Dear Gary,

I just read your interesting *Science* article. I was intrigued by your comments about cerebral cortex that "its basic logic remains unknown" and that "there is little evidence that such uniform architectures can capture the diversity of cortical function in simple mammals". I was also struck by your comments that cortical models "might include circuits for shifting the focus of attention, for encoding and manipulating sequences, and for normalizing the ratio between the activity of an individual neuron and a set of neurons…and for working memory storage, decision-making, storage and transformation of information via population coding…, alongside machinery for hierarchical pattern recognition". You also commented about such matters as "temporal synchrony among neural ensembles…to precisely controlled recurrent interactions between the prefrontal cortex and basal ganglia…"

Actually, there is an emerging unified laminar cortical theory that embodies all of these properties, and that has been used to provide unified explanations and predictions about psychological, anatomical, neurophysiological, biophysical, and some biochemical data. This theory, whose various component models are often unified under the general heading of LAMINART theory, has been getting rapidly developed since the first article about it appeared in *Trends in Neurosciences* in 1997.

The name LAMINART acknowledges the synthesis of concepts about the design of laminar cortical architectures with more long-standing principles and mechanisms of Adaptive Resonance Theory, or ART, which began as a cognitive and neural theory of how the brain autonomously learns to categorize, recognize, and predict objects and events in a changing world. As illustrated in the review article Grossberg (2012, http://cns.bu.edu/~steve/ART.pdf), ART is arguably the currently most highly developed cognitive and neural theory available, with the broadest explanatory and predictive range. It has been getting progressively developed since I introduced it in 1976 to propose a solution to the classical *stability-plasticity dilemma*. This proposed solution enables ART to carry out fast, incremental, and self-stabilizing unsupervised and supervised learning in response to a changing world.

ART specifies mechanistic links between processes of consciousness, learning, expectation, attention, resonance, and synchrony during both unsupervised and supervised learning. ART provides functional and mechanistic explanations of such diverse topics as laminar cortical circuitry; invariant object and scenic gist learning and recognition; prototype, surface, and boundary attention; gamma and beta oscillations; learning of entorhinal grid cells and hippocampal place cells; computation of homologous spatial and temporal mechanisms in the entorhinal-hippocampal system; vigilance breakdowns during autism and medial temporal amnesia; cognitive-emotional interactions that focus attention on valued objects in an adaptively timed way; item-order-rank working memories and learned list chunks for the planning and control of sequences of linguistic, spatial, and motor information; conscious speech percepts that are influenced by future context; auditory streaming in noise during source segregation; and speaker normalization. Brain regions that are functionally described include visual and auditory neocortex; specific and nonspecific thalamic nuclei; inferotemporal, parietal, prefrontal, entorhinal, hippocampal, parahippocampal, perirhinal, and motor cortices; frontal eye fields; supplementary eye fields; amygdala; basal ganglia: cerebellum; and superior colliculus. Due to the complementary organization of the brain, ART does not describe many spatial and motor behaviors whose matching and learning laws differ from those of ART.

Given Randy O-'Reilly's comments about Leabra, it is also of historical interest that I introduced the core equations used in Leabra in the 1960s and early 1970s, and they have proved to be of critical importance in all the developments of ART.

To illustrate how LAMINART illustrates the type of laminar cortical theory that your *Science* article discusses, let me refer interested readers to a few archival articles. LAMINART proposes how all cortical areas combine bottom-up, horizontal, and top-down interactions, thereby beginning to functionally clarify why all granular neocortex has a characteristic architecture with six main cell layers, and how these laminar circuits may be specialized to carry out different types of biological intelligence. In particular, this unification shows how variations of a shared laminar cortical design can be used to explain and simulate psychological and neurobiological data about vision, speech, and cognition:

Vision. The 3D LAMINART model integrates bottom-up and horizontal processes of 3D boundary formation and perceptual grouping, surface filling-in, and figure-ground separation with top-down attentional matching and oscillatory dynamics in cortical areas such as V1, V2, and V4 (Cao and Grossberg, 2005; Fang and Grossberg, 2009; Grossberg, 1999; Grossberg and Raizada, 2000; Grossberg and Swaminathan, 2004; Grossberg and Versace, 2008; Grossberg and Yazdanbakhsh, 2005; Raizada and Grossberg, 2001). It is arguably the currently most highly developed vision model with the broadest explanatory and predictive range, laminar or not. This model, as well as the other models listed below, also makes multiple predictions about the functional roles that are played by identified cortical cells in all of these visual processes.

Speech. The cARTWORD model proposes how bottom-up, horizontal, and top-down interactions within a hierarchy of laminar cortical processing stages, modulated by the basal ganglia, can generate a conscious speech percept that is embodied by a resonant wave of activation that occurs between acoustic features, acoustic item chunks, and list chunks (Grossberg and Kazerounian, 2011). Chunk-mediated gating allows speech to be heard in the correct temporal order, even when what is consciously heard depends upon using future context to disambiguate noise-occluded sounds, as occurs during phonemic restoration.

Cognition. The LIST PARSE model how bottom-up, horizontal, and top-down interactions within the laminar circuits of lateral prefrontal cortex may carry out working memory storage of event sequences within layers 6 and 4, how unitization of these event sequences through learning into list chunks may occur within layer 2/3, and how these stored sequences can be recalled at variable rates that are under volitional control by the basal ganglia (Grossberg and Pearson, 2008). In particular, the model uses variations of the same circuitry to quantitatively simulate human cognitive data about immediate serial recall and immediate free recall, delayed free recall, and continuous distracter free recall; and monkey neurophysiological data from the prefrontal cortex obtained during sequential sensory-motor imitation and planned performance.

Prefrontal-basal ganglia interactions. In addition to the thalamcortical interactions embodied in the above models, neocortical interactions with other subcortical structures have been developed as part of this emerging theory, notably cognitive-emotional interactions, reinforcement learning, and gating of plans and movements. These are also reviewed in Grossberg (2012). Here I will just mention one of these models that focuses on the kinds of prefrontal-basal ganglia interactions that you mentioned in your *Science* article.

The lisTELOS model builds upon, and unifies, the working memory and basal ganglia circuits of the LIST PARSE and TELOS models. In particular, Silver et al. (2011) have incorporated an item-order-rank spatial working memory into a comprehensive model of how sequences of eye movements, which may include repetitions, may be planned and performed. Similar mechanisms may be expected to control other types of sequences as well, for reasons that are reviewed in Grossberg (2012). The lisTELOS model's name derives from the fact that it unifies and further develops concepts from LIST PARSE about how item-order-rank working memories store lists of items, and of how TELOS model properties of the basal ganglia (Brown et al., 1999, 2004) help to balance reactive vs. planned movements by selectively gating sequences of actions through time.

Shunting dynamics and ratio processing. The kind of shunting dynamics that enables automatic computation of activity ratios has been a critical component of all models that my colleagues and I have developed since my foundational article (Grossberg, 1973) first mathematically proved how this works in both non-recurrent and recurrent networks. Indeed, ART models may be viewed as self-organizing production systems that carry out a novel kind of probabilistic hypothesis testing and decision-making that is designed to work in response to big non-stationary data bases.

New computational paradigms. These examples illustrate an emerging unified theory of how variations of a shared laminar neocortical design can carry out multiple types of biological intelligence. Semi-classical models, such as deep learning, have been very useful in technology, but have little to offer in explaining how our brains have evolved to control autonomous adaptive behaviors. This weakness of deep learning is partly explained by the fact that these laminar cortical models embody revolutionary new computational paradigms that I have called Laminar Computing and Complementary Computing, which underlie natural computational realizations for biological systems that have evolved to autonomously and stably adapt in real time to a rapidly changing and unpredictable world.

Indeed, LAMINART embodies a new type of hybrid between *feedforward* and *feedback* computing, and also between *digital* and *analog* computing for processing distributed data. These properties go beyond the types of Bayesian models that are so popular today. They underlie the fast but stable self-organization that is characteristic of cortical development and life-long learning. Their circuits "run as fast as they can": they behave like a real-time probabilistic decision circuit that operates as quickly as possible, given the evidence. There is thus a trade-off between certainty and speed. They operate in a fast feedforward mode when there is little uncertainty, and automatically switch to a slower feedback mode when there is uncertainty. Feedback selects a winning decision that enables the circuit to speed up again, since activation amplitude, synchronization, and processing speed both increase with certainty.

LAMINART also embodies a novel kind of hybrid computing that simultaneously realizes the stability of digital computing and the sensitivity of analog computing. The coherence that is derived from synchronous storage in interlaminar and intercortical feedback loops provides the stability of digital computing—the feedback loop exhibits hysteresis that can preserve the stored pattern against external perturbations— while preserving the sensitivity of analog computation.

I should add that the new models are also of interest in technology, and indeed have been embodied in the software and hardware applications of many companies during the past few decades. A great deal of additional exciting research remains to be done to develop a unified software and hardware platforms for multiple types of autonomous adaptive intelligence. These promise to revolutionize computer science in general, and the design of autonomous adaptive mobile robots in particular.

References

Brown, J., Bullock, D., and Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience, 19*, 10502-10511.

Brown, J.W., Bullock, D., and Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, *17*, 471-510.

Cao, Y. and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, *18*, 515-578.

Fang, L. and Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, *22,* 45-82.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *52*, 213-257.

Grossberg, S., and Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: A neural model of phonemic restoration using subsequent context in noise. *Journal of the Acoustical Society of America*, *130*, 440-460.

Grossberg, S., and Pearson, L. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review, 115*, 677-732 .

Grossberg, S., and Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40***,** 1413-1432.

Grossberg, S., and Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention and bistability. *Vision Research*, *44*, 1147-1187.

Grossberg, S., and Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, *1218*, 278-312.

Grossberg, S., and Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, *45*, 1725-1743.

Raizada, R. and Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, *13*, 100-113.

Silver, M.R., Grossberg, S., Bullock, D., Histed, M.H., and Miller, E.K. (2011). A neural model of sequential movement planning and control of eye movements: Item-order-rank working memory and saccade selection by the supplementary eye fields. *Neural Networks*, *26*, 29-58.