (formed on the flanking layer 4 cells), these do not register the triggering coincidence event. A similar, complementary, arrangement (not shown) could be used to proofread "anticoincidences", reflecting close post-pre spike pairing underlying long-term depression, and we think this alternate arrangement is that actually used.

5 From Mammals to Humans: Neural Sex

In this account, all mammals, possessing a neocortex, could learn to understand aspects of their world. Such ability ("insight") is the hallmark of intelligence, and would be uniquely conferred by neocortical proofreading. However, it seems only humans can do this systematically. The problem is of course that the necessary slowness of learning, which as explained stems from the inevitability of synaptic crosstalk (even though greatly mitigated by neocortical proofreading) means that little deep understanding can be achieved in an individual lifespan, given the limited sampling of necessary high-order statistics. Clearly human culture and language somehow overcome this difficulty. While novel insight fragments could be generated in individual brains by the process described above (incredibly accurate learning driven by higher-order correlations), they cannot accumulate without culture and language. Our new account of cortical learning leads to an unexpected parallel between this rather conventional view of culture and language, and recent understanding of the role of sex in Darwinian evolution[2]. Only eukaryotes have the necessary machinery to engage in true sexual reproduction, which is essentially, like language, a species- agreed protocol for the exchange of (genetic) information. Crucially, it appears that sex alleviates the Eigen error threshold. Thus the human per generation mutation rate is around 10^{-8}, tenfold higher than the genome length, which is in turn ten times greater than the reciprocal per-base error rate. This high level, mostly due to successful sperm-delivery by older men, is far above the error threshold. But sexual recombination blunts Muller's ratchet, which would otherwise lead to mutational meltdown. Bacteria, without sex, are forced to live well below their error threshold, and never evolved complex forms.

Most human learning is not based on individual discovery (driven by subtle correlations in an apparently random input data stream, requiring extreme synaptic accuracy, as just described) but by much more robust, banal, supervised learning insights of others. Language/culture converts the very difficult, slow, process of individual discovery to the rather trivial problem of copying available solutions; as noted above, analysis shows that if one can initially get close to the correct solution, a quite high degree of crosstalk allows one to perfect this, based on experience. More concretely, sex allows various alleles, individually near-neutral, to accumulate in a population, and provides a way they can be systematically and synergistically be combined, either negatively (and eliminated) or positively (and spread). This is achieved without an intolerable increase in the mutation rate (which is the only way that near-neutral alleles can accumulate in an asexual population). Likewise, humans can individually discover new idea fragments (such as those outlined in this paper) but only the collective process of combination and appraisal called Science allows their diffusion. Much of the human massive cortical expansion underpins the protocols that allow such "brain-sex", but this requires the core, generic underlying neocortical proofreading process, in much the same way that sex is underpinned by mutation, and requires elaborate special machinery.

6 Summary

Although this work covers many technical details at various levels and fields of analysis, our thesis is simple, naïve and we hope powerful: the mysterious and quasimiraculous states of matter we call "Life" and "Mind" are the result of the intensive repetition of elementary selective amplification processes such as base-copying and synapse-strengthening. The outcome of such straightforward processes is remarkable because the selectivity is extraordinarily high: in the case of base-copying no constant in physics has a lower error. But extraordinary selectivity requires extraordinary machinery. For DNA, that machinery involves an elaborate protein complex whose key component is a proofreading step that enormously boosts accuracy. Our contribution to this emerging picture, which explains unexpected "effects" in terms of elemental "causes", has 2 parts. First, we (and others) show mathematically that learning from higher order correlations, probably necessary for any form of understanding (and thus "mind"), breaks down, in a fixed-point bifurcation, unless synaptic adjustment accuracy is extremely high. Second, we propose that the unique elaborate circuitry of the neocortex (which seems to at least facilitate intelligence) performs the proofreading operation necessary for such accuracy. Intriguingly, both these ideas have strong parallels in Darwinian evolution, suggesting that life and mind are closely related phenomena. But these ideas lie firmly within the existing scientific framework: we are NOT proposing new and outlandish principles. Instead, we believe that very careful analysis of the implications of current ideas and facts, has and can lead to significant progress. Mind would be extraordinary because it uses extraordinary, though understandable, machinery.

References

1. Maynard Smith, J., Szathmary, E.: The Major Transitions in Evolution. Freeman (1995)
2. Otto, S.P.: The Evolutionary Enigma of Sex. American Naturalist 174, S1–S114 (2009)
3. Cox, K.J.A., Adams, P.: Hebbian crosstalk prevents nonlinear unsupervised learning. Front. Comput. Neurosci. 3, 11 (2009), doi:10.3389/neuro.10.011.2009 (published online September 24, 2009)
4. Elliott, T.: Cross-Talk Induces Bifurcations in Nonlinear Models of Synaptic Plasticity. Neural Computation (in press)
5. Adams, P.R., Cox, K.J.A.: A neurobiological perspective on building intelligent devices. The Neuromorphic Engineer 3, 2–8 (2006), http://www.ine-news.org/view.php?source=0036-2006-05-01