Computational Contributions of the Thalamus to Learning and Memory

Randall C. O'Reilly & Thomas E. Hazy
Departments of Psychology and Computer Science, eCortex, Inc.
Center for Neuroscience
University of California, Davis
1544 Newton Ct
Davis, CA 95616
oreilly@ucdavis.edu

October 19, 2021

Abstract:

The higher order thalamus (e.g., pulvinar) is widely thought to play a critical role in its interactions with the neocortex, but identifying precisely what that role is has been somewhat challenging. Here, we describe how a computational approach to understanding the nature of learning and memory in the neocortex suggests three distinct, well-defined contributions of the thalamus: 1) Attention, which is perhaps the most widely discussed function of the pulvinar, is supported by a pooled inhibition dynamic involving the thalamic reticular nucleus; 2) Predictive learning, where the pulvinar serves as a kind of screen on which predictions are projected, and a temporal difference between predictions and subsequent outcomes can drive error-driven learning throughout the thalamocortical system; and 3) Executive function in the circuits involving the frontal cortex, where mediodorsal (MD) thalamus is largely similar anatomically to the pulvinar, and could thus support similar attentional and predictive learning functions, while ventral thalamic nuclei receive inhibitory modulation from the basal ganglia that support a gating function to regulate action based on a strong Go vs. NoGo competition informed by reinforcement learning. Taken together, these important modulatory and learning contributions of the thalamus suggest that a full computational understanding of the neocortex is significantly incomplete without integrating the thalamic circuitry.

Keywords: Attention, Predictive Learning, Basal Ganglia, Prefrontal Cortex, Gating, Computational Models

From an anatomical perspective, the thalamus should be in a position to make essential contributions to the overall function of the mammalian neocortex, given its extensive interconnectivity with every cortical area, and role as an intermediary between extensive subcortical systems. However, despite many promising leads and well-developed theories, the critical role of the thalamus arguably remains significantly underrepresented in the broader literature, all-too-often still relegated to its well-worn *relay* role. In this chapter, we provide a more thalamus-centric perspective, informed by a gradual appreciation for the many critical roles of the thalamus. This appreciation has emerged from the intersection between neural and computational necessity in the course of attempting to create large-scale systems-neuroscience-based computational models of the brain, to understand the neural basis of a variety of cognitive functions.

The scope of functions we consider are as follows:

Draft Manuscript: Do not cite or quote without permission.

- **Attention.** One of the most widely-discussed and supported functions of the thalamus is its role in modulating attention across the cortex, at multiple different scales (Bender & Youakim, 2001; Fiebelkorn & Kastner, 2019; Halassa & Kastner, 2017; LaBerge & Buchsbaum, 1990; Robinson & Petersen, 1992; Saalmann & Kastner, 2011; Snow, Allen, Rafal, & Humphreys, 2009; Zhou, Schafer, & Desimone, 2016). The influential *attentional searchlight* idea of Crick (1984) included a critical role for the thalamic reticular nucleus (TRN) in driving the focus of this attentional signal. However, the bursting-based mechanisms leveraged in that paper have been shown to operate mainly in the anesthetized and sleeping brain (and *in vitro* preparations), and are specifically associated with reductions in arousal (Halassa et al., 2011; Lewis et al., 2015). Nevertheless, one of the simplest possibilities was perhaps too quickly rejected by Crick (1984): the TRN can provide *pooled inhibition* across multiple different spatial scales, without the need for more complex bursting dynamics. Pooled inhibition aggregates the excitatory signals across a given pool of neurons, and feeds that "averaged" excitation uniformly back to those same neurons, such that only the most strongly excited neurons can remain strongly active in the face of the inhibition.

  The TRN constitutes a shell of inhibitory neurons surrounding and projecting topographically to the underlying thalamic nuclei, putting it in a unique position to support pooled inhibition across widely separated areas in the neocortex, where direct corticocortical pooling over such long distances would not be feasible. Distances in the thalamus are much smaller than in the neocortex because it is a compact structure without significant internal connectivity, compared to the complex networks present in the cortex. This form of inhibition can replicate the influential, abstract *normalization* model of Reynolds and Heeger (2009), and also fits with many elements of the *folded feedback* model of Grossberg (1999). Functionally and computationally, the broad-scale inhibition between pathways in the brain can enable dynamic switching between different ways of processing information, potentially playing a critical role in cognitive flexibility.

- **Learning.** The central, integrative role of the higher-order thalamic areas such as the *pulvinar* and *mediodorsal* (MD) nuclei was highlighted in the early *blackboard* model of Mumford (1991), and was further supported by the extensive analysis of this thalamocortical connectivity (Sherman & Guillery, 2006; Shipp, 2003; Usrey & Sherman, 2018). Likewise, more recent work has highlighted the contributions of these thalamic areas to integrating and coordinating processing across distributed brain areas (Fiebelkorn & Kastner, 2019; Fiebelkorn, Pinsk, & Kastner, 2018; Halassa & Kastner, 2017; Saalmann, Pinsk, Wang, Li, & Kastner, 2012). Building on these ideas and data, we hypothesized that the distinctive pattern of thalamocortical connectivity in these circuits, consisting of separate *driver* vs. *modulatory* pathways (Sherman & Guillery, 2006; Usrey & Sherman, 2018), may also play a critical role in learning (O'Reilly, Russin, Zolfaghar, & Rohrlich, 2021; O'Reilly, Wyatte, & Rohrlich, 2014, 2017). Specifically, the numerous top-down corticothalamic (CT) projections can generate a *prediction* of what will happen next over the pulvinar and other higher-order thalamic areas, while the strong bottom-up driver pathways provide a *ground truth* signal representing what actually did happen. This pulvinar-based predictive learning mechanism provides a biologically-plausible solution to the much-discussed, unresolved challenge of understanding what drives something like error backpropagation learning in the neocortex (Lillicrap, Santoro, Marris, Akerman, & Hinton, 2020; Whittington & Bogacz, 2019). Predictive learning provides an essentially unbounded, continuous supply of prediction error signals, which can drive the neocortex to develop sophisticated internal models of the environment. Thus, the thalamus may play a central role in one of the most important functions in the brain.

- **Executive Function.** Building on the earlier foundation of its well-established role in motor control (Mink, 1996), there is an emerging consensus that the basal ganglia circuitry, which centrally involves

the thalamus, plays a critical role in *gating* activity in the frontal cortex in the service of higher-level cognitive functions (Brown & Marsden, 1990; Chatham, Frank, & Badre, 2014; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Dahlin et al., 2008; Frank, 2005; Frank, Loughry, & O'Reilly, 2001; Graybiel, 1995; Houk, 2005; Middleton & Strick, 2000; O'Reilly & Frank, 2006; Pasupathy & Miller, 2005; Rac-Lubashevsky & Frank, 2020; Voytek & Knight, 2010). In the classical basal ganglia (BG) circuit, the thalamus serves as the interface between the BG output and the frontal cortex, where tonic inhibition from the globus pallidus internal segment (GPi) or substantia nigra pars reticulata (SNr) blocks strong recurrent activation through the corticothalamic loops. When the direct (Go) pathway in the BG fires, it disinhibits this thalamocortical loop. By analogy with motor action initiation (Mink, 1996), these same kinds of mechanisms could also serve to initiate the updating of working memory in the PFC. This function is known as *gating* in existing models of dopamine modulation of PFC (Braver & Cohen, 2000) and in a widely-used computational model, LSTM (Hochreiter & Schmidhuber, 1997). Gating is similar in many ways to attentional modulation, suggesting a commonality across these functions, but it takes on a more significant functional role in the context of the specialized dopamine-based learning mechanisms present in the BG. Whereas our original models of this gating dynamic used highly simplified representations of the thalamus (O'Reilly, 2006; O'Reilly & Frank, 2006), we are now exploring the diversity of thalamic interconnections with the frontal cortex, and discovering some important new functional roles as a result.

These three functions are synergistic: prediction operates on the focus of attention (not everything can be predicted), and prediction helps guide attention toward anticipated outcomes and spatial locations (Richter & de Lange, 2019). Both are driven by the same top-down corticothalamic projections. Further, executive function depends critically on manipulating the levers of attention in posterior cortical areas (including via direct control projections into the TRN via frontal area 46), and predictive learning likely plays a critical role in shaping executive function abilities. Taken together, these mechanisms can also be seen as driving some of the most important aspects of our subjective, conscious experience. For example, the fact that we perceive a stable and coherent external world, despite receiving constantly changing sensory signals, depends on predictive learning. To see this, you can gently nudge the bottom of one of your eyeballs with your finger — the world moves around with this manipulation, but it is otherwise rock steady despite constant saccadic eye movements. We have learned to predict the effects of our planned eye movements (Cavanagh, Hunt, Afraz, & Rolfs, 2010; Duhamel, Colby, & Goldberg, 1992; Wurtz, 2008), but have no such learning experience for the manually-induced motion. This principle extends well beyond eye movements, and likely underlies the various perceptual *constancies* such as size constancy, color constancy, etc: the brain learns about what is solid and reliably predictable (the external world), not the idiosyncratic and fleeting raw sensory input.

In light of the long history and large amount of empirical data documenting the dependence of conscious alertness and arousal on thalamic input to the cerebral cortex, we can usefully relate each of these three functions to the nature of *conscious awareness*. Specifically, many theoretical frameworks for understanding the nature of consciousness center around the notion of a mental workspace or blackboard, producing an integrated information processing system (Baars, 1988; Dehaene & Naccache, 2001; Edelman & Tononi, 2001; Tononi, 2004). Based on extensive neural and behavioral data, Lamme (2006) attributed a critical role for recurrent processing in supporting conscious awareness, and the extensive recurrent nature of corticothalamic connectivity suggests that it should also play a critical role, especially in focusing attention across areas in a coordinated way (Fiebelkorn & Kastner, 2019; Halassa & Kastner, 2017; Saalmann & Kastner, 2011). Ward (2011) offers a more elaborated theory of consciousness associated with thalamic function, along with a review of relevant literature. At a more basic level, the modulation of overall cortical arousal by nonspecific, intralaminar nuclei has long been recognized (Schiff, 2008), and the recurrent connections involving the deep cortical layers have now been shown to be critical in sustaining arousal states (Redinbaugh et al., 2020).

If these computational contributions of the thalamus are correct, then it is hard to imagine a more conse-quential non-cortical substrate for understanding overall cortical function. The thalamus is the modulatory and training master of the neocortex. While the neocortex is where all the learning and knowledge is actually encoded, without the thalamus it would be an undisciplined, untutored wild beast! Indeed, that colorfully characterizes the behavior of our computational models in the absence of the various thalamic mechanisms described above. In the remaining sections, we review each of the above-mentioned computational contri-butions of the thalamus.

Attentional Focusing Through TRN Pooled Inhibition

The importance of thalamic attentional mechanisms can be illustrated by the limitations of computational models of the visual system that lack such mechanisms. For example, we developed models that captured the overall hierarchy of superficial-layer cortical processing up through invariant object categorization neurons in inferotemporal (IT) cortex (Kobatake & Tanaka, 1994), but which lacked thalamic attentional mechanisms (O'Reilly, Wyatte, Herd, Mingus, & Jilk, 2013). These models showed that bidirectional excitation between cortical areas could drive appropriate bottom-up and top-down *constraint satisfaction* processing, which is important for resolving the many ambiguities in visual inputs (Wyatte, Curran, & O'Reilly, 2012). However, because these models only simulated superficial-layer (i.e., cortical lamina 2-3) pathways, they suffered significantly when faced with complex real-world visual scenes containing multiple different objects. All of the different features across the different objects merged and collided with each other, creating a big messy jumble.

Interestingly, this jumbled perception is exactly what people who have damage to various attentional pathways in the brain report as their subjective experience, including those with damage to the pulvinar nucleus of the thalamus (Karnath, Himmelbach, & Rorden, 2002; Petersen, Robinson, & Morris, 1987). If they are faced with single discrete objects or letters, they can process those just fine, but once multiple different objects or a normal page of text is presented, they cannot handle the sensory overload (known as *simultanagnosia* — the inability to process multiple simultaneously-present stimuli (Farah, 1990)). Thus, it is clear that the brain needs some kind of attentional mechanism to focus processing on a small subset of the external world at a time (Buschman & Kastner, 2015; Desimone & Duncan, 1995; Reynolds & Desimone, 2003; Reynolds & Heeger, 2009; Treisman, 1993). These attentional mechanisms are known to involve thalamocortical circuits (Halassa & Kastner, 2017), in addition to larger-scale brain pathways in the frontal and parietal lobes (Bisley & Goldberg, 2010; He et al., 2007). At a computational level, introducing these thalamically-based attentional circuits into our models has likewise enabled the processing of complex visual scenes.

Interestingly, most machine-learning approaches to vision have managed to get by without including at-tentional mechanisms (LeCun, Bengio, & Hinton, 2015), but these models have great difficulty recognizing some *adversarial* images that people consider obvious, and extensive analysis shows that these models are using texture and other aggregate patterns to recognize object categories, and are not really processing indi-viduated objects per se (Sinz, Pitkow, Reimer, Bethge, & Tolias, 2019). When spatial attention mechanisms are introduced into these models, these adversarial recognition problems are greatly reduced (Luo, Boix, Roig, Poggio, & Zhao, 2015), consistent with the idea that the mammalian visual system depends critically on attention for supporting robust object recognition.

To understand the neural basis of such attentional mechanisms, we begin with the influential *biased competition* model (Desimone & Duncan, 1995), which posits that attended neurons receive some kind of additional excitation or bias (which can come from any number of sources, including bottom-up salience or top-down control signals), which then enables them to out-compete other neurons, which become less active as a consequence. The frontal and parietal attentional networks, involving the systems subserving

overt eye movements, including the frontal eye fields (FEF) and the lateral inferior parietal area (LIP), play a particularly strong role in driving spatially-based attentional biases (Buschman & Kastner, 2015; He et al., 2007). This biased competition model has been widely supported (Reynolds & Desimone, 2003), and has been captured mathematically in the *normalization* model of attention shown in Figure 1, which shows how a competitive process of normalization driven by the *pooling* of activity over retinotopically-organized spatial regions can account for a range of attentional phenomena (Reynolds & Heeger, 2009).

It is this critical role of inhibitory pooling where the thalamus can potentially play a unique role in the larger network of brain circuits involved in attention. This is because, with the strong and topographically organized inhibitory projection of the TRN to all of the excitatory thalamic nuclei, the thalamus is in a unique position to exert a net inhibitory pooling effect across a wider range of downstream cortical area than would be possible through direct corticocortical connections within cortex itself. Thus, whereas direct corticocortical pathways can support pooling and normalization within localized cortical areas, the thalamus may be particularly important for mediating competition between different brain areas. This idea is consistent with the data from Saalmann et al. (2012) showing a central role for the pulvinar in coordinating activity across brain areas, and in particular a cross-modal attention role for the TRN (Wimmer et al., 2015), which we revisit after first introducing a more detailed anatomical picture for how the thalamus fits in with the overall biased competition and normalization dynamic.

Although the Reynolds and Heeger (2009) model is abstract and not directly tied to specific neural mechanisms, we found that it provided a useful framework for understanding the different contributions of the cortical lamina and the thalamus, in a manner that is consistent overall with the *folded feedback* model of Grossberg (1999) (Figure 1b). This model is anchored by the finding that layer 6 corticothalamic (6CT) neurons exert a multiplicative or gain-field effect on neural activations in the superficial network (Bortone, Olsen, & Scanziani, 2014; Dougherty, Cox, Ninomiya, Leopold, & Maier, 2017; Olsen, Bortone, Adesnik, & Scanziani, 2012; van Kerkoerle et al., 2014), which corresponds to the net modulatory effect of the *suppressive drive* signal from the normalization model. Thus, where activations are strong in the suppressive drive layer, the corresponding superficial layer activations will remain strong, but where they are weaker, the superficial layer activations will be reduced, as if multiplied by a gain factor less than 1.

To walk through the model shown in Figure 1 in detail: bottom-up sensory inputs project to layer 4, which may provide a more faithful representation of their bottom-up strength differences (which are equal in the example shown). Initially, these equal-strength bottom-up signals activate the superficial layers equally, but the top-down attentional signal (e.g., from LIP in this case) causes the attended stimulus to become more strongly activated, especially as it converges on the layer 5 neurons (with some top-down modulation of superficial layers likely as well). The broad lateral connectivity among these deep layer 5 and 6 cortico-cortical (CC) neurons, serves to integrate across discrete sensory features to create a more broadly-tuned spatial activity pattern reflecting both bottom-up and top-down signals. This is then reflected in the 6CT (corticothalamic) projections to the thalamic relay cells (TRCs) and the TRN. The TRN connectivity causes significant pooling across many of these descending projections, such that the TRN projects back a broadly pooled inhibitory signal to the TRCs.

Critically, the top-down attentional bias effectively increases the strength of this overall pooled inhibitory signal, which in turn more strongly inhibits the *unattended* TRCs that are not also receiving the direct excitatory benefits from this top-down signal. Thus, the net effect of this circuit is that TRCs in the attended region remain robustly active, but those in the unattended region are relatively inhibited, and these differential activity levels then feed back up into the cortex, where the attended region thus has stronger thalamic drive. The TRCs then provide the equivalent of the multiplicative attentional filter in the Reynolds and Heeger (2009) model. In addition, there are direct intracortical, interlaminar circuits (Bortone et al., 2014; Frandolig et al., 2019) driven by 6CT collaterals that provide focused, local center-surround excitation and inhibition throughout the cortical column, which serve to further augment the multiplicative-like effect of
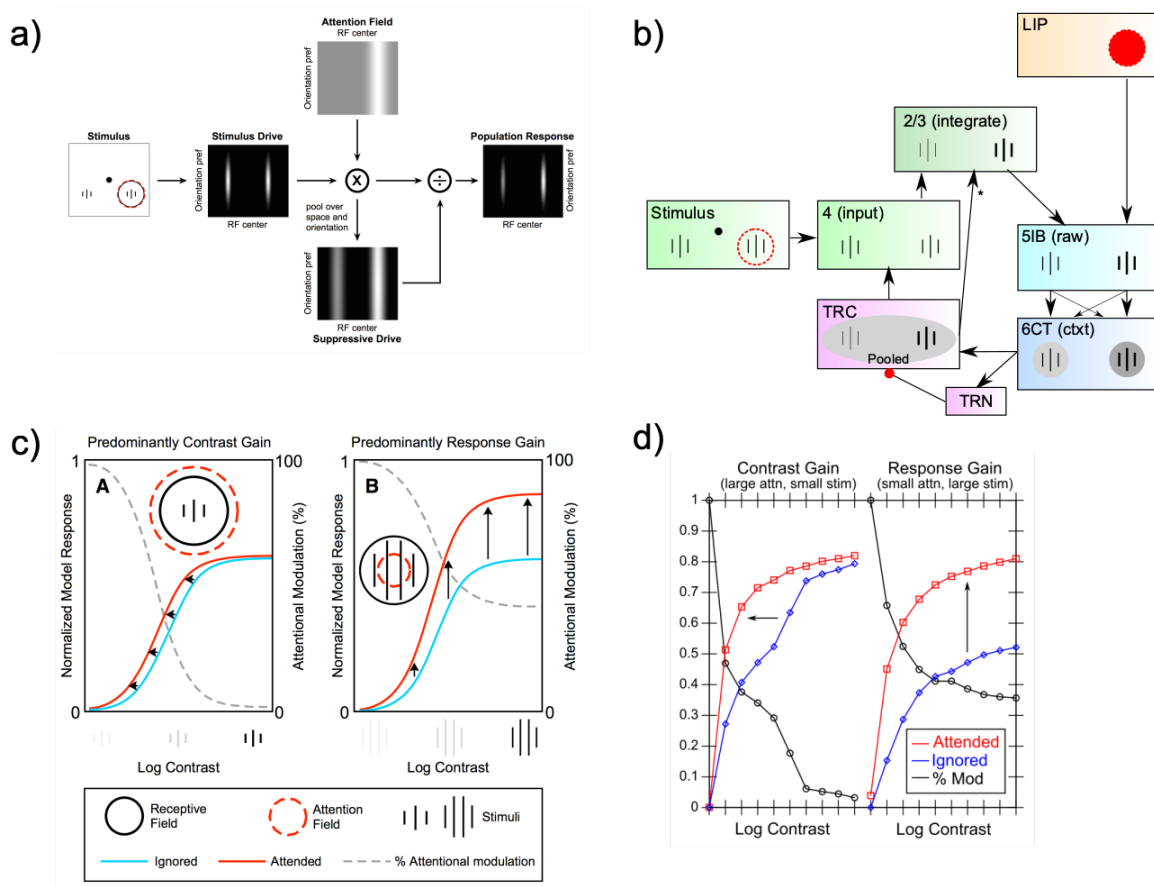
Figure 1: a) The Reynolds & Heeger (2009) mathematical model of pooling and normalization processes in attention. b) How attentional modulation is computed across the deep layers in our biological model, in response to a top-down attentional focus (as encoded in LIP of parietal cortex). Layer 4 receives bottom-up sensory input (initially equally weighted), which then drives superficial layers (2/3), which initially do not reflect the attentional modulation (not shown). The deep 5IB neurons integrate deep-to-deep top-down attentional inputs from LIP plus the local stimulus features from 2/3, to produce the "raw" deep output, prior to the contextual normalization process. The 6CC neurons integrate across the 5IB activations (context integration or pooling), within the limits of corticocortical connectivity, resulting in a 6CT signal that drives TRC and TRN, with the TRN pooling across a wider area, resulting in a more fully normalized activity over the TRC, decreasing the unattended location activity. The TRC projections have a net multiplicative effect on other cortical lamina. c) Key data accounted for by Reynolds & Heeger (2009) model: two qualitatively different types of attentional modulation resulting from differences in size of attentional spotlight relative to stimulus size, which directly reflect pooling and normalization processes. d) Results from our thalamically-equipped model driven by large LIP attentional top-down spotlight relative to a small bottom-up stimulus (left) versus a small LIP spotlight relative to a larger stimulus, reproducing same qualitative effects (from O'Reilly et al, 2017).

the TRC feedback into the cortex.

Figure 1d shows that our model of these laminar circuits through the deep layers and thalamus captures the same key data as the Reynolds and Heeger (2009) model, where the relative balance of the enhancing vs. suppressive effects of attentional modulation can shift depending on the relative sizes of the attentional spotlight and the stimulus input (and as a function of stimulus contrast), producing the shift from *contrast gain* to *response gain* effects of attention. Thus, although there is much more work to be done here to explore the full range of attentional dynamics, this provides a solid foundation building on the well-established Reynolds and Heeger (2009) model. Furthermore, our model is related to the *folded-feedback* model of Grossberg (1999) (see Raizada & Grossberg, 2003 for a more elaborated version), which also posits this

same kind of attentional modulation dynamic between layer 6 and the superficial layers. Interestingly, top-down attentional signals, like those coming from LIP down to lower-level visual pathways, are preferentially communicated via a network of deep-to-deep projections (Markov et al., 2014; van Kerkoerle et al., 2014; von Stein, Chiang, & König, 2000).

A critical question for the effects of attention in the abstract normalization model (Reynolds & Heeger, 2009) is the spatial scope over which these normalization effects can occur. Anatomically, it is likely that local inhibitory interneurons within the neocortex could support some local amount of pooled inhibitory competition, but to cover the full visual field in an area the size of V1 in a primate, the TRN contribution to pooled inhibition is likely to be critical. Thus, more selective tests of thalamic contributions to attention could focus on measuring across a wide range of spatial scales, with the prediction that the broader scales are differentially affected.

Predictive Learning in the Pulvinar

The central question of how learning is able to shape the networks of neurons in the neocortex to support our incredibly powerful and adaptive human cognitive abilities remains one of the most important unresolved scientific challenges. This question has been addressed from many different angles at many different levels of analysis, and the recent advances in deep neural network (DNN) machine-learning algorithms (LeCun et al., 2015) based on error-backpropagation (Rumelhart, Hinton, & Williams, 1986) have reignited interest into how this form of learning might operate in the neocortex (Lillicrap et al., 2020; Whittington & Bogacz, 2019). Our models are based on an early proposal for how bidirectional connectivity between cortical areas could communicate error signals in a way that mathematically approximates error backpropagation (O'Reilly, 1996; O'Reilly & Munakata, 2000). The main principles of this form of error-driven learning are as follows:

- Error signals take the form of *changes over time* in the firing patterns of neurons, in particular parts of the network that first encode an expectation or guess, and then experience the correct answer relative to that prior expectation. This is known as a *temporal difference* form of error signal, as contrasted with an alternative possibility of an explicitly-coded error signal where the firing of individual neurons directly reflects the error. In simple artificial networks (and current-day DNNs), there are "output layers" that represent things such as the category label for an object presented to the visual "input layer" of the network. The network's expectation is the output layer activity in response to the visual input, which is then followed by the correct answer being presented, producing the temporal-difference sequence of activations. Current DNNs use huge datasets of hand-labeled images to provide such correct answers to drive output-layer targets.

- The ubiquitous bidirectional connectivity in the neocortex allows individual neurons everywhere to be influenced by these changes over time in the firing of output-layer neurons, and the math shows that, with certain assumptions such as symmetrical connectivity (which can be significantly relaxed in practice), the local changes in activity state in an individual neuron accurately reflect the error-backpropagation derivative that is needed to drive learning to minimize the error on the output layer of the network.

This framework for neocortical error-driven learning avoids some of the most glaring forms of biological implausibility associated with the error-backpropagation algorithm (Crick, 1989), and a large number of models developed over several years demonstrate that a wide range of cognitive functions can be accounted for within this general framework (O'Reilly, Hazy, & Herd, 2016; O'Reilly & Munakata, 2000; O'Reilly, Munakata, Frank, Hazy, & Contributors, 2012). However, there remains a critical unresolved question: what
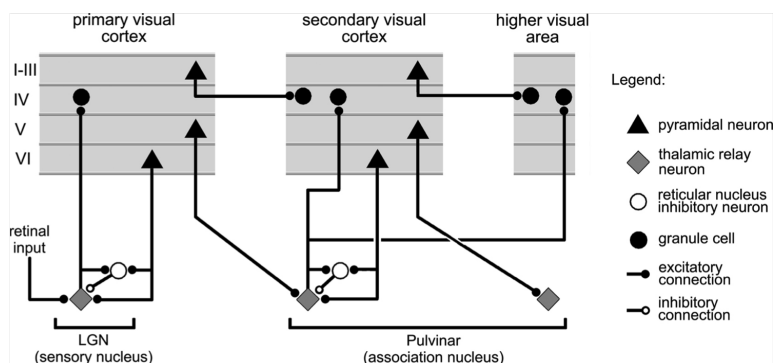
Figure 2: Summary figure from Sherman & Guillery (2006) showing the strong feedforward *driver* projection emanating from layer 5IB cells in lower layers (e.g., V1), and the much more numerous feedback "modulatory" projection from layer 6CT (corti-cothalamic) cells. We interpret these same connections as providing a prediction (6CT) vs. outcome (5IB) activity pattern over the pulvinar.

are the primary ecologically-plausible sources of the error signals needed to drive error backpropagation learning? Unlike the artificial models, developing humans and other animals are not provided with massive labeled datasets to drive their learning. Somehow, we must learn in a more "self-organizing", naturalistic fashion.

One widely-discussed solution to this problem is that we learn by predicting what will happen next, which goes back to Helmholtz in his *recognition by synthesis* proposal (von Helmholtz, 1867), and has been widely embraced in a range of different frameworks (Clark, 2013; Dayan, Hinton, Neal, & Zemel, 1995; J. Elman et al., 1996; J. L. Elman, 1990; Friston, 2005; Kawato, Hayakawa, & Inui, 1993; Mumford, 1992; Rao & Ballard, 1999). In this *predictive learning* framework, the raw sensory input itself can generate the error signal as a "ground truth" relative to a prior prediction, avoiding the need for any form of human labeling, and providing an essentially unlimited and automatic source of error-signals to drive learning. This abstract computational idea turns out to be synergistic with the circuitry of higher-order thalamic areas (primarily the pulvinar and MD), providing another important example where the thalamus comes to the rescue in our computational understanding of overall cortical functioning (O'Reilly et al., 2021).

Specifically, Figure 2 shows the critical elements of this thalamocortical circuitry as emphasized by Sherman and Guillery (2006), which features two separate sources of input into the higher-order thalamic relay cells: the strong, focal, bottom-up *driver* inputs, and the much more numerous but weaker "modulatory" top-down inputs. These two inputs provide a natural organization for predictive learning, where the top-down inputs are responsible for generating the prediction or expectation, and the bottom-up drivers impose the "ground truth" outcome that actually occurred relative to this prediction.

The sources of these driver inputs are the layer 5 intrinsic bursting (5IB) neurons, which fire discrete bursts with intrinsic dynamics having a period of roughly 100 ms between bursts (Connors, Gutnick, & Prince, 1982; Franceschetti et al., 1995; Larkum, Zhu, & Sakmann, 1999; Saalmann et al., 2012; Silva, Amitai, & Connors, 1991). These are thought to drive the widely-studied *alpha* frequency of ~ 10 Hz that originates in cortical deep layers and has important effects on a wide range of perceptual and attentional tasks (Buffalo, Fries, Landman, Buschman, & Desimone, 2011; Clayton, Yeung, & Kadosh, 2018; Jensen, Bonnefond, & VanRullen, 2012; K. Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; VanRullen & Koch, 2003). Critically, unlike many other such bursting phenomena, this 5IB bursting occurs in awake animals (Luczak, Bartho, & Harris, 2009, 2013; Sakata & Harris, 2009, 2012), consistent with the presence of alpha in awake behaving states.

From a computational perspective, the burst firing of these driver inputs provides the necessary timing
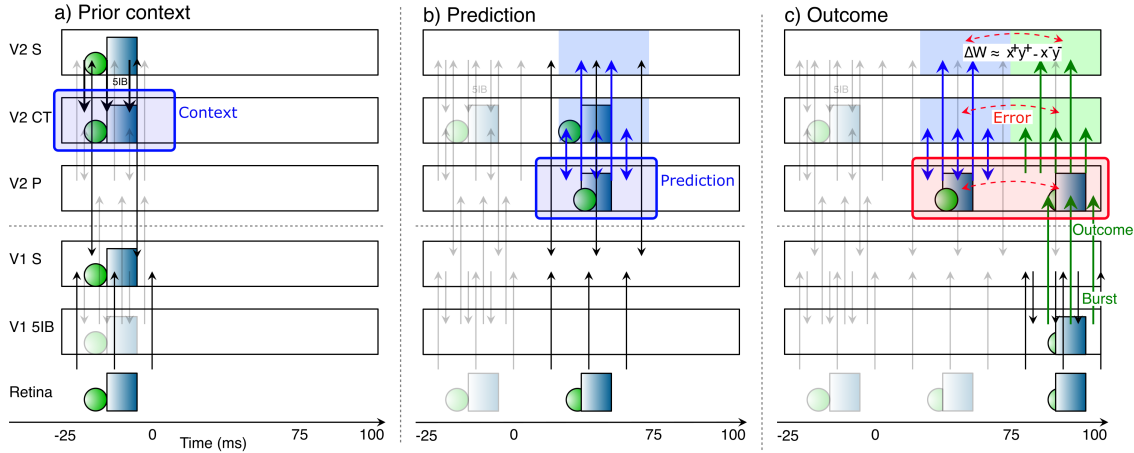
Figure 3: Corticothalamic information flow under our predictive learning hypothesis, shown as a sequence of movie frames (Retina), illustrating the three key steps taking place within a single 125 ms time window, broken out separately across the three panels: **a)** prior context is updated in the V2 CT layer; **b)** which is then used to generate a prediction over the pulvinar (V2 P); **c)** against which the outcome, driven by bottom-up 5IB bursting, represents the prediction error *as a temporal difference between the prediction and outcome states over the pulvinar*. Changes in synaptic weights (learning) in all superficial (S) and CT layers are driven from the *local* temporal difference experienced by each neuron, using a form of the contrastive hebbian learning (CHL) term as shown, where the '+' superscripts indicate outcome activations, and '−' superscripts indicate prediction. CHL approximates the backpropagated prediction error gradient experienced by each neuron (O'Reilly, 1996), reflecting both direct pulvinar error signals, and indirect corticocortical error signals as well. In specific: **a)** CT context updating occurs via 5IB bursting (not shown) in higher layer (V2) during prior alpha (100 ms) cycle — this context is maintained in the CT layer and used to generate predictions. **b)** The prediction over pulvinar is generated via numerous top-down CT projections. This prediction state also projects up to S and CT layers, and from S to all other S layers via extensive bidirectional connectivity, so their activation state reflects this prediction as well. **c)** The subsequent outcome drives pulvinar activity bottom-up via V1 5IB bursting, and is likewise projected to S and CT layers, ensuring that the relevant temporal difference error signal is available locally in cortex. The difference in activation values across these two time points, in S and CT layers throughout the network, drives learning to reduce prediction errors. Note that the single most important property of the 5IB bursting is that these driver cells are *not* active during the prediction phase — the bursting itself may also be useful in the driving property, but that is a secondary consideration to the critical feature of having a time when the prediction alone can be projected onto the pulvinar.

for a temporal-difference form of error signal as in our earlier models. Specifically, when these 5IB neurons are not active in between their burst phases, the top-down inputs to the TRC neurons can generate a prediction, which is then followed immediately in time by the bottom-up driver outcome signal. The extensive projections from the thalamus back into cortex Shipp (2003) can convey this temporal-difference error signal into the neocortical neurons, driving local synaptic changes that will end up minimizing the prediction error (O'Reilly et al., 2021).

Figure 3 illustrates the temporal evolution of activity states according to this predictive learning theory, over a single 125 ms time window of a 100 ms alpha cycle (the actual timing is likely to be more dynamic). The activity state in pulvinar TRC neurons, representing a prediction, as driven by the top-down 6CT projections, should develop during the first ~ 75 ms, when the 5IB neurons are paused between bursting. Then the final ~ 25 ms largely reflects the strong 5IB bottom-up ground-truth driver inputs when they burst. Thus, the prediction error signal is reflected in the temporal difference of these activation states as they develop over time. In other words, our hypothesis is that the pulvinar is directly representing either the top-down prediction or the bottom-up outcome at any given time, and the temporal difference between these states implicitly encodes a prediction error. While the deep 6CT layer is involved in generating a top-down prediction over the pulvinar, the superficial layer neurons continuously represent the current state, simultaneously incorporating bottom-up and top-down constraints via their own connections with other areas. To ensure that the prediction is not directly influenced by this current state representation (i.e., "peeking at the right answer"),

it is important that the 6CT neurons encode temporally delayed information, consistent with available data (Harris & Shepherd, 2015; Sakata & Harris, 2009; Thomson, 2010).

In addition to the extensive primary sources of anatomical and electrophysiological evidence cited above, there is a wide range of other neural data consistent with this theory of higher-order thalamic contributions to predictive learning, reviewed in O'Reilly et al. (2021). For example, there is an extensive literature on alpha-frequency entrainment emerging from deep neocortical layers and the pulvinar, and how these organize cortical processing over time in a manner consistent with this model (Fiebelkorn et al., 2018; Klimesch, 2011; Makeig et al., 2002; K. E. Mathewson et al., 2012; Saalmann et al., 2012; Spaak, de Lange, & Jensen, 2014). Furthermore, this framework makes a number of testable predictions, outlined in O'Reilly et al. (2021), that could be conducted using available techniques, to more directly test this theory. Two such tests are now under way, so hopefully we will have further relevant data soon.

We have shown how this predictive error-driven learning mechanism, based directly on the biology of the corticothalamic loops, can drive the learning of abstract object-category representations in higher layers of a simulated visual cortex, based solely on short movies of 3D objects moving and rotating through space (O'Reilly et al., 2021). This model did not have any input telling it how to categorize or label the objects — the learning was purely "self-organizing" based on the raw visual inputs from the movies. Thus, this model provides an initial proof-of-concept that predictive error-driven learning can drive the formation of more abstract "conceptual" representations, avoiding the implausible use of hand-labeled training data. Furthermore, comparison models based on current machine-learning algorithms did not develop these abstract, conceptual representations, indicating the importance of some of the other biologically-motivated properties of this model. These properties include the bidirectional connections needed to drive the biological form of error-driven learning in the first place, which also have the effect of enabling higher-level representations to influence lower levels in the network through top-down connectivity. Furthermore, extensive inhibitory connectivity is required to balance and control the bidirectional excitatory connections, and the resulting competitive dynamics can also shape representations in useful ways to promote categorical learning (Kohonen, 1989; Rumelhart & Zipser, 1985).

This thalamocortically-mediated temporal-difference mechanism for predictive error-driven learning contrasts with widely-discussed explicit error-coding ideas for predictive learning (Bastos et al., 2012; Friston, 2005, 2010; Kawato et al., 1993; Lotter, Kreiman, & Cox, 2016; Ouden, Kok, & Lange, 2012; Rao & Ballard, 1999), where a dedicated population of neurons (typically suggested to be superficial-lamina cortical neurons) explicitly codes for the error signal in their direct firing rates, via an inhibitory-mediated subtraction-like process between the bottom-up outcome signal and a top-down prediction. Despite many attempts to identify such explicit error-coding neurons in the cortex, no substantial body of unambiguous evidence has been discovered (Kok & de Lange, 2015; Kok, Jehee, & de Lange, 2012; Lee & Mumford, 2003; Summerfield & Egner, 2009; Walsh, McGovern, Clark, & O'Connell, 2020). Thus, the thalamic mechanism may provide a more biologically-supported framework.

Thalamic Gating of Frontal Function

To more completely understand learning in the brain, it is essential to understand the significant role that goals, motivation, and reward play in shaping the learning process (O'Reilly, 2020). The thalamus plays a central role in this domain as well, through its interactions with the frontal cortex. In our initial computational models of these frontal / thalamic circuits (Frank et al., 2001), we considered how the basal ganglia (BG) disinhibition of thalamus could lead to the *gating* of prefrontal cortex active maintenance of task and goal representations, in the same way that it was thought to affect the selection of overt actions in motor frontal cortex (Mink, 1996) (Figure 4). Anatomically, the same kinds of excitatory thalamocortical loops are present in frontal cortex as discussed above in posterior cortex, but unlike in posterior cortex, much
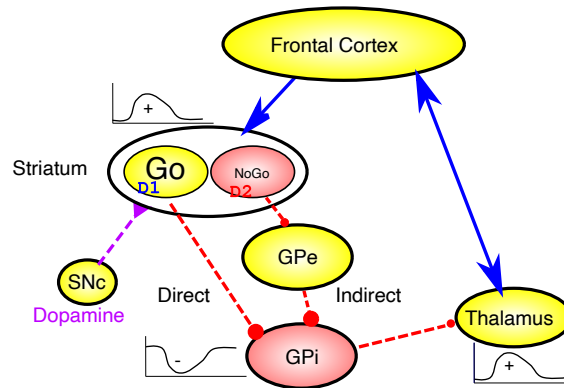
Figure 4: Classical basal ganglia (BG) circuit, producing a disinhibitory, *gating*-like effect on frontal activity to influence when new task / goal states are updated in frontal cortex, in the same way that the BG is widely thought to support action initiation in motor cortical areas. The thalamus provides the critical interface upon which this disinhibitory control can operate, leveraging the bidirectional excitatory thalamocortical loops.

of the frontally-projecting thalamus is under tonic inhibitory control by the output nuclei of the BG (the GPi and SNr). According to the classic understanding of the BG circuitry (Nambu, 2008), when the direct BG pathway fires, it disinhibits the thalamocortical loop, which then enables a new surge of cortical activation that could drive the updating of task and goal states.

Furthermore, the neuromodulatory effects of phasic dopamine on the BG direct and indirect pathways (Gerfen & Surmeier, 2011; Shen, Flajolet, Greengard, & Surmeier, 2008) are ideally situated to sculpt adaptive learning to reinforce such updates associated with good outcomes, and punish those associated with less good ones, providing a computationally effective overall account of frontal executive function (Frank, 2005; Hazy, Frank, & O'Reilly, 2006, 2007; O'Reilly & Frank, 2006). These models have shown that many complex working memory tasks can be learned from trial-and-error experience, including the 1-2-AX and phonological loop (O'Reilly & Frank, 2006), ID/ED dynamic categorization (O'Reilly, Noelle, Braver, & Cohen, 2002), WCST (Rougier & O'Reilly, 2002), N-back (Chatham et al., 2011), task switching, the Stroop task (Herd et al., 2014), hierarchical rule learning (Badre & Frank, 2012), and the reference-back-2 task (Rac-Lubashevsky & Frank, 2020).

Consistent with the theme of a progressive appreciation of the importance of the thalamus over time, our computational models have only recently begun to incorporate a more complete understanding of the anatomical and physiological data on thalamic interconnectivity with the frontal cortex, and the potentially significant functional implications thereof. For example, these data indicate that the BG gating dynamic shown in Figure 4 only applies to a subset of frontal thalamocortical circuits, predominantly those involving the ventral anterior and medial thalamic areas (VA, VM). There are significant other circuits, particularly involving the mediodorsal (MD) thalamus that appears to have only sparse or patchy disinhibitory BG inputs, and are similar in many ways to the pulvinar circuitry discussed above in the context of predictive learning (Giguere & Goldman-Rakic, 1988; Rovó, Ulbert, & Acsády, 2012).

Thus, consistent with computational models of pulvinar, the MD might function as a predictive learning and attentional control system for the frontal cortex (Rikhye, Gilra, & Halassa, 2018), while the ventral thalamic circuits support the gating dynamics captured in prior models. Furthermore, the differential patterns of termination in these pathways suggest that the MD may drive updating of the NMDA-dependent working memory circuits in frontal layer 3 in response to new sensory inputs (Arnsten, Wang, & Paspalas, 2012; Kritzer & Goldman-Rakic, 1995; Lisman, Fellous, & Wang, 1999; Sanders, Berends, Major, Goldman, & Lisman, 2013), while the BG to VA & VM gating pathway may be more responsible for the activation of layer 5 output neurons. Thus, at a computational level, we can think of the MD-based pathways as

supporting an attentionally-modulated *input / maintenance* gating function, while the BG may be more focused on gating the cortical *output* pathways (Chatham et al., 2014).

In the following section, we review a range of data relevant to these multiple fronto-thalamic circuits, which provide a more fine-grained and high-dimensional picture relative to a simple MD vs. ventral dichotomy: some areas and patches of MD are more similar to those in the ventral nuclei, and vice-versa. Furthermore, modern updates to the classical distinction between *core* vs. *matrix* thalamic pathways (Jones, 1998a, 1998b, 2007), provide relevant ways of understanding the functional differences between the input / maintenance vs. output gating types. Despite these additional sources of complexity, it does appear that across these areas, the BG disinhibitory control over the frontal cortex is conveyed predominantly through a distinct type of corticothalamic pathway, having broad patterns of connectivity terminating preferentially in cortical layer 1, consistent with the matrix-type of thalamic neurons. This would then imply that the BG output-gating control over the cortex is at a very coarse-grained level, such that it can influence the *timing* of action initiation, but it is unlikely to be useful for providing more fine-grained control over the detailed encoding of information in frontal cortex.

Thus, contrary to our original hypothesis that the BG could provide more fine-grained control over working-memory updating, it now seems that the attentionally-gated corticothalamic pathways predominantly in the MD are critical for that function. This is consistent with the frontal areas preferentially targeted by the MD projections as well, which include the primate-specific dorsolateral PFC cognitive control and working memory areas. We are currently developing new computational models to incorporate these new anatomically-motivated ways of understanding frontal / thalamocortical function.

Details of Frontothalamic Pathways

The distinction between core and matrix type projections is one of the early foundational ideas about thalamic organization (Jones, 1998a, 1998b, 2007). Core-type projections have more focal connectivity and target the central cortical lamina, including 3 and 4 most strongly. In contrast, matrix-type send broad, diffuse connections spanning roughly entire cortical areas, and preferentially target lamina 1, where the apical tufts of pyramidal cells from lamina 2, 3, and 5b reside, the thick tufts of subcortically-projecting layer 5b being particularly prominent (Harris & Shepherd, 2015; Larkum, Petro, Sachdev, & Muckli, 2018; Ramaswamy & Markram, 2015). More recent anatomical data including gene expression analyses support a more complex, high-dimensional set of thalamic cell and projection types (Clascá, Rubio-Garrido, & Jabaudon, 2012; Phillips et al., 2019), which include elements of the original core vs. matrix distinction. At the broadest level, three different categories of thalamic cells can be identified: primary sensory / motor core-type (e.g., LGN for vision and VL primary motor cortex), higher-order types (including pulvinar, MD, VA, and VM), and the highly diffuse matrix-type associated with the intralaminar and midline nuclei (Phillips et al., 2019). Within each such category, there is significant variability.

Within the higher-order types of most relevance for our computational models, the pulvinar and MD predominantly have cells labeled as a *focal* matrix-type, which target layer 3 and also 1 and receive cortical driver inputs, while VA and VM predominantly have broader *multiareal* matrix-type cells that receive BG disinhibitory inputs and project most strongly to layer 1 and also to 3 (Clascá et al., 2012). In the context of the classical core vs. matrix type dichotomy, the focal matrix-type cells are more core-like in their focal connectivity and strong termination in layer 3, while the multiareal matrix-type are more clearly matrix-like in their broad connectivity and strong termination in layer 1. We adopt the terminology of *focal-type* vs. *broad-type* in what follows. In MD, the focal nature of the thalamic projections has been characterized as being organized at around the 1 mm scale (Giguere & Goldman-Rakic, 1988), which would provide relatively fine-grained influence of these thalamic cells on corresponding frontal activity states.

Figure 5 shows a schematic diagram integrating a number of different findings across the literature, for how the focal vs. broad distinction aligns with the different frontal thalamocortical and BG circuits, in
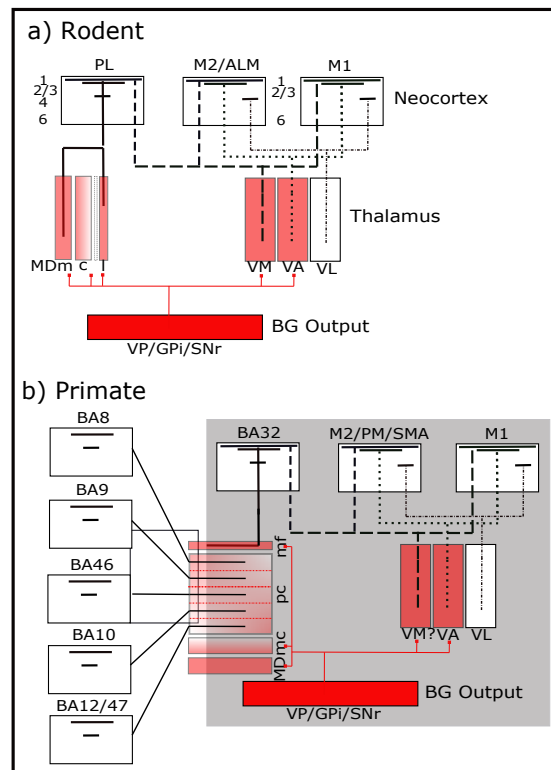
Figure 5: Focal vs. broad thalamocortical connectivity and basal ganglia (BG) modulation in rodent vs. primate prefrontal cortex. **a)** Rodent prelimbic cortex (PL) is interconnected with both the mediodorsal (MD) and ventral motor thalamus (VM & VA). MD projects more focally to middle cortical laminae, most prominently from its lateral segment (l), while VM & VA project more broadly and more to layer 1. VA & VM also project broadly to secondary (M2/ALM) and primary (M1) motor cortices. In contrast, VL projects focally to the middle cortical laminae of motor cortices, and uniquely receives cerebellar inputs (not shown), with only minimal BG input, while VA & VM receive the densest BG input. BG projections to MD are of intermediate density and exhibit a patchy connectivity pattern. The central part of MD (c) projects to orbital frontal cortex (not shown). **b)** Primate PFC connectivity essentially replicates the rodent pattern as a subset (gray box) although some homologous areas typically have different labels, e.g., primate BA32 is homologous to the PL in rodents, and PM / SMA are secondary motor areas. Primates have several additional PFC areas (left side of diagram), all of them interconnected with segments of an essentially new part of a greatly enlarged MD nucleus, the parvocellular middle segment (MDpc). BG output nuclei are VP/GPi/SNr (ventral pallidum / globus pallidus, internal segment / substantia nigra, pars reticulata); BA are Brodmann areas; VM? = ventromedial thalamic nucleus, not prominent in primates and sometimes considered part of the midline nuclei; MDmc/pc/mf are magnocellular / parvocellular / multiforme segments of MD.

both rodent and primate. Critically, two major ventral thalamic areas (VM, VA) predominantly send broad-type projections (Kuramoto et al., 2009, 2015), and also receive dense BG projections (Ilinsky, Jouandet, & Goldman-Rakic, 1985; Kuramoto et al., 2009, 2015; Tanibuchi, Kitano, & Jinnai, 2009a). In contrast, the MD receives more patchy BG projections, and it also has patchy distributions of broad-type cells (Münkle, Waldvogel, & Faull, 2000; Phillips et al., 2019), which are sufficiently aligned as to be consistent with the overall idea that the BG disinhibitory control over the thalamus is primarily conveyed through broad-type thalamocortical pathways.

A complementary distribution of cell types is present with respect to the focal-type neurons, with the MD having a much higher concentration of these, with more patchy distributions of focal-type cells in VM and VA (Kuramoto et al., 2009). Interestingly, as shown in Figure 5, the MD focal-type connectivity predominates in the prefrontal cortical (PFC) areas that are present in primates but not rodents. These areas include the same areas that have specialized layer 3 NMDA-based working memory circuits not found in rodents (Arnsten et al., 2012; Kritzer & Goldman-Rakic, 1995; Lisman et al., 1999; Sanders et al., 2013),

which is consistent with the focal-type projections terminating strongly in layer 3. The more focal, clumpy, topographically-organized nature of these projections (Giguere & Goldman-Rakic, 1988) would enable them to have more fine-grained control over updating and maintenance of working memory information in these PFC areas, which was the original computational motivation for working-memory gating in our models (O'Reilly & Frank, 2006), based on the earlier and widely-used LSTM (long short-term memory) machine-learning algorithm (Gers, Schmidhuber, & Cummins, 2000; Hochreiter & Schmidhuber, 1997). These ideas are also consistent with direct evidence linking MD and active maintenance in frontal cortex (Rikhye et al., 2018; Tanibuchi, Kitano, & Jinnai, 2009b; Watanabe & Funahashi, 2012; Watanabe, Takeda, & Funahashi, 2009; Wyder, Massoglia, & Stanford, 2004).

Although the widespread effects of broad-type output gating initially may not sound very useful, it actually could have important benefits. Specifically, any given frontal area can only activate a single motor output at any given point in time (e.g., you can saccade and move your limbs at the same time, but you cannot execute two different saccades at the same time). Thus, perhaps broad-type output gating enforces this singular constraint, where only one output per broader area can be gated at a time. In this case, it would make sense for there to be a *sequential* consideration of different possible output actions, and the broad-type gating fires whenever one such action is determined to be over threshold, via the Go vs. NoGo pathway competition within the BG (Herd, Krueger, Nair, Mollick, & O'Reilly, 2019; O'Reilly, Nair, Russin, & Herd, 2020). Furthermore, the broad impact of the gating signal may also serve to reset the circuit to facilitate moving on to the next step in a sequence of actions: neural recordings of delayed action tasks inevitably show a rapid reset of maintenance activation coincident with motor actions.

The broad vs. focal distinction applies to the output projections from the thalamus, but the driver inputs to these neurons are equally important in determining their functional contributions, as we saw in the case of predictive learning above. A recent technique for mapping the sources and types of driver inputs across the primate thalamus (Rovó et al., 2012) showed that MD predominantly has cortical-type driver inputs similar to those found in the pulvinar, which is consistent with the idea that it provides similar attentional and predictive learning functions to the frontal cortex. In contrast, VM and VA have *no* strong drivers, and instead receive GABAergic inhibitory inputs from the BG output nuclei (Kuramoto et al., 2009, 2015). VL, which interconnects with primary and secondary motor areas, receives subcortical drivers primarily from the deep cerebellar nuclei. This VL pathway would enable the cerebellum to train the motor cortex to produce better motor commands, which is consistent with a substantial literature showing that the dysmetria associated with cerebellar damage is significant, it is relatively less impactful after motor development has stabilized, but remains crucial for adapting to novel trajectories (Chen, Hua, Smith, Lenz, & Shadmehr, 2006; Manto, 2009).

Finally, it is important to note that even if we adopt this clear distinction between BG-mediated output gating and cortically / attentionally-mediated maintenance gating, there remains considerable ambiguity about the potential timecourse of these respective gating signals. For example, it is possible that BG-gated disinhibition of the 5b neurons could drive a *preparatory* form of pre-output activity, in advance of actually executing an action. For example, there is evidence in the mouse for a BG-mediated preparatory form of gating, through the VM thalamic nucleus (K. Guo, Yamawaki, Svoboda, & Shepherd, 2018; Z. V. Guo et al., 2017). Likewise, there is extensive evidence in people of a *proactive* engagement of frontal control and action representations, which may well be associated with BG preparatory gating of output-pathway neurons (Braver, Paxton, Locke, & Barch, 2009). Furthermore, the purely maintenance form of gating may largely be restricted to primates (Elston, 2003; Wang et al., 2013), where it may provide a more "offline" form of active maintenance of information that may be useful for subsequent actions, but falls short of actually driving output-level action preparation. In short, there are numerous important functional questions about the nature of gating that remain to be explored, in conjunction with critical experimental tests that can directly inform these models.

Conclusions

In summary, across all of these potential contributions of the thalamus, there is a core *modulatory* role, where thalamic activity modulates corresponding neocortical activity, whether that is characterized as being an attentional or gating modulation. Furthermore, the thalamus may play a *training* role, conveying prediction error signals that shape neural representations in the cortex to develop a systematic predictive model of the world. Across the thalamus, the source of these training signals arises from the distinctive *driver* inputs on thalamic neurons — the map of such drivers across the thalamus (Rovó et al., 2012) thus provides a window into the nature of training signals for different cortical areas.

Despite its truly "core" (and matrix) connectivity, one could argue that the thalamus itself doesn't really *do* anything — it just happens to be in the right place to connect other circuits in an effective way. However, it is not clear that any part of the brain can avoid this argument: neurons mainly just pass the buck around to other neurons, making their own modest contributions along the way. What thalamic neurons may lack in such processing contributions, they more than make up for by their unique connectivity that results in such critical functional dynamics, including the attention, predictive learning, and executive function mechanisms discussed here.

*

References

Arnsten, A. F. T., Wang, M. J., & Paspalas, C. D. (2012, October). Neuromodulation of thought: Flexibilities and vulnerabilities in prefrontal cortical network synapses. *Neuron*, *76*(1), 223–239. doi: 10.1016/j.neuron.2012.08.038

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.

Badre, D., & Frank, M. J. (2012, March). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21693491`

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012, November). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23177956`

Bender, D. B., & Youakim, M. (2001, January). Effect of attentive fixation in macaque thalamus and cortex. *Journal of neurophysiology*, *85*, 219–234. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11152722`

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1–21. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/20192813`

Bortone, D. S., Olsen, S. R., & Scanziani, M. (2014, April). Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron*, *82*. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/24656931`

Braver, T. S., & Cohen, J. D. (2000, December). On the Control of Control: The Role of Dopamine in Regulating Prefrontal Function and Working Memory. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII* (pp. 713–737). Cambridge, MA: MIT Press.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009, May). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences USA*, *106*(18), 7351–7356.

Brown, R. G., & Marsden, C. D. (1990, February). Cognitive function in Parkinson's disease: From description to theory. *Trends in Neurosciences*, *13*, 21–29. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/1688671`

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011, July). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11262–11267. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21690410`

Buschman, T. J., & Kastner, S. (2015, October). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. doi: 10.1016/j.neuron.2015.09.017

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010, April). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153. doi: 10.1016/j.tics.2010.01.007

Chatham, C. H., Frank, M., & Badre, D. (2014, January). Corticostriatal output gating during selection from working memory. *Neuron*, *81*(4), 930–942.

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R. C., & Friedman, N. P. (2011, November). From an executive network to executive control: A computational model of the n-back task. *Journal of Cognitive Neuroscience*, *23*, 3598–3619. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21563882`

Chen, H., Hua, S. E., Smith, M. A., Lenz, F. A., & Shadmehr, R. (2006, October). Effects of human cerebellar thalamus disruption on adaptive control of reaching. *Cerebral Cortex*, *16*(10), 1462–1473. doi: 10.1093/cercor/bhj087

Clark, A. (2013, June). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23663408`

Clascá, F., Rubio-Garrido, P., & Jabaudon, D. (2012). Unveiling the diversity of thalamocortical neuron subtypes. *European Journal of Neuroscience*, *35*(10), 1524–1532. doi: 10.1111/j.1460-9568.2012.08033.x

Clayton, M. S., Yeung, N., & Kadosh, R. C. (2018). The many characters of visual alpha oscillations. *European Journal of Neuroscience*, *48*(7), 2498–2508. doi: 10.1111/ejn.13747

Connors, B. W., Gutnick, M. J., & Prince, D. A. (1982, December). Electrophysiological properties of neocortical neurons in vitro. *Journal of Neurophysiology*, *48*(6), 1302–1320. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/6296328`

Crick, F. (1984, July). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *81*, 4586–4590. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/6589612`

Crick, F. (1989, February). The recent excitement about neural networks. *Nature*, *337*, 129–132. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/2911347`

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008, June). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510–1512. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/18556560`

Dayan, P., Hinton, G. E., Neal, R. N., & Zemel, R. S. (1995, January). The Helmholtz machine. *Neural Computation*, *7*(5), 889-904.

Dehaene, S., & Naccache, L. (2001, February). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1-2), 1–37. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11164022`

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. doi: 10.1146/annurev.ne.18.030195.001205

Dougherty, K., Cox, M. A., Ninomiya, T., Leopold, D. A., & Maier, A. (2017, February). Ongoing alpha activity in v1 regulates visually driven spiking responses. *Cerebral Cortex*, *27*(2), 1113–1124. doi:

10.1093/cercor/bhv304

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992, April). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90–92. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/1553535`

Edelman, G. M., & Tononi, G. (2001). *A Universe Of Consciousness: How Matter Becomes Imagination*. New York, NY: Basic Books.

Elman, J., Bates, E., Karmiloff-Smith, A., Johnson, M., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Elman, J. L. (1990, January). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Elston, G. N. (2003). Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cerebral Cortex*, *13*(11), 1124–1138. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/14576205`

Farah, M. J. (1990). *Visual Agnosia*. Cambridge, MA: MIT Press.

Fiebelkorn, I. C., & Kastner, S. (2019, February). A rhythmic theory of attention. *Trends in Cognitive Sciences*, *23*(2), 87–101. doi: 10.1016/j.tics.2018.11.009

Fiebelkorn, I. C., Pinsk, M. A., & Kastner, S. (2018, August). A dynamic interplay within the frontoparietal network underlies rhythmic spatial attention. *Neuron*, *99*(4), 842-853.e8. doi: 10.1016/j.neuron.2018.07.038

Franceschetti, S., Guatteo, E., Panzica, F., Sancini, G., Wanke, E., & Avanzini, G. (1995, October). Ionic mechanisms underlying burst firing in pyramidal neurons: Intracellular study in rat sensorimotor cortex. *Brain Research*, *696*(1–2), 127–139. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/8574660`

Frandolig, J. E., Matney, C. J., Lee, K., Kim, J., Chevée, M., Kim, S.-J., ... Brown, S. P. (2019, September). The synaptic organization of layer 6 circuits reveals inhibition as a major output of a neocortical sublamina. *Cell Reports*, *28*(12), 3131-3143.e5. doi: 10.1016/j.celrep.2019.08.048

Frank, M. J. (2005, January). When and when not to use your subthalamic nucleus: Lessons from a computational model of the basal ganglia. In A. K. Seth, T. J. Prescott, & J. J. Bryson (Eds.), *Modelling Natural Action Selection: Proceedings of an International Workshop* (pp. 53–60). Sussex: AISB.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001, January). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, *1*, 137–160. Retrieved from `https://www.ncbi.nlm.nih.gov/pubmed/12467110`

Friston, K. (2005, April). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, *360*(1456), 815–836. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/15937014`

Friston, K. (2010, February). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/20068583`

Gerfen, C. R., & Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine. *Annual Review of Neuroscience*, *34*, 441–466. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21469956`

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000, November). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*, 2451–2471. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11032042`

Giguere, M., & Goldman-Rakic, P. S. (1988). Mediodorsal nucleus: Areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. *Journal of Comparative Neurology*, *277*(2), 195–213. doi: 10.1002/cne.902770204

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia.

Current Opinion in Neurobiology, *5*(6), 733–741. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/8805417`

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial vision*, *12*. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/10221426`

Guo, K., Yamawaki, N., Svoboda, K., & Shepherd, G. M. G. (2018, October). Anterolateral motor cortex connects with a medial subdivision of ventromedial thalamus through cell type-specific circuits, forming an excitatory thalamo-cortico-thalamic loop via layer 1 apical tuft dendrites of layer 5b pyramidal tract type neurons. *Journal of Neuroscience*, *38*(41), 8787–8797. doi: 10.1523/JNEUROSCI.1333-18.2018

Guo, Z. V., Inagaki, H. K., Daie, K., Druckmann, S., Gerfen, C. R., & Svoboda, K. (2017, May). Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, *545*(7653), 181–186. doi: 10.1038/nature22324

Halassa, M. M., & Kastner, S. (2017, December). Thalamic functions in distributed cognitive control. *Nature Neuroscience*, *20*(12), 1669. doi: 10.1038/s41593-017-0020-1

Halassa, M. M., Siegle, J. H., Ritt, J. T., Ting, J. T., Feng, G., & Moore, C. I. (2011, September). Selective optical drive of thalamic reticular nucleus generates thalamic bursts and cortical spindles. *Nature Neuroscience*, *14*(9), 1118–1120. doi: 10.1038/nn.2880

Harris, K. D., & Shepherd, G. M. G. (2015, February). The neocortical circuit: Themes and variations. *Nature Neuroscience*, *18*(2), 170–181. doi: 10.1038/nn.3917

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006, April). Banishing the homunculus: Making working memory work. *Neuroscience*, *139*, 105–118. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/16343792`

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007, August). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1485), 1601–1613. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/17428778`

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., & Corbetta, M. (2007, March). Breakdown of Functional Connectivity in Frontoparietal Networks Underlies Behavioral Deficits in Spatial Neglect. *Neuron*, *53*(6), 905–918. doi: 10.1016/j.neuron.2007.02.013

Herd, S. A., Krueger, K., Nair, A., Mollick, J., & O'Reilly, R. (2019). Neural Mechanisms of Human Decision-Making. *Cognitive Affective and Behavioral Neuroscience*. Retrieved from `Preprint: https://arxiv.org/abs/1912.07660`

Herd, S. A., O'Reilly, R. C., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014, April). A neural network model of individual differences in task switching abilities. *Neuropsychologia*. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/24791709`

Hochreiter, S., & Schmidhuber, J. (1997, January). Long Short-Term Memory. *Neural Computation*, *9*, 1735–1780.

Houk, J. C. (2005, June). Agents of the mind. *Biol Cybern*, *92*(6), 427–437. Retrieved from `http://dx.doi.org/10.1007/s00422-005-0569-8`

Ilinsky, I. A., Jouandet, M. L., & Goldman-Rakic, P. S. (1985). Organization of the nigrothalamocortical system in the rhesus monkey. *Journal of Comparative Neurology*, *236*(3), 315–330. doi: 10.1002/cne.902360304

Jensen, O., Bonnefond, M., & VanRullen, R. (2012, April). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences*, *16*(4), 200–206. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/22436764`

Jones, E. G. (1998a). A new view of specific and nonspecific thalamocortical connections. *Advances in Neurology*, *77*, 49–71. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/9709817`

Jones, E. G. (1998b, April). Viewpoint: The core and matrix of thalamic organization. *Neuroscience*, *85*(2), 331–345. doi: 10.1016/S0306-4522(97)00581-2

Jones, E. G. (2007). *The Thalamus* (Second ed., Vol. 2 volume set). Cambridge: Cambridge University Press.

Karnath, H. O., Himmelbach, M., & Rorden, C. (2002, February). The subcortical anatomy of human spatial neglect: Putamen, caudate nucleus and pulvinar. *Brain : a journal of neurology*, *125*, 350–360. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11844735`

Kawato, M., Hayakawa, H., & Inui, T. (1993, January). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems*, *4*(4), 415–422. doi: 10.1088/0954-898X_4_4_001

Klimesch, W. (2011, August). Evoked alpha and early access to the knowledge system: The P1 inhibition timing hypothesis. *Brain Research*, *1408*, 52–71. doi: 10.1016/j.brainres.2011.06.003

Kobatake, E., & Tanaka, K. (1994, January). Neuronal selectivities to complex object features in the ventral visual pathway. *Journal of Neurophysiology*, *71*(3), 856–867.

Kohonen, T. (1989). *Self-organization and associative memory*. Springer-Verlag New York, Inc. New York, NY, USA.

Kok, P., & de Lange, F. P. (2015). Predictive Coding in Sensory Cortex. In *An Introduction to Model-Based Cognitive Neuroscience* (pp. 221–244). Springer, New York, NY. doi: 10.1007/978-1-4939-2236-9_11

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012, July). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. doi: 10.1016/j.neuron.2012.04.034

Kritzer, M. F., & Goldman-Rakic, P. S. (1995, August). Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *359*(1), 131–143. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/8557842`

Kuramoto, E., Furuta, T., Nakamura, K. C., Unzai, T., Hioki, H., & Kaneko, T. (2009, September). Two types of thalamocortical projections from the motor thalamic nuclei of the rat: A single neuron-tracing study using viral vectors. *Cerebral cortex*, *19*(9), 2065–2077. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19174446`

Kuramoto, E., Ohno, S., Furuta, T., Unzai, T., Tanaka, Y. R., Hioki, H., & Kaneko, T. (2015, January). Ventral medial nucleus neurons send thalamocortical afferents more widely and more preferentially to layer 1 than neurons of the ventral anterior–ventral lateral nuclear complex in the rat. *Cerebral Cortex*, *25*(1), 221–235. doi: 10.1093/cercor/bht216

LaBerge, D., & Buchsbaum, M. S. (1990, March). Positron emission tomographic measurements of pulvinar activity during an attention task. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *10*, 613–9. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/2303863`

Lamme, V. A. F. (2006, January). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. doi: 10.1016/j.tics.2006.09.001

Larkum, M. E., Petro, L. S., Sachdev, R. N. S., & Muckli, L. (2018). A perspective on cortical layering and layer-spanning neuronal elements. *Frontiers in Neuroanatomy*, *12*. doi: 10.3389/fnana.2018.00056

Larkum, M. E., Zhu, J. J., & Sakmann, B. (1999, March). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, *398*(6725), 338–341. doi: 10.1038/18686

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May). Deep learning. *Nature*, *521*(7553), 436–444. doi: 10.1038/nature14539

Lee, T. S., & Mumford, D. (2003, July). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America*, *20*(7), 1434–1448. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/12868647/`

Lewis, L. D., Voigts, J., Flores, F. J., Schmitt, L. I., Wilson, M. A., Halassa, M. M., & Brown, E. N. (2015, October). Thalamic reticular nucleus induces fast and local modulation of arousal state. *eLife*, *4*, e08760. doi: 10.7554/eLife.08760

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020, June). Backpropagation and the brain. *Nature Reviews Neuroscience*, *21*(6), 335–346. doi: 10.1038/s41583-020-0277-3

Lisman, J. E., Fellous, J. M., & Wang, X. J. (1999, April). A role for NMDA-receptor channels in working memory. *Nature Neuroscience*, *1*, 273-275. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/10195158`

Lotter, W., Kreiman, G., & Cox, D. (2016, May). Deep predictive coding networks for video prediction and unsupervised learning. *arXiv:1605.08104 [cs, q-bio]*. Retrieved 2017-08-11, from `http://arxiv.org/abs/1605.08104`

Luczak, A., Bartho, P., & Harris, K. D. (2009, May). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, *62*(3), 413–425. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19447096`

Luczak, A., Bartho, P., & Harris, K. D. (2013, January). Gating of sensory input by spontaneous cortical activity. *The Journal of Neuroscience*, *33*(4), 1684–1695. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23345241`

Luo, Y., Boix, X., Roig, G., Poggio, T., & Zhao, Q. (2015, November). Foveation-based Mechanisms Alleviate Adversarial Examples. *arXiv:1511.06292 [cs]*. Retrieved 2016-11-04, from `http://arxiv.org/abs/1511.06292`

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002, January). Dynamic Brain Sources of Visual Evoked Responses. *Science*, *295*, 690–693.

Manto, M. (2009, April). Mechanisms of human cerebellar dysmetria: Experimental evidence and current conceptual bases. *Journal of NeuroEngineering and Rehabilitation*, *6*(1), 10. doi: 10.1186/1743-0003-6-10

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., … Kennedy, H. (2014, January). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex: Cortical counterstreams. *Journal of Comparative Neurology*, *522*(1), 225–259. doi: 10.1002/cne.23458

Mathewson, K., Gratton, G., Fabiani, M., Beck, D., & Ro, T. (2009). To see or not to see: Prestimulus alpha phase predicts visual awareness. *The Journal of Neuroscience*, *29*(9), 2725–2732.

Mathewson, K. E., Prudhomme, C., Fabiani, M., Beck, D. M., Lleras, A., & Gratton, G. (2012, August). Making waves in the stream of consciousness: Entraining oscillations in EEG alpha and fluctuations in visual awareness with rhythmic visual stimulation. *Journal of Cognitive Neuroscience*, *24*(12), 2321–2333. doi: 10.1162/jocn_a_00288

Middleton, F. A., & Strick, P. L. (2000, May). Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain research. Brain research reviews*, *31*(2-3), 236–250. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/10719151`

Mink, J. W. (1996, March). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, *50*(4), 381–425. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/9004351`

Mumford, D. (1991, June). On the computational architecture of the neocortex. *Biological Cybernetics*, *65*(2), 135–145. doi: 10.1007/BF00202389

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/1540675`

Münkle, M. C., Waldvogel, H. J., & Faull, R. L. M. (2000, July). The distribution of calbindin, calretinin and parvalbumin immunoreactivity in the human thalamus. *Journal of Chemical Neuroanatomy*, *19*(3),

155–173. doi: 10.1016/S0891-0618(00)00060-0

Nambu, A. (2008, December). Seven problems on the basal ganglia. *Current Opinion in Neurobiology*, *18*(6), 595–604. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19081243`

Olsen, S., Bortone, D., Adesnik, H., & Scanziani, M. (2012, February). Gain control by layer six in cortical circuits of vision. *Nature*, *483*(7387), 47–52.

O'Reilly, R. C. (1996, January). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*(5), 895–938. doi: 10.1162/neco.1996.8.5.895

O'Reilly, R. C. (2006, October). Biologically based computational models of high-level cognition. *Science*, *314*(5796), 91–94. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/17023651`

O'Reilly, R. C. (2020, April). Unraveling the Mysteries of Motivation. *Trends in Cognitive Sciences*. doi: 10.1016/j.tics.2020.03.001

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/16378516`

O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2016). The Leabra cognitive architecture: How to play 20 principles with nature and win! In S. Chipman (Ed.), *Oxford handbook of cognitive science.* Oxford University Press. Retrieved 2015-05-15, from `http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199842193.001.0001/oxfordhb-9780199842193-e-8`

O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2012). *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition, URL: http://ccnbook.colorado.edu. Retrieved from `http://ccnbook.colorado.edu`

O'Reilly, R. C., Nair, A., Russin, J. L., & Herd, S. A. (2020, March). How Sequential Interactive Processing Within Frontostriatal Loops Supports a Continuum of Habitual to Controlled Processing. *Frontiers in Psychology*, *11*, 380. doi: 10.3389/fpsyg.2020.00380

O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002, February). Prefrontal cortex and dynamic categorization tasks: Representational organization and neuromodulatory control. *Cerebral cortex (New York, N.Y. : 1991)*, *12*, 246–257. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/11839599`

O'Reilly, R. C., Russin, J. L., Zolfaghar, M., & Rohrlich, J. (2021). Deep predictive learning in neocortex and pulvinar. *Journal of Cognitive Neuroscience*, *33*(6), 1158–1196.

O'Reilly, R. C., Wyatte, D., Herd, S. A., Mingus, B., & Jilk, D. J. (2013). Recurrent Processing during Object Recognition. *Frontiers in Psychology*, *4*(124). Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23554596`

O'Reilly, R. C., Wyatte, D., & Rohrlich, J. (2014, July). Learning Through Time in the Thalamocortical Loops. *arXiv:1407.3432 [q-bio]*. Retrieved 2015-05-15, from `http://arxiv.org/abs/1407.3432`

O'Reilly, R. C., Wyatte, D. R., & Rohrlich, J. (2017, September). Deep predictive learning: A comprehensive model of three visual streams. *arXiv:1709.04654 [q-bio]*. Retrieved 2017-09-15, from `http://arxiv.org/abs/1709.04654`

Ouden, H. E. M., Kok, P., & Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*(548). Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23248610`

Pasupathy, A., & Miller, E. K. (2005, January). Different time courses for learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*, 873–876. Retrieved from `http://www.nature.com/nature/journal/v433/n7028/full/nature03287.html`

Petersen, S. E., Robinson, D. L., & Morris, J. D. (1987, January). Contributions of the pulvinar to visual spatial attention. *Neuropsychologia*, *25*(1), 97–105. doi: 10.1016/0028-3932(87)90046-7

Phillips, J. W., Schulmann, A., Hara, E., Winnubst, J., Liu, C., Valakh, V., ... Hantman, A. W. (2019, November). A repeated molecular architecture across thalamic pathways. *Nature Neuroscience*, *22*(11), 1925–1935. doi: 10.1038/s41593-019-0483-3

Rac-Lubashevsky, R., & Frank, M. J. (2020, December). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. *bioRxiv*, 2020.12.21.423791. doi: 10.1101/2020.12.21.423791

Raizada, R. D. S., & Grossberg, S. (2003, January). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral cortex*, *13*(1). Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/12466221`

Ramaswamy, S., & Markram, H. (2015). Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron. *Frontiers in Cellular Neuroscience*, *9*. doi: 10.3389/fncel.2015.00233

Rao, R. P., & Ballard, D. H. (1999, January). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Andryk, S., Dooley, G. L., ... Saalmann, Y. B. (2020, April). Thalamus modulates consciousness via layer-specific control of cortex. *Neuron*, *106*(1), 66-75.e12. doi: 10.1016/j.neuron.2020.01.005

Reynolds, J. H., & Desimone, R. (2003, March). Interacting roles of attention and visual salience in V4. *Neuron*, *37*, 853–863. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/12628175`

Reynolds, J. H., & Heeger, D. J. (2009, January). The normalization model of attention. *Neuron*, *61*(2), 168–185. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19186161/`

Richter, D., & de Lange, F. P. (2019, August). Statistical learning attenuates visual activity only for attended stimuli. *eLife*, *8*, e47869. doi: 10.7554/eLife.47869

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018, December). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience*, *21*(12), 1753–1763. doi: 10.1038/s41593-018-0269-z

Robinson, D. L., & Petersen, S. E. (1992, June). The pulvinar and visual salience. *Trends in neurosciences*, *15*, 127–32. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/1374970`

Rougier, N. P., & O'Reilly, R. C. (2002, January). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cognitive Science*, *26*, 503–520. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog2604_4`

Rovó, Z., Ulbert, I., & Acsády, L. (2012, December). Drivers of the Primate Thalamus. *Journal of Neuroscience*, *32*(49), 17894–17908. doi: 10.1523/JNEUROSCI.2815-12.2012

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986, January). Learning representations by back-propagating errors. *Nature*, *323*(9), 533–536.

Rumelhart, D. E., & Zipser, D. (1985, January). Feature discovery by competitive learning*. *Cognitive Science*, *9*(1), 75–112. doi: 10.1207/s15516709cog0901_5

Saalmann, Y. B., & Kastner, S. (2011, July). Cognitive and perceptual functions of the visual thalamus. *Neuron*, *71*(2), 209–223. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/21791281`

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., & Kastner, S. (2012, August). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science*, *337*(6095), 753–756. doi: 10.1126/science.1223082

Sakata, S., & Harris, K. D. (2009, November). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron*, *64*(3), 404–418. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19914188`

Sakata, S., & Harris, K. D. (2012). Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Frontiers in neural circuits*, *6*. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23267317`

Sanders, H., Berends, M., Major, G., Goldman, M. S., & Lisman, J. E. (2013, January). NMDA and GABAB (KIR) Conductances: The "Perfect Couple" for Bistability. *Journal of Neuroscience*, *33*(2), 424–429. doi: 10.1523/JNEUROSCI.1854-12.2013

Schiff, N. D. (2008). Central Thalamic Contributions to Arousal Regulation and Neurological Disorders of Consciousness. *Annals of the New York Academy of Sciences*, *1129*(1), 105–118. doi: 10.1196/annals.1417.029

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008, August). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, *321*(5890), 848–851. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/18687967`

Sherman, S. M., & Guillery, R. W. (2006). *Exploring the Thalamus and Its Role in Cortical Function*. Cambridge, MA: MIT Press. Retrieved from `http://www.scholarpedia.org/article/Thalamus`

Shipp, S. (2003, October). The functional logic of cortico-pulvinar connections. *Philosophical Transactions of the Royal Society of London B*, *358*(1438), 1605–1624. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/14561322`

Silva, L. R., Amitai, Y., & Connors, B. W. (1991, January). Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. *Science*, *251*(4992), 432–435. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/1824881`

Sinz, F. H., Pitkow, X., Reimer, J., Bethge, M., & Tolias, A. S. (2019, September). Engineering a Less Artificial Intelligence. *Neuron*, *103*(6), 967–979. doi: 10.1016/j.neuron.2019.08.034

Snow, J. C., Allen, H. A., Rafal, R. D., & Humphreys, G. W. (2009, March). Impaired attentional selection following lesions to human pulvinar: Evidence for homology between human and monkey. *Proceedings of the National Academy of Sciences*, *106*(10), 4054–4059. doi: 10.1073/pnas.0810086106

Spaak, E., de Lange, F. P., & Jensen, O. (2014, March). Local Entrainment of Alpha Oscillations by Visual Stimuli Causes Cyclic Modulation of Perception. *Journal of Neuroscience*, *34*(10), 3536–3544. doi: 10.1523/JNEUROSCI.4385-13.2014

Summerfield, C., & Egner, T. (2009, September). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. doi: 10.1016/j.tics.2009.06.003

Tanibuchi, I., Kitano, H., & Jinnai, K. (2009a, November). Substantia nigra output to prefrontal cortex via thalamus in monkeys. I. Electrophysiological identification of thalamic relay neurons. *Journal of Neurophysiology*, *102*(5), 2933–2945. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19692504`

Tanibuchi, I., Kitano, H., & Jinnai, K. (2009b, November). Substantia nigra output to prefrontal cortex via thalamus in monkeys. II. Activity of thalamic relay neurons in delayed conditional go/no-go discrimination task. *Journal of Neurophysiology*, *102*(5116), 2946–2954. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/19692503`

Thomson, A. M. (2010). Neocortical layer 6, a review. *Frontiers in Neuroanatomy*, *4*(13). Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/20556241`

Tononi, G. (2004, November). An information integration theory of consciousness. *BMC Neuroscience*, *5*, 42. doi: 10.1186/1471-2202-5-42

Treisman, A. (1993, January). The perception of features and objects. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 5–35). Oxford: Oxford University Press.

Usrey, W. M., & Sherman, S. M. (2018). Corticofugal circuits: Communication lines from the cortex to the rest of the brain. *Journal of Comparative Neurology*, *0*(0). doi: 10.1002/cne.24423

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014, October). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences U.S.A.*, *111*(40), 14332–14341. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/25205811`

VanRullen, R., & Koch, C. (2003, May). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/12757822`

von Stein, A., Chiang, C., & König, P. (2000, December). Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14748–14753. doi: 10.1073/pnas.97.26.14748

von Helmholtz, H. (1867). *Treatise on Physiological Optics, Vol III*. Courier Corporation.

Voytek, B., & Knight, R. T. (2010, October). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proceedings of the National Academy of Sciences USA*, *107*(42), 18167–18172. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/20921401`

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020, March). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. doi: 10.1111/nyas.14321

Wang, M., Yang, Y., Wang, C.-J., Gamo, N. J., Jin, L. E., Mazer, J. A., ... Arnsten, A. F. T. (2013, February). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron*, *77*(4), 736–749. doi: 10.1016/j.neuron.2012.12.032

Ward, L. M. (2011, June). The thalamic dynamic core theory of conscious experience. *Consciousness and Cognition*, *20*(2), 464–486. doi: 10.1016/j.concog.2011.01.007

Watanabe, Y., & Funahashi, S. (2012, January). Thalamic mediodorsal nucleus and working memory. *Neuroscience & Biobehavioral Reviews*, *36*(1), 134–142. doi: 10.1016/j.neubiorev.2011.05.003

Watanabe, Y., Takeda, K., & Funahashi, S. (2009, June). Population vector analysis of primate mediodorsal thalamic activity during oculomotor delayed-response performance. *Cerebral Cortex*, *19*, 1313–1321. Retrieved from `http://cercor.oxfordjournals.org/cgi/content/abstract/19/6/1313`

Whittington, J. C. R., & Bogacz, R. (2019, March). Theories of error back-propagation in the brain. *Trends in Cognitive Sciences*, *23*(3), 235–250. doi: 10.1016/j.tics.2018.12.005

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., & Halassa, M. M. (2015, October). Thalamic control of sensory selection in divided attention. *Nature*, *526*(7575), 705–709. doi: 10.1038/nature15398

Wurtz, R. H. (2008, September). Neuronal mechanisms of visual stability. *Vision Research*, *48*(20), 2070–2089. doi: 10.1016/j.visres.2008.03.021

Wyatte, D., Curran, T., & O'Reilly, R. C. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/22905822`

Wyder, M. T., Massoglia, D. P., & Stanford, T. R. (2004, June). Contextual modulation of central thalamic delay-period activity: Representation of visual and saccadic goals. *Journal of Neurophysiology*, *91*(6), 2628–2648. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/14762161`

Zhou, H., Schafer, R. J., & Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron*, *89*, 209–220.