

# Neural Predictive Mechanisms and Their Role in Cognitive Incrementalism

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

June 18, 2012

## Abstract

Many neuroscientists view prediction as one of the core brain functions, particularly due to its support of fast movements in complex environments. This leads to the natural question of whether this predictive knowledge forms the cornerstone of our common-sense understanding of the world. However, there is little consensus as to the exact nature of predictive information and processes, nor the neural mechanisms that realize them. This paper compares procedural versus declarative notions of prediction, examines how the brain appears to perform them, and discusses to what degree, and at what level, these neural mechanisms support *cognitive incrementalism*: the notion that high-level cognition stems from sensorimotor behavior.

## 1 Introduction

Several prominent neuroscientists have recently championed the ability to predict as a fundamental function of all brains, from insects to humans (Llinas, 2001; Hawkins, 2004). The basic argument for prediction's unique importance begins with sensorimotor behavior, where the processing speed of sensory inputs simply cannot keep pace with the rate at which the coupled agent-environment system changes in many (often intense and/or life-threatening) situations. In calculating the next action, the nervous system does not have time to interpret all sensory data. Instead, it uses predictions of future states to guide action, and only when predictions conflict with reality must the system realign its predictive machinery with the sensory present. Neuroscientists (Wolpert, Miall, & Kawato, 1998; Llinas, 2001) generally agree that the brain needs predictive abilities similar to those found in control theory (such as Kalman filters) to handle the different timescales involved in sensing, moving, subconscious action choice and conscious decision making. They cite potential predictive mechanisms in areas such as the cerebellum (Wolpert et al., 1998), basal ganglia (Houk, Davis, & Beiser, 1995), hippocampus (Gluck & Myers, 1989) and neocortex (Hawkins, 2004) as central to this endeavor.

As we discuss in (K. L. Downing, 2005), these predictive facilities may underlie our common-sense understanding of the world and may provide support for *cognitive incrementalism* (Clark, 2001) - 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 (Squire & Zola, 1996) (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.

In (K. L. Downing, 2009), we find interesting predictive architectures, as proposed by experimental and computational neuroscientists, in five different brain systems: cerebellum, basal ganglia, hippocampus, neocortex and thalamocortical loop, with the former two performing procedural prediction and the latter three acting more declaratively. The key difference between the procedural and declarative predictive forms resides in the explicit awareness (or lack thereof) of the connections between spatiotemporal states that embody predictive knowledge. For example, a golfer can consciously (explicitly) predict the flight of a ball based on current wind conditions, but she may not have explicit awareness of the effects of hip rotation upon that trajectory; and yet, she may instinctively (implicitly)

*feel* that a shot will hook or slice even before the club hits the ball, based on subconscious somatosensory feedback from her hips.

The neural evidence summarized in (K. L. Downing, 2009) indicates that procedural and declarative predictions seem to involve two distinct, yet general, neural architectures. However, these structures reappear in different parts of the brain, indicating that prediction could indeed be a primary brain process, both at conscious and subconscious levels.

This article presents two general, neural network models for prediction, one procedural and one declarative, based on the review and synthesis in (K. L. Downing, 2009). It also considers hippocampal place cells, predictive relationships between them, and their dual role in navigation and higher-level cognition. Together, these neural mechanisms give hints as to how prediction could constitute a core element of cognitive incrementalism.

## 2 Prediction in the Brain

The dictionary (Prediction, 2009) defines *predict* as *to declare or indicate in advance*, where *to declare* is to *make known formally, officially or explicitly*; and *to indicate* is to *be a sign, symptom or index of*. Thus, the declarative form of prediction is more concrete and direct, while the indicative form is more indirect and implicit. For example, one may declaratively predict a win for his favorite team by boasting about the upcoming championship game to friends, or he may only indicate such a prediction by betting large sums of money on-line, buying champagne for a victory party, etc.

To understand prediction in the brain, we begin with Figure 1 and a neural perspective on *recognition*, defined as attaining a brain state, S, that has previously exhibited a strong correlation with the (now familiar) experience, e.g. lightning. Informally, S is the brain state that is both a) most likely to arise under the given experience, 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. 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.

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.

## 3 Neural Architectures for Procedural Prediction

The two primary loci of procedural prediction in the mammalian brain appear to be the cerebellum and basal ganglia. Their anatomical circuitry is beyond the scope of this article (see (K. L. Downing, 2009) for more details), but the coarse model of Figure 3 summarizes the main neural components.

The entry point is a layer of *context detectors*, which are neurons that a) tend to fire only when a large number of their upstream neighbors (a.k.a. afferents) are active, and b) tend to inhibit one another when they do fire. Thus, these neurons *compete to classify* the activation patterns in the immediate upstream layer, with only those that are most highly stimulated being able to overcome the inhibition from the other detectors.

Context detectors that survive the intra-layer competition send strong outputs to downstream layers. The number and type of these layers varies between the cerebellum and basal ganglia, but these share the basic properties of a) being feed-forward, with little or no excitatory, intra-layer recurrence - thus, spreading activation within a layer to *complete a pattern* has little architectural support - and b) having firing patterns that correlate well with physical actions. In short, the contexts map, through a series of intermediate layers, to actions and/or conscious or subconscious plans to perform actions. These circuits are the basis of skill acquisition and performance, with those of the cerebellum handling more immediate stimulus-response situations, while the basal ganglia seem to detect more

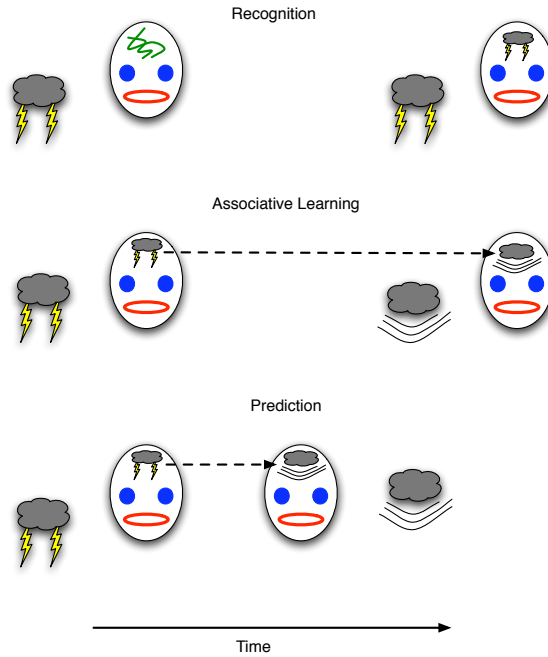


Figure 1: (Above) Recognition depicted as the formation of a brain state (drawn as lightning on the forehead) that becomes correlated with a physical event (lightning). (Middle) Learning the association between one event (lightning) and its successor (thunder) by linking the brain states that correlate 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.

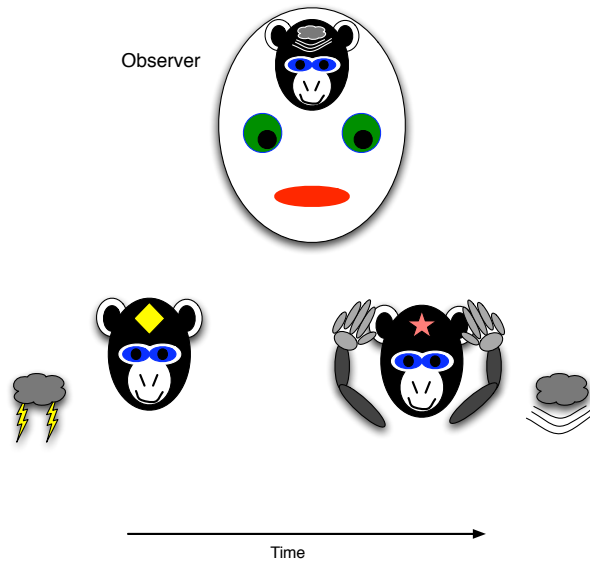


Figure 2: Procedural prediction, wherein the agent's actions *indicate* specific knowledge of a future world state, even though the agent (monkey) has no explicit brain state that strongly correlates with the world state. The agent's ear-covering behavior can easily lead an observer to infer that the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

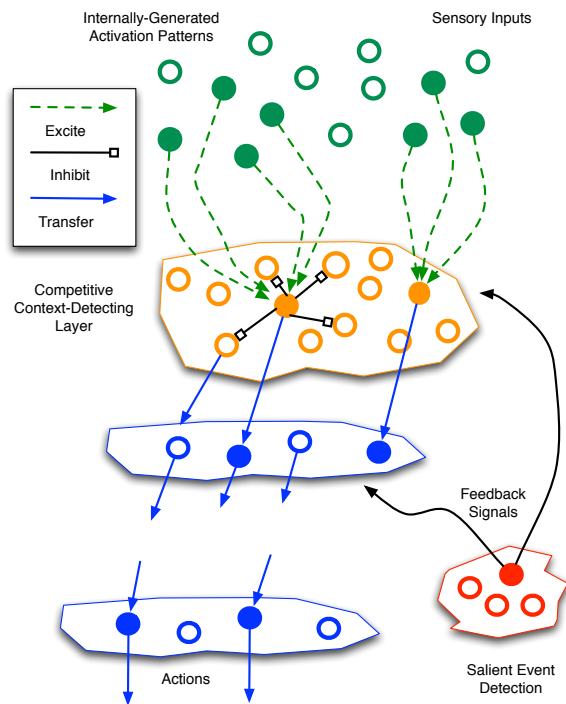


Figure 3: General topology of a neural network for procedural prediction. Context detectors compete to classify activation patterns of upstream neurons (top) representing sensory inputs as well as internally-generated conscious or subconscious thoughts. Winning contexts then transfer their influence - via a series of excitatory and/or inhibitory connections - through several layers to action-specific neurons. A salient-event detector, triggering on sensory input patterns (arrows not shown), then provides feedback to the context-action circuit either in the form of neural signals along dedicated synapses or via the diffuse secretion of neuromodulators.

complex contexts (due to afferents that both encode higher-level concepts and have a higher fan-in to the competitive neurons) and to control behavior over longer time frames.

Learning of useful mappings between contexts and actions involves feedback signals from salient-event detectors in both regions. In the cerebellum, there are many such detectors located throughout the body, for registering pain and detecting other *deviations*, such as retinal image slippage (as part of the vestibulo-ocular reflex (Ito, 1982)). These detectors send electrical signals directly, along dedicated neural lines, from the local detection point to a few synapses in the circuit that links the current context to the offending action, resulting in very local synaptic change and embodying a primitive form of supervised learning (Doya, 1999; Wolpert et al., 1998). In contrast, the saliency detector in the basal ganglia reacts to generally *surprising* events and secretes neuromodulator chemicals, over a potentially wide area, as a feedback response. This signal affects many synapses to realize a form of reinforcement learning (Doya, 1999; Houk, Adams, & Barto, 1995).

To understand prediction within these circuits, the temporal aspects of both the context-to-action map and learning must be considered. In general, cross-sections of the past (embodied as contexts) determine present (action) decisions about future behaviors. As depicted in Figure 4, 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. That is, based on the world state at time  $T - \delta_1 - \delta_2$  that is perceived at time  $T - \delta_1$ , an action is chosen/initiated at time  $T$ . Its effects occur in the world at time  $T + \delta_3$  and are detected by the agent itself at time  $T + \delta_3 + \delta_4$ . The brain must therefore have the ability to apportion credit and blame to choices made  $\delta_3 + \delta_4$  milliseconds in the past.

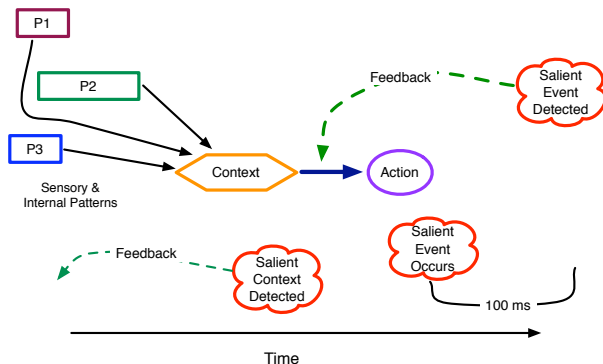


Figure 4: The temporal scope of procedural decision making. Beginning on the left, neurons activated by either sensory inputs or internal states stimulate context-detecting neurons which, in turn, stimulate motor and pre-motor neurons leading to an action (oval), with each neural population activated later in time. The action’s consequence, labeled as a *salient event*, occurs thereafter but is not detected by the organism for another 100 ms. Detection initiates feedback signals that can promote changes to neural-network components that were active earlier (as shown by the curved right-to-left arrows).

To maintain an approximate record of what channels were active, and when, and thus what synapses are most deserving of modification, the cerebellum, basal ganglia, and other brain areas utilize a complex biochemical process that essentially yields a synapse most receptive to modification (triggered by a feedback signal) about 100 msec after high transmission activity (Kettner et al., 1997). This *eligibility trace*, in the parlance of reinforcement learning theory (Sutton & Barto, 1998), 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.

Considering that the human cerebellum consists of millions of context-action links, 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 (synaptic long-term depression). Importantly, contexts reflect states of the world at  $T - \delta_1 - \delta_2$ , but actions that they recommend should be those most appropriate for states of the body and world at  $T + \delta_3$ . Rules 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.

The added predictive power of the basal ganglia lies in its ability to learn connections between **context detectors** and the saliency-detecting neurons, as shown at the bottom of Figure 4. Essentially, in feedback situations, the basal ganglia not only strengthens a context-action rule,  $C \rightarrow A$ , but also strengthens a synapse linking C’s detector neuron  $N_C$  to the saliency detectors. This fortified synapse can then trigger feedback signals prior to the detection of the salient world state, W. Due to the complex biochemical dynamics mentioned above, this implies that contexts that are active prior to C (by about 100 milliseconds) will have high eligibility traces when this  $N_C$ -triggered feedback signal arrives. Thus, they too will have their links to both an action and the saliency detector strengthened. Via backward chaining of credit assignment, the basal ganglia learns to **predict** salient world states using salient contexts that precede them. Since, in theory, there is no limit to the number of salient contexts that can be chained together, the time delay between the first such predictive context and W can be of any length. This contrasts with the cerebellum, whose predictive rules appear more strictly bound to a temporal gap of  $\delta_1 + \delta_2 + \delta_3$ , although delay lines between the sensory inputs and context detectors can modify this to some small degree.

In summary, procedural prediction is achieved by neural dynamics and architectures that adapt (in both evolutionary and lifetime time scales) to the inherent delays in sensory perception and motor action of the organism. The context-action rules that these systems embody have implicit predictive knowledge, since they can a) choose actions for world states occurring at one time point that will be most appropriate for some future time point, and b) be tuned to indicate future salient events.

## 4 Neural Architectures for Declarative Prediction

Declarative prediction requires machinery that can associate two patterns, both of which have strong correlations with external states. The hippocampus, cortex and thalamocortical regions of the brain all appear to include variations on a common connection scheme known as the Generic Declarative Prediction Network (GDPN) (K. L. Downing, 2009).

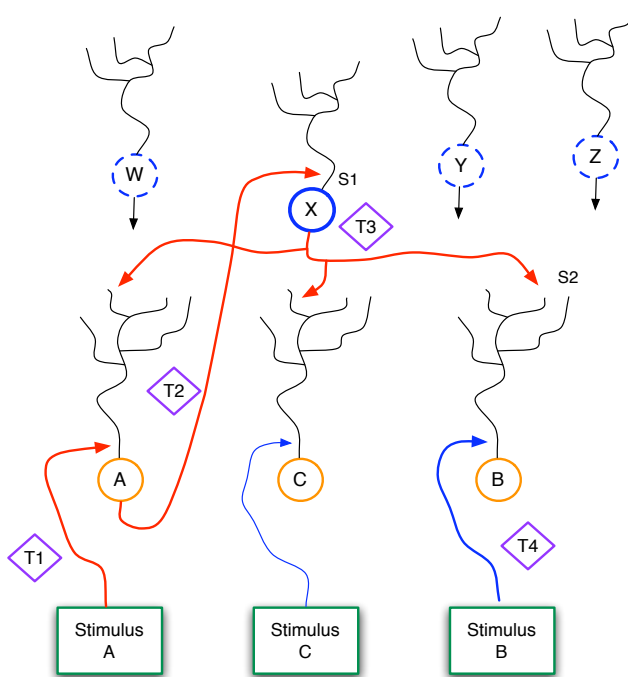


Figure 5: The Generic Declarative Prediction Network (GDPN). Neurons A, B and C serve as low-level detectors for stimuli A, B and C, while W-Z represent neurons at a higher level. Only the axonal projections from X are shown, though W, Y and Z have similar links to the lower level. The T1 - T4 diamonds represent time steps, while S1 and S2 denote important synapses, as further discussed in the text.

Figure 5 sketches the basic GDPN framework. One of its most notable features is that low-level inputs to higher levels occur proximally, i.e. close to the soma (circles), whereas top-down signals, such those from X to A, B and C, enter via distal dendrites. Hence, low-level signals can more easily *drive* the activity of their high-level neighbors than vice versa. In contrast to the above, feed-forward, procedural model, declarative architectures tend to be replete with *recurrence*: combinations of bottom-up and top-down links.

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 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 coincidentally coactivate with A and have their proximal synapses (from A) 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.

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 (Kandel, Schwartz, & Jessell, 2000). 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 (that are not simultaneously activated by bottom-up signals). This results from electrophysiological mechanisms (Artola, Brocher, & Singer, 1990) that produce long-term potentiation (LTP) when a synapse has been very active, but long-term depression (LTD) when it has only been mildly active.

Thus, X simply becomes a dedicated **predictor** link between A and B. After the repeated presentation of many sequential patterns, LTP and LTD gradually convert a blanket of bet-hedging anticipatory links into a smaller population of dedicated connections between associated pattern-detecting neurons.

In the neocortex, the GDPN is clearly manifest, with the individual neurons of Figure 5 replaced by cortical columns, each functioning as a processing modules (Hawkins, 2004; Fuster, 2003). Bottom-up sensory interpretation involves cascades of neural firing from the back (sensory) areas of the brain to the front (executive) areas, while top-down predictive and bet-hedging activity moves front to back.

Several elegant neurocomputational models (Rodriguez, Whitson, & Granger, 2004; Granger, 2006) illustrate GDPN circuitry and behavior in the thalamocortical loop. However, rather than linking sequences of real-world events, these networks associate sequential neural states of perceptual processing, wherein the initial states tend to involve the most salient features, with other aspects registering in later steps. This work blurs the borders between prediction and conventional association, since any perception can now be interpreted as a time series of partial interpretations of sensory input, each predicting the next.

Finally, an interesting GDPN-like model of the hippocampus (Wallenstein, Eichenbaum, & Hasselmo, 1998) provides an intricate mechanism (based on recurrent, distal, bet-hedging connections) for predictive learning that can connect events separated by significant and varying temporal delays. Basically, the extensive hippocampal recurrence enables neurons such as X in Figure 5 to remain active (via membership in self-stimulating recurrent loops) for long periods of time, sending out top-down anticipatory signals all the while and linking up to any lower-level neurons that activate during that time window.

## 5 Procedural versus Declarative Predictive Topologies

The architectures above indicate differential predictive functionality for procedural and declarative circuits. Assuming that a good deal of our explicit, conscious representations (i.e., those that we can reason about) consist of perceptual information, 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 topographic nature of many of these perceptual regions, the patterns probably reside near one another and even share active neurons. Thus, these two patterns would probably reside within a tight mesh of intra-layer and intra-region connections: a highly recurrent sub-circuit.

Conversely, the cerebellum and basal ganglia have many parallel tracts, but little intra-layer, excitatory recurrence. These procedurally predictive areas are therefore hard pressed to link representation R1 for world-state 1 to R2, 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.

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 targets synapses **between** one area and its downstream neighbor region. Conversely, 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, which is often more complicated than tuning strictly feed-forward links. This suggests that the earlier evolutionary emergence of the cerebellum and basal ganglia, compared to the hippocampus and cortex, was no coincidence (Striedter, 2005; Allman, 1999).

Several researchers have pointed out these different topologies and speculated as to their functionalities (Edelman & Tononi, 2000; Doya, 2000; Deacon, 1998), with Doya (Doya, 1999, 2000) providing the most comprehensive analysis, wherein the cerebellum, basal ganglia and cortex each manifest a different type of learning: supervised, reinforced and unsupervised, respectively. He emphasizes the mounting evidence that none of these areas are exclusively dedicated to motor nor cognitive tasks: each plays vital roles in both endeavors. Doya's work is an important motivator of our own investigation, where we see prediction as an essential component of both motor and cognitive activity, and where predictive knowledge can be acquired in each of these brain regions, using each of the learning methods outlined by Doya. However, we share Squire's (Squire & Zola, 1996) concern over the procedural-declarative distinction and argue that - although cortex, cerebellum and basal ganglia (and hippocampus) are clearly linked via complex recurrent topologies - the predictive knowledge housed within the cerebellum or basal ganglia is not analyzable by nor easily transferable to more declarative areas such as the prefrontal cortex. In short, although action and cognition may share a good deal of both neural architecture and mechanisms for learning, it is a separate question as to whether predictions acquired for motoricity can be reused for explicit cognitive reasoning.

## 6 Place Cells and Prediction in the Hippocampus

In general, the hippocampus exhibits the most recurrence of any brain region (Rolls & Treves, 1998; Kandel et al., 2000). Neuroscientists generally agree that this is essential for the pattern storage and completion/retrieval that underlies associative learning (Rolls & Treves, 1998). When these associations include a temporal component, they become predictive; and indeed, the hippocampus is also touted as a premier predictive area of the brain, particularly with respect to navigation (Burgess & O'Keefe, 2003; Gluck & Myers, 1989).

This is probably best exemplified by the well-documented phenomenon of *phase precession* in hippocampal place cells (Burgess & O'Keefe, 2003), whereby a neuron that codes for location L begins to fire (predictively) at locations prior to L along a familiar path. The formation of these predictive links between place cells coding for successive locations along an often-travelled route is convincingly explained by Mehta (Mehta, 2001), who shows that standard spike-timing dependent plasticity (STDP) on the synapses between place cells can form asymmetric place fields, such that a neuron is highly active prior to arrival at its place field, but inactive immediately afterwards.

Note that place cells are a particularly tacit example of declarative representation in that specific neurons fire when the animal resides in a particular location (L). Phase precession is an equally compelling example of declarative prediction, since the place cell fires on the approach to L. Furthermore, phase precession (and thus the predictions underlying it) may play a much deeper role in cognition, due, in part, to the dual roles of the hippocampus in both navigation and general memory formation.

As shown in Figure 6, as a rodent moves along the corridor from locations A to G, STDP could easily lead to the formation of predictive connections between the place cells for each location, particularly when the intervals between



arrival at successive locations are in the range of 0 - 50 msec (the time window for STDP).

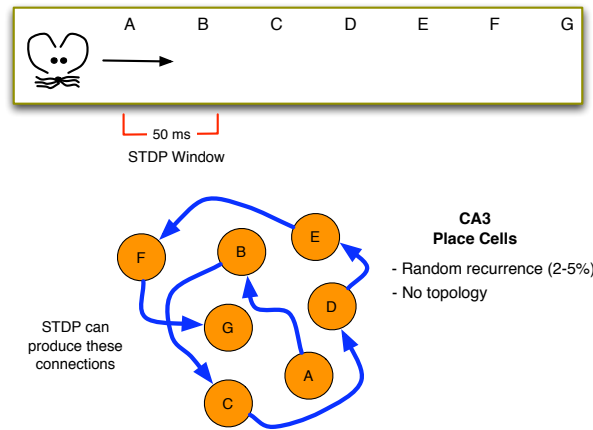


Figure 6: (Top) A corridor along which a mouse runs, with landmarks (A - G) encountered approximately every 50 msec. (Bottom) A hypothetical connection pattern, formed via STP, among CA3 place cells for landmarks A-G.

Once formed, these synapses can be activated in sequence, with gaps much smaller than 50 msec. Lisman and Redish (Lisman & Redish, 2009) have shown that gamma waves (40-100 Hz) elicit this fast replay, with these high-frequency oscillations riding atop the slower (6-10 Hz) theta waves that characterize hippocampal activity. As shown at the top of Figure 7, during one theta cycle, a good many *successive* place cells can be activated, each by a gamma peak. At the peak of each theta cycle, the current location determines the active place cell, but throughout the remainder of the cycle, succeeding place cells activate in a predictive manner.

This rapid sequencing brings several place-cell firings within the 0-50 msec window of STDP. So, for example, after cell A fires due to the current sensory input (as shown on the far left of the wave in Figure 7), cells B, C, D and E will fire in rapid succession via gamma stimulation. STDP then dictates that synapses from A to B, C, D and E will all experience LTP, thus forming cell assemblies that manifest information chunking.

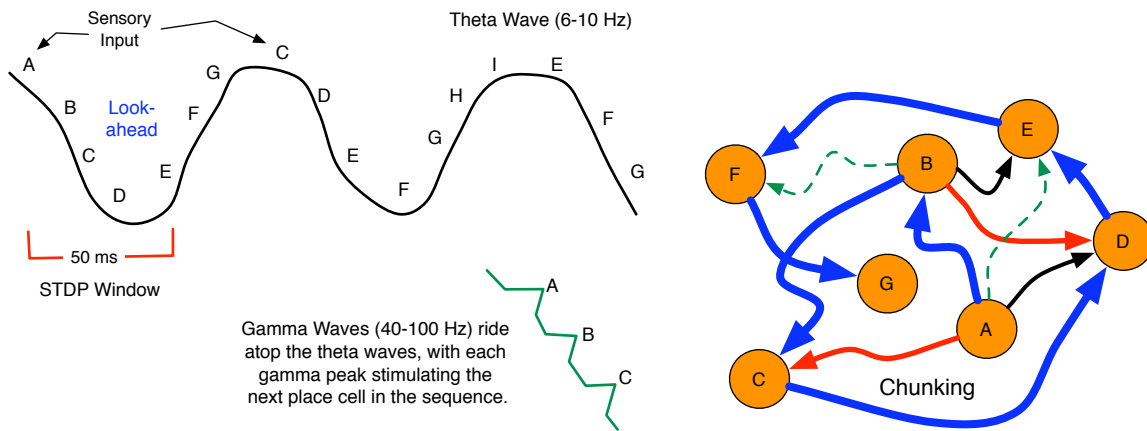


Figure 7: (Left) Gamma waves riding atop theta waves stimulate place cells in rapid succession. (Right) Rapid stimulation facilitates predictive chunking of temporally-related contexts. Stronger connections are drawn with thicker lines.

The dual role of the hippocampus in both navigation and general memory formation raises the obvious question of whether this place-cell chunking could manifest general information binding and integration. So the aggregated sequences could represent locations, steps in a procedure, words in a memorable phrase or song melody, or even

the successive interpretations of a perceived pattern (as discussed above in relation to GDPN-like models of the thalamocortical loop). In all such cases, the same basic predictive machinery (grounded in the dynamics of STDP) combines with gamma-induced replay (and further STDP) to produce tightly-linked neurons and neural firing patterns which may represent episodes or concepts.

The leap from place cells to concepts is elaborated by Buzsaki (Buzsaki, 2006), who begins by describing an interesting property of place-cell learning: when mice move back and forth along a corridor, the same location (depicted as "C" in Figure 8) binds to two different place cells ( $C_1$  and  $C_2$ ) depending upon the direction of travel. Hence, to the mouse, these are two different locations. Conversely, in an open arena, without the constraints to movement imposed by corridor walls, the place cells tend to be omnidirectional: the same cell fires, regardless of the angle of approach. Initially, they are unidirectional, as in the corridor, but with continued exploration of the arena, a unique place cell begins to represent the same location, without directional bias. Dragol et. al. (Dragol, Harris, & Buzsaki, 2003) explains this situation with evidence of the continuous mapping and re-mapping of place cells to spatial fields (via LTP); repeated trials in an arena environment could easily stimulate this re-mapping. Thus, many approach *episodes* become bound to the same place cell, a process fitting of the term *generalization*.

Buzsaki (Buzsaki, 2006) likens this generalization over experienced episodes to concept formation, wherein the invariants of many specific scenarios are eventually distilled into a general-purpose representation: a concept, which can be metaphorically described as a thought arrived at from many different paths of reasoning.

Since the hippocampus plays a key role in both navigation and memory consolidation (Andersen, Morris, Amaral, Bliss, & O'Keefe, 2007; McClelland, McNaughton, & O'Reilly, 1994), place cells and their interconnections could provide the substrate for both a) spatial recognition, look-ahead and chunking, and b) general concept formation. And in both cases, the predictive links formed by STDP provide a fundamental starting point. So if Buzsaki's analogy is correct, the hippocampus could be a critical junction between an advanced form of sensorimotor behavior (i.e. navigation), and some of the highest cognitive faculties: abstraction and concept formation.

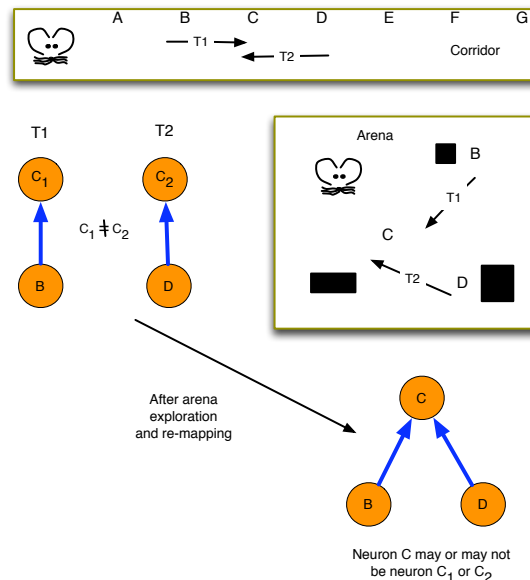


Figure 8: (Left) Unidirectional place-cell formation in a corridor environment. (Right) Transition of unidirectional to omnidirectional place cells in an arena environment.  $T_1$  and  $T_2$  are distinct temporal intervals, and B,C and D are locations.

## 7 The Predictive Basis of Cognitive Incrementalism

A naive argument for the predictive basis of high-level cognition is the following:

1. Complex movement requires an ability to predict the immediate future.

2. Procedural prediction involves various mappings between neural ensembles representing contexts and those linked to actions.
3. These mappings constitute basic common sense, a key foundation for high-level reasoning.
4. Thus, the demands of movement necessitate prediction, which then provides the basis for cognition.

Unfortunately, as discussed in (K. Downing, 2005; K. L. Downing, 2007) and implied by the two distinct topological frameworks above, procedural abilities do not translate directly into declarative knowledge. The evolutionarily later, more declarative, brain regions such as the hippocampus and cortex did not arise to help bring procedural processes (in the cerebellum and basal ganglia), and the predictions that they implicitly embody, to conscious awareness. Rather, their evolutionary arrival enhanced intelligence by adding more complex machinery for completing patterns, attending to patterns, and associating patterns, the latter of which embodies declarative forms of prediction.

Based on evidence from neuroscience, the more likely scenario involves a relatively clean separation between procedural and declarative predictive knowledge, with the former being a) implicit in the temporal delays and context-to-action wiring patterns of largely feed-forward circuits in the cerebellum and basal ganglia, and b) cognitively impenetrable. Conscious, declarative actions are involved in putting the body in situations where procedural skills can be honed (e.g. the driving range or putting green of a local golf course), but the implicit predictive knowledge learned through this training is not directly accessible to conscious analysis. To figure out *what the body has learned*, the conscious agent would probably have to watch a videotaped version of herself hitting a golf ball and speculate as to the exact nature of the predictive *rules* employed by the cerebellum and basal ganglia.

So procedural knowledge is finely tuned for skills but relatively useless with respect to a key element of cognition: a common-sense understanding of the physical world. Yet, it obviously plays a key supporting role in the brain-body-environment coupling, a complex linkage that is essential to the acquisition of declarative predictive knowledge - which resides in different brain areas and is acquired by different means than the procedural variety.

However, at a level of intelligence just above pure sensorimotoricity, that of navigation, a new cognitive demand arises: recognizing and anticipating spatial locations. This may constitute the key predictive gateway to high-level cognition. As discussed above, declarative concept formation may exapt mechanisms for place-cell mediated navigation, namely a) STDP-based predictive linkage, b) gamma-cycle-driven replay, and c) the consequent chunking. In short, the predictive sequencing and aggregation underlying navigation may also manifest one of the most advanced cognitive faculties: abstraction. Thus, the hippocampus may constitute the key juncture between prediction and cognition.

The basal ganglia also play a key role in predictive sequence learning (Houk, 1995; Kandel et al., 2000), of both a physical and cognitive nature, but possibly in a more rote manner. The basal ganglia appear instrumental in memorizing the words of a song, and even multiplication tables (Lakoff & Nunez, 2000), but the knowledge embodied therein has a very *unidirectional* characteristic: one cannot reach it from many different reasoning angles. For example, the fourth line of a song can often only be recalled by singing the first three lines. So multi-directional access to information may be a unique property of the more declarative cerebral areas, and such access seems essential for the sophisticated cognitive manipulations (e.g., symbol processing) that are the trademark of higher intelligence.

In short, the link between the predictive machinery and common-sense knowledge is not as obvious as it may seem. Knowledge in the brain has many forms, only some of which are cognitively penetrable, with a smaller subset permitting omnidirectional access. Yet both implicit and explicit forms constitute our complete predictive competence. Proponents of cognitive incrementalism cannot lay claim to all sensorimotor-enhancing predictions as key stepping-stones to high-level cognition. But vigilant monitoring of neuroscientific progress in declarative areas such as the hippocampus may eventually reveal the reuse of relatively low-level predictive mechanisms (concerning basic spatiotemporal relationships) for our most advanced reasoning and learning skills. Though less aesthetically pleasing than the naive argument above, this alternate framework of reuse still serves as a strong indicator of the evolutionary emergence of higher cognition from the basic predictive mechanisms necessary to survive in the world.

## References

- Allman, J. (1999). *Evolving brains*. New York, NY: W.H. Freeman and Company.
- Andersen, P., Morris, R., Amaral, D., Bliss, T., & O'Keefe, J. (2007). *The hippocampus book*. New York, NY: Oxford University Press.

- Artola, A., Brocher, S., & Singer, W. (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature*, *347*, 69-72.
- Burgess, N., & O'Keefe, J. (2003). Hippocampus: Spatial models. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (p. 539-543). Cambridge, MA: The MIT Press.
- Buzsaki, G. (2006). *Rhythms of the brain*. New York, NY: Oxford University Press.
- Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. Cambridge, MA: The MIT Press.
- Deacon, T. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: W.W. Norton and Company.
- Downing, K. (2005). Tantrix: A minute to learn, 100 (genetic algorithm) generations to master. *Genetic Programming and Evolvable Machines*, *6*(4), 381-406.
- Downing, K. L. (2005). The predictive basis of situated and embodied artificial intelligence. In *Gecco '05: Proceedings of the 2005 conference on genetic and evolutionary computation* (pp. 43-50). New York, NY, USA: ACM. doi: <http://doi.acm.org/10.1145/1068009.1068016>
- Downing, K. L. (2007). Neuroscientific implications for situated and embodied artificial intelligence. *Connection Science*, *19*(1), 75-104.
- Downing, K. L. (2009). Predictive models in the brain. *Connection Science*, *21*(1), 39-74.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex? *Neural Networks*, *12*, 961-974.
- Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinions in Neurobiology*, *10*(6), 732-739.
- Dragol, G., Harris, K., & Buzsaki, G. (2003). Place representation within hippocampal networks is modified by long-term potentiation. *Neuron*, *39*(5), 843-853.
- Edelman, G., & Tononi, G. (2000). *A universe of consciousness*. New York, NY: Basic Books.
- Fuster, J. (2003). *Cortex and mind: Unifying cognition*. Oxford: Oxford University Press.
- Gluck, M., & Myers, C. (1989). *Gateway to learning: An introduction to neural network modeling of the hippocampus and learning*. Norwell, Massachusetts: Kluwer Academic Publishers.
- Granger, R. (2006). Engines of the brain: The computational instruction set of human cognition. *Artificial Intelligence Magazine*, *27*(2), 15-32.
- Hawkins, J. (2004). *On intelligence*. New York: Henry Holt and Company.
- Houk, J. (1995). Information processing in modular circuits linking basal ganglia and cerebral cortex. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 3-9). Cambridge, MA: The MIT Press.
- Houk, J., Adams, J., & Barto, A. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (p. 249-270). Cambridge, MA: The MIT Press.
- Houk, J., Davis, J., & Beiser, D. (1995). *Models of information processing in the basal ganglia*. Cambridge, MA: The MIT Press.
- Ito, M. (1982). Cerebellar control of the vestibulo-ocular reflex – around the flocculus hypothesis. *Annual Review of Neuroscience*, *5*, 275-297.
- Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principles of neural science*. New York, NY: McGraw-Hill.
- Kettner, R., Mahamud, S., Leung, H., Sitkoff, N., Houk, J., & Peterson, B. (1997). Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *Journal of Neurophysiology*, *77*(4), 2115-2130.
- Lakoff, G., & Nunez, R. (2000). *Where mathematics comes from*. New York: Basic Books.
- Lisman, J., & Redish, A. (2009). Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society of London*, *364*(1521), 1193-1201.
- Llinas, R. R. (2001). *i of the vortex*. Cambridge, MA: The MIT Press.
- McClelland, J., McNaughton, B., & O'Reilly, R. (1994, March). *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. No. PDP.CNS.94.1). Carnegie Mellon University.
- Mehta, M. (2001). Neural dynamics of predictive coding. *Neuroscientist*, *7*(6), 490-495.
- Prediction. (2009). *Merriam-webster's collegiate dictionary*. Merriam-Webster Inc.

- Rodriguez, A., Whitson, J., & Granger, R. (2004). Derivation and analysis of basic computational operations of thalamocortical circuits. *Journal of Cognitive Neuroscience*, *16*(5), 856–877.
- Rolls, E., & Treves, A. (1998). *Neural networks and brain function*. New York: Oxford University Press.
- Squire, L., & Zola, S. (1996). Structure and function of declarative and nondeclarative memory systems. *Genetic Programming and Evolvable Machines*, *93*, 13515 – 13522.
- Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, Massachusetts: Sinauer Associates.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Wallenstein, G., Eichenbaum, H., & Hasselmo, M. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neuroscience*, *21*(8), 317–323.
- Wolpert, D., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338-347.