Predictive Models in the Brain

Keith L. Downing
The Norwegian University of Science and Technology
Trondheim, Norway
keithd@idi.ntnu.no

September 30, 2008

Abstract

Many neuroscientists view prediction as one of the core brain functions. However, there is little consensus as to the exact nature of predictive information and processes, nor the neural mechanisms that realize them. This paper reviews a host of neural models believed to underlie the learning and deployment of predictive knowledge in a variety of brain regions: neocortex, hippocampus, thalamus, basal ganglia and cerebellum. These are compared and contrasted in order to codify a few basic aspects of neural circuitry and dynamics that appear to be the heart of prediction.

1 Introduction

Keen predictive abilities have long been recognized as special talents. Those who can consistently determine what the future brings often enjoy high salaries and elevated social standing. In early human civilizations, the well-being of an entire tribe was contingent upon the ability to foresee a rough winter or a pending enemy attack, while today, in a world governed by a perplexing interplay between complex systems such as climate, politics and international markets, predictive prowess is at an absolute premium.

Yet despite its well-respected role in society, prediction often goes unappreciated as a fundamental component of intelligence. Recently, several prominent scientists [34, 23] have championed the primacy of prediction in cognition. In *On Intelligence*, computer scientist and founder of the Redwood Neuroscience Institute, Jeffrey Hawkins [23] argues for a more prediction-centered view of intelligence:

> Intelligence and understanding started as a memory system that fed predictions into the sensory stream. These predictions are the essence of understanding. To know something means that you can make predictions about it .. We can now see where Alan Turing went wrong. Prediction, not behavior, is the proof of intelligence.. (pp. 104-105)

In Hawkins' view, the brain is constantly predicting future states, and these expectations combine with sensory inputs to produce our perceived reality.

In *i of the Vortex* [34], the renowed neuroscientist Rodolfo Llinas states:

> The capacity to predict the outcome of future events - critical to successful movement - is, most likely, the ultimate and most common of all global brain functions...(pg. 21)

Llinas [34] begins with the earliest mobile lifeforms and their demands for accurate sensorimotor control in a world whose tempo often exceeds the maximum speeds of neurally-controlled perception and action. Without prediction, mobile animals cannot choose, at time t, the proper actions for time $t + \triangle_a$, based on the state of the world at time $t - \triangle_p$, where $\triangle_p$ and $\triangle_a$ are the delays for sensory processing and motor activation, respectively. Control theorists are well aware of the problems imposed by these types of delays [22], and mechanisms for predicting future system states, such as Kalman filters [30], are a common solution. Neuroscientists [52, 34] generally agree that the brain needs similar predictive abilities, and they cite areas such as the cerebellum [52], basal ganglia [27], hippocampus [19] and neocortex [23] as central to this endeavor.

As we discuss in [15, 14], these predictive facilities may underlie our common-sense understanding of the world and may provide support for *cognitive incrementalism* [10] - the view that cognition arises directly from sensorimotor activity - which, in turn, is a motivating philosophy of situated and embodied artificial intelligence (SEAI). However, we also point to the pronounced differences between procedural and declarative knowledge [47] (and the brain areas that appear to facilitate them), which leave considerable doubt as to whether a single corpus of predictive information could support both sensorimotor activity and higher-level cognition.

This paper continues our quest to better understand the role of prediction in the brain. We examine a host of neural subsystems and associated computational models to distill a set of basic anatomical and physiological factors that support predictive behavior. Although Hawkins [23] focuses on the cortex, and Llinas [34] on the cerebellum, we find interesting predictive architectures, as proposed by experimental and computational neuroscientists, in five different systems: cerebellum, basal ganglia, hippocampus, neocortex and thalamocortical. The former two embody procedural predictions, while the latter three have a more declarative nature.

The key difference between the procedural and declarative predictive forms resides in the explicit awareness of the connections between spatiotemporal states that embody predictive knowledge. For example, a basketball player is explicitly aware of the fact that a strong rebound and quick outlet pass often predict a fast break, but she may not know what shooting movements can predict a successful shot, even though she can *feel* whether a shot will hit or miss the instant it leaves her fingertips.

As we will see, these two types of prediction, the explicit and implicit, require different architectures. However, within separate regions of the brain, the same types of architectures for procedural and declarative prediction, respectively, seem to reoccur. This apparent duplication of prediction-supporting machinery supports claims that prediction is a fundamental brain process, both at the conscious and subconscious level.

We begin by defining prediction for our neuroscientific purposes. Procedural prediction in the cerebellum and basal ganglia is then explored, with detailed anatomy and physiology of both regions presented and analyzed. We then move on to declarative prediction, where the hippocampus, neocortex and thalamocortical system are dissected, as are a collection of computational models, all by different researchers. We discovered a very interesting commonality across these models, which served as the prime motivation for this article. This common abstraction is summarized in our Generic Declarative Prediction Network (GDPN) prior to the presentation of the individual models. Next, we compare the 5 predictive systems to find a) key similarities between those supporting the same type of prediction and b) key differences between those underlying different predictive modes. Finally, we conclude with general remarks on the role of prediction in brain science.

2 Defining Prediction

From a psychological or social perspective, prediction denotes a wide range of abilities, many of which involve the capacity to learn temporal correlations among events or world states. One can predict the consequences of actions by using acquired associations between those acts and the world states that have, in the past, immediately succeeded them. One can predict the world state that normally follows another world state, where, presumably the earlier state includes some hints as to the key processes governing the state change. When viewing a snapshot of a baseball player running full speed across the warning track, we can predict that one of the following states involves the same player crashing into the outfield wall. We connect the two states via the action, hard running, so clearly evident in the first state.

The dictionary [1] gives two primary definitions of *predict*:

1. to declare or indicate in advance, and
2. to foretell on the basis of observation.

Here, the latter definition essentially supplements the former by implying that observations are the basis for the advance declaration or indication. For our purposes, the definitions of *declare* and *indicate* have significance as well.

To declare is to *make known formally, officially or explicitly*, while to indicate is to *be a sign, symptom or index of* [1]. In short, the declarative form of prediction is more concrete and direct, while the indicative form is more indirect and implicit. As a simple example, one may declaratively predict an upcoming sunny day by explicitly stating, "Tomorrow will be a sunny day." Conversely, one can indicate that prediction by various preparatory acts such as purchasing sunblock, retrieving the lawnmower from the deep recesses of the garage, etc.

Interestingly, these two forms of prediction map quite well to distinct neural structures. The explicit variant seems to coincide with cortical and hippocampal activity, while the implicit type maps to commonly-proported cerebellar and basal gangliar functions, which are often described as

procedural or *non-declarative* [46]. Consequently, the terms *declarative* and *procedural* will be used to denote the explicit and indicative forms of prediction, respectively.

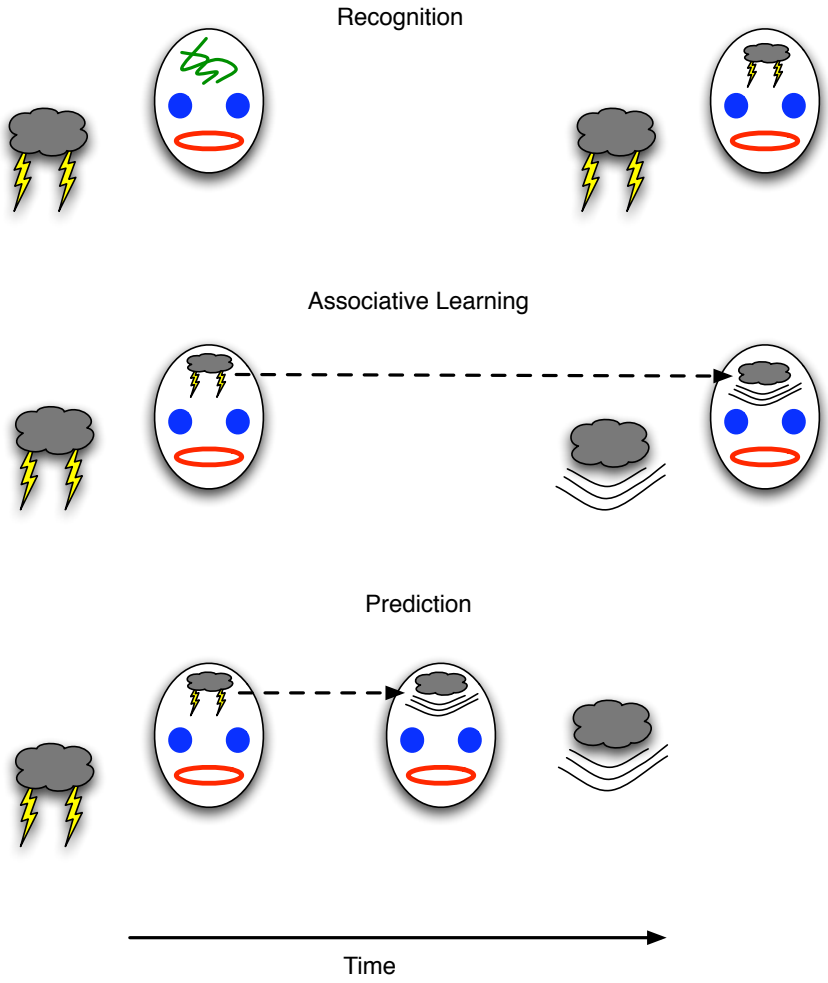

Figure 1: (Above) Recognition depicted as the formation of a brain state (drawn as lightning on the forehead) that becomes correlated with a physical event (lightning). (Middle) Learning the association between one event (lightning) and its successor (thunder) by linking the brain states that correlate with each. (Below) Declarative prediction entails recognizing one event (lightning) and forming the succeeding brain state for thunder prior to (or even in the absence of) the real-world event with which it correlates.

Figure 1 illustrates the basic conception of declarative prediction from a neural perspective. First, *recognition* is defined as attaining a brain state, S, that has previously exhibited a strong correlation with the (now familiar) experience (e.g., object, event or state of the agent itself). Informally, S is that brain state that is both a) most likely to arise under the given experience, e.g. lightning, and b) not likely to arise under other conditions. Along these same lines, declarative predictive knowledge involves two such correlations between brain states and experiences plus a link between the two brain states such that one can trigger the other prior to (or even in the complete absence of) the latter's associated experience. This link need not be bi-directional, so the experience of lightning

may lead to the brain state corresponding to the experience of thunder, but not necessarily vice versa.

In the framework of Figure 1, predictive learning is essentially a special case of associative learning in which the related items represent events having at least a small temporal deviation such that the start of event A precedes that of event B. Then, during the interim between the two starts, prediction *proves its worth* by indicating B (and thus enabling the animal to prepare for B) prior to B's occurrence. This *preparatory window* gives prediction a survival advantage above and beyond that of non-temporal association. In the latter, an antelope can link the sight of a tiger to fear (and its consequences such as hiding or fleeing), but without the ability to associate events over time, the antelope could not link the rustling of bushes at time t with the appearance of a tiger at time t+d, much to the antelope's detriment.

Figure 2 depicts procedural prediction. Here, the agent (a monkey) has acquired a link between a brain state that weakly correlates with lightning (a diamond) and one that weakly maps to thunder (a star). These are weak in the sense that they may not be completely specific for these events, such that any flashing light would trigger the diamond state, and any loud noise would trigger the star state. Thus, it is difficult to claim that the monkey declaratively predicts thunder. In contrast to a declarative representation, the general, weakly correlated state would not stimulate other conscious *thunderstorm thoughts* such as the association with dark rain clouds, the potential dangers, examples of destructive effects, etc. However, an observer may easily interpret the monkey's procedural act of covering its ears as an explicit prediction of thunder. When the agent's actions, but not its brain state, appears to foretell a specific event, the prediction is procedural.

The difference between the two predictive forms is probably most easily discernible in humans, since our communicated descriptions of future states often indicate a declarative component, whereas many physical situations require us to act quickly and appropriately, but without forming clear neural correlates of the next situation. For example, in watching the slow-motion replay of a tennis serve, a coach may predictively describe where and how the ball will land, but in playing the return shot herself, she would simply run to the appropriate spot and adjust her body to the speed, angles and spin expected of the incoming serve. Only a naive outside observer would infer that she had explicit knowledge of those parameters.

In the sections that follow, five neural systems, all of which have been posited as centers of predictive activity by several authors, are examined both in general and with respect to prediction. In many cases, the focus is on computational models of those systems, as these typically provide more thorough - albeit unproven - mechanistic explanations than do the more traditional neuroscientific findings. These systems are the cerebellum and basal ganglia, both viewed as procedurally predictive engines, and the hippocampus, neocortex, and thalamocortical loop, each of which shows strong declarative tendencies.

3 Procedural Prediction in the Cerebellum

The cerebellum has a well-established role in the learning and control of complex motions [31, 6], and many believe that this involves the use of predictive models [52, 2]. A brief anatomical overview

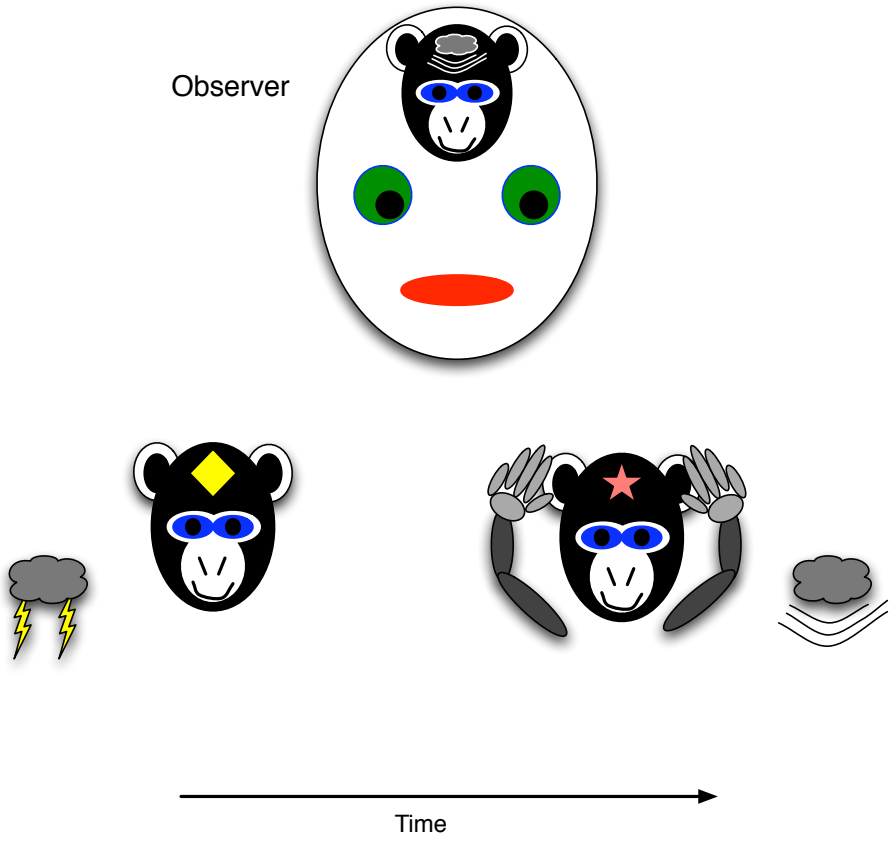

Figure 2: Procedural prediction, wherein the agent's actions *indicate* specific knowledge of a future world state, even though the agent (monkey) has no explicit brain state that strongly correlates with the world state. The agent's ear-covering behavior can easily lead an observer to infer that the agent has the strongly-correlated brain state, i.e., explicit knowledge of the upcoming thunder.

(based on [6]) of the cerebellum appears in Figure 3, which indicates the highly ordered structure of this region.

As shown in Figures 3 and 4, the cerebellar input layer, the granular cells, receive a variety of peripheral sensory and cortical signals via mossy fibers stemming from the spinal cord and brainstem. These signals experience differential delays before converging upon the granular cells, with an average of 4 such inputs per cell [44] The large number of such cells, approximately 10^{11} in humans [31], combined with their tendency to laterally inhibit one another, via the interspersed golgi cells, indicates that the granular cells serve as sparse-coding detectors of relatively simple (i.e. involving just a few integrated stimuli) contexts [44]. Since delay times vary along the mossy fibers, each context has both temporal and spatial extent.

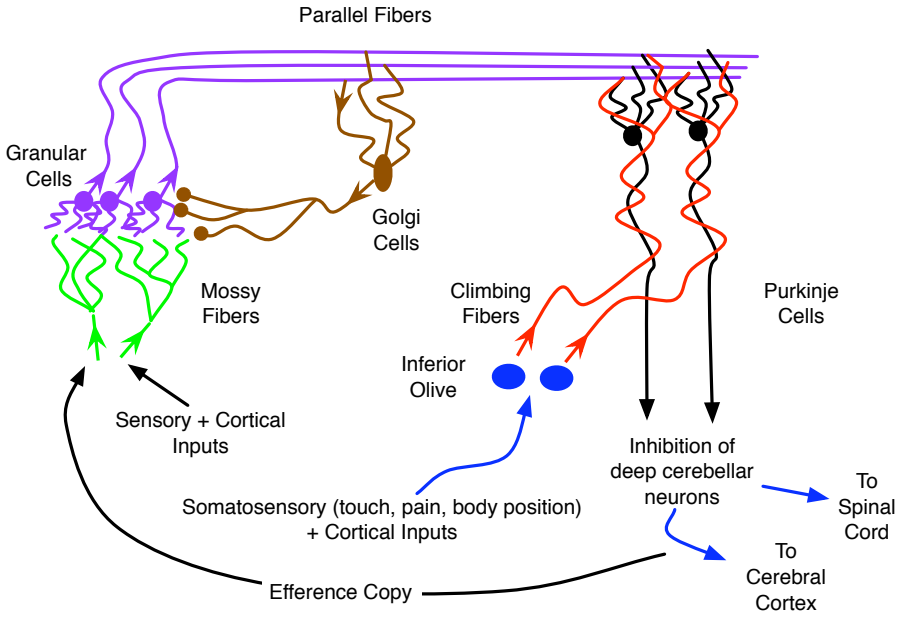

Figure 3: The basic organization of the cerebellum, an abstraction and combination of more complex diagrams in Bear et al. [6], originally appearing in [15].

One parallel fiber emanates from each granular cell and synapses onto the dendrites of many Purkinje cells, each of which may receive input from 10^5 to 10^6 parallel fibers [31]. Since the Purkinje outputs are the cerebellum's ultimate contribution to the control of motor (and possibly cognitive) activity, the plethora granular inputs to each Purkinje cell would appear to embody a complex set of preconditions for the generation of any such output. Since the PF-PC synapses are modifiable [31, 44], these preconditions are subject to learning/adaptation.

As shown in Figure 3, climbing fibers from the inferior olive send signals to the PF-PC synapses. The climbing fibers transfer pain signals from the muscles and joints controlled by those fibers' corresponding Purkinje cells, and these affect long-term depression (LTD) of the neighboring PF-PC synapses [31, 44]. Thus, the climbing fibers provide a primitive form of supervised learning [16] wherein the combination of parallel fibers that cause a Purkinje cell to fire (and thus promote a muscular movement resulting in discomfort) will be less likely to excite the same PC in the future.

In short, the feedback from the inferior olive and clmbing fibers helps to filter out inappropriate contexts (embodied in the parallel fibers) for particular muscle activations.

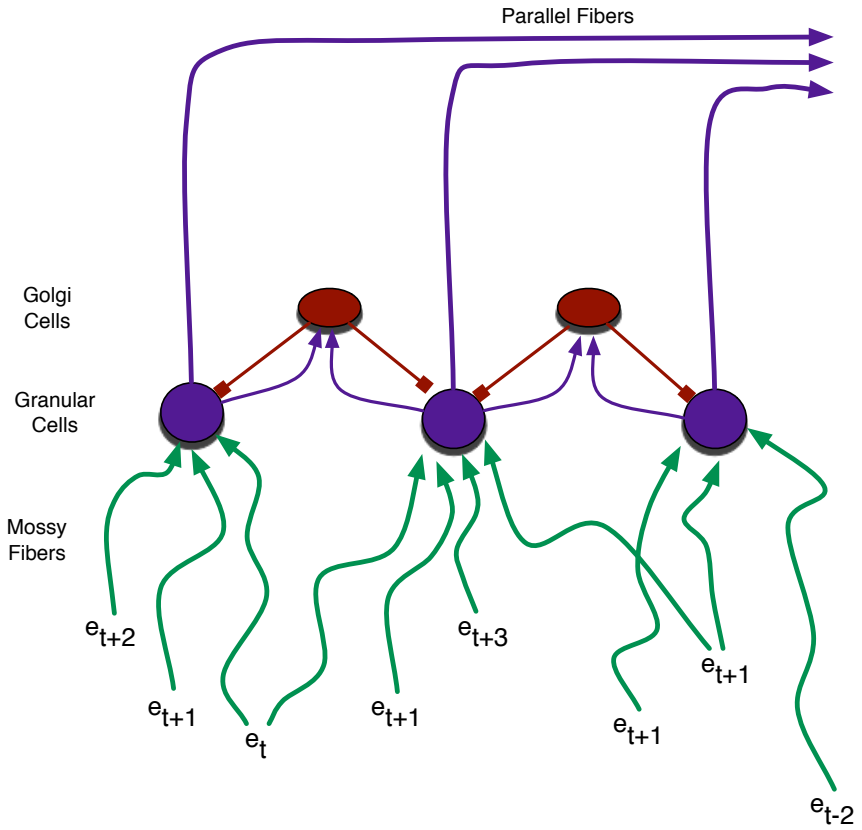

Figure 4: Granular cells realize sparse coding for temporally-blended contexts. Events (e) of various temporal origins (denoted by subscripts) simultaneously activate granular cells due to differential delays - roughly depicted by line length, with longer lines denoting events that occurred further in the past -along mossy fibers.

Plasticity at the PF-PC synapse relies on post-synaptic long-term depression (LTD). When a CF forces a PC to fire strongly, those PC dendrites that were recently activated by parallel fibers undergo chemical changes that reduce their sensitivity to glutamate (the neurotransmitter used by PFs). Hence, the influence of those PFs on the PC declines [6].

Somewhat counterintuitively, the simplest behaviors often require the most complex neural activity patterns. For example, it takes a much more intricate combination of excitatory and (particularly) inhibitory signals to wiggle a single finger (or toe) than to move all five. Hence, the tuning of PC cells to achieve the appropriate inhibitory mix is a critical factor in basic skill learning.

Figure 5 gives a hypothetical example of a behavioral rule implemented by a cerebellar tract. A baseball outfielder receives a variety of sensory inputs with different temporal delays, shown here as converging on the same granular cell. The granular output then affects several Purkinje cells, including those whose ultimate effect is to adjust the player's orientation and leg angle in the

attempt to rapidly accelerate toward the projected destination of the ball.

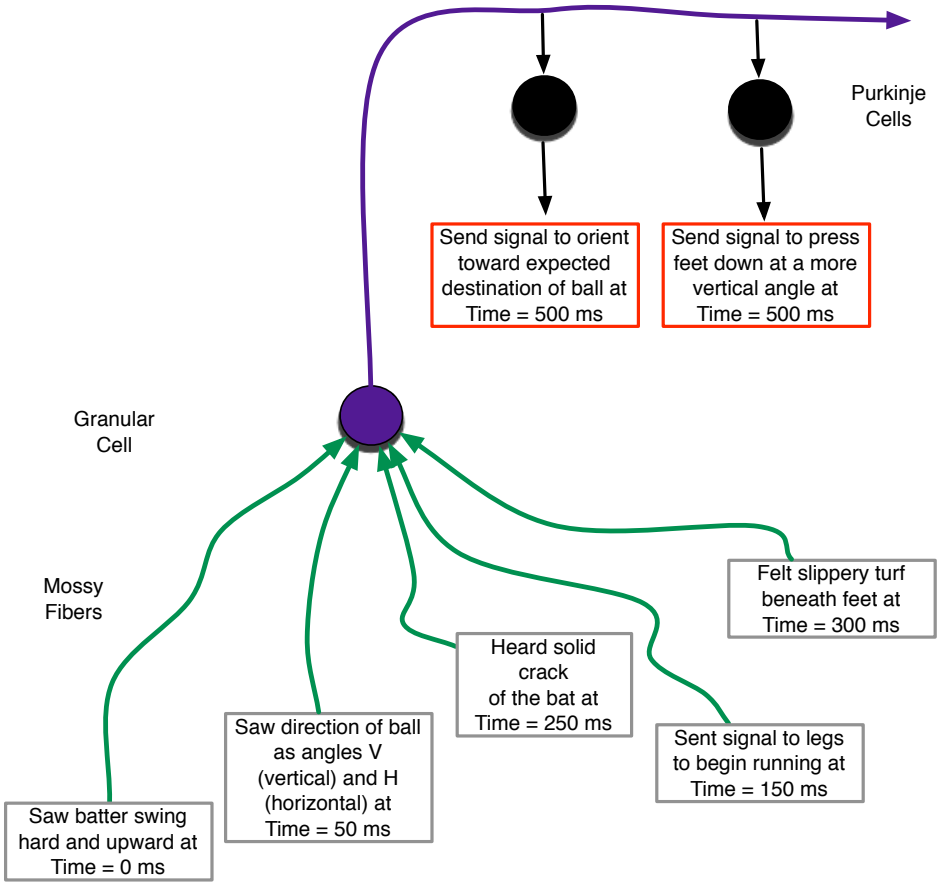

Figure 5: Temporally-mixed sensory and proprioceptive experiences of a baseball outfielder. These form a context for increasing vertical foot plant while accelerating to catch a fly ball.

The *predictiive* nature of this and similar rules involves the integration of sensory stimuli, whose temporal relationships are highly salient, to determine proper actions. Thus, cross-sections of the past determine present decisions about future behaviors. As depicted in Figure 6, the detection of any salient consequences or errors comes even later, due to sensory-processing delays. That error signal should then provide feedback regarding the decisions made earlier.

To maintain an approximate record of what channels were active, and when, and thus what synapses are most *eligible* for modification, the cerebellum and many other brain areas utilize a complex biochemical process that essentially yields a synapse most receptive to LTP or LTD about 100 msec after high transmission activity (as discussed in [32, 26]). This *eligibility trace*, in the parlance of reinforcement learning theory [50], helps compensate for the time delays of sensory processing and motor activation. Eligibility dynamics have probably coevolved with the sensory, motor and proprioceptive apparatus to support optimal learning. Figure 7 shows the eligibility traces associated with several context-action pairs, with those occurring within a narrow time window prior to error detection having the highest values.

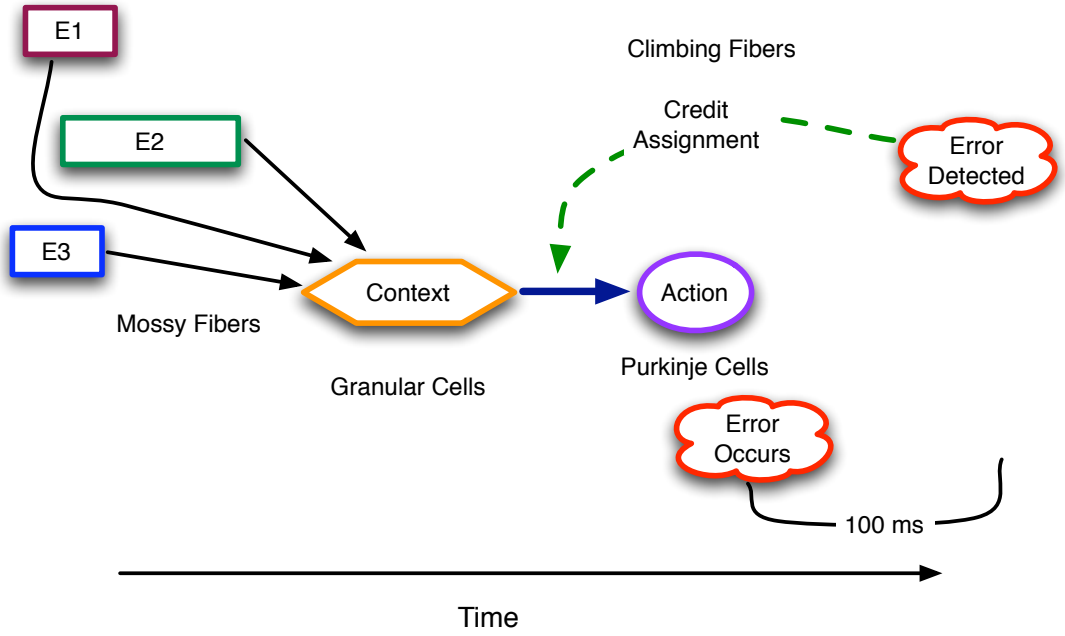

Figure 6: The temporal scope of cerebellar decision making. Context from the past affects present action choices whose actions are realized in the future and whose consequences are perceived even further in the future.

Considering that the human cerebellum consists of over a million parallel fibers, each of which embodies a context-action association, physical skill learning may consist of the gradual tuning and pruning of this immense rule set. Links of high utility should endure, while others will fade via LTD. Importantly, since contexts reflect states of the world prior to action choice and action performance - again, due to inherent sensory-processing delays - the actions that they recommend should be those most appropriate for states of the body and world at some future time (relative to the contexts). Recommendations that lack this predictive nature will produce inferior behavior and be weakened via LTD. By trial and error, the cerebellum learns to support the most salient predictions, which are those that properly account for the inherent delays in sensory processing and motor realization.

From the viewpoint of an outside observer, the cerebellum's actions would appear to involve explicit knowledge as to future states, such as L, the location of the baseball 3 seconds after contact with the bat. However, the cerebellar rule need only embody the behavior that will eventually move the player to that spot, without an explicit representation of the spot itself. With respect to our definition of declarative prediction (as drawn in Figure 1), there is not necessarily a brain state that correlates with L, and even if there is, it need not be stimulated by the cerebellar activity that helps move the player to L.

For example, in the eye-tracking simulations and primate trials of Kettner et. al., [32], both monkeys and computer models anticipate future points along complex visual trajectories by shifting gaze to the appropriate locations. In describing these systems as predictive, the authors refer to overt behaviors that indicate, to the outside observer, explicit knowledge of future locations. However,

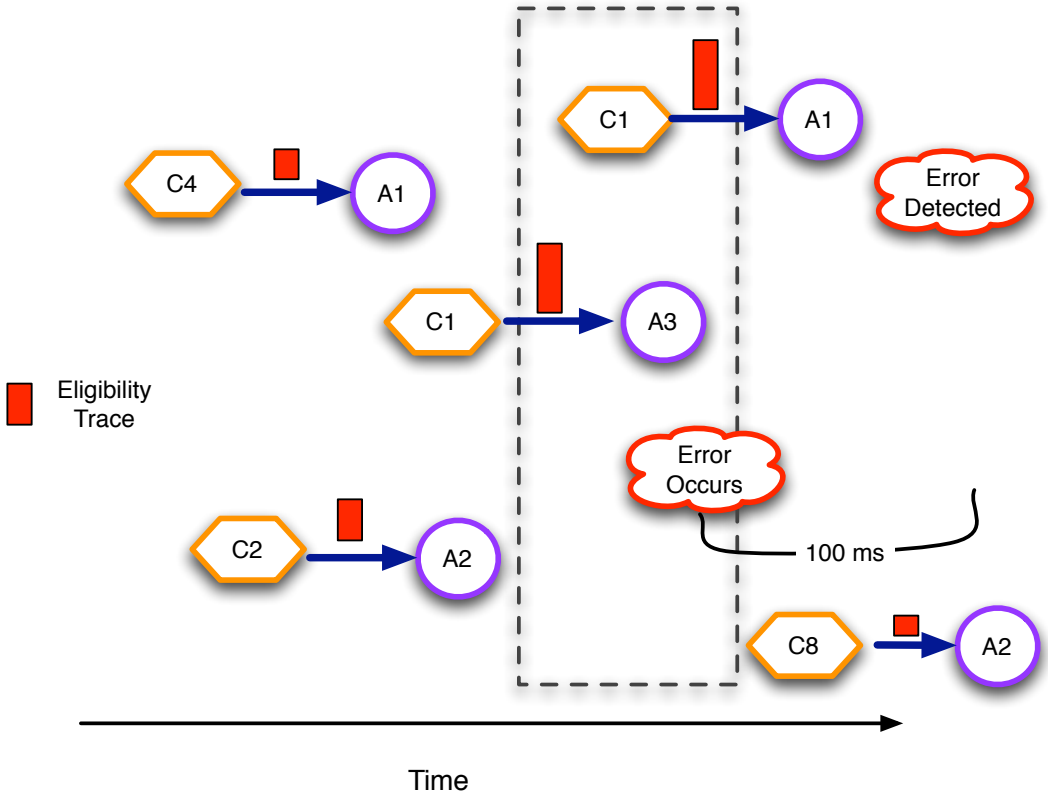

Figure 7: Cerebellar eligibility traces, drawn as rectangles on the condition-action arcs, with taller rectangles denoting higher eligibility. Synapses are most eligible for modification approximately 100 milliseconds after they transmit an action potential.

neither system is claimed to explicitly house representations (i.e. correlated brain states) for those sites. The predictive knowledge is purely procedural. Knowing how and when to *look* at a location is a lot different than explicitly knowing *about* that spot.

4 Regressive Procedural Prediction in the Basal Ganglia

In the basal ganglia, prediction arises in the course of reinforcement learning (RL) [50], which many researchers view as a central capability of this region [16, 26, 39]. RL systems learn associations between environmental (and bodily) states and various rewards or punishments (i.e. reinforcements) that those states may incur, either immediately or at some time in the future. Thus, the system learns to predict the reinforcement from the state. Naturally, RL provides a survival advantage, since it enables organisms to behave proactively instead of merely reactively.

However, the extent of prediction in RL is somewhat suspect: animals are not necessarily foretelling future states in any great detail. Instead, they may only possess basic intuitions about impending pleasure or pain. As shown in Figure 8, the selective advantage stems from recognizing these reinforcements based on earlier, and often more subtle, clues. Thus, the predictive ability *regresses* in time. For example, a monkey that anticipates thunder at the first sight of atmospheric light can more consistently protect its ears than one dependent upon the sight of a lightning bolt, which, when close to the observer, strikes almost simultaneously with the thunder blast.[1]

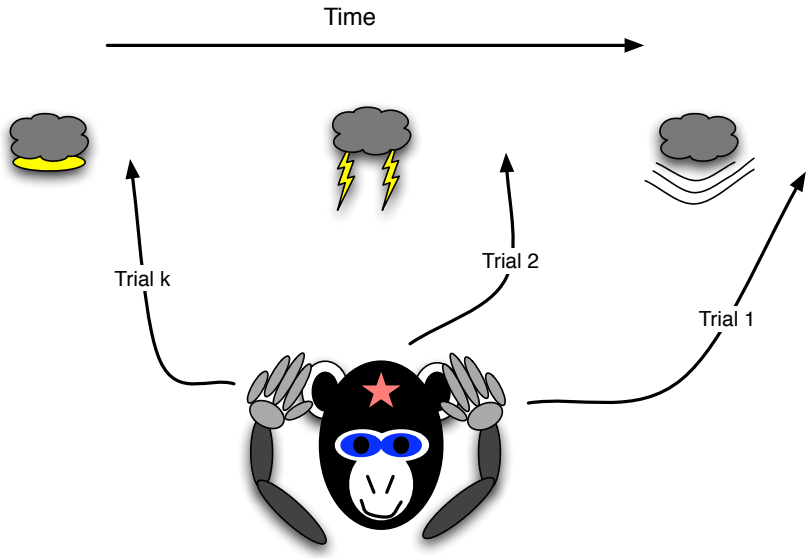

Figure 8: Regressive prediction: the agent recognizes earlier and earlier indicators of the emotive event.

Sketched in Figures 9 and 10, the BG are large midbrain structures that receive convergent inputs

[1]What follows is a modified version of section 5.2 in [15], enhanced to incorporate a slightly wider range of anatomical information and to highlight the predictive role of the basal ganglia.

from many cortical areas onto the striatum (consisting of caudate nucleus and putamen) and the subthalamic nucleus (STN). The striatal cells appear to function as a layer of competitive context detectors [25], since a) each neuron receives inputs from circa 10,000 cortical neurons, b) their electrochemical properties are such that they only fire if many of those inputs are active, and c) they have intra-layer inhibitory connections.

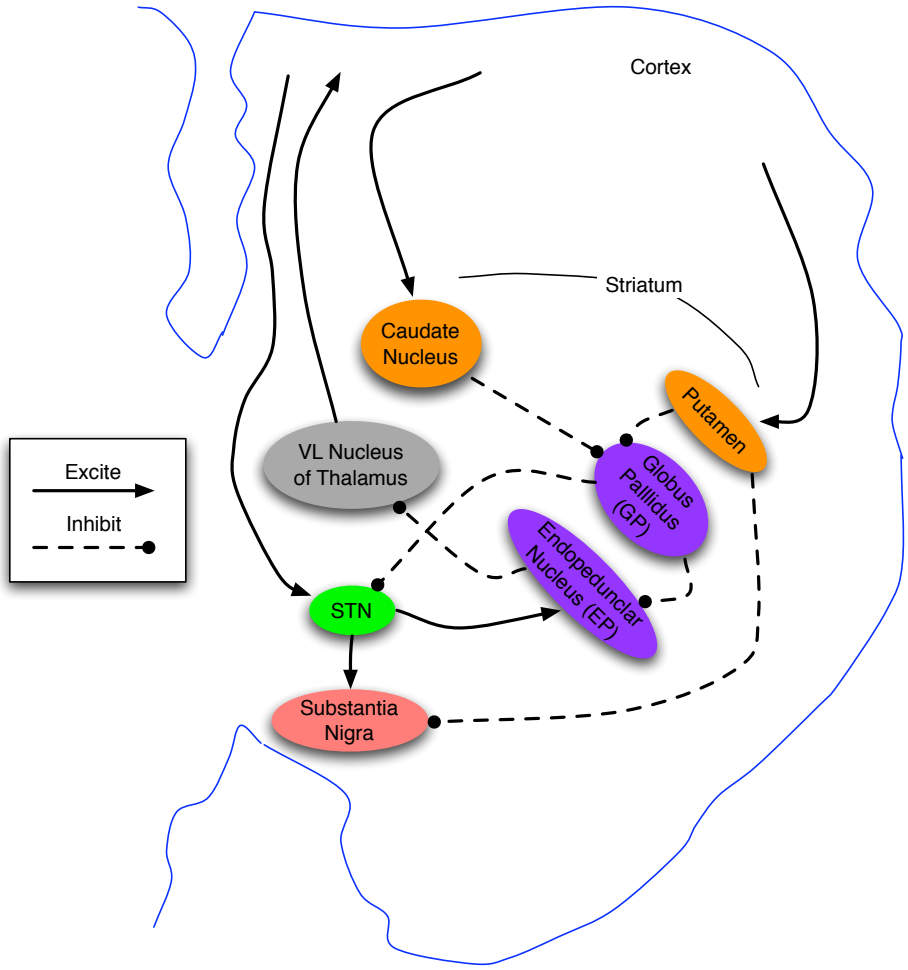

Figure 9: Basic anatomy of the basal ganglia in one hemisphere, shown as a coronal cross-section. Based on diagrams in [6, 40]

Strong evidence [48, 21] indicates that the BG are arranged in parallel loops wherein a striatal cell's inputs come from a region of a particular cortex, such as the motor cortex (MC). Their outputs to the substantia nigra pars compacta (SNc), substantia nigra pars reticulata (SNr) and entopeduncular nucleus (EP) are eventually channeled back to the MC in the form of both action potentials (via the thalamus) and the neuromodulator dopamine. A great majority of these loops appear to involve the prefrontal cortex (PFC)[25, 31, 48], thus indicating BG contributions to attention, possibly as the mechanism for gating new patterns into working memory [21, 38].

Accounts of GB functional topology vary considerably [25, 26, 40, 21, 20], but several similarities do exist. First, the the striatum appears to consist of two main neuron types: striosomes and

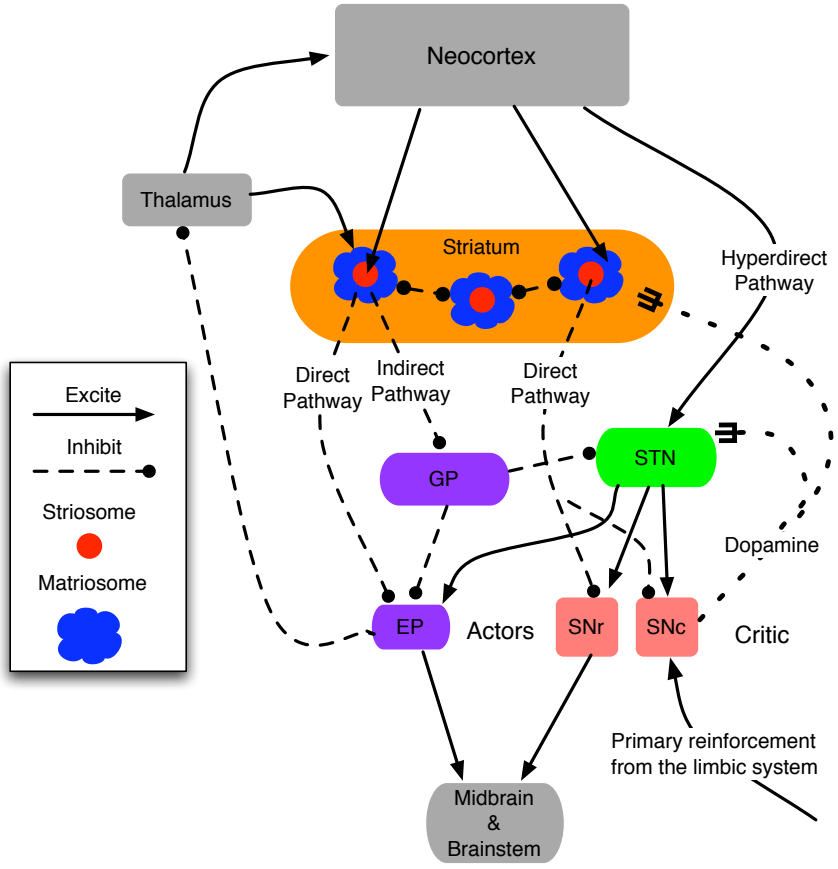

Figure 10: Functional topology of the basal ganglia and their main inputs, derived from text and diagrams in [25, 26, 40, 21]. The actor denotes the direct outputs of the BG: EP and SNr, while the critic consists of the diffuse neuromodulatory output from SNc. Matriosomes are primarily gateways to the actor circuit, while striosomes have direct-pathway links to both actors and critics.

matriosomes, where the former are surrounded by the latter. Several prominent researchers [5, 26] characterize the BG as a combination of actor and critic, with the matriosomes and pallidal neurons (EP and SNr) as the actor's input and output ports, respectively, while the striosomes and SNc comprise the critic. Although this characterization is not completely consistent with other sources, such as [28], the matriosomes and striosomes are often characterized as respectively supporting action selection and state assessment (via dopamine signalling from SNc). See [27, 21] for overviews of the empirical data and theoretical models.

From an abstract perspective, the BG map contexts to actions. When a context-detecting matriosome fires, it inhibits a few downstream pallidal (GP and EP) neurons. In stark contrast to the striatum, the EP consists of low-fan-in neurons, most of which are constantly firing and thereby inhibiting their downstream counterparts in the thalamus [25]. When a striatal cell inhibits a pallidal neuron, this momentarily disinhibits the corresponding thalamic neuron, which then excites a cortical neuron, often in the PFC. The cortical excitation links back to the thalamus, creating a positive feedback loop that sustains the activity of both neurons, even though pallidal disinhibition may have ceased. Thus, the striatal-pallidal actor circuit momentarily gates in a response whose trace may reside in the working memory of the PFC for many seconds or minutes [25, 39].

Since the PFC is the highest level of motor control [18], its firing patterns often influence activity in the pre-motor (PMC) and motor (MC) cortices, while the MC sends signals to the muscles via the spinal cord. In addition, the sustained PFC activity provides further context for the next round(s) of striatal firing and pallidal inhibition that embody context detection and action selection, respectively. Via this recurrent looping, the basal ganglia execute high-level action sequences. The situation-action rules housed within the BG may comprise significant portions of our common sense understanding of body-environmental interactions, whether consciously or only subconsciously accessible.

The BG learns salient contexts via dopamine (DA) signals from the SNc, which influence the synaptic plasticity of regions onto which they impinge [31]. DA acts as a second messenger that strengthens and prolongs the response elicited by the primary messenger. For example, when a striatal neuron, S, is fired via converging inputs from the cortex, the primary messenger is the neurotransmitter from the axons of the cortical neurons (C) that recently fired. The immediate response of those S' dendrites (D) connected to the active axons is to transmit an action potential (AP) toward S's cell body. The summation of these D inputs will lead to S's production of a new AP. If dopamine enters these dendrites shortly after AP transmission, a series of chemical (and sometimes physical) changes occur which make those dendrites more likely to generate an AP (and a stronger one) the next time its upstream axon(s) produce neurotransmitters. Since the chemicals involved in this strengthening process are conserved, those dendrites that did not receive neurotransmitter may become less likely to fire an AP later on, even when neurotransmitters reach them. Thus, in the future, when the C neurons fire, the likelihood of S firing will have increased, whereas other cortical firing patterns will have less chance of stimulating S. In short, S has become a detector for the context represented by C. Without the dopamine infusion, S develops no bias toward C and may later fire on many diverse cortical patterns.

In unfamiliar situations, the SNc fires upon receiving stimulation from various limbic structures, such as the amygdala (the seat of emotions [33]), which triggers on painful or pleasurable experiences. The ensuing dopamine signal encourages the striatum to remember the context that elicited

those emotions - the stronger the emotion, the greater the learning bias. Due to the biochemical temporal dynamics [26], the striatal neurons that become biased (i.e., learn a context) are those that fired approximately 100 ms **prior** to the emotional response. Hence, the BG learns a context (C) that **predicts** the reinforcing situation (R).

Since dopamine signaling is diffuse, the matriosomes and striosomes in a striatal module are both stimulated to learn. Hence, the critic not only learns to predict important states, but assists in the learning of proper situation-action pairs by the actor circuit.

Figure 11 portrays the changes in a single context-action link in the basal ganglia as initiated by a reinforcement signal and modulated by an eligibility trace. The key difference between this and the situation in the cerebellum (Figure 7) involves the connections from the STN to the SNc. By strengthening these links, the basal ganglia allow contexts to directly predict rewards, as shown by the arrow from C3 to the internal reward signal at the bottom of Figure 11. This, in turn, allows earlier contexts (C2) to predict the same reward during later trials. Hypothetically speaking, a similar functionality in the cerebellum would require tuneable direct links from granular cells (the context detectors) to the inferior olive (the source of feedback signals).

Again, descriptions of the critical topological elements differ - see [28] for a review - but many experts name two paths from the cortex to SNc [21, 40]. The first, often called the *hyperdirect pathway*, bypasses the striatum and directly excites the subthalamic nucleus (STN), which, in turn, excites SNc. The second, termed the *direct pathway*, involves a strong inhibitory link from striatum to SNc. The hyperdirect pathway is quick but excites SNc for only a short period. Conversely, the direct pathway is slower, but it inhibits SNc for a much longer period.

This timing difference between excitation and inhibition enables these predictions (of reinforcement based on context) to regress backwards in time such that very early clues can prepare an organism for impending pleasure or pain. As pointed out by Joel et. al. [28], physiological evidence indicates that the excitatory and inhibitory signals to SNc cannot both come from the striatum, but more likely from the prefontal cortex (via the hyperdirect pathway) and the striatum, respectively.

Consider the simplified scenario of Figure 12, in which an animal experiences a temporal series of sensorimotor contexts: X, Y and Z before attaining the reinforcing state R. When this sequence first occurs, the attainment of R will be the first indicator of success, and the limbic reward signal will excite SNc and STN, causing dopamine-induced learning of context Z in both. The BG has learned a predictive rule that Z **eventually leads to** R: $Z \rightsquigarrow R$.

On a later trial, the occurrence of Z will initially stimulate SNc, and the ensuing dopamine will assist learning of a salient context immediately prior to Z, which is Y. Thus, another striosome is recruited to recognize a new context and notify SNc upon its detection. The system has thus learned $Y \rightsquigarrow Z \rightsquigarrow R$, but only **implicitly** in the sense that from Y, the agent knows the actions needed to attain Z, without necessarily knowing **of** Z and its relationship to Y. An outside observer might infer declarative knowledge of the $Y \rightsquigarrow Z \rightsquigarrow R$ sequence, but its true nature need only be procedural.

When R is attained, the limbic system still signals SNc, but by that time, Z's inhibitory signal has reached SNc, thereby preventing further dopamine dissemination. Neuroscientists [31, 27] have long

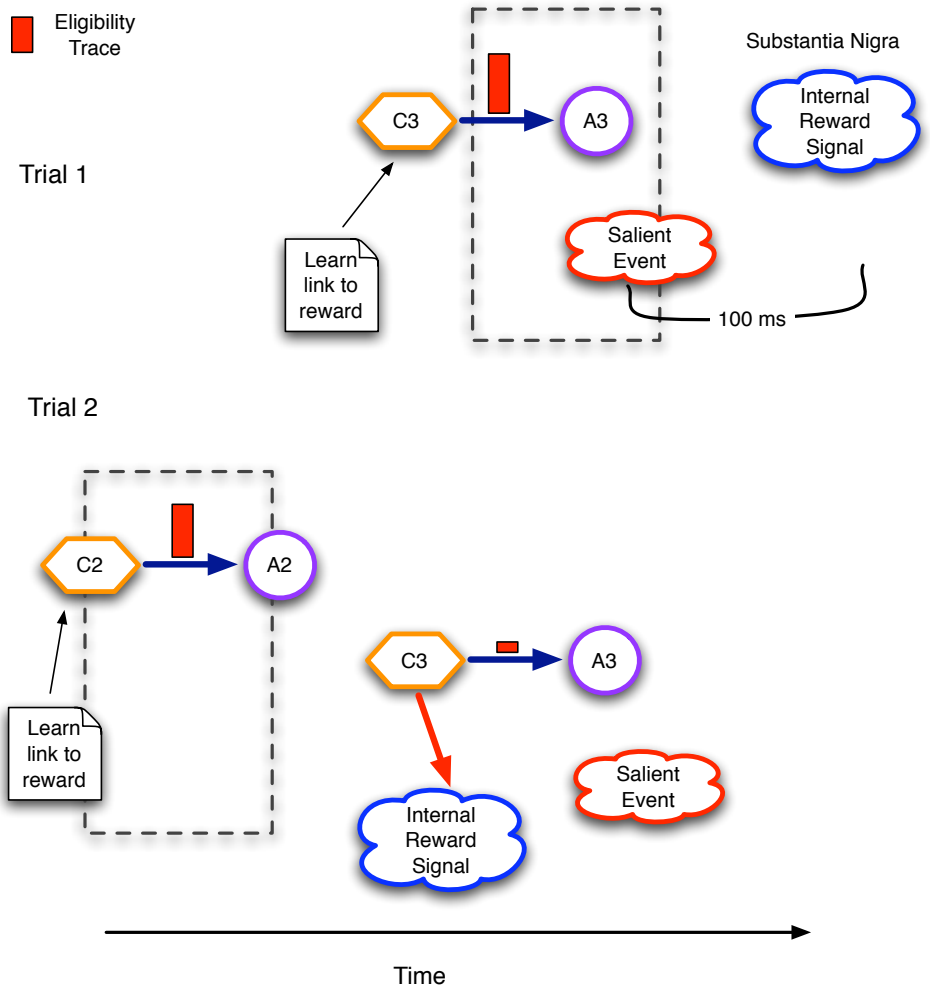

Figure 11: Temporal regression of predictive competence using eligibility traces and reinforcement signals in the basal ganglia. C2 and C3 are contexts detected by the striatum and STN, while A2 and A3 are accepted/chosen versions of C2 and C3, respectively.

known that dopamine signals only occur when a reinforcement is not expected, i.e., not predicted by a prior context. The temporal aspects of the biochemistry and the critic-circuit topology provide a clear explanation: when a context predicts a reward, its latent inhibition of SNc blocks subsequent attempts to stimulate it.

Finally, on a still later trial, the occurrence of Y will stimulate SNc, causing X to be encoded by a striosome and $X \rightsquigarrow Y \rightsquigarrow Z \rightsquigarrow R$ to be implicitly learned.

In the end, what predictive information does this model of the BG produce? An outside observer might infer that, indeed, the sequence $X \rightsquigarrow Y \rightsquigarrow Z \rightsquigarrow R$ is now explicit knowledge of the system. However, this model indicates that only the following associations have been acquired, in approximately the order shown:

1. $Z \rightsquigarrow R$, so state Z is a good state to attain.
2. When in Z, perform action a_z to attain the reward state R.
3. $Y \rightsquigarrow R$, so state Y is a good state to attain.
4. When in Y, perform action a_y, which just happens to put the system in state Z, although the system itself has no direct knowledge of this $Y \rightsquigarrow Z$ connection.
5. $X \rightsquigarrow R$, so state X is a good state to attain.
6. When in X, perform action a_x, which just happens to put the system in state Y.

Sequence learning is often posited as a key faculty of the basal ganglia [27, 40, 21], but the above description implies that the BG only learns how to **get** from one element of a sequence to the next, just as the outfielder's cerebellum helps him **get** to the ball. Actual knowledge of the links between sequence elements need not be explicitly represented anywhere in the system.

5 Predictive Topologies

Despite their many anatomical and (apparent) functional differences, the neural networks of the cerebellum and basal ganglia share several important features:

1. The entry points to each - granular and striatal cells, respectively - have high fan-in, strongly inhibit one another, and appear to serve as detectors of contexts with significant temporal extent.
2. The downstream pathways from these context cells are parallel tracts, with little integration.
3. Outputs have direct effects upon physical actions (cerebellum) or *planning* states that prepare the agent for action (basal ganglia).

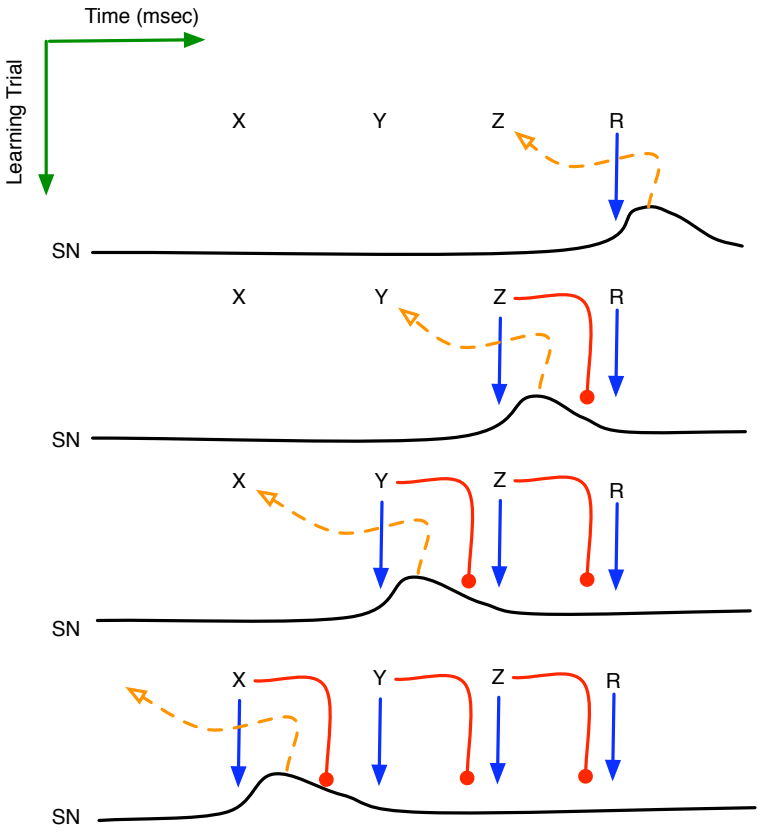

Figure 12: The implicit reinforcement learning of sequence $X \rightsquigarrow Y \rightsquigarrow Z \rightsquigarrow R$. Horizontal plots are the time series activation levels for the substantia nigra pars compacta (SNc). Solid arrows denote excitatory effects upon the SNc, while round heads represent the delayed inhibition. Dashed arrows portray the learning of a new context governed by the SNc's dopamine signal.

4. Synaptic biochemistry embodies *eligibility traces* with maximum values appearing along pathways that were active just prior (i.e. 100 msec) to the supervisory/reinforcement signal.

As described earlier, the tuning of the plethora context-sensitive *rules* is driven by error or reward signals, and modulated by eligibility traces. It yields procedurally predictive controllers that are adapted to the inherent sensory and motor delays of the organism.

The contextual input neurons for precedural prediction appear to **detect** complex multi-modal patterns, but intra-layer competition among these neurons and the parallel tracts of their efferents seem to preclude the actual **representation** of complex contexts in any manner that would support explicit reasoning about them. It permits contexts to serve as atomic triggers for action, but little else.

Declarative prediction requires different machinery, i.e. that which can associate two patterns, both of which have strong correlations with external states. The review, in the three upcoming sections, of several neural models of declarative prediction reveals a common connection scheme, which forms the basis of our Generic Declarative Prediction Network (GDPN).

Figure 13 sketches the basic GDPN framework, in which a set of sensory inputs map directly to a set of low-level detector neurons (A,B, and C). Above these lies a second (higher) level of neurons (W,X,Y and Z). This topology provides one, relatively simple, mechanism for learning temporal correlations among events, such as the fact that stimulus A is normally followed by stimulus B.

In this diagram, it is important to note that low-level inputs to higher levels occur proximally, i.e. close to the soma,, whereas top-down signals, such those from X to A, B and C, enter via distal dendrites. In general, this means that low-level signals can more easily *drive* the activity of their high-level neighbors, while the high-level signals have a much weaker effect upon lower levels.

Consider a situation in which stimulus A precedes stimulus B. The following series of events explains how the network learns to *predict* B when A occurs in future situations.

First, at time t1, stimulus A has a strong effect upon neuron A, via its proximal synapse. Neuron A then fires and sends *bottom-up* signals to W,X,Y and Z. At this level, as in all levels of the brain, neurons fire randomly, with probabilities depending upon their electrochemical properties and those of their surroundings. Assume that neuron X happens to fire during, or just after (i.e. within 100 msec of) neuron A. Assuming that synapse S1 is modifiable, the A-X firing coincidence will lead to a strengthening of S1, via standard Hebbian learning. In reality, several such high-level neurons may coactivate with A and have their proximal synapses modified as well.

When X fires, it sends signals horizontally and to both higher and lower levels. These latter *top-down* signals have a high fanout, impinging upon the distal dendrites of neurons A, B and C. Since entering distally, along unrefined synapses, these signals have only weak effects upon their respective soma, so at time t3, neurons B and C are receiving only mild stimulation. At this stage, we can metaphorically say that a) X is *waiting* for B and C (and thousands or millions of other low-level neurons) to fire, and b) X *hedges its bets* by investing equally and weakly in each potential outcome.

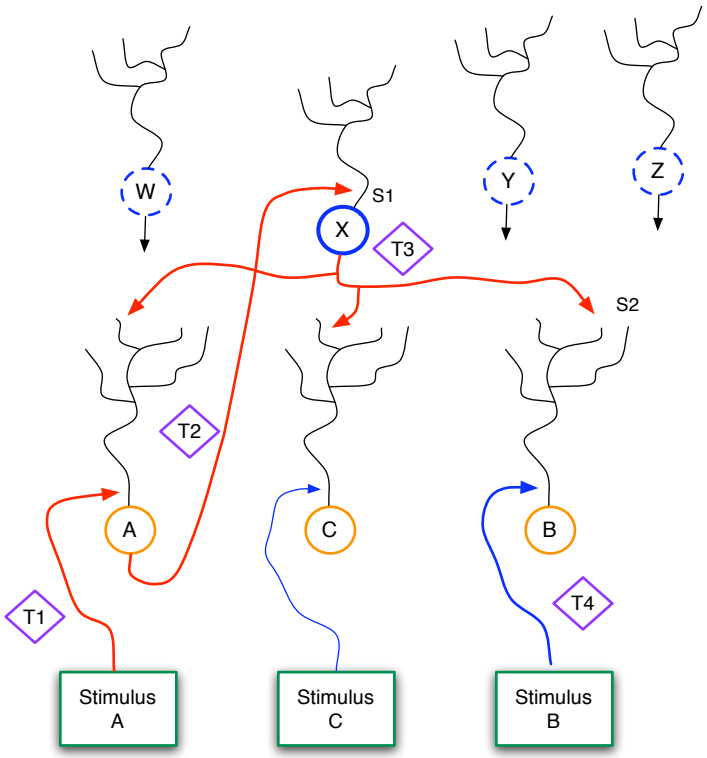

Figure 13: The Generic Declarative Prediction Network (GDPN). Neurons A, B and C serve as low-level detectors for stimuli A, B and C, while W-Z represent neurons at a higher level. Only the axonal projections from X are shown, though W, Y and Z have similar links to the lower level. The T1 - T4 diamonds represent time steps, while S1 and S2 denote important synapses, as further discussed in the text.

At time t4, when event B occurs, neuron B will fire hard due to the proximal stimulation from below. This will cause further bottom-up signalling, as when A fired, but the critical event for our current purposes involves the LTP that occurs at synapse S2. Previously, stimulation from X alone was not sufficient to fire neuron B. But if synapse S2 houses NMDA receptors, as do many dendrites throughout the brain, then the coincidence of B firing and S2 being (even mildly) active in the 100-msec time window prior to t4 will lead to strengthening of S2 [31]. Thus, in the future, the firing of X will send stronger signals across S2, possibly powerful enough to fire neuron B *without help* from stimulus B.

Through one or several A-then-B stimulation sequences, S1 and S2 can be modified to the point that an occurrence of stimulus A will fire neuron A, as before, but this will then directly cause X to fire, which in turn will fire neuron B. Thus, stimulus A will *predict* stimulus B.

Over time, neuron X ceases to hedge its bets and achieves a significant bias toward neuron B. This stems from both the strengthening of S2 and the weakening, via long-term depression (LTD), of X's synapses upon other low-level neurons, as explained below. Thus, X simply becomes a dedicated link between A and B. In a larger system, X and other neurons in its level, would become links between a pattern of activation in the lower level, P1, and a subsequent pattern, P2.

From the viewpoint of synaptic electrophysiology, the acquisition of declarative predictive models within this hierarchical network has a very plausible explanation based on bi-modal thresholding. As illustrated in Figure 14, Artola et al. [4] have shown that *weak* stimulation of neurons (in the visual cortex) leads to long-term depression (LTD) of the synapses that were active during this stimulation, while stronger stimulation incurs long-term potentiation (LTP) of the active synapses.

Three learning cases are worth considering with respect to a) a particular neuron, N, b) its low-level sensory inputs, S, with **proximal** synapses onto N, and c) its high-level predictive inputs, P, with **distal** synapses onto N.

First, if S is active but P is not, then the effects of S on N will produce a high enough firing rate in N to incite LTP of the S-to-N proximal synapse. Hence, N will learn to recognize certain low-level sensory patterns.

Second, if both P and S provide active inputs to N, then an even higher firing rate of N can be expected, so LTP of both the S-to-N and P-to-N synapses should ensue. In essence, the predictive and sensory patterns create a meeting point at N by tuning the synapses there to respond to the P-and-S conjunction. In fact, after repeated co-occurences of S and P, the synapses in N may strengthen to the point of responding to the P-or-S disjunction as well, in effect saying that it *trusts* the prediction P even in the absence of immediate sensory confirmation.

In the third case, when only P is active, the distal contacts of the P axons may only suffice to weakly stimulate N, thus leading to LTD: a weakening of the P-to-N synapses. Hence, future signals from P will not suffice to fire N, and thus P's predictions will not propagate through N in the absence of verification from S. In short, the system learns that P is not a good predictor of S.

These three scenarios provide a very simple mechanism for the synaptic tuning that gradually converts a blanket of bet-hedging anticipatory links into a few dedicated connections between

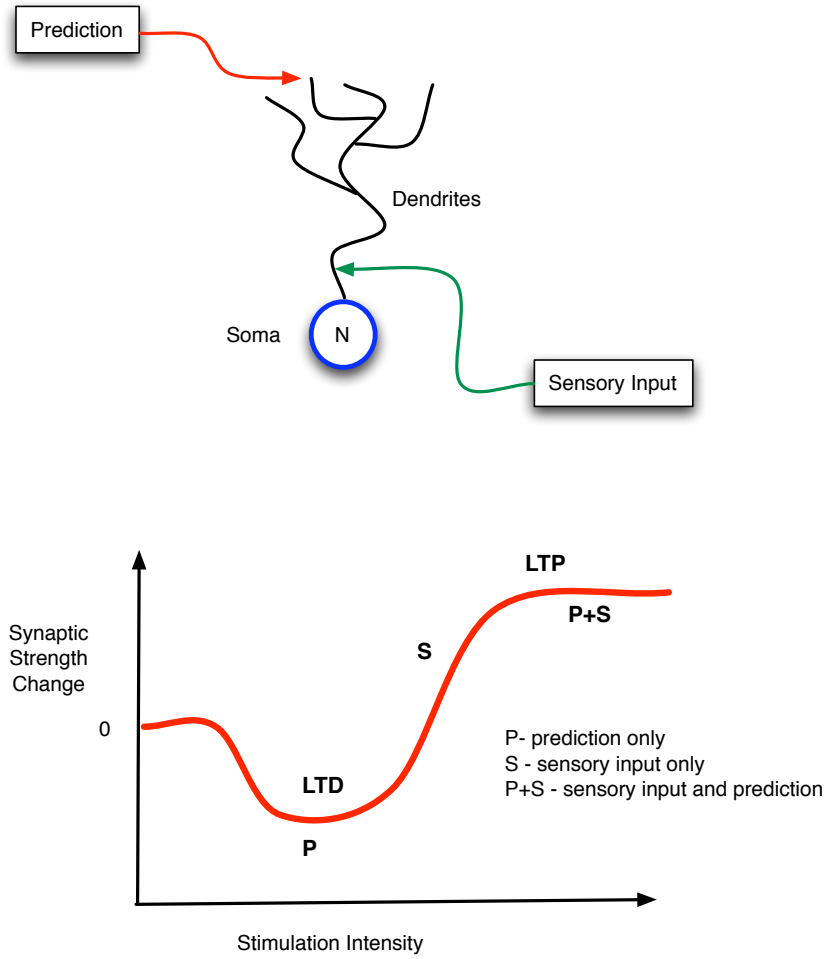

Figure 14: (Above) Top-down, predictive, distal and bottom-up, sensory, proximal inputs to a neuron. (Below) Changes in synaptic strength as a function of post-synaptic stimulation intensity.

associated neural patterns.

In the following three sections, the neocortex, hippocampus and thalamocortical system are analyzed with respect to prediction, with each showing clear evidence of the bet hedging and refinement so characteristic of the GDPN.

6 Declarative Prediction in the Neocortex

The neocortex provides a very straightforward instantiation of the GDPN, with individual neurons replaced by cortical columns. The neocortex is the thin outer surface of the cerebrum, only a few millimeters in thickness and composed of 6 cell layers. These cells appear to be grouped into vertical columns [36] which are often viewed as processing modules [23, 18]. In mammals, it is convenient, and reasonably accurate, to view cortical columns near the back of the head as processors of primitive sensory information, particularly visual, while the more anterior columns process and represent higher-level concepts [18, 31] Under this abstraction, bottom-up sensory-driven interpretation processes involve cascades of activation patterns moving from the back to the front of the neocortex. Conversely, top-down, memory-biased processing moves front to back, as shown in Figure 15.

Within each cortical column, the neurons capable of emitting the strongest and most influential signals to other columns are large pyramidals with cell bodies residing in the lower layers, 5 and 6 [36]. Axons emanating from these two layers tend to synapse on lower-level cortical columns [23, 44], particularly motor neurons, and the thalamus, a subcortical structure known as a key relay station for sensory signals and believed to play a key role in integrating information [45]. The dendrites of these large pyramidal cells extend up to layer 1, which is essentially a mat of axons coming from both higher-level cortical columns and subcortical structures such as the thalamus, hippocampus and basal ganglia. Signals from layer 1 reach the large pyramidals either directly, via the latter's dendrites, or indirectly via small neurons in layers 2 and 3.

Incoming axons from other columns can synapse with the large pyramidals at just about any point along their dendrites, from layer 4 up to layer 1. Proximal synapses (i.e., those close to the cell body, such as in layer 4) typically have a stronger effect upon the pyramidal's firing activity than will distal synapses at layer 1 or relay pathways through layers 2 and 3 [36].

Of critical importance to understanding predictive-model learning in the neocortex is the fact that the axonal inputs from lower-level (i.e., posterior) cortical columns tend to enter higher-level cortical columns in layer 4, with some synapses also forming at layer 6 [23, 36, 31] . Thus, the low-level inputs form synapses near the cell bodies of the large pyramidals, whereas the inputs from higher-level (i.e., anterior) columns normally connect via layer 1. The immediate implication is that low-level sensory signals, which essentially represent the organism's current sensation of *reality*, have a stronger influence upon a cortical column than do the high-level thoughts (i.e. predictions) that often bias perception.

Hawkins emphasizes the branching factors of bottom-up versus top-down pathways [23] in the neocortex. In general, the number of cortical columns decreases in moving up the processing

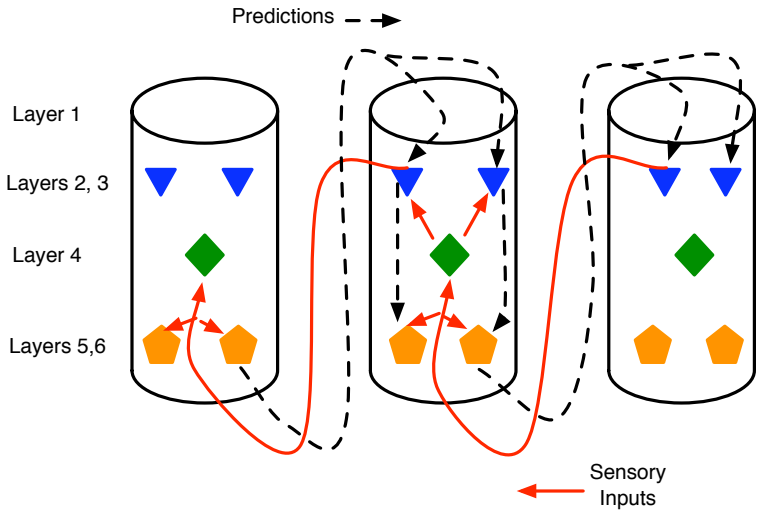

Figure 15: Abstract view of cortical columns and top-down versus bottom-up information flow. Bottom-up flow (solid lines) goes from layers 2 and 3 of the sending column to layer 4 of the higher-level column, but with additional synapses onto the large pyramidals in layers 5 and 6 (pentagons). In the top-down pathway (dotted line), large pyramidals output to layer 1 of lower-level columns, with the signal eventually reaching layers 5 and 6 via either the layer 2-3 relays or directly via long dendrites from the large pyramidals.

hierarchy. Hence, bottom-up pathways appear convergent in that many primitive sensory neurons feed into the same higher-level neurons. Likewise, top-down pathways appear divergent, with one associative neuron signalling many lower-level columns. Hence, a *predictive* high-level neuron, P, may initially supply many lower-level neurons, in effect encoding a bet-hedging expectation that many different sensory patterns will be active when P fires. Through experience, many of these divergent connections will be pruned as their synapses weaken due to unfulfilled expectations and the resulting LTD.

Learning of a temporal correlation between stimulus states A and B follows the basic GDPN protocol in the cortex. Low-level cortical columns serve as detectors for specific features and (in moving up the hierarchy) combinations of features. Assuming column C_A detects stimulus A, its layer 2-3 neurons will fire, sending signals to the proximal dendrites of layer-5 neurons in higher-level columns. Any of those neurons that randomly fire in that same time frame will thus have their synapses from C_A strengthened. Assume X is one such layer-5 pyramidal. It will send divergent axons to layer 1 of many lower-level columns, thus hedging all bets and waiting for the next stimulus detector to fire. When stimulus B arrives, its detector column, C_B, activates (i.e. its layer-5 pyramidals fire) and the links between X and layer 1 of C_B are enhanced. After one or several occurrences of A-then-B, C_A will fire X, which will then fire C_B, even in the absence of stimulus B.

Another key aspect of this model concerns temporal relationships. Assume an initial sensory scenario, S_1 at time t_1. This will propagate up the cortical hierarchy but will also evoke top-down predictive signals in layers that house expectations associated with S_1. For these expectations to

propagate further down the hierarchy, they will need to match new sensory data (thereby firing the layer-5 pyramidals that send signals to lower levels). Due to the inherent time lag in neural signalling pathways, the predictions related to S_1 will need to match up with ascending sensory data from situation S_2, which occurred at a slightly later time, t_2. Hence, the natural time delays in the system will insure the learning of predictions between states at time t_1 and time t_2, thus neatly corresponding with our general conception of *causal* knowledge.

Over time, a network of cortical columns with the topology and learning mechanisms described above will adapt its synaptic strengths to form a system that can both interpret sensory data and use top-down expectations to a) complete partial sensory states, and b) predict future states. In fact, under the general view of a context as an amalgamation of related information with some degree of temporal scope, the completion of a partial context could essentially involve the recall of future states associated with states closer to the present. Hence, many acts of pattern completion embody prediction as well, and the cortex seems particularly adept at this task.

One final aspect of Hawkins' theory of cortical function [23] deserves mention. He postulates that sensory inputs will propagate up the cortical hierarchy until they reach a level at which expectations/predictions (presumably based on previous inputs and/or brain states) match the *reality* embodied in the current inputs. The correspondence between prediction and reality will then block further upward progress. In short, *surprise* travels upwards until it ceases to be unexpected. If it confounds all predictions and reaches the higher associational areas, it then feeds into the hippocampus and (assuming significant emotional content) spurs learning, which helps insure that it will not be such a surprise on its next occurrence. Our discussion of the thalamocortical loop (in a later section) sheds further light on these ideas.

7 Declarative Prediction in the Hippocampus

The hippocampus (HC) resides in the temporal lobe and is commonly viewed as the seat of long-term memory *formation*, but not necessarily of storage [46, 19, 35]. Anatomically, it resembles a horn [3], as shown at the top of Figure 16. A wide variety of high-level associational areas send inputs to HC, most of which are funneled through the entorhinal cortex (EC). As implied at the bottom of Figure 16, the HC and surrounding regions employ high convergence to compress information between the neocortex and area CA3, and a complementary expansion (via divergence) on the return path through CA1 and Subiculum [44, 3]. The topology of the HC proper is a main loop with several shortcuts from the EC to CA3 and CA1 [3].

Only CA3 contains extensive recurrence, with each neuron connected to 1 to 4% of the others [44, 3]. This indicates that CA3 performs associative learning by standard Hebbian means: *neurons that fire together wire together* [24]. The high convergence from a diverse array of neocortical areas onto CA3 hints of the holistic nature of these patterns. In rats, individual neurons in CA3 and CA1 are known as place cells [7], since they fire only when the animal is at a particular location, while in monkeys, they are called view cells, since they fire when the primate merely looks at such a location [44]. Discovery of these cells has motivated a plethora of artificial neural network (ANN) models of HC-based navigation, as summarized in [7, 41]. Many of these involve implicit predictive knowledge in CA3 and CA1, wherein place cells fire before the animal arrives at the corresponding

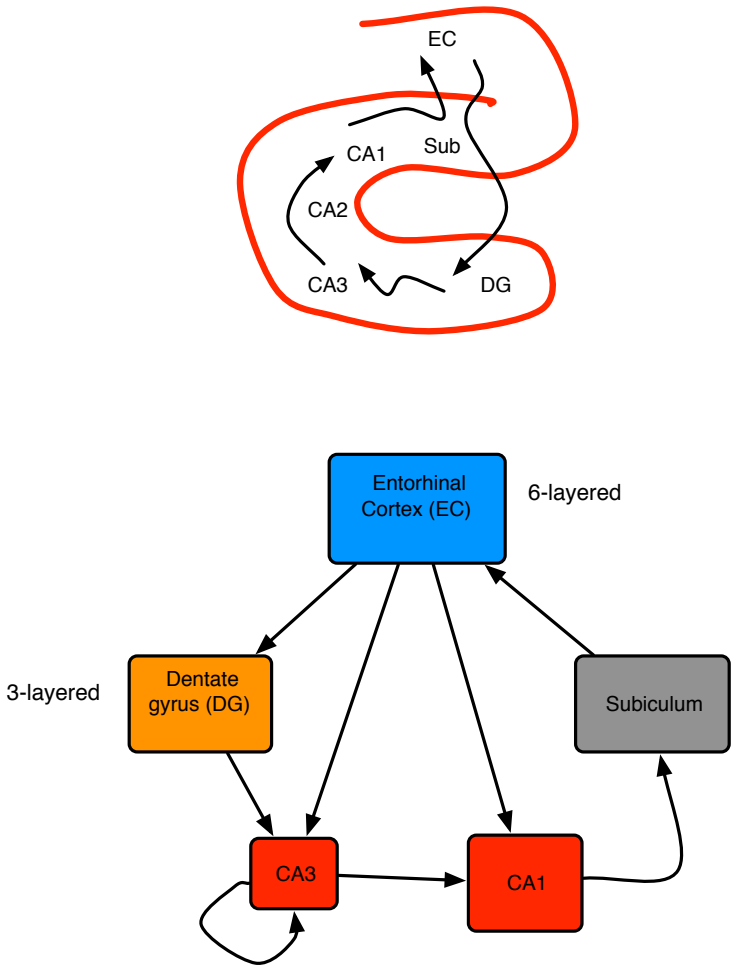

Figure 16: (Above) Basic anatomy of the mammalian hippocampus. (Below) Primary hippocampal areas and their connectivity. Box dimensions roughly illustrate relative sizes of neural populations in each area; all connections are excitatory. Based on pictures and diagrams in [3, 37, 44].

location.

In one of the more popular models, Burgess et al. [37] propose a layer of goal cells (possibly in the subiculum) that receive inputs from many CA1 place cells. Goals represent very salient locations, often those involving reinforcements. As a simulated rat moves about, the goal cells fire at frequencies correlated with the rat's proximity to their target fields, so navigation is achieved by choosing movements that increase the firing of focal goal cells.

Another interesting variant [29] posits CA3 as the site of predicted situations and CA1 as the site of real situations (via direct inputs from EC). Mismatches between the two drive learning in CA3 and thus improve the accuracy of future predictions.

The model that we now consider in detail is that of Wallenstein et. al. [51]. It provides an intricate mechanism for predictive learning that is a) centered largely in CA3, b) able to connect events separated by significant temporal delays, and c) quite similar to our GDPN.

Their model utilizes the key differences between proximal and distal contacts to CA3 pyramidal cells. As shown in Figure 17, afferents from the dentate gyrus have proximal targets on CA3 neurons, while CA3's recurrent collaterals have distal synapses. Hence, DG inputs, when active, tend to drive CA3 pyramidals, enforcing external conditions (via EC and its cortical afferents) upon them. However, the authors cite the well-known 4-10 Hz theta oscillations [31, 8] as a simple switching mechanism between afferent-driven and collateral-driven CA3 activation. Hence, CA3 can receive DG inputs during one phase of a theta wave and exploit internal computations (via recurrence) during the opposite phase.

The following detailed example, based on the Wallenstein et. al. model, reveals very clear GDPN dynamics within CA3, but without the need for a hierarchy of layers.

Assume that the system will learn the association between two temporally distinct events, E1 and E2, where E1 involves two sensory features, 1 and 3, while E2 involves features 2 and 6. In the initial phase, shown in Figure 18, features 1 and 3 enter CA3 via DG. Each DG granular cell projects to a small number (around 15) CA3 cells [3]; this is abstractly depicted as a 1-1 relationship in our figures, where the DG is drawn with 6 output *ports*.

When neurons 1 and 3 fire, they send signals to all neurons with which they have recurrent connections. In the hippocampus, this would be 1-5% of all CA3 neurons. Hence, many CA3 neurons receive weak, distal stimulation.

As is common in CA3 and many other brain areas, neurons randomly burst (i.e. produce action potentials) at all times, although normally at lower rates than those of neurons that receive many inputs. In the upper left of Figure 18, neurons 4 and 5 happen to be bursting within approximately the same 100 msec window as the arrival of stimuli 1 and 3. Hence, neurons 4 and 5 fire while receiving distal inputs from 1 and 3. The NMDA receptors on these distal dendrites will then detect this coincidence (of presynaptic input and postsynaptic firing) and strengthen the distal synapses. The connection between 1-3 and 4-5 would thereby strengthen.

Neurons 4 and 5 are referred to as *context neurons*. As shown below, they form the *glue* between

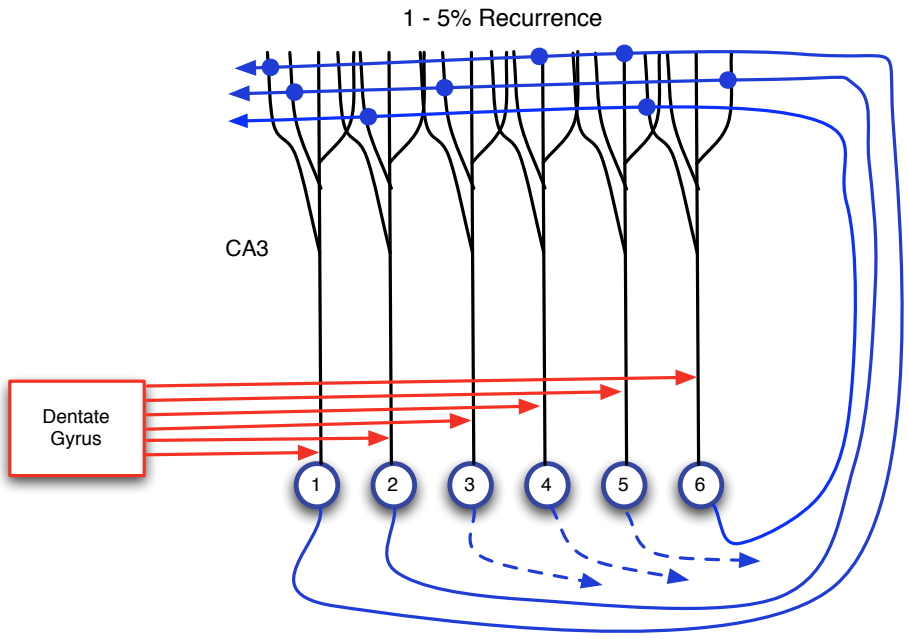

Figure 17: Two main sources of input to CA3 neurons: proximal connections from the dentate gyrus and distal, recurrent inputs from CA3 itself. Each CA3 neuron has recurrent links to 1-5% of the others [44, 3].

E1 and E2, even when many seconds (or minutes) separate the two events.

At this point, CA3 has learned the association between 1-3 and this context. It must now learn to connect the context to E2. It is therefore important that the context remains active while simultaneously sending out *monitoring* signals to other CA3 neurons, in anticipation of future inputs from DG. The upper right of Figure 18 shows this state, wherein the context sends distal signals to many neurons, including 4 and 5. Note that since 4 and 5 are firing hard during the arrival of these self-monitoring inputs, their distal synapses will strengthen, as shown on the bottom left of the figure, making the context an autocatalytic activation pattern, and thus one that can remain active for long periods of time. Neurons 1 and 3 are now inactive, since a) event E1 is completed, and b) the system experiences the second half of a theta oscillation, the half in which internal dynamics dominates extrinsic influences.

After a delay of seconds or even minutes, event E2 occurs, sending signals for features 2 and 6 into CA3 via EC and DG. These proximal inputs force neurons 2 and 6 to fire, but, as shown on the bottom left of Figure 18, the key learning now occurs on the *distal* dendrites to 2 and 6, which were active during monitoring. This LTP strengthens the links between context neurons 4-5 and neurons 2 and 6. Although not discussed in [51], the distal links from context to neurons 1 and 3 could weaken by LTD, since the monitoring inputs did not coincide with postsynaptic activity in those 2 neurons. Note that all of this assumes that the context and, more importantly, its monitoring signals, remain active during the E1-E2 delay. Contexts are able to achieve this prolonged activation if they consist of enough neurons and can quickly strengthen autocatalytic connections.

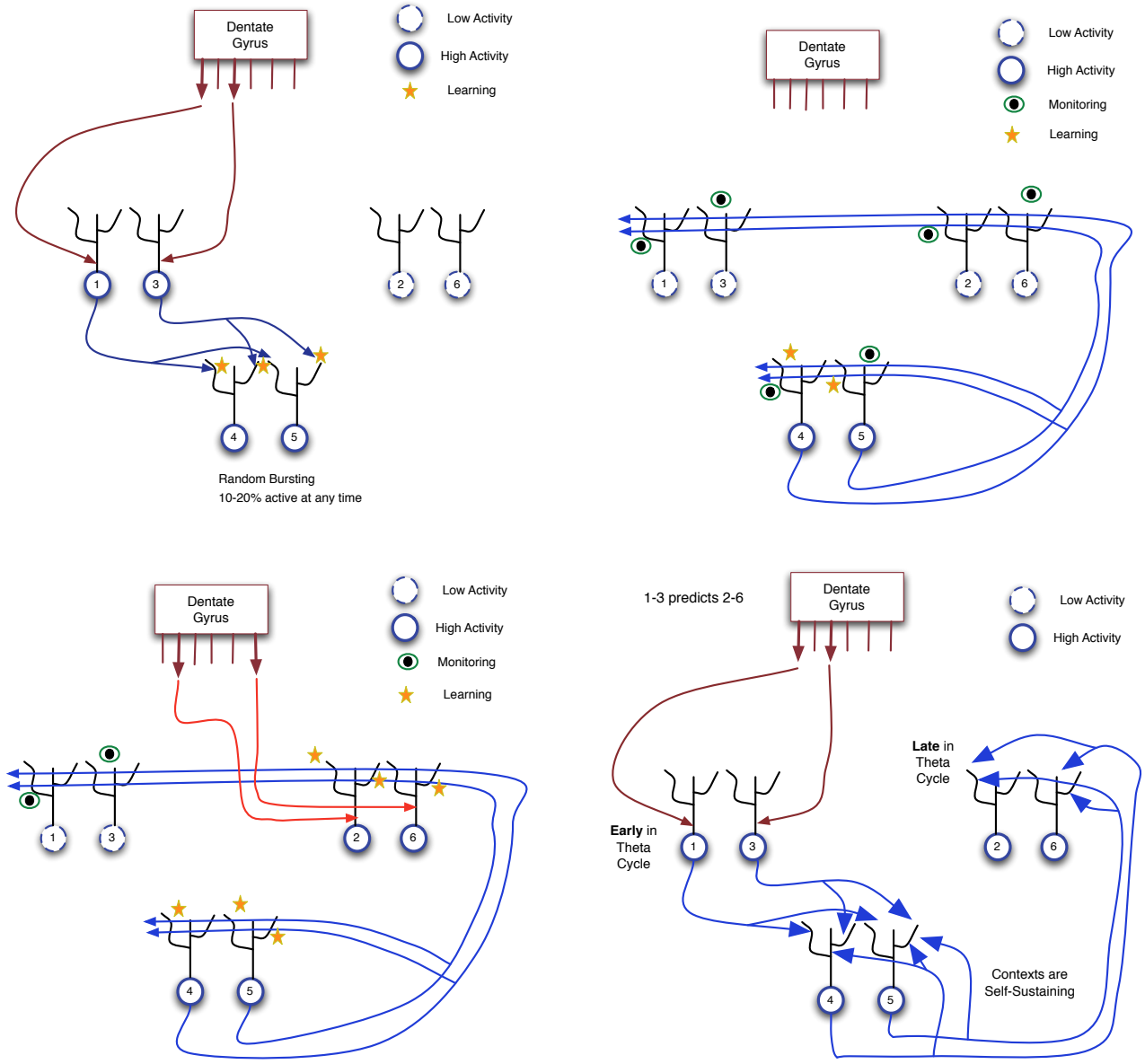

Figure 18: Learning a predictive association in the hippocampus. In these diagrams, a) large arrows extending from DG indicate active output *ports*, and b) CA3 neurons are separated into 3 groups for explanatory purposes only. (Upper left) Input of pattern 1-3 from the dentate gyrus via proximal synapses onto CA3 pyramidals. (Upper right) Context neurons send distal monitoring signals to their recurrent connections and to themselves. This causes weak activation of distal dendrites throughout CA3 and incites autocatalytic learning within the 4-5 context. (Lower left) distal monitoring inputs from the 4-5 context coincide with DG-forced firing of neurons 2 and 6 (as a consequence of event E2). This leads to LTP at the distal synapses of neurons 2 and 6. (Lower right) Using the learned association between events E1 and E2: when E1 occurs, 4-5 context neurons and then neurons 2 and 6 fire, thus predicting the future occurrence of E2.

As shown by the large arrows on the bottom right of Figure 18, the links from E1 detectors to context and then to E2 detectors have all been enhanced. Thus, when E1 occurs, its distal contacts suffice to fire the context neurons, and their distal contacts can then a) keep the context active, and b) excite neurons 2 and 6. This constitutes a *prediction* that E2 will occur.

Since inputs to HC come from high-level associational cortices, they need not have direct links to immediate sensory and proprioceptive activity. For example, sequences of activation states in these cortices may represent different steps in a reasoning process. Hence, the predictive learning in CA3 can also encompass associations between any mental states, but particularly when they have high emotional content, since emotions trigger neuromodulators that enhance activation and learning in areas such as CA3 [33]. For instance, when pondering the events of a mystery novel, thoughts of the butler may initiate a chain of inferences ending in the conclusion that he must be the murderer. The emotional content of this deduction may lead to a strong link in CA3 between butler and murderer, with many of the intervening (emotion-free) inferences eventually being forgotten.

8 Declarative Prediction in the Thalamocortical Loop

Although neuroscientists previously viewed the thalamus as merely a transfer point for signals from sensory periphery to cortex, it is now known that only about 20% of thalamic inputs are ascending pathways (i.e. upward along the spinal cord) from the senses, while most of its remaining afferents descend from the cortex [31, 45]. Hence, the thalamus is now seen as a key component and integrator of several brain functions, including predictive/sequential learning [43].

As shown in figure 19, the thalamus is divided into several nucleii, each of which relays sensory signals to and receives top-down feedback from a specific cortical region, such as auditory, visual or multi-modal cortices. Two key neuron types in each nucleus are the *core* and *matrix* cells. The former are large and have sparse, topographic connections to layer-4 neurons in the corresponding cortical area, while the latter are smaller and send diffuse projections to the layer-1 dendritic mats of many cortical columns within a region.

Inputs to core cells come from a) ascending sensory pathways, b) cortical feedback, and c) the nucleus reticularis, a strong inhibitory module. Conversely, matrix cells receive the vast majority of their afferents from layer 5 of the cortex [45, 43]. Thus, they are important for thalamocortical feedback but not for the initial sensory relay.

The nucleus reticularis (NR) maps topographically to core cells, with excited NR cells strongly inhibiting their core counterparts. As shown in Figure 20, descending pathways from cortical layer 6 link to the corresponding NR and core cells. Since NR neurons have proximal links to core neurons, their inhibitory effect is very strong, tending to override sensory input from ascending pathways. As pointed out by Granger [20], the chemistry of excitatory glutamate versus inhibitory GABA synaptic potentials is such that the former persist for a mere 15-20 msec, while inhibition last from 80-150 msec. This, combined with the transmission pathways of Figure 20, implies that a sensory stimulus will briefly excite a core thalamus cell, causing further activation of the corresponding cortical column; but feedback from this column, via NR, will then silence the core neuron for a considerably longer period.

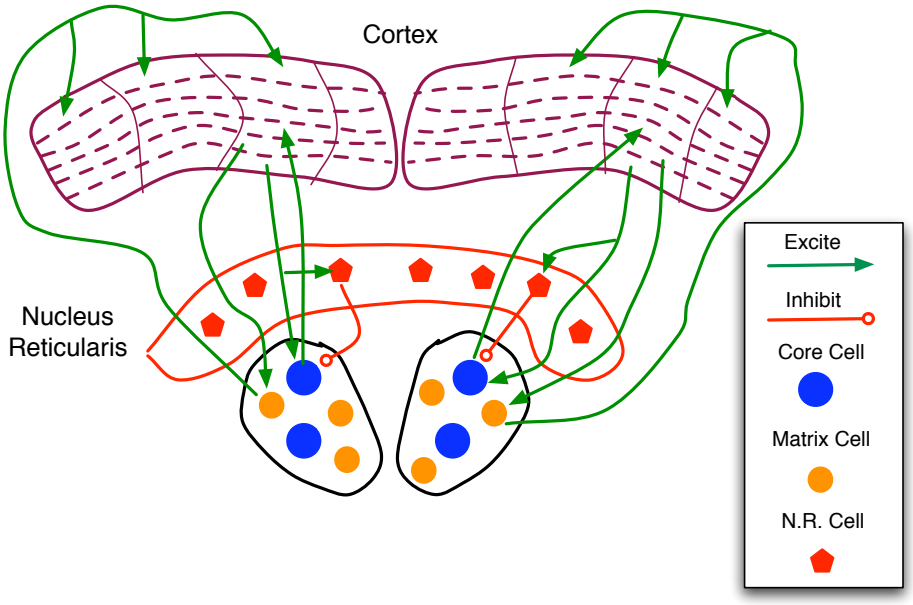

Figure 19: Basic anatomy of the thalamocortical system, based on similar drawings in [45]. The cortex is divided into regions, vertical columns, and the 6 well-known horizontal layers. Each core cell maps to layer 4 of a specific cortical column and receives feedback from layer 6. Matrix cells receive cortical afferents from layer 5 and send signals to the layer-1 dendritic mats of many columns.

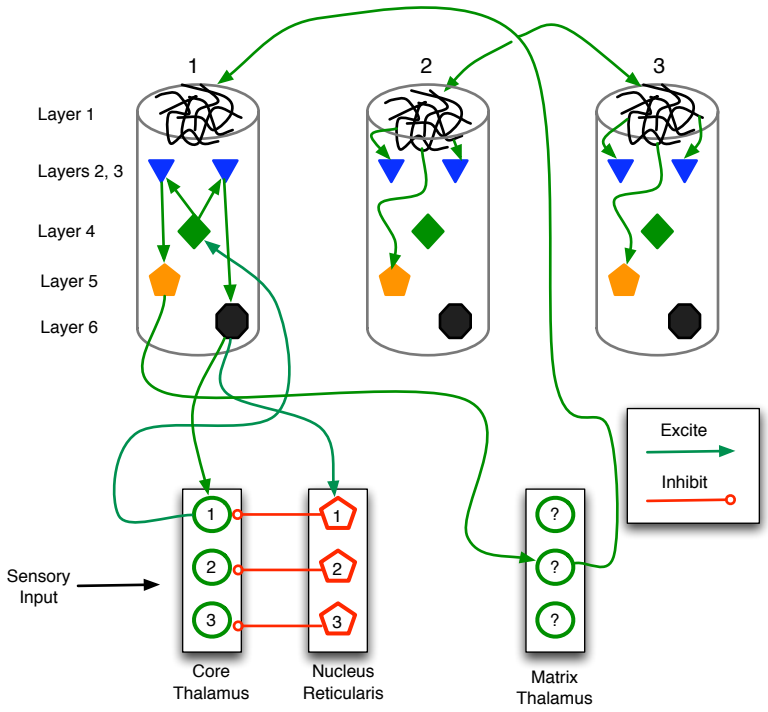

Figure 20: Sketch of thalamocortical loops for three columns of a cortical region. Core, matrix and nucleus reticularis (NR) cells are separated into three modules for illustrative purposes. The main types (but not all instances) of connections are drawn. Within a column, the key connections are a) entry layer 4 links to layers 2 and 3, b) layer 2-3 stimulation of layers 5 and 6, and c) layer 1 excitation of layers 2, 3 and 5. In general, the column is considered *active* when layer 5 and 6 neurons fire at high frequency.

The models of Rodriguez et. al.[43] and Granger [20] provide excellent insights into the emergence of predictive associations within thalamocortical loops. As sketched in Figure 21, two stimuli are linked via thalamic and cortical activation, monitoring and learning. Figure 21d then illustrates the recall process wherein stimulus 1 leads to the expectation of stimulus 2.

Initially (Figure 21a), stimulus 1 enters the core thalamus via ascending pathways, firing core neuron 1, which then relays to layer 4 of column 1. This excites layers 2 and 3, which then stimulate layers 5 and 6. Layer 6 then sends positive feedback to both the original core neuron and its antagonistic NR cell, while layer 5 excites matrix thalamus neurons. Those matrix cells that happen to randomly burst within this time frame (drawn as solid circles) accrue stronger synapses via LTP.

Next, (Figure 21b), the active matrix cell (now linked to column 1 via LTP) sends *monitoring* signals to the layer-1 dendritic mats of many cortical columns. In addition, the active NR cell (pentagon 1) inhibits its core thalamus counterpart (circle 1), despite the potentially continuous presence of stimulus 1. This gradually de-activates cortical column 1.

Stimulus 2 now activates the core thalamus (Figure 21c), which relays excitation to column 2. The coincident activation of deep-layer neurons and layer-1 dendrites lead to NMDA-driven LTP in column 2, thus forging a strong link between matrix neuron 1 and column 2. Inhibition of core neuron 1 remains in effect, so even if stimuli 1 and 2 are co-present, only 2 has an effect at this stage.

Figure 21d illustrates the recall of the newly-formed predictive association. The occurrence of stimulus 1 fires core neuron 1, which stimulates cortical column 1. In turn, this excites matrix neuron 1, which then drives column-2 activity. The firing of layer 5 and 6 neurons in column 2 embodies *recall* of stimulus 2 as a consequence of stimulus 1.

Although they serve our purpose well, these models were originally designed to show the role of thalamocortical activity in perceptual processing. In that case, stimulus 1 represents the set of most salient features in a percept, typically known as the *invariants*: the most common features among instances of a perceptual class. For example, in the elephant category, these might include size and the presence of a trunk. Once recognized, the effects of these features should be damped such that other, more specific, attributes can be processed in order to perform fine-grained discrimination, when necessary. The negative feedback loop from cortical level 6 to the nucleus reticularis to core thalamus performs this function. It allows stimulus 2 (which represents a set of secondary features) to dominate processing for a few fractions of a second.

The link formed between stimulus 1 and stimulus 2 is indeed anticipatory, but for this classification task, the prediction is between one stage of perceptual processing (represented by a neural activity pattern) and another, with no direct connection to temporally sequential events in the external world: the trunk and large body of the elephant do not actually enter our visual field before the other features, they are merely **processed** first.

Clearly, a sensorimotor system benefits from this incremental processing, since immediate reactions, when necessary, can be mobilized on the basis of only the most salient features, and thus with minimal delay. Thus, evolution would favor such an approach over an all-perception-before-action scheme reminiscent of early artificial intelligence and robotics [42]. More general predictive

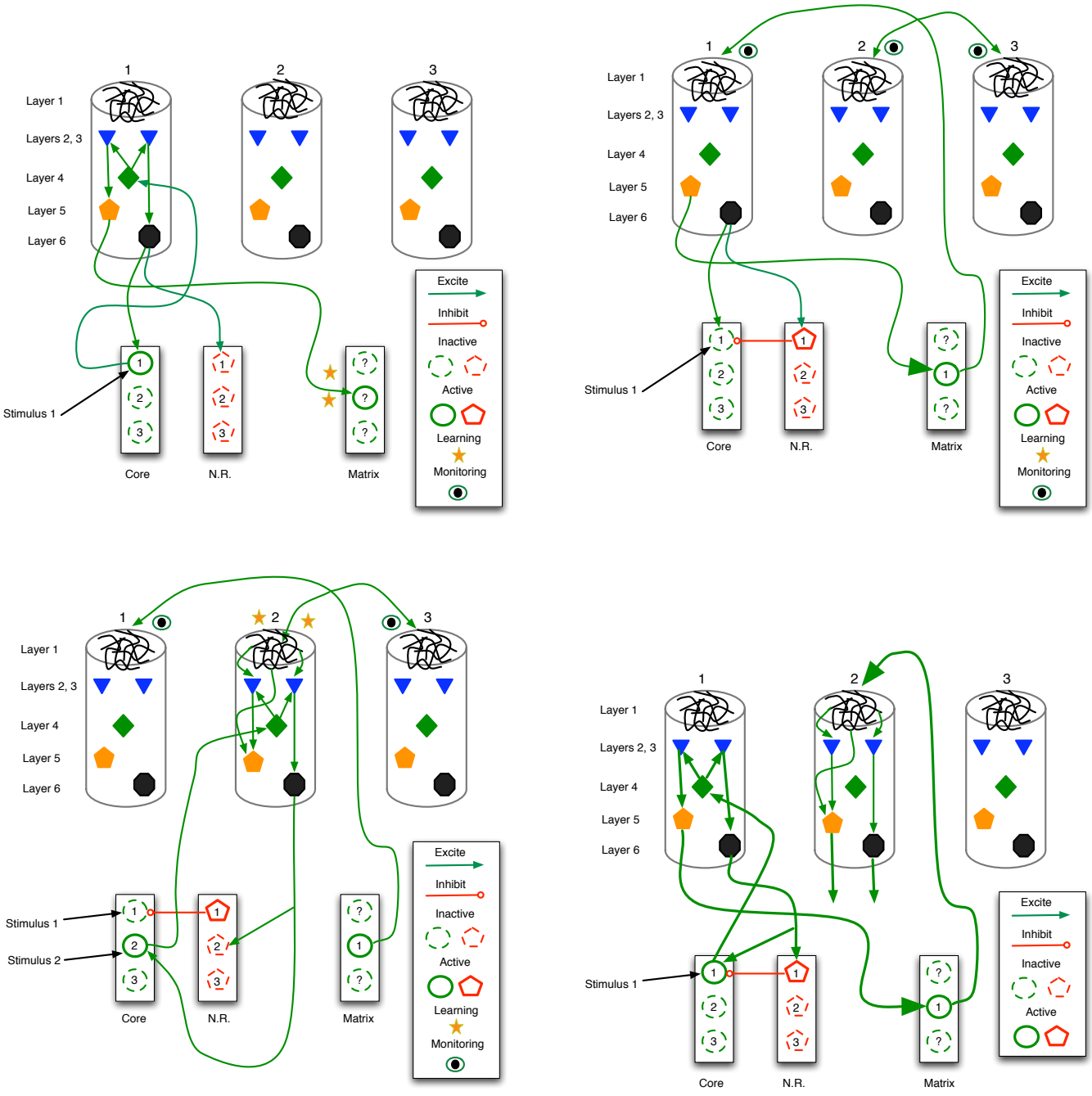

Figure 21: Learning of a predictive sequence in the thalamocortical circuit. Only the main active connections are shown. (a)(upper left) Entry of stimulus 1 stimulates the corresponding core thalamus neuron and cortical column, as well as a random matrix thalamus neuron. (b) (upper right) Matrix stimulation leads to distal monitoring of diverse cortical columns. (c) (lower left) Entry of stimulus 2 adds proximal stimulation to (already distally stimulated) cortical column 2, producing distal synaptic strengthening. (d)(lower right) In future trials, stimulus 1 leads to excitation of column 1, then 2, thus embodying a prediction of stimulus 2.

knowledge, involving temporally-contiguous real-world events, could then capitalize on the same basic thalamocortical system. The only significant difference is in Figure 21c, where only stimulus 2 would appear, not both 1 and 2. In short, our predictive capacities stem directly from our ability to incrementally classify, or vice versa.

Finally, the thalamocortical model helps explain Hawkins' theory [23] that surprise propagates up the cortical hierarchy. As in Figure 21d, assume that stimulus 1 has *primed* column 2, which now fully *expects to see* stimulus 2. If it arrives, it does so via the core thalamus and then level 4 of column 2. This further stimulates layers 2-3 and then layers 5 and 6. Since layer 6 fires very hard, this sends strong signals to NR, which swiftly inhibits core neuron 2, thus removing the excitatory input to level 4, which in turn removes a strong excitatory input to layers 2 and 3. Since layers 2 and 3 are the main output ports for propagation up the cortical hierarchy, further ascent is blocked or significantly reduced.

Conversely, if stimulus 2 arrives in column 2 without the simultaneous presence of expectation-driven firing (beginning in layer 1), then layer 6 may not fire hard enough to completely inhibit core neuron 2. Thus, column 2 would remain active, sending signals up and down the cortical hierarchy.

This conflicts somewhat with our original description of stage 1 of thalamocortical predictive learning in Figures 21a and 21b, in which the NR neurons are stimulated (and the corresponding core neurons blocked). However, the dynamics are probably more continuous than discrete, such that NR excitation reduces core and then cortical activity *to varying degrees*, depending upon NR firing levels.

This view hints of a learning model wherein, upon seeing an object for the first time, cortical columns corresponding to the most salient features fire, but at levels too weak to fully stimulate NR. Hence, secondary features go largely unnoticed during the early trials as primary features dominate cortical activity. However, with repeated presentation of the object, the salient columns begin to fire harder, due to synapses strengthened by simple Hebbian means during earlier trials. Thus, NR becomes a more active player, helping to shut down salient core neurons and columns, thereby allowing access to secondary and eventually tertiary stimuli. In short, we learn the most important features first and require repeated trials to fully absorb the details. Stated differently, the learning of a feature set S entails the ability to make predictions about S, and as these predictions become more accurate, less processing time is required (and used) for S, and more can be devoted to other features.

Along the same lines, a predictive sequence cannot be learned in its entirety, but piecemeal, with links between temporally later events forming only after earlier segments have become familiar.

9 General Features of Predictive Circuitry

A high-level topological comparison indicates the differential predictive functionality of the procedural and declarative circuits. Under the reasonably standard view that our explicit representations (i.e., those that we can attend to, reason about, etc.) consist of a good deal perceptual information

(i.e., that based on past or present sensations), it makes sense that a predictive association between two such representations involves connections within the more perceptually-oriented areas of the brain. If these patterns represent similar world states, then, due to the topological nature of much of the brain, they probably reside near one another and may even have many shared active neurons. A birds-eye view of these two patterns and the active synapses that embody the predictive link would indicate a tight mesh of intra-layer and intra-region connections: high recurrence.

Conversely, the cerebellum and basal ganglia have little recurrence. They exhibit parallel tracts that map sensorimotor contexts to actions. The basal ganglia appears slightly more cognitive in that it may be *gating* sensorimotor contexts - that embody perceptions plus intended actions - into prefrontal areas, where they can then influence future motor acts and reasoning.

The procedurally predictive areas are therefore hard pressed to link representation R1 for world-state 1 to R2, the representation for world-state 2. However, they can learn to map R1 to actions and action plans that are *appropriate* for world-state 2. And in a fast-moving world, this is often all that is required, or permitted.

As shown in Figure 22, a key difference between the procedural and declarative predictive mechanisms involves space. In the procedural areas, activation patterns move along parallel tracks, and the learning initiated by a salient event (such as an error or reinforcement signal) targets synapses **between** one area and the next. For example, in the cerebellum, the competing contexts stem from diverse brain regions whose axons converge upon the granular cells, while the winning contexts are those whose granular cells can maintain activity in the face of inhibitory signals from other granular cells. By linking the granular cells to the Purkinje cells, the parallel fibers provide a distinct spatial location for the transfer from contexts to actions. The basal ganglia house similar parallel tracks, although the direct connection between any BG area and action is less obvious, since most BG outputs target high-level cortical areas.

In learning declarative predictions, the brain must link contexts to contexts, and these often reside in the same brain region. Hence, learning involves a modification of recurrent arcs, as shown at the bottom of Figure 22, and as detailed by the previous models of the cortex, hippocampus and thalamocortical system.

As reviewed by Dominey [13], artificial neural network (ANN) experiments indicate that recurrence alone will not suffice to learn pattern sequences, since the delayed neuromodulatory feedback has a credit assignment problem: difficulty targeting the relevant activation rounds that accounted for an action. However, as Dominey et. al. [12] show, an ANN with a combination of a) leaky integrator neurons with a diverse range of time constants, and b) synaptic modification restricted to a small portion of the network, can learn temporal sequences.

The models in this article indicate other factors that may be critical for learning in recurrent regions, including eligibility traces (which seem to be a standard component of biologically-plausible learning rules [17]), a wealth of available neurons to serve as links between two stimulus patterns, phasic toggling between external driving and recurrent signalling, and the general *monitoring* mechanism involving distal dendrites.

Evolution has clearly endowed the brains of higher mammals with considerable recurrence, and

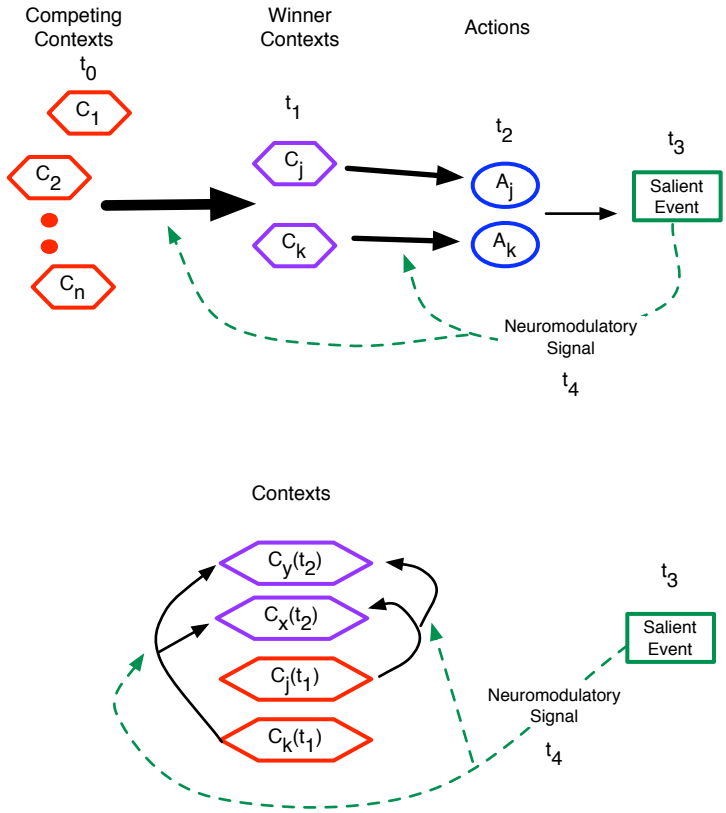

Figure 22: Abstract comparison of the procedural (top) and declarative (bottom) predictive mechanisms, with relative time points given as t_0 - t_4 and vertical cell columns indicating brain regions. (Top) In procedural prediction, neural patterns corresponding to competing contexts, winning contexts, and proposed actions have distinct spatial locations in the brain. Salient events and the ensuing feedback (via neuromodulators) then alter synapses between these regions, such as between granular cell contexts of the cerebellum and the action-regulating Purkinje cells. Similar spatial localization occurs in the basal ganglia. (Bottom) In declarative areas such as the hippocampus, cortex and thalamocortical system, active patterns have considerably more spatial overlap such that neuromodulatory feedback effects recurrent (often distal) connections **within** a region. The competition among contexts thus occurs *in place*.

thus with more sophisticated declarative predictive abilities. The hippocampus was the first major step [49], with both a) high recurrence within CA3, and b) the loop organization of the entire hippocampal region, with most signals entering and leaving via EC. The neocortex, with its massive intra- and inter-region feedback elaborates that trend. With increased evolutionary size, the cortex also sends a greater density of axons to diverse neural regions, both higher and lower. Thus, the most advanced predictive machinery attains a higher degree of control in the mammalian, and particularly primate, brain.

A finer grained analysis of the five systems above reveals several features that many share. These properties involve neurons, synapses and circuitry.

Obviously, LTP and LTD are essential for strengthening relevant and weakening irrelevant connections between neurons. When Hebbian LTP governs synaptic change, the random bursting becomes a vital characteristic of predictive networks as well. In all of the models above, randomly bursting neurons form convergence points for the active neurons of a salient context. This is evident in granuler cells of the cerebellum, striatal cells of the basal ganglia, CA3 cells of the hippocampus, and matrix cells of the thalamus. Also, in cortical columns, neurons of layers 4 and 5 exhibit higher degrees of bursting than those of other layers [36]; as the primary input port for convergent lower-level signals, layer 4 also appears to perform a context-detecting role.

In discussing their hippocampus model, Wallenstein et. al. [51] point to *asymmetric connectivity* as a critical topological factor. If neurons (or nuclei or columns) A and B are connected, then asymmetry entails that the connections are not bi-directional, or they may be bi-directional but differ in density or synaptic location (i.e. proximity to the soma, layer of entry, etc.).

Completely symmetric connections may confound the learning of uni-directional predictive sequences, since, if the neurons of event 1's activity pattern (E1) have symmetric bi-directional couplings to those of E2, then E1 will predict E2 just as often as E2 predicts E1. Since recurrence in CA3 is only 1-5% (which is still very high compared to the cortex, when viewed as a single module), there is little chance of symmetry. Hence, it becomes unlikely that, while enhancing the connections from E1 to E2, the hippocampus simultaneously strengthens those from E2 to E1.

The density and synaptic-location aspects of asymmetry are apparent in the neocortex. First, the distribution of connections between lower and higher cortical areas is far from bi-directionally equivalent. For example, the general view that the bottom-up signally pathways are convergent, while the top-down are divergent implies that a low-level cortical module may only send axons to a few higher modules but receive axons from a diverse array of such modules. Also, there is a clear trend of increasing top-down control in the brains of higher organisms [49, 11], and this coincides with the presence of far more top-down than bottom-up connections [23]. Second, as discussed earlier, top-down axons tend to have distal targets, whereas bottom-up projections typically synapse close to the soma of large pyramidals [31, 23]. Thus, low-level activity will have different effects upon high-level activity than the latter will have on the former: a lower level can often drive a higher level, while the latter often has a more controlling or *monitoring* effect upon the former. This monitoring ability is central to the GDPN and its manifestation in the hippocampal, cortical and thalamacortical systems.

Several authors [51, 43, 20] cite mode-switching as essential to the proper functioning of their

focal networks. In the hippocampus, the primary influences upon CA3 neurons must alternate between external (from DG) and internal (via recurrent links). The theta cycle (4-12 Hz) appears to govern this switching. In the thalamortical loop, thalamic and cortical inhibitory neurons are active periodically, with frequencies in delta (1-4 Hz), theta (4-12 Hz) and even gamma (30-40 Hz) bands. Excitatory inputs are received and propagated during the down phases of these periods.

These phases agree with the general view that neural networks for declarative memory require at least two modes of activity: storage and retrieval. In the former, sensory inputs should govern activity such that the scenario is mainly remembered *as is*, without a lot of embellishment. During retrieval, partial clues often require the pattern completion provided by spreading activation. Hence, the network needs a mechanism for modulating the expanding wave of excitation.

The thalamotical modellers [43, 20] also mention timing differences as critical: while excitatory post-synaptic potentials (PSPs) have durations of 15-20 msec, inhibitory PSPs last for 100-150 msec. This insures that the effects of early (and/or most salient) stimuli are muted while other signals arrive (even when the early stimuli remain present), thus allowing unique elements of a sequence (or features of a scenario) to be handled relatively independently, and thereby avoiding blending and loss of information.

Similarly, in the basal ganglia, the quick but fleeting excitatory effects of context detectors upon the SNc (via the hyperdirect pathway) lead to a timely jolt of dopamine, which is then muted for the (possibly lengthy) remainder of the context sequence by slow, but persistent, inhibition of SNc. This insures that only the most newly-recognized salient context is learned, and not a less-informative blend of several sequential contexts.

Timing appears in other guises as well. In the basal ganglia model, the chemistry of LTP in the striatum insures that contexts at time t are associated with reinforcements at approximately time t+100 ms. In contrast, the CA3 model allows the association of events across a wide range of temporal gaps, due to the auto-catalytic nature of the context/monitoring neurons. In the cerebellar model, each context includes events with different time stamps, due to the differential delays along mossy fibers. However, these delays are on the order of 5-120 msec, not seconds, as in the CA3 model. However, the authors [32] do include eligibility traces on the parallel-Purkinje synapses, wherein they remain eligible for modification (via LTD) for up to a half second after activation.

Finally, the more general phenomena that *surprise stimulates learning* appears in both the procedural and declarative predictive mechanisms. Hawkins [23] proposes that sensory signals propagate up the cortical hierarchy but stop at the level where they *agree with* expectations, as described in the above discussion of the thalamocortical loop. Those signals that confound all predictions spread to the higher association layers and into the hippocampus, where they can begin to be learned and (in the future) predicted. Similarly, the basal ganglia, particularly the SNc, does not respond to expected reinforcements nor to well-known cues to reinforcement, only to relatively new (i.e. still somewhat surprising) cues. Since the SNc response indirectly stimulates Hebbian synaptic change via dopamine signalling, the connection between surprise and learning is again evident.

Adaptive Resonance Theory [9] and its corresponding ANNs have been used to model a wide variety of cerebral circuitry, including the neocortex and basal ganglia. Its cornerstone *match-based*

learning algorithm triggers on inputs that are either completely new/surprising or very similar to expectations, with the former being quickly and forcefully imprinted onto the network, while the latter promote only small changes.

None of the above characteristics are unique to the predictive functionality of the brain. However, since prediction pervades many conscious and unconscious activities, they may actually constitute essential aspects of sensorimotor behavior and cognition. In general, they are ubiquitous factors in neuroscience. One important reason for summarizing them here is to alert neural modelers to several components that may prove pivotal to their systems, particularly when prediction is a primary goal.

10 Conclusion

Although relatively well-understood at the conscious behavioral level as an ability to foresee the future, the concept of prediction becomes significantly more diffuse when one searches for neural correlates. Readers who accept the general definitions of declarative and procedural prediction presented above, along with the key mechanisms of the real and artificial neural systems discussed throughout, should come to the general conclusion that prediction is not localized to any one part of the brain but is distributed, in various forms, to many cortical and sub-cortical areas.

Researchers such as Llinas [34] and Hawkins [23] elevate prediction to the pinnacle of cerebral functionality - the essence of simple and sophisticated intelligences alike. Their arguments do carry a lot of weight, but a deeper investigation paints a more complex picture in which prediction seems to play an important role, but probably as one of many vital functions. However, this says little about the *origins* of complex brains, and how, as Llinas argues, they may have arisen to satisfy the predictive needs of motion. In that view, other cerebral functions have exapted the predictive machinery, and, from an evolutionary perspective, indeed, all neural functionality derives from prediction. Ultimately, this evolutionary interpretation may rule the day, but in looking at the myriad complex neural mechanisms, it is often beyond current neuroscientific knowledge to ascertain their potentially predictive origins.

Unfortunately, neither Llinas nor Hawkins gives a clear description of the predictive neural process, although Hawkins does present a detailed cortical sketch that captures many of the key structures and some of the behavioral dynamics. In attempting to further clarify the links between structure and function, this research has summarized several systems and mechanisms that appear to embody prediction at the neural level. The ubiquity of these mechanisms in the brain could support either of the following conclusions:

1. Prediction is the basis of all neural functionality and the driving force behind neural evolution.
2. The basic elements of neural behavior are employed for a wide variety of tasks, including prediction, but it has no special status with respect to brain evolution.

We adopt an intermediate stance by postulating that the predictive *perspective* is a very useful

one to take when analyzing many areas of the brain, both in terms of their origins and current functionality. This viewpoint may provide more leverage than the more general interpretation of the brain as a pattern-association machine. Thus, although we cannot completely isolate prediction from other functions, *predictive glasses*, as worn in the writing of this article, may reveal improved insights into many aspects of brain and behavior.

As briefly discussed earlier, predictive learning is a special case of associative learning in which the linked brain states correspond to real-world events whose starting points have at least a small amount of temporal disparity. The preparatory window provided by prediction lends an additional survival advantage to that accrued by other forms of associative learning. Thus, from an evolutionary perspective, it makes sense to at least consider predictive interpretations of neural behavior when analyzing a brain region, since prediction could indeed constitute the raison d'etre of that structure.

In terms of basic neurophysiology, the delays inherent in synaptic plasticity (that manifest eligibility traces) as well as those of neural signal transmission, naturally facilitate the linkage of non-simultaneous events. Furthermore, since the bandwidth limits of sensory processing can preclude immediate, detailed snapshot capture of events, the brain must link temporally sequenced inputs (for example, from visual saccades) into a holistic perception. In this case, the inputs are linked together via associations that essentially encode the prediction of *what aspect of the image the organism will next perceive*, as described earlier in the perceptual paradigm supported by the thalamocortical models of Rodriguez et. al. [43] and Granger [20].

If correct, those models indicate that the basic mechanisms for predictive learning help to overcome perceptual bottlenecks and give the important illusion that the mental correlates of single real-world events can be simultaneously active in the brain. More generally, they imply that predictive associations are a necessary support for memories of complex events even when those events lack salient temporal aspects such as causal associations. Consequently, vanilla associative learning may not scale up to complex events without predictive associations (about one's own perceptual-processing sequence). Prediction may be fundamental to associative learning.

In the case of the thalamus, predictive goggles highlight a potentially fundamental process: linking successive sensory-processing states, which provides a mechanism for sequential learning of **both** external events and internal brain states. This would help support the emerging view [45] that the thalamus is much more than a relay station for sensory input.

These goggles are firmly in place throughout Hawkins' analysis of the neocortex [23]; we have merely elaborated a few of the physiological details while summarizing his basic model, most of which fits nicely into our Generic Declarative Prediction Network (GDPN) framework, particularly his emphasis on distal layer-1 synapses as the carriers of expectations. It is these relatively weak, yet influential, connections that are vital to many of the models. From alternate perspectives, these connections might be overlooked, but predictive glasses magnify them significantly.

In the hippocampus, the predictive perspective helps expand the concepts of memory and memory formation beyond that of snapshots, despite the fact that this brain region appears to realize quick imprinting [3, 46]. Wallenstein et. al. [51] show that events separated by substantial time windows can still be linked (again with the help of distal signalling) in CA3, a hippocampal region often

viewed as the brain's ultimate pattern associater [44].

The phenomena of phase precession in hippocampal place cells [7, 37], whereby a neuron that codes for location L begins to fire at locations prior to L along an often-travelled path, seems quite logical if prediction is seen as a vital aspect of navigation. Since strong evidence links the hippocampus to both memory formation and spatial navigation [3], the reasonably obvious potential contribution of prediction to the latter could further implicate it as an aid to the former.

In the more procedural regions, such as the cerebellum and basal ganglia, predictive roles are often posited, but we argue that the correct interpretation is that these areas aid the organism in *doing the right thing* in the immediate future without necessarily having an explicit representation of future world states. Thus, the distinctions between procedural and declarative predictions appear very real, and the architectural differences between these two types of regions indicate divergent functionality. In this case, the predictive viewpoint aids in the clarification of declarative versus procedural processing, since they can be further differentiated with respect to their contributions to the task of prediction: foretelling future states versus proactively selecting the actions to best handle those states.

In the end, we hope that this article helps to make prediction a less nebulous concept within neuroscience, where it is often tossed around without formal grounding. We have sought these principled underpinnings in neuroscience itself, not philosophy nor psychology, since modern-day understanding of neural circuitry does facilitate bottom-up investigations, though they are still plagued with constrained speculation.

In *Rhythms of the Brain*, Gyorgy Buzsaki [8] begins by stating:

> The short punch line of this book is that brains are foretelling devices and their predictive powers emerge from the various rhythms they perpetually generate (pg. vii)

He then goes on to show the importance of brain oscillations for producing the stability and linearity required for expectations and predictive processes. Along the way, he emphasizes that many of the standard cognitive scientific concepts such as attention, intentionality and reasoning map only poorly to neural mechanisms. The useful metaphors for understanding the brain may lie elsewhere, in control- or dynamic-systems theory, he suggests. And while these fields do provide nice tools for mathematically analyzing complex-system behavior, your average phase diagram says very little about functionality.

Evolution, on the other hand, has a lot to say about functionality, and it seems quite natural to infer that the basic neural mechanisms arose and evolved to serve a very primitive, but useful, purpose in early organisms - a purpose such as prediction. Although Llinas' claim that prediction is the brain's ultimate function [34] can be endlessly debated, if predictive circuitry can be verified in a cross-section of brain regions (such as those discussed in this article), then the general primacy of this useful capability will be hard to ignore.

References

[1] www.merriam-webster.com.

[2] N. S. A.G. Barto, A. H. Fagg and J. Houk, *A cerebellar model of timing and prediction in the control of reaching*, Neural Computation, 11 (1999), pp. 565–594.

[3] P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, *The Hippocampus Book*, Oxford University Press, New York, NY, 2007.

[4] A. Artola, S. Brocher, and W. Singer, *Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex*, Nature, 347 (1990), pp. 69–72.

[5] A. Barto, *Adaptive critics and the basal ganglia*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 215–232.

[6] M. Bear, B. Conners, and M. Paradiso, *Neuroscience: Exploring the Brain*, Lippincott Williams and Wilkins, Baltimore, MD, 2 ed., 2001.

[7] N. Burgess and J. O'Keefe, *Hippocampus: Spatial models*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 539–543.

[8] G. Buzsaki, *Rhythms of the Brain*, Oxford University Press, New York, NY, 2006.

[9] G. Carpenter and S. Grossberg, *Adaptive resonance theory*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 87–90.

[10] A. Clark, *Mindware: An Introduction to the Philosophy of Cognitive Science*, The MIT Press, Cambridge, MA, 2001.

[11] T. Deacon, *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton and Company, New York, 1998.

[12] P. Dominey, T. Lelekov, J. Ventre-Dominey, and M. Jeannerod, *Dissociable processes for learning the surface structure and abstract structure of sensorimotor sequences*, Journal of Cognitive Neuroscience, 10 (1998), pp. 734–751.

[13] P. F. Dominey, *Sequence learning*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 1027–1030.

[14] K. L. Downing, *The predictive basis of situated and embodied artificial intelligence*, in GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation, New York, NY, USA, 2005, ACM, pp. 43–50.

[15] ——, *Neuroscientific implications for situated and embodied artificial intelligence*, Connection Science, 19 (2007), pp. 75–104.

[16] K. Doya, *What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex?*, Neural Networks, 12 (1999), pp. 961–974.

[17] Y. Fregnac, *Hebbian synaptic plasticity*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 515–522.

[18] J. Fuster, *Cortex and Mind: Unifying Cognition*, Oxford University Press, Oxford, 2003.

[19] M. Gluck and C. Myers, *Gateway to Learning: An Introduction to Neural Network Modeling of the Hippocampus and Learning*, Kluwer Academic Publishers, Norwell, Massachusetts, 1989.

[20] R. Granger, *Engines of the brain: The computational instruction set of human cognition*, Artificial Intelligence Magazine, 27 (2006), pp. 15–32.

[21] A. M. Graybiel and E. Saka, *The basal ganglia and the control of action*, in The Cognitive Neurosciences III, M. S. Gazzaniga, ed., The MIT Press, Cambridge, MA, 2004, pp. 495–510.

[22] F. Haugen, *Anvendt Reguleringsteknikk*, Tapir Forlag, Trondheim, Norway, 1992.

[23] J. Hawkins, *On Intelligence*, Henry Holt and Company, New York, 2004.

[24] D. Hebb, *The Organization of Behavior*, John Wiley and Sons, New York, NY, 1949.

[25] J. Houk, *Information processing in modular circuits linking basal ganglia and cerebral cortex*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 3–9.

[26] J. Houk, J. Adams, and A. Barto, *A model of how the basal ganglia generate and use neural signals that predict reinforcement*, in Models of Information Processing in the Basal Ganglia, J. Houk, J. Davis, and D. Beiser, eds., Cambridge, MA, 1995, The MIT Press, pp. 249–270.

[27] J. Houk, J. Davis, and D. Beiser, *Models of Information Processing in the Basal Ganglia*, The MIT Press, Cambridge, MA, 1995.

[28] D. Joel, Y. Niv, and E. Ruppin, *Actor-critic models of the basal ganglia: New anatomical and computational perspectives*, Neural Networks, 15 (2002), pp. 535–547.

[29] S. Kali and P. Dayan, *The involvement of recurrent connections in area ca3 in establishing the properties of place fields: a model*, Journal of Neuroscience, 20 (2000), pp. 7463–7477.

[30] R. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering, 82 (1960), pp. 35–45.

[31] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*, McGraw-Hill, New York, NY, 2000.

[32] R. Kettner, S. Mahamud, H. Leung, N. Sitkoff, J. Houk, and B. Peterson, *Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement*, Journal of Neurophysiology, 77 (1997), pp. 2115–2130.

[33] J. LeDoux, *Synaptic Self: How Our Brains Become Who We Are*, Penguin Books, Middlesex, England, 2002.

[34] R. R. Llinas, *i of the vortex*, The MIT Press, Cambridge, MA, 2001.

[35] J. Mcclelland, B. McNaughton, and R. O'Reilly, *Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory*, Tech. Rep. PDP.CNS.94.1, Carnegia Mellon University, Mar. 1994.

[36] V. Mountcastle, *Perceptual Neuroscience: The Cerebral Cortex*, Harvard University Press, Cambridge, Massachusetts, 1998.

[37] M. R. N. Burgess and J. O'Keefe, *A model of hippocampal function*, Neural Networks, 7 (1994), pp. 1065–1083.

[38] R. O'Reilly, *Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm*, Neural Computation, 8 (1996), pp. 895–938.

[39] R. C. O'Reilly and Y. Munakata, *Computational Explorations in Cognitive Neuroscience*, The MIT Press, Cambridge, Massachusetts, 2000.

[40] T. Prescott, K. Gurney, and P. Redgrave, **basal ganglia*, in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed., The MIT Press, Cambridge, MA, 2003, pp. 147–151.

[41] A. D. Redish, *Beyond the Cognitive Map: From Place Cells to Episodic Memory*, MIT Press, Cambridge, Massachusetts, 1999.

[42] E. Rich, *Artificial Intelligence*, McGraw-Hill Book Company, New York, NY, 1983.

[43] A. Rodriguez, J. Whitson, and R. Granger, *Derivation and analysis of basic computational operations of thalamocortical circuits*, Journal of Cognitive Neuroscience, 16 (2004), pp. 856–877.

[44] E. Rolls and A. Treves, *Neural Networks and Brain Function*, Oxford University Press, New York, 1998.

[45] S. Sherman and R. Guillery, *Exploring the Thalamus and its Role in Cortical Function*, The MIT Press, Cambridge, MA, 2006.

[46] L. Squire and E. Kandel, *Memory: From Mind to Molecules*, Henry Holt and Company, New York, 1999.

[47] L. Squire and S. Zola, *Structure and function of declarative and nondeclarative memory systems*, Genetic Programming and Evolvable Machines, 93 (1996), pp. 13515 – 13522.

[48] P. L. Strick, *Basal ganglia and cerebellar circuits with the cerebral cortex*, in The Cognitive Neurosciences III, M. S. Gazzaniga, ed., The MIT Press, Cambridge, MA, 2004, pp. 453–461.

[49] G. F. Striedter, *Principles of Brain Evolution*, Sinauer Associates, Sunderland, Massachusetts, 2005.

[50] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 1998.

[51] G. Wallenstein, H. Eichenbaum, and M. Hasselmo, *The hippocampus as an associator of discontiguous events*, Trends in Neuroscience, 21 (1998), pp. 317–323.

[52] D. Wolpert, R. C. Miall, and M. Kawato, *Internal models in the cerebellum*, Trends in Cognitive Sciences, 2 (1998), pp. 338–347.