Biological Correlates of Machine-Learning

W. Gordon Kruberg, M.D.

February 11, 2021

Abstract

The greatest successes of artificial intelligence are intelligent machines founded on models of how neurons interact with each other. In creating these models, machine-learning modelers divide intelligent behavior into separate learning and operating experiences: training and inference. We, the public, see machine-learning engines while they are operating, in inference mode, as they interpret our requests and images. To make them operational, their neural nets are trained using algorithms that are widely tested, optimized, and which even compete against each other regularly. The most successful training algorithms use back-propagation to train their neural networks.

What has been learned about intelligence from these models is largely absent in biological models: that creating the memories underlying intelligent behavior occurs independently of network operations and requires network-level functions. This paper recasts memory research in the context of those two requirements and outlines novel biological correlates for training and inference modes, vector spaces and error terms. Specific biological machinery is identified as holding the key to understanding memory creation: the operation of tripartite synapses and how astrocytes act as normalization operators to manage synaptic plasticity.

Introduction

The back-propagation algorithm was devised to train the synaptic weights of a multi-layer artificial neural network. Using a gold-standard collection of characterized samples, the algorithm specifies how to change synaptic weights in each network layer guided by the influence each synapse has on errors. One summary [1] highlights both its value and the reasons it is considered implausible in nature, underscoring the persistent interest in constructing biological models using the algorithm.

Biological systems have short-term and long-term memories. There is an elegant congruence between the accepted model of biological memory and the training-inference paradigm used to create artificial neural networks. In the analogy this paper uncovers the biological tools for gradient-descent.

- During wakefulness, the hippocampus receives entorhinal and amygdala signals encoding the brains inputs and outputs, which it uses to identify novel features and patterns. These are stored temporarily as a cache by modifying synaptic weights and function as our short-term memory.
- During sleep, conversely, replay of that hippocampus cache provides the target output-vector used for gradient descent in the neocortex. Thalamic neurons use a side-channel to communicate, burst-mode, to encode reference input-vectors using the same synapses used for inference.
- Short-term and long-term memory are compartmentalized; however, they simultaneously demand real-time integration of their vector-spaces. This can only occur if transposition is implemented in real-time between compartments.
- Astrocytes implement synaptic and network topology changes using an astonishing array of molecular tools. The intersection of astrocyte network and neuronal network at tripartite glutamate synapses drives network calculations such as normalization.

Training and Inference In Biological Memory

Making memories in a brain requires two steps. The original report of patient Henry Molaison [2] inspired theories that separate the creation of explicit memory from the use of existing memory. In the two-steps, a short-term memory of an event or item is created during wakefulness and then later consolidated into long-term memory. The consolidation is driven by spindles during sleep [3] and managed locally by different brain regions [4]. This process divides brain operation into two modes, where short-term memory identifies new patterns and long-term learning happens during sleep.

The two memory stages, short-term cache and long-term memory, are not colocated. Short-term memory is in the hippocampus and long-term memory distributed throughout the neocortex. Memory consolidation during replay is the transfer of pattern-matching and pattern-completion operations from the hippocampus into the neocortex. While awake, though, these separate memory compartments maintain real-time mapping which transposes the hippocampus output-space onto its own input-space.

During wakefulness, while the rest of the cerebrum is engaged as an inference engine, dentate gyrus granule cells discover new patterns. The identification of novel features in these patterns has been shown to occur on a behavioral time-scale[5]. Both separating [6] and completing [7] patterns extracted from hippocampus input, they project engrams outside the hippocampus where amygdala and entorhinal cortex cells form part of an observable engram ensemble. These engram-ensembles emerge after some form of competition between neurons for participation in the synaptic weight changes[8]

Hippocampal Replay Is The Neocortical Target Output-Vector

During sleep, the synaptic weight changes supporting these pattern operators are pushed to the rest of the brain and consolidated into long-term memory. The hippocampus cells repeat firing patterns from recent experiences, reactivating cells in the cerebral cortex. This cellular replay recapitulates activity during NREM sleep [9] to underpin high-level semantic-trees [10] – even extending to integrating motor memories[11]. It contributes to consolidation but also underlies navigation planning, and appears to be the expression or recall of short-term memory[12].

Our search for biological correlates of machine-learning requires a target output-vector, the desired activity-levels we can use as a gold standard to guide changes in synaptic weights. Hippocampal replay, projecting recent behavior to the cortex, is the prime candidate for expressing the target output-vectors needed for a gradient-descent strategy in the neocortex. During sleep, biology's training session, the neocortex uses the activity saved as target patterns in the hippocampus to compare against untrained activity.

Thalamocortical Bursts Are The Neocortical Input-Vectors

To understand what the neocortex uses for comparison against hippocampal replay, we need to recast the role of the thalamus from mere sensory relay [13] to authoritative reference of sensory information. During wakefulness, the primary thalamic sensory nuclei form the basis vectors of the sensory input-vector space. In the context of back-propagation within the neocortex, this vector-space is preserved but used for training. By extension, higher-order thalamic relays from cingulate and motor cortices provide input-vectors for forward-pass error-calculations in other regions. A critical question is what signaling supports the neocortex recognizing thalamocortical training from inference. Thalamocortical cells support two signaling modes [14], tonic-mode and phasic-, or burst-, mode, where tonic mode encodes sensory signals directly and burst-mode is characteristic of sleep [15] This dichotomy supports the idea that phasic bursts form a side-channel to post-synaptic neurons or possibly astrocytes to support consolidation during replay. This side-channel uses the same neuronal network, and synapses, as the tonic channel uses during inference mode; using these signals as input in back-propagation satisfies the requirement for symmetry.

There are two thalamic regions of special architectural note. First, the thalamic reticular nucleus has a role broadly across all thalamic nuclei and likely manages thalamic toggling between inference and training modes[4]. Second, anterior thalamic nuclei are critical for consolidation [16] and known for burst properties[17]. They are the endpoint of pathways from the hippocampus via the mammillary bodies, and project widely including to the cingulate cortex and from there have wide distribution and impact[18].

Managing Synaptic Weight Changes

During memory consolidation, using burst-mode thalamocortical projections for input and hippocampal replay as target output, synaptic weight changes require network-level coordination. Evidence for synaptic strengthening and weakening in response to neuronal activity has been investigated and well-described elsewhere[19]. Recent works have indicated the importance of non-

neuronal actors, specifically astrocytes, to synaptic plasticity[20]. These developments open a critical door to the possibility of network-level operators that is shut to neuron-only architectures.

Astrocytes located throughout the brain have long been known to distribute energy to neurons [21]. They are now known to participate in tripartite synapses and have been implicated directly in memory consolidation [22] [23]. One human astrocyte can encompass as many as 2 million so-called tripartite synapses [24] and extends its influence by interlinking [25] to other astrocytes using gap junctions. Intracellular astrocyte Ca^{++} concentration is a strong candidate as a synapse-level activity measure that propagates along astrocyte processes the extent of an astrocyte's region.

An astrocyte's arsenal for synaptic strength modification during replay is substantial, in the presynaptic bouton, synaptic cleft, and post-synaptic density. Astrocytes support several secretory peptide responses [26] which are active in synaptogenesis [27] [28] [29] and relevant to the fear-conditioning response [30]. Astrocytic interferon receptors have been related to spatial learning[31], their TGF- β release triggers downstream Usp9X that regulates ankyrin-G stability [32] and they are known to provide direction for axon growth cones[33]. It is known that glia contribute cholesterol compounds to memory construction[34] and that ApoE is a functional presynaptic signal from astrocytes to neurons [35] involved in synaptogenesis. The control signals and relative strengths of all these mechanisms require characterization, but are of sufficient magnitude to influence network characteristics. System response time-constants range from milliseconds for calcium-based intracellular signals to hours for proteogenesis. They are in the apparently unique position of monitoring activity of unconnected neurons while directing arborization. Whether they contribute significantly to neuronal activity during inference mode may be in question[36]; however, they are certainly active participants in the management of neuronal networks.

Astrocytes Are Normalization Operators

Tripartite synapses are the crossroads of communication systems using glutamate, D-serine, sodium, and both astrocyte and neuron intracellular calcium. Electrical signals are rapidly transmitted from the presynaptic to postsynaptic neuron; orthogonally, calcium-mediated diffusion along astrocyte processes communicates synaptic activity between tens of thousands of synapses. Signals across the synapse involve at least three channels: one from axon to postsynaptic AMPAR Na^+ channel, one from axon to astrocyte via GLT-1 transport, and one jointly from axon (glutamate) and astrocyte (D-serine) to postsynaptic NMDAR Ca^{++} channel.

This multifunctional junction-box embeds a mechanism for gating post-synaptic responses to errors: the voltage-dependency of NMDA receptors. NMDA receptors require post-synaptic depolarization, act with long time constants, and are known to be involved in LTP[37]. The voltage-dependence has been described as coincidence-detection and Hebbian in nature but the D-serine switch managed by astrocytes [38] really gives the astrocyte status as controller.

Astrocytic D-serine release is modulated by Ca^{++} concentration, itself a function of calcium diffusion from other synapses along the astrocyte process. In astrocytes, where D-serine release is mediated by Ca^{++} , neighborhood synaptic activity increases or decreases local concentrations. This normalizes an astrocyte's neighborhood D-serine release into synapses. This form of normalization is needed during the engram-collective assignment competition in the hippocampus, where astrocytes contribute to learning [22]. It can signal presynaptic non-winners of the engram competition to stand-down and not undergo synaptic strengthening.

In the neocortex, an additional signal is available to astrocytes where GLT-1 uptake [39] may transduce thalamic burst-mode signaling. Data encoded in both burst-mode and tonic-mode is comparable [40] but sits on different baseline activity: tonic-firing releases glutamate with baseline activity greater than burst-mode firing. The astrocytic recycling response impacts local Ca^{++} levels and, including Ca^{++} diffusion along the astrocyte processes from neighborhood synapses, local D-serine release.

Finally, astrocytes can use the data sifted from network-wide monitoring of synapses to change the topology of the inhibitory neurons as well. GABAergic neurons have been shown to respond with axonal topology changes including morphology, arborization, and synaptogenesis [41].

Discussion

Neither pattern-recognition caching nor long-term memory can occur without a network management system to implement network-level weight changes. Hebb's idea that "*persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability...*" has not found a place in machine learning because training is a network function: it occurs not where two cells fire at the same time, but in situations where they explicitly do not fire at the same time yet should. The core principle of gradient descent is calculation across the matrix of the impact of weight changes needed to recognize specific input features.

Biology appears to have both architectural features and sufficient network-operators to consider memory consolidation in the neocortex an implementation of gradient-descent. Biological formation of long-term memory is operationally congruent to the training cycles of machine learning and presents a plausible approach to biological gradient descent. Biological gradient descent is plausible because (1) inference is separated from forward and backward error and training passes by the bistable sleep-wake cycle, (2) error terms are provided in replay content, (3) transpose operators are evident in the bidirectional mapping between cerebral cortex and hippocampus compartments, and (4) the astrocyte network is available to implement network operators like normalization and gradient calculations.

The significant role of astrocytes is not the transient direct modulation of neuronal activity, but the modulation of synaptic strength and network topology with durable impact. These changes are informed by monitoring, at individual and network-levels, neuronal activity and synaptic strength. Much is to be learned about the specifics of astrocyte signaling with neurons during memory formation, and how tripartite synapses differ between hippocampus, neocortex, and cerebellum to enable short versus long time-constant changes.

This operational model of a bicameral memory in a bimodal brain is an architecture amenable to the theoretical underpinnings of machine learning that draws on observations well-described in neurophysiological literature. The tripartite glutamatergic synapse is implicated as the specific mediator of synaptic weight change and network-level management. The model allows us to differentiate more clearly those aspects of neurophysiology which are topological from those which are algorithmic. My hope is that the congruence between models will assist in unifying our approach to complex behavior and in forming a foundation for general intelligence.

References

- [1] **Lillicrap, T. P.; Santoro, A.; Marris, L.; Akerman, C. J. and Hinton, G. (2020).** *Backpropagation and the brain*, Nature Reviews Neuroscience 21 : 335-346.
- [2] **Scoville, W. B. and Milner, B. (1957).** *Loss Of Recent Memory After Bilateral Hippocampal Lesions* , Journal of Neurology, Neurosurgery & Psychiatry 20 : 11-21.
- [3] **Peyrache, A. and Seibt, J. (2020).** *A mechanism for learning with sleep spindles*, Philosophical Transactions of the Royal Society B: Biological Sciences 375 : 20190230.
- [4] **Vantomme, G.; Osorio-Forero, A.; Lüthi, A. and Fernandez, L. M. J. (2019).** *Regulation of Local Sleep by the Thalamic Reticular Nucleus*, Frontiers in Neuroscience 13.
- [5] **Bittner, K. C.; Milstein, A. D.; Grienberger, C.; Romani, S. and Magee, J. C. (2017).** *Behavioral time scale synaptic plasticity underlies CA1 place fields*, Science 357 : 1033-1036.
- [6] **Leutgeb, J. K.; Leutgeb, S.; Moser, M.-B. and Moser, E. I. (2007).** *Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus*, Science 315 : 961-966.
- [7] **Guzowski, J. F.; Knierim, J. J. and Moser, E. I. (2004).** *Ensemble Dynamics of Hippocampal Regions CA3 and CA1*, Neuron 44 : 581-584.
- [8] **Josselyn, S. A. and Tonegawa, S. (2020).** *Memory engrams: Recalling the past and imagining the future*, Science 367 : eaaw4325.
- [9] **Oyarzún, J. P.; Mors, J.; Luque, D.; de Diego-Balaguer, R. and Fuentemilla, L. (2017).** *Targeted Memory Reactivation during Sleep Adaptively Promotes the Strengthening or Weakening of Overlapping Memories*, The Journal of Neuroscience 37 : 7748-7758.
- [10] **Mao, D.; Neumann, A. R.; Sun, J.; Bonin, V.; Mohajerani, M. H. and McNaughton, B. L. (2018).** *Hippocampus-dependent emergence of spatial sequence coding in retrosplenial cortex*, Proceedings of the National Academy of Sciences 115 : 8015-8018.
- [11] **Mander, B. A.; Zhu, A. H.; Lindquist, J. R.; Villeneuve, S.; Rao, V.; Lu, B.; Saletin, J. M.; Ancoli-Israel, S.; Jagust, W. J. and Walker, M. P. (2017).** *White Matter Structure in Older Adults Moderates the Benefit of Sleep Spindles on Motor Memory Consolidation*, The Journal of Neuroscience 37 : 11675-11687.
- [12] **Ólafsdóttir, H. F.; Bush, D. and Barry, C. (2018).** *The Role of Hippocampal Replay in Memory and~Planning*, Current Biology 28 : R37-R50.
- [13] **Guillery, R. and Sherman, S. (2002).** *Thalamic Relay Functions and Their Role in Corticocortical Communication*, Neuron 33 : 163-175.
- [14] **Sherman, S. M. (2001).** *Tonic and burst firing: dual modes of thalamocortical relay.*, Trends in neurosciences 24 : 122-126.
- [15] **Steriade, M.; McCormick, D. A. and Sejnowski, T. J. (1993).** *Thalamocortical oscillations in the sleeping and aroused brain.*, Science (New York, N.Y.) 262 : 679-685.
- [16] **Garden, D. L. F.; Massey, P. V.; Caruana, D. A.; Johnson, B.; Warburton, E. C.; Aggleton, J. P. and Bashir, Z. I. (2009).** *Anterior thalamic lesions stop synaptic plasticity in retrosplenial cortex slices: expanding the pathology of diencephalic amnesia.*, Brain : a journal of neurology 132 : 1847-1857.
- [17] **Bokor, H.; Frère, S. G.; Eyre, M. D.; Slézia, A.; Ulbert, I.; Lüthi, A. and Acsády, L. (2005).** *Selective GABAergic Control of Higher-Order Thalamic Relays*, Neuron 45 : 929-940.
- [18] **Aggleton, J. P.; O'Mara, S. M.; Vann, S. D.; Wright, N. F.; Tsanov, M. and Erichsen, J. T. (2010).** *Hippocampal-anterior thalamic pathways for memory: uncovering a network of direct and indirect actions*, European Journal of Neuroscience 31 : 2292-2307.
- [19] **Asok, A.; Leroy, F.; Rayman, J. B. and Kandel, E. R. (2019).** *Molecular Mechanisms of the Memory Trace*, Trends in Neurosciences 42 : 14-22.
- [20] **Allen, N. J. and Eroglu, C. (2017).** *Cell Biology of Astrocyte-Synapse Interactions*, Neuron 96 : 697-708.
- [21] **Sofroniew, M. V. (2020).** *Astrocyte Reactivity: Subtypes, States, and Functions in CNS Innate Immunity*, Trends in Immunology 41 : 758-770.

- [22] **Adamsky, A.; Kol, A.; Kreisel, T.; Doron, A.; Ozeri-Engelhard, N.; Melcer, T.; Refaeli, R.; Horn, H.; Regev, L.; Groysman, M.; London, M. and Goshen, I. (2018).** *Astrocytic Activation Generates De Novo Neuronal Potentiation and Memory Enhancement*, *Cell* 174 : 59-71.e14.
- [23] **Kol, A.; Adamsky, A.; Groysman, M.; Kreisel, T.; London, M. and Goshen, I. (2020).** *Astrocytes contribute to remote memory formation by modulating hippocampal–cortical communication during learning*, *Nature Neuroscience* 23 : 1229-1239.
- [24] **Oberheim, N. A.; Takano, T.; Han, X.; He, W.; Lin, J. H. C.; Wang, F.; Xu, Q.; Wyatt, J. D.; Pilcher, W.; Ojemann, J. G.; Ransom, B. R.; Goldman, S. A. and Nedergaard, M. (2009).** *Uniquely Hominid Features of Adult Human Astrocytes*, *Journal of Neuroscience* 29 : 3276-3287.
- [25] **Arizono, M.; Inavalli, V. V. G. K.; Panatier, A.; Pfeiffer, T.; Angibaud, J.; Levet, F.; Veer, M. J. T. T.; Stobart, J.; Bellocchio, L.; Mikoshiba, K.; Marsicano, G.; Weber, B.; Oliet, S. H. R. and Nägerl, U. V. (2020).** *Structural basis of astrocytic Ca²⁺ signals at tripartite synapses*, *Nature Communications* 11.
- [26] **Jha, M. K.; Kim, J.-H.; Song, G. J.; Lee, W.-H.; Lee, I.-K.; Lee, H.-W.; An, S. S. A.; Kim, S. and Suk, K. (2018).** *Functional dissection of astrocyte-secreted proteins: Implications in brain health and diseases*, *Progress in Neurobiology* 162 : 37-69.
- [27] **Eroglu, Ç.; Allen, N. J.; Susman, M. W.; O'Rourke, N. A.; Park, C. Y.; Özkan, E.; Chakraborty, C.; Mulinyawe, S. B.; Annis, D. S.; Huberman, A. D.; Green, E. M.; Lawler, J.; Dolmetsch, R.; Garcia, K. C.; Smith, S. J.; Luo, Z. D.; Rosenthal, A.; Mosher, D. F. and Barres, B. A. (2009).** *Gabapentin Receptor $\alpha 2\delta$ -1 Is a Neuronal Thrombospondin Receptor Responsible for Excitatory CNS Synaptogenesis*, *Cell* 139 : 380-392.
- [28] **Kucukdereli, H.; Allen, N. J.; Lee, A. T.; Feng, A.; Ozlu, M. I.; Conatser, L. M.; Chakraborty, C.; Workman, G.; Weaver, M.; Sage, E. H.; Barres, B. A. and Eroglu, C. (2011).** *Control of excitatory CNS synaptogenesis by astrocyte-secreted proteins Hevin and SPARC*, *Proceedings of the National Academy of Sciences* 108 : E440-E449.
- [29] **Xie, Y.; Kuan, A. T.; Wang, W.; Herbert, Z. T.; Mosto, O.; Olukoya, O.; Adam, M.; Vu, S.; Kim, M.; Gómez, N.; Tran, D.; Charpentier, C.; Sorour, I.; Tolstorukov, M. Y.; Sabatini, B. L.; Lee, W.-C. A. and Harwell, C. C. (2020).** *Astrocyte-neuron crosstalk through Hedgehog signaling mediates cortical circuit assembly*, .
- [30] **Mazaré, N.; Oudart, M.; Moulard, J.; Cheung, G.; Tortuyaux, R.; Mailly, P.; Mazaud, D.; Bemelmans, A.-P.; Boulay, A.-C.; Blugeon, C.; Jourdain, L.; Crom, S. L.; Rouach, N. and Cohen-Salmon, M. (2020).** *Local Translation in Perisynaptic Astrocytic Processes Is Specific and Changes after Fear Conditioning*, *Cell Reports* 32 : 108076.
- [31] **Hosseini, S.; Michaelsen-Preusse, K.; Grigoryan, G.; Chhatbar, C.; Kalinke, U. and Korte, M. (2020).** *Type I Interferon Receptor Signaling in Astrocytes Regulates Hippocampal Synaptic Plasticity and Cognitive Function of the Healthy CNS*, *Cell Reports* 31 : 107666.
- [32] **Yoon, S.; Parnell, E. and Penzes, P. (2020).** *TGF- β -Induced Phosphorylation of Usp9X Stabilizes Ankyrin-G and Regulates Dendritic Spine Development and Maintenance*, *Cell Reports* 31 : 107685.
- [33] **Hung, J. and Colicos, M. A. (2008).** *Astrocytic Ca²⁺ Waves Guide CNS Growth Cones to Remote Regions of Neuronal Activity*, *PLoS ONE* 3 : e3692.
- [34] **Mauch, D. H. (2001).** *CNS Synaptogenesis Promoted by Glia-Derived Cholesterol*, *Science* 294 : 1354-1357.
- [35] **Huang, Y.-W. A.; Zhou, B.; Nabet, A. M.; Wernig, M. and Südhof, T. C. (2019).** *Differential Signaling Mediated by ApoE2, ApoE3, and ApoE4 in Human Neurons Parallels Alzheimer's Disease Risk*, *The Journal of Neuroscience* 39 : 7408-7427.
- [36] **Sloan, S. A. and Barres, B. A. (2014).** *Looks Can Be Deceiving-- Reconsidering the Evidence for Gliotransmission.pdf*, *Neuron* 84 : 1112-1115.
- [37] **Henneberger, C.; Papouin, T.; Oliet, S. H. R. and Rusakov, D. A. (2010).** *Long-term potentiation depends on release of d-serine from astrocytes*, *Nature* 463 : 232-236.

- [38] **Neame, S.; Safory, H.; Radzishevsky, I.; Touitou, A.; Marchesani, F.; Marchetti, M.; Kellner, S.; Berlin, S.; Foltyn, V. N.; Engelender, S.; Billard, J.-M. and Wolosker, H. (2019).** *The NMDA receptor activation by d-serine and glycine is controlled by an astrocytic Phgdh-dependent serine shuttle*, Proceedings of the National Academy of Sciences 116 : 20736-20742.
- [39] **Price, J. L.; Ko, A. I.; Wade, M. J.; Tsou, S. K.; McKeel, D. W. and Morris, J. C. (2001).** *Neuron Number in the Entorhinal Cortex and CA1 in Preclinical Alzheimer Disease*, Archives of Neurology 58 : 1395.
- [40] **Reinagel, P.; Godwin, D.; Sherman, S. M. and Koch, C. (1999).** *Encoding of Visual Information by LGN Bursts*, Journal of Neurophysiology 81 : 2558-2569.
- [41] **Hughes, E. G.; Elmariah, S. B. and Balice-Gordon, R. J. (2010).** *Astrocyte secreted proteins selectively increase hippocampal GABAergic axon length, branching, and synaptogenesis*, Molecular and Cellular Neuroscience 43 : 136-145.