Hierarchical Error Representation: A Computational Model of Anterior Cingulate and Dorsolateral Prefrontal Cortex

William H. Alexander
william.alexander@ugent.be
Department of Psychological and Brain Sciences, Indiana University,
Bloomington, IN 47405, U.S.A., and Ghent University, Department
of Experimental Psychology, B-9000 Gent, Belgium

Joshua W. Brown
jwmbrown@indiana.edu
Department of Psychological and Brain Sciences, Indiana University,
Bloomington, IN 47405, U.S.A.

Anterior cingulate and dorsolateral prefrontal cortex (ACC and dlPFC, respectively) are core components of the cognitive control network. Activation of these regions is routinely observed in tasks that involve monitoring the external environment and maintaining information in order to generate appropriate responses. Despite the ubiquity of studies reporting coactivation of these two regions, a consensus on how they interact to support cognitive control has yet to emerge. In this letter, we present a new hypothesis and computational model of ACC and dlPFC. The error representation hypothesis states that multidimensional error signals generated by ACC in response to surprising outcomes are used to train representations of expected error in dlPFC, which are then associated with relevant task stimuli. Error representations maintained in dlPFC are in turn used to modulate predictive activity in ACC in order to generate better estimates of the likely outcomes of actions. We formalize the error representation hypothesis in a new computational model based on our previous model of ACC. The hierarchical error representation (HER) model of ACC/dlPFC suggests a mechanism by which hierarchically organized layers within ACC and dlPFC interact in order to solve sophisticated cognitive tasks. In a series of simulations, we demonstrate the ability of the HER model to autonomously learn to perform structured tasks in a manner comparable to human performance, and we show that the HER model outperforms current deep learning networks by an order of magnitude.

1 Introduction

Working memory for task sets and strategies is thought to depend heavily on the dorsal lateral prefrontal cortex (dlPFC; Nee & Brown, 2012). While it

is known that dorsolateral prefrontal cortex (dlPFC) cells show sustained activity that is thought to underlie working memory (Niki & Watanabe, 1979), little is known about how dlPFC representations are formed (Seger & Miller, 2010) or how they self-organize into hierarchical representations (Badre & D'Esposito, 2009; Badre & Frank, 2012; Badre, Kayser, & D'Esposito, 2010; Collins & Frank, 2013; Koechlin, Ody, & Kouneiher, 2003; Nee, Jahn, & Brown, 2013). Furthermore, it is not clear how these self-organized hierarchical representations learn exactly when to store information related to task set, for how long to maintain the information, and when to release the information. Some models have addressed some of these issues (Collins & Frank, 2013; O'Reilly & Frank, 2006), but currently it is unclear how working memory representations of task-related stimuli can efficiently self-organize into hierarchies and at the same time learn when and for how long to maintain specific information that is needed for optimal task performance. Here we propose a model of the interactions of the dlPFC and anterior cingulate cortex (ACC) that addresses both of these issues. The new model, based on results from empirical and theoretical neuroscience, is able to autonomously learn complex cognitive tasks in a manner comparable to humans, as well as providing substantial performance improvements over existing machine learning methods.

Coactivation of ACC and dlPFC is commonly observed in neuroimaging studies of cognitive control. Although the regions are thought to interact to support top-down control (MacDonald, 2000), a quantitative account of such interactions remains elusive. Qualitative interpretations suggest that ACC may signal the need for control that is then implemented by dlPFC (Botvinick, Braver, Barch, Carter, & Cohen, 2001), that ACC indicates the need for shifts in attention to external stimuli or task sets associated with behavior (Bush, Luu, & Posner, 2000; Donoso, Collins, & Koechlin, 2014; Kolling, Behrens, Mars, & Rushworth, 2012; Shima & Tanji, 1998), which are then executed by dlPFC, or that ACC selects action plans executed by dlPFC and monitors the results of actions (Holroyd & Yeung, 2012). Despite these proposals, efforts to develop computational models that are able to describe and predict interactions are still in the early stages.

A possible reason for the lack of computational accounts regarding such interactions may be the lack of a consensus regarding the computational function of each region independent of its involvement in a larger cognitive control network. dlPFC is generally considered to be a region concerned with representing rules or task sets needed to perform a task (Miller & Cohen, 2001), though not necessarily in representing stimuli per se (Riggall & Postle, 2012), especially for tasks in which such information may need to be maintained over a period of time in order to successfully guide behavior ("working memory"; Nee, Jahn, & Brown, 2013; Niki & Watanabe, 1979). dlPFC also appears to be involved in representing higher-order task-information (Koechlin, Ody, & Kouneiher, 2003). Recent evidence indicates that dlPFC may signal a state prediction error derived from model-based

reinforcement learning (RL) algorithms (Gläscher, Daw, Dayan, & O'Doherty, 2010). Finally, dlPFC is thought to be organized along a rostrocaudal abstraction gradient, with more rostral regions representing successively more abstract task information (Badre & D'Esposito, 2007; Koechlin et al., 2003; Nee & Brown, 2013; Reynolds, O'Reilly, Cohen, & Braver, 2012).

The function of ACC has also been the subject of intense debate. The activity of ACC in cognitive control tasks has variously been attributed to error signaling and processing (Gehring, Coles, Meyer, & Donchin, 1990; Holroyd & Coles, 2002), conflict detection and resolution (Botvinick et al., 2001; Yeung, Cohen, & Botvinick, 2004), learning action values (Rudebeck et al., 2008; Walton, Devlin, & Rushworth, 2004), reward prediction (Amador, Schlag-Rey, & Schlag, 2000), predicting the likelihood of error (Brown & Braver, 2005), environmental volatility (Behrens, Woolrich, Walton, & Rushworth, 2007), and many more.

Despite this proliferation of competing accounts regarding ACC activity, a recent computational model of ACC and the surrounding medial prefrontal cortex (mPFC) has provided a comprehensive account of the role of ACC that unifies observations on ACC from fMRI, EEG, behavioral, and single-unit neurophysiology studies and across a variety of subfields in neuroscience, including cognitive, social, and affective neuroscience (Alexander & Brown, 2011, 2014). The PRO model states that ACC/mPFC learns predictions of the likely outcomes of actions, regardless of affective valence, and signals surprising deviations from expected outcomes. Given the success of the PRO model in accounting for existing data and providing an explanatory framework for studies conducted after its publication, it is possible that the two key signals attributed to ACC/mPFC by the PRO model, prediction of action outcomes and prediction error, can provide insight into the manner in which ACC/mPFC interacts with dlPFC.

The PRO model begins with several key concepts that form the basis for the expanded model below. First, predictions at the neural level are represented by neurons that are trained to reliably increase their firing just before a corresponding event is likely to occur. Events that are more likely will elicit stronger firing, and the activity peaks around the time when the event is most likely to occur. Second, outcome signals at the neural level are represented by neurons that are reliably activated as a consequence of some event. Outcomes in general may depend on prior actions taken by the agent. Finally, prediction errors (or surprise signals) are computed as the difference between a prediction and an outcome signal. In the PRO model, the negative surprise (nonoccurrence) prediction error is computed by cells that are excited by predictions and inhibited by outcomes. These neural signals collectively are the fundamental building blocks of the models described below. The neural origins of these signals have been described in the original PRO model paper (Alexander & Brown, 2011) and are also depicted in Figure 1A and equations 2.1 and 2.2 below.

How might the prediction and surprise signals generated by the PRO model inform our understanding of dlPFC? First, in the PRO model, predictions regarding the likely outcomes of actions are learned on the basis of current stimuli. Although this explains ACC activity for simple tasks in which only a single stimulus-action-outcome sequence is observed on a trial, it does not address ACC involvement in tasks, such as the 1-2AX task (Krueger & Dayan, 2009; O'Reilly & Frank, 2006), in which the outcomes of an action are highly dependent on previously observed stimuli. This constrains the possible interactions with regions outside the cingulate in that, in order for predictions generated in ACC to be appropriately contextualized, at least some region should provide inputs to ACC that in some manner modulate the predictions that are generated by ACC. Put simply, the ACC must know about not only the present but also the past events in order to generate an appropriate prediction, and empirical evidence suggests that it does in fact have information about past events (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006). Second, the error signal generated by the PRO model is used to update prediction weights in order to refine predictions, and this appears to occur within ACC without the involvement of other regions (Jahn, Nee, Alexander, & Brown, 2014). As noted above, however, the model itself learns only simple associations of a stimulus with its outcome. Still, the error signal generated by ACC may serve additional roles beyond the cingulate pertaining to the allocation of additional neural resources for contexts in which simple stimulus-action-outcome associations are insufficient to meet task demands.

Accordingly, in this letter, we propose a new hypothesis and model of ACC/dlPFC interaction. The error representation hypothesis proposes that multidimensional error signals generated by ACC are used to train error prediction representations in dlPFC, and, once trained, error representations maintained by dlPFC are a basis of working memory signals that suffice to guide task performance. Specifically, the dlPFC learns to maintain representations of stimuli that reliably co-occur with prediction errors (Friston, 2010). In turn, the dlPFC error prediction representations are deployed by ACC to refine predictions about the likely outcomes of actions. Here we formalize the error representation hypothesis in a new computational model of ACC/dlPFC, the hierarchical error representation (HER) model.

The HER model proposes that within a hierarchical organization of layers in prefrontal cortex, higher layers learn to predict the expected error of predictions made by lower layers and that predictions generated by lower layers are modulated by the error predictions generated by higher layers. Generally the two computational roles of the model, prediction error and prediction maintenance, are associated with mPFC and dlPFC, respectively. The HER model extends the PRO model of mPFC by suggesting a hierarchical organization of mPFC in which multiple regions, arranged in discrete layers following a rostrocaudal gradient (Taren, Venkatraman, & Huettel, 2011), learn predictions of progressively more abstract outcomes

and signal surprising deviations from those outcomes. Additionally, the HER model proposes that learning in dlPFC is driven by error prediction signals originating in mPFC and that dlPFC influences activity in mPFC through modulation of specific predictions generated by mPFC. In the HER model, dlPFC learns to maintain specifically the information that predicts errors. Of course, if the information is sufficient to predict that a response will be an error, then it may also be useful in biasing behavior, either toward a correct response or away from responses associated with error feedback. In this way, the dlPFC learns to represent (and maintain as sustained activity in working memory) exactly the information that is necessary to guide behavior.

2 Methods

2.1 Model Description

2.1.1 General Modeling Approach. In our previous model of ACC/mPFC (Alexander & Brown, 2011), we adopted a reinforcement learning (RL) framework based on temporal difference learning (Sutton, 1990; Sutton & Barto, 1990) in which discrete iterations of the model corresponded with passage of a certain amount of real time. This approach is distinct from classical models of associative learning (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla, 1971) in which stimuli and their consequences are effectively presented simultaneously, that is, trial-level models. In this level, we adopt an intermediate approach in which salient sensory and behavioral events are modeled sequentially (effectively an event-level model). On each iteration of the model, a stimulus is presented and values, such as working memory representations and outcome predictions, within the model are updated accordingly. After this update, a response is generated by the model (described in detail below), and feedback related to the outcome is delivered to the model, with attendant updates in adjustable weights within the model. Decay parameters, where applicable, are applied following this sequence, and the next model iteration begins. Functionally, this approach to modeling working memory components is comparable to models with real-time dynamics (O'Reilly & Frank, 2006), which nonetheless interpret working memory representations as being essentially on or off for the duration in which the item is in working memory.

The HER model (see Figure 1) is organized hierarchically in layers: functionally segregated computational blocks that interact with other layers in a highly constrained manner (see Figures 1A and 1B). Generally the hierarchical organization of the model corresponds to an increasing caudal to rostral abstraction gradient hypothesized to exist in prefrontal cortex. The exact number of discrete regions associated with this theoretical organization of PFC in a hierarchy varies between authors and papers, although observed activity in lateral PFC tends to indicate three to five hierarchical
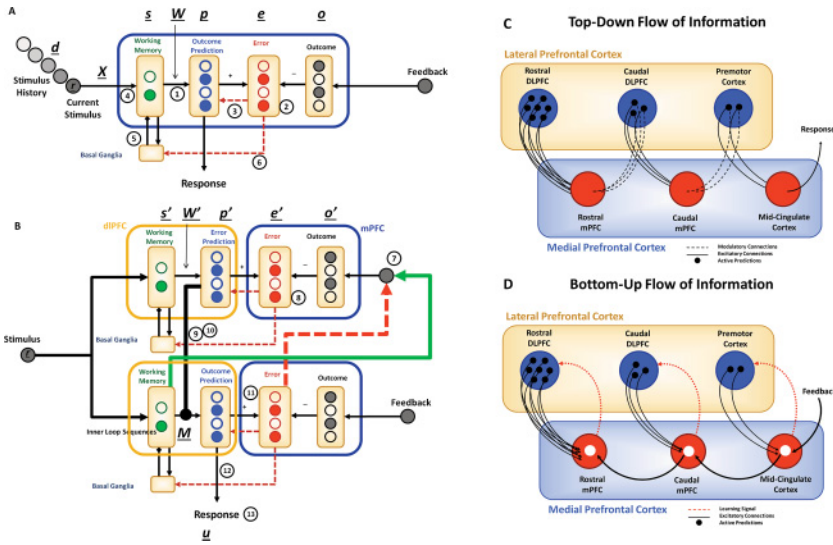
Figure 1: The HER model: Circled numbers indicate corresponding equations in the main text. Bold, underlined, italicized letters reflect equation variables. Note that the dimensions of the network depicted do not correspond to the simulations described in this letter. (A) A single layer of the HER model has the same form as the PRO model of anterior cingulate in that a stimulus representation produces predictions of multiple possible response-outcome (RO) conjunctions and is trained by a vector-valued error signal. While the PRO model formed predictions based on any stimuli that were presented to the model, the HER model includes a gating mechanism that determines whether an external stimulus will be granted access to internal model representations. The architecture of a single layer forms a motif that is repeated at additional hierarchical layers. (B) Interaction between layers in the model follows bottom-up and top-down paths. Prediction errors generated at an inferior layer serve as proxy outcomes for superior layers, which then form predictions regarding future errors that are likely to be observed given the contents of working memory (WM) at the superior layer. Such predictions are then used to modulate predictions generated by inferior layers. (C) Top-down flow of information mapped to prefrontal cortex. Rostral mPFC selects from maintained representations in rostral dlPFC in order to modulate the activity of predictions maintained in caudal dlPFC. This pattern repeats in successively posterior regions until it terminates in regions of mPFC associated with generating behavior responses. (D) The bottom-up flow of information begins with task-related feedback. Error signals are computed in regions of mPFC, which are used to refine predictions in lateral PFC. Errors calculated in posterior regions are combined with stimulus information maintained in WM (connections not shown for clarity) and used as feedback for successively anterior regions.

layers (Badre, 2008; Koechlin et al., 2003; Nee et al., 2013). In the simulations we describe, we use a version of the HER model with three layers, though in principle it is possible to add an arbitrary number of additional layers, notwithstanding hardware limitations. Although the nomenclature used to describe regions within lateral PFC also tends to vary among research groups, the discrete layers in the three-layer model may correspond roughly to the regions identified by Koechlin et al. (2003), in which the lowest layer of the HER model (layer 1) is identified with dorsal premotor cortex (PMd), and successive layers correspond to caudal and rostral dlPFC. In addition to a rostrocaudal organization, the model has a medial to lateral organization (see Figures 1C and 1D), with medial regions engaging in computation of prediction error (mPFC) and selection of currently applicable predictions, and lateral regions maintaining predictions of outcome error (dlPFC). The first layer of the hierarchy is functionally equivalent to the PRO model (Alexander & Brown, 2011); it learns to predict action-outcomes and generates error signals when a predicted outcome fails to occur. Error signals generated at each level have three functional roles (depicted in Figure 1B as dashed red arrows). First, as in the initial PRO model, error signals are used to train predictions of future outcomes within the same level of the hierarchy. Second, error signals are used to update weights in the basal ganglia (BG) gating mechanism for updating working memory using a modified Widrow-Hoff learning rule (Widrow & Hoff, 1960). The BG gating mechanism determines if stimuli presented to the model are stored in WM, similar to the mechanism proposed by O'Reilly and Frank (2006). Third, the conjunction of the error signal with the attended stimulus is used as an outcome signal for higher layers of the hierarchy. Each layer of the hierarchy is functionally similar to the PRO model; the lowest hierarchical layer learns predictions of response-outcome conjunctions, while at higher layers, the "current" stimulus is taken to be the current contents of working memory at each level, and the outcome at higher levels is the outer product of the vector-valued error signal and the stimulus vector from the previous level. Such conjunctive representations as those formed by the outer product have been observed in single cells previously, for example, the gain fields of cells in the parietal cortex (Andersen, Essick, & Siegel, 1985). By using the error signal generated at lower layers as a proxy outcome, which is then associated with items stored in WM, higher-order layers in the model learn a prediction of the likelihood of subsequently observing a given pattern of errors generated by the lower-order layer. Error signals generated by higher-order layers are derived by comparing the observed error signal from the lower-order layer and comparing it to the expected error signal. Thus, each higher level computes an "error of errors" from the lower level—hence, the "hierarchical error" component of the model's name.

In the HER model, each layer has a WM module that functions separately from WM modules at other layers of the hierarchy. Each WM module is

able to store a single stimulus at a time; in order to store a new item at a given hierarchical layer, any items already stored are removed. Initially, the "decision" made by the BG mechanism as to whether an item should be stored in WM is random; as learning progresses, the error signal generated at a particular hierarchical layer trains BG weights connecting a stimulus representation to the WM modules. Weights corresponding to items that are reliably associated with subsequent errors generated by a lower-order layer are selectively strengthened, leading to preferential storage of that item in WM. Once stored in WM, items are maintained until a new stimulus is stored.

2.1.2 Model Components. Although the model is organized into distinct hierarchical layers, each layer contains functionally identical "moving parts" that constitute a repeated motif for each of the three hierarchical layers. In order to describe overall model function, we first describe the components that are computationally identical for all layers, followed by equations governing how layers interact with one another. In the equations, we adopt the following notational conventions: bold, capitalized letters indicate matrices, bold lowercase letters indicate vectors, italicized lowercase letters indicate scalars, model parameters are indicated by Greek letters, and a superscript $^{\mathrm{T}}$ indicates the transpose of a matrix or vector. The model description focuses on the model equations and their function. A step-by-step description of how the model learns a simple hierarchical task is provided in the supplementary material, available online at http://www.mitpressjournal.org/doi/suppl/10.1162/NECO_a_00779.

2.1.3 A Note on Nomenclature. In the HER model, each layer learns to predict likely outcomes and computes errors as the difference between an expected and actual outcome. While this statement of the model's function is straightforward, some confusion is possible when considering the interactions between hierarchical layers: the outcomes to be predicted at layer 2 are the prediction errors of layer 1, and thus the terms *prediction error* and *outcome* may refer to the same signal. Similarly, predictions generated at each layer serve multiple roles: predictions are used to compute an error signal for training associative weights within each layer, as well as an outcome signal for superior layers, while they are also used to modulate predictions at inferior layers in the hierarchy (or, in the case of the first hierarchical layer, generate model responses). In the remainder of the letter, we generally adhere to the convention that the signals generated in the HER model are referred to in their role within a given layer rather than in their interlayer roles.

2.1.4 Outcome Prediction. As in the PRO model, the HER model learns predictions of the likely outcomes of actions based on current stimuli

and signals deviations from expectations. In our event-level model (see Figure 1A), this function is implemented as a simple feedforward network,

$$\mathbf{p} = \mathbf{W^T r}, \tag{2.1}$$

where $\mathbf{p}$ is a vector of predictions of outcomes, $\mathbf{r}$ is a vector of task stimulus representations, and $\mathbf{W}$ is a weight matrix associating $\mathbf{r}$ and $\mathbf{p}$. Errors are calculated as

$$\mathbf{e} = \mathbf{a}(\mathbf{o} - \mathbf{p}), \tag{2.2}$$

where $\mathbf{o}$ is the vector of observed outcomes and $\mathbf{a}$ is a filter that is 0 for outcomes corresponding to unselected actions and 1 everywhere else. This essentially prevents learning about the relationship between task stimuli and the outcomes associated with particular actions that were not chosen. Weights in the model are updated according to

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \alpha(\mathbf{e}_t \mathbf{r}_t^T), \tag{2.3}$$

where α is a learning rate parameter and t indicates the current model iteration.

2.1.5 Working Memory Gating. At each level of the hierarchy, external stimuli that have been presented to the model may be stored in WM based on the learned value of storing that stimulus versus maintaining currently active WM representations. This notion is similar to the basal ganglia gating mechanism used in the O'Reilly and Frank (2006) PBWM model, although since our goal is not to faithfully model basal ganglia, the mechanism used in the HER model abstracts away many of the neurobiological details of their model. Additionally, at each level of the hierarchy, we posit a capacity limit such that only one active WM representation may be maintained at any time. Although the WM capacity at each layer is limited to a single item, each layer may store different items in WM at each moment. The total WM capacity of the model, then, is dependent on the number of layers. In the simulations described in this letter, three layers are used, and so the total task WM capacity of the model is three. We also note that WM is not a unitary construct. Empirical studies show that WM for task rules may be decodable from the dorsolateral prefrontal cortex (Nee & Brown, 2012), but WM for sensory representations may be decodable from the sensory cortices (Riggall & Postle, 2012). Here we treat only WM for task rules, so in general, human WM capacity would include sensory WM as well and would thus have a capacity greater than three.

External stimuli are represented as a vector $\mathbf{s}$ (and are distinct from internal representations of stimuli denoted in equations 2.1 and 2.3 as $\mathbf{r}$). Upon

presentation of a stimulus, the value of **s** at a given position corresponding to a fixed, arbitrary, one-to-one mapping of task stimuli to **s**, is set to 1, indicating a currently observed stimulus. The task stimuli signified by **s** and **r** can be thought of as external and internal stimulus representations, respectively: **s** is the appearance of a letter on the screen, for example, while **r** is the neural representation elicited by that letter. The value of storing the stimulus represented by **s** in WM versus maintaining current WM representation **r** is determined as follows:

$$\mathbf{v} = \mathbf{X}^{\mathrm{T}}\mathbf{s}, \tag{2.4}$$

where **X** is a matrix of weights associating external stimuli (**s**) with corresponding WM representations (**r**). Unless otherwise specified, **X** fully connects **s** and **r**; each element in **s** is connected to every element in **r** The use of two separate stimulus representations in the model reflects the idea that the mere presentation of a stimulus does not necessarily guarantee that an individual will encode it. In order to influence behavior, the existence of an external stimulus (**s**) must first be registered by the brain (**r**). Equation (2.4) yields a vector **v** that contains the value of storing the current stimulus in WM relative to the value of maintaining the current contents of WM. Since the model maintains only one WM representation at any time for each level, the value of storing stimulus $s_i(v_i)$ is compared to the value of maintaining the current contents, r_j, of WM (v_j). These values are compared using a softmax function,

$$\text{probability of storing } s_i = \frac{(\exp^{\beta v_i} + bias)}{(\exp^{\beta v_i} + bias) + \exp^{\beta v_j}}, \tag{2.5}$$

where β is a gain parameter governing the probability of selecting the highest-valued stimulus to store in WM. If there are no active WM representations, stimuli that are currently active will automatically be stored. The bias term in equation 2.5 regulates how likely the model is to update WM at each layer; high-bias values encourage WM updates regardless of the value of **v**, while WM updates for low-bias values are based primarily on **v**

Associative weights **X** are updated through backpropagation of the error obtained in equation 2.3 (but see the supplementary material for a more biologically plausible demonstration based on reinforcement learning). Rather than training weights using the stimulus vector **s**, however, an eligibility trace vector **d** is used instead (Barto, Sutton, & Anderson, 1983; Klopf, 1972). When a stimulus i is presented, the value of d_i is set to 1, indicating a currently observed stimulus. At each iteration of the model, **d** is multiplied by a constant decay rate parameter, λ indicating gradually decaying

eligibility traces. The weight matrix $\mathbf{X}$ is then updated using the backprop-agated error:

$$\mathbf{X}_{t+1} = \mathbf{X}_t + (\mathbf{e}_t^{\mathbf{T}}\mathbf{W}_t \cdot \mathbf{r}_t)\mathbf{d}_t^{\mathbf{T}}. \tag{2.6}$$

Here the $\cdot$ operator indicates element-wise multiplication, and the term $(\mathbf{e}_t^{\mathbf{T}}\mathbf{W}_t \cdot \mathbf{r}_t)$ is the error signal backpropagated from the prediction units at each layer. The purpose of using eligibility traces (rather than currently visible task stimuli) to train weights in the gating mechanism is to allow the model to learn information regarding temporally distant events. By train-ing the gating mechanism on eligibility traces, it effectively learns when an item in WM ought to be purged in favor of storing a current stimulus. When a stimulus is not stored in WM and prediction errors are consistently observed following the failure to store the stimulus, negative weights de-velop between the stimulus and any other items that were stored in WM at the time of the prediction error. The eligibility trace allows the repre-sentation of the item that was not stored to persist over time so as to be associated with subsequent errors. Note that eligibility traces themselves cannot be stored in working memory; stimuli can be stored only on the model iteration in which they are presented. Eligibility traces of this sort have been used extensively to explain learning in conditions in which a con-ditioned stimulus (CS) and an unconditioned stimulus (US) are temporally noncontiguous (Sutton & Barto, 1990). Biologically, eligibility traces may correspond to persistent, decaying activity in neurons that enables learning but is not sufficient to evoke postsynaptic activity (Seo, Barraclough, & Lee, 2007). Alternately, an eligibility trace may be realized by a transient increase in synaptic plasticity following the activation of a neuron, which decays to baseline over time (Izhikevich, 2007)

2.2 Interactions between Levels of the Hierarchy. Having covered the components that operate at each layer independent of their involvement in a hierarchy, we now turn to describing how layers in the HER model interact with one another. As noted previously, prediction errors generated by mPFC at a given level of the hierarchy serve as "outcomes" at the next highest level, and predictions generated by the mPFC and maintained by dlPFC modulate predictions that are generated by the next lower level. We can generally refer to these two roles as bottom-up and top-down processing. As a notational convention, we will use a superscript single quote to refer to variables for hierarchical layers that are one level higher than nonsuperscripted variables (e.g., $\mathbf{p}$ and $\mathbf{p}'$ refer to predictions generated by layers that are adjacent in the hierarchy).

2.2.1 Bottom-Up Processes. For a given layer, the error signal from a lower layer serves as the "outcome" signal. At the lowest layer, outcomes reflect

conjunctions of actions and outcomes, and error is considered only for selected actions (see equation 2.2); distinct events (action and feedback) are interpreted as a single compound event, and the number of possible events is the number of possible actions multiplied by the number of possible types of feedback. Similar conjunctions occur for the outcome signals at higher layers, which are computed as the outer product of the error vector $\mathbf{e}$ and stimulus representation vector $\mathbf{r}$ at the lower level to obtain a matrix of compound events:

$$\mathbf{O}' = \mathbf{r}\mathbf{e}^{\mathsf{T}}. \tag{2.7}$$

In the case of the second layer of the hierarchy, these events are conjunctions of stimuli, actions, and outcomes. For computational convenience, this matrix is reshaped into a vector $\mathbf{o}'$. Prediction error at each layer in the hierarchy is computed using the same form as equation 2.2,

$$\mathbf{e}' = \mathbf{a}'(\mathbf{o}' - \mathbf{p}'). \tag{2.8}$$

2.2.2 *Top-Down Processes.* For a given layer, the prediction signal $\mathbf{p}'$ additively modulates stimulus-specific predictions $\mathbf{p}$, which are generated by the lower layer. Note that the number of elements in $\mathbf{p}'$ equals the number of elements in the weight matrix used to generate predictions at the lower layer. This follows from computing outcome $\mathbf{o}'$ as $\mathbf{r}\mathbf{e}^{\mathsf{T}}$. A single element in $\mathbf{p}'$ corresponds with a single element in $\mathbf{W}$ whose value denotes the strength of the association of a particular stimulus in $\mathbf{r}$ with a particular outcome in $\mathbf{o}$. In order to modulate predictive activity in lower hierarchical layers, $\mathbf{p}'$ is reshaped into a matrix ($\mathbf{P}'$) and added to $\mathbf{W}$ in order to generate a modulated prediction of likely outcomes:

$$\mathbf{m} = (\mathbf{W} + \mathbf{P}')^{\mathsf{T}}\mathbf{r}. \tag{2.9}$$

When more than two layers are included in the hierarchical structure of the model, $\mathbf{p}'$ may itself be modulated by the predictions of superior layers $\mathbf{p}''$, and thus equation 2.9 would become

$$\mathbf{m} = (\mathbf{W} + \mathbf{M}')^{\mathsf{T}}\mathbf{r}. \tag{2.10}$$

Similarly, when the predictions generated by a layer are modulated by superior layers, the errors are calculated as the difference between the current, modulated prediction and the observed outcome. Equation 2.8 then becomes

$$\mathbf{e} = \mathbf{a}(\mathbf{o} - \mathbf{m}). \tag{2.11}$$

Equation 2.11 is used to train associative weights **W** within each layer of the model; however, the error term in equation 2.2 is always used as the proxy outcome for superior layers.

2.2.3 Response Generation. In our simulated tasks, actions could result in correct or error feedback, which are learned as response-outcome conjunctions at the lowest layer of the hierarchy. In order to generate responses, for each candidate response, the model compares the (modulated) prediction of correct feedback to the prediction of error feedback,

$$u_{response} = m_{Response/Correct} - m_{Response/Error}, \tag{2.12}$$

which is then used in a softmax function to determine a response:

$$Prob(u_i) = \frac{\exp^{\gamma u_i}}{\sum \exp^{\gamma \mathbf{u}}}, \tag{2.13}$$

where γ is a gain parameter. Because the presentation of feedback to the model is in the form of RO conjunctions, the predictions generated at the lowest hierarchical layer contain information regarding both affectively positive ("Correct") and negative ("Error") outcomes, as well as the responses that are likely to result in each. It is an assumption of the model that behaving agents will tend to avoid negative events, even in the absence of an affectively positive alternative, and pursue positive events; thus equation 2.12 incorporates both affective components of the prediction in determining responses.

Each layer of the HER model contains four free parameters, plus an additional global parameter for response selection gain. Parameters were hand-tuned to yield acceptable performance, where "acceptable" is an arbitrary, modeler-defined metric indicating that the modeler has stumbled on a region in the parameter space for which the model appears to learn a simulated task more quickly and reliably than was previously observed by the modeler.

2.2.4 Flat Model. In order to assess the importance of the hierarchical structure of the model in learning, a flat version of the model was implemented. In this version of the model, the modules described above were each connected directly to the response-outcome units, and error feedback for all modules was the observed outcome minus the predicted outcome. Unlike the hierarchically organized model, individual modules in the flat model did not interact with each other except through the outcome predictions generated by each module, which were summed to obtain an overall prediction, generate model responses, and compute error. Other than these

Table 1: Model Parameters for the 1-2AX Simulations.

			Value for:		
Parameter	Description	Equation	Layer 1	Layer 2	Layer 3
α	Learning rate	3	0.075	0.075	0.075
λ	Eligibility trace decay	6	0.1	0.5	0.99
β	Working memory update gain	5	15	15	15
Bias	Bias for updating working memory	5	1	0.1	0.01
γ	Response selection temperature	13	15	NA	NA

Table 2: Model Parameters for the Generic Hierarchical Task.

			Value for:		
Parameter	Description	Equation	Layer 1	Layer 2	Layer 3
α	Learning rate	3	0.05	0.02	0.02
λ	Eligibility trace decay	6	0.3	0.5	0.9
β	Working memory update gain	5	12	14	14
bias	Bias for updating working memory	5	0	0	0
γ	Response selection temperature	13	12	NA	NA

changes, all equations were identical to those described above, and parameters were as reported in Table 2.

2.3 Model Simulations. For all model simulations, a version of the HER model with three layers (parameterized as indicated in Tables 1 and 2) was used, with a total of 13 parameters.

2.3.1 Simulation 1: 1-2AX Continuous Performance Task. The AX and 1-2AX continuous performance task (CPT) are frequently used in the study of WM and in diagnosing behavioral and cognitive deficits related to WM dysfunction (Carter et al., 1998; Servan-Schreiber, Cohen, & Steingard, 1996). The 1-2AX CPT is a hierarchically organized task (see Figure 2A) in which a subject's response to a target cue (X or Y) is governed by both a "pattern" cue that immediately preceded it (A or B), as well as a "context" cue (1 or 2), that indicates which pattern cue-target sequence (AX or BY) is valid at any given time. Sequences of stimuli may be thought of as being organized in inner and outer loops (see Figure 2B), where inner loops are composed of two-stimulus sequences with A or B followed by X or Y, and outer loops are the sequence of inner loops followed by the presentation of a context cue. We simulated the HER model on a version of the 1-2AX task as described in O'Reilly and Frank (2006) in which each outer loop consisted of one to four inner loops and the probability of observing a valid sequence for each
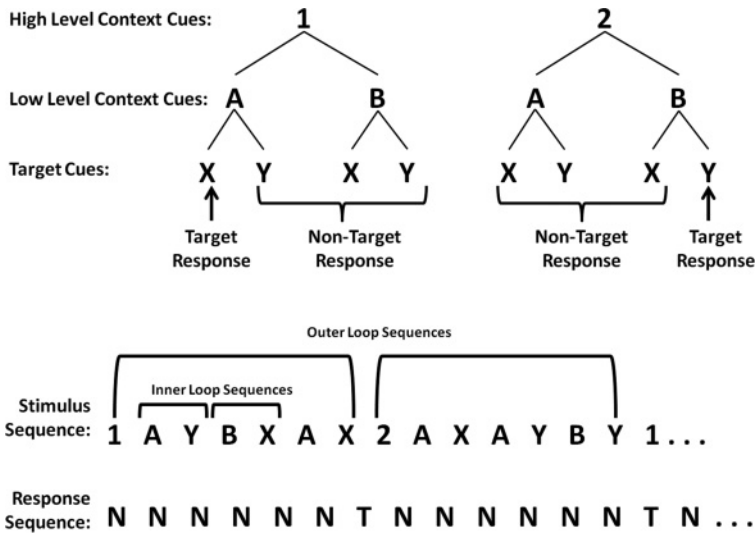
Figure 2: 1-2AX CPT.

inner loop was 0.25. There were eight inputs to the model, corresponding to the six relevant cues in the task, as well as two distractor cues that had no task relevance (corresponding to C and Z cues as implemented in O'Reilly & Frank, 2006). At each cue, the model made a response to indicate whether the current stimulus was a target. In order to perform the task correctly, target responses should be made only at the presentation of a valid target cue; all other cues should result in nontarget responses. Feedback to the model indicated correct or incorrect performance. One thousand simulations of the HER model performing the 1-2AX task were conducted for 4000 outer loop sequences (equivalent to 160 training epochs in O'Reilly & Frank, 2006, or approximately 24,000 individual cue presentations). The model was considered to have successfully learned the task on the first of 1000 consecutive cue presentations (approximately seven epochs) in which no response errors were made.

2.3.2 *Simulation 2: Atemporal Task Structure.* In the 1-2AX CPT, task structure is partially enforced by the sequential presentation of individual cues. On each presentation of a single task stimulus, the model may elect (or not) to gate only the current item into WM at each layer of the model, potentially simplifying the problem of learning to represent task structure. Structured tasks in which information related to every hierarchical layer is presented simultaneously (Badre et al., 2010) pose an additional problem in that the model needs to learn in parallel (rather than serially) the mapping of information for the entire task structure to the appropriate model layer.

Generally, hierarchically structured tasks (either temporal or atemporal) reported in the literature involve presentation of a compound stimulus composed of at least two dimensions, with all dimensions being relevant to identifying an appropriate response. Stimulus dimensions are typically treated as categorical, even in cases in which a continuous interpretation is possible. The orientation of an image, for example, may be continuous, but generally only a few, well-differentiated orientations would be used in a task, and the spatial relation between two orientations has no bearing on the appropriate response. Accordingly, in our structured task simulations, stimulus dimensions are treated as categorical, and the number of categories of each dimension is manipulated to derive different structured tasks. While compound stimuli can in some cases be treated as noncompound (Kehoe, 1986), in our simulations, values for a particular stimulus dimension are treated as being independent.

An additional point in discussing task structure is to question what is meant by structure in the first place. For structured tasks in which multiple stimulus dimensions inform an eventual response, one dimension is generally regarded as signifying an abstract rule that governs how one responds to other dimensions. In tasks with a temporal component, such as the 1-2AX task, this relationship is partially enforced by the order in which stimulus components are presented, with more abstract stimulus dimensions preceding concrete dimensions. In the absence of such temporal structure, it is unclear whether and how one stimulus dimension may come to function as a rule and another function as the concrete stimulus. One possibility is that this relationship among stimulus dimensions is determined by the properties of the dimensions themselves: stimulus dimensions that may take a greater number of values relative to other stimulus dimensions may be more or less likely to function as the rule for a given task. An alternative possibility is that the relationship of each stimulus dimension with associated responses determines whether a particular dimension functions as a rule or concrete cue in a structured task. In order to investigate the ability of the model to successfully acquire structured tasks in which information is presented without the additional benefit of temporal structure as well as how the model learns to map stimulus dimensions to hierarchical layers, we simulated the model on a class of generic structured tasks (see Figure 3) in which two notional stimulus dimensions (see Figure 3A), each having multiple possible values, were used. The number of distinct values for each stimulus dimension was varied from two to seven, resulting in 36 (6×6) distinct structured tasks. For each task, the number of possible responses was determined by the dimension with the highest dimensionality, and response mappings were permuted a number of times equal to the dimensionality of the second stimulus (while ensuring that response mapping permutations were nonoverlapping).

In the notional task depicted in Figure 3, each compound stimulus consists of two dimensions, shape and orientation, with dimensionalities of two
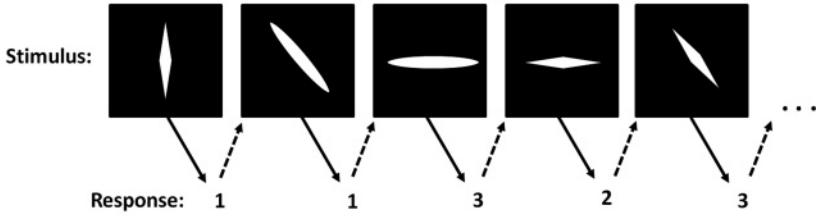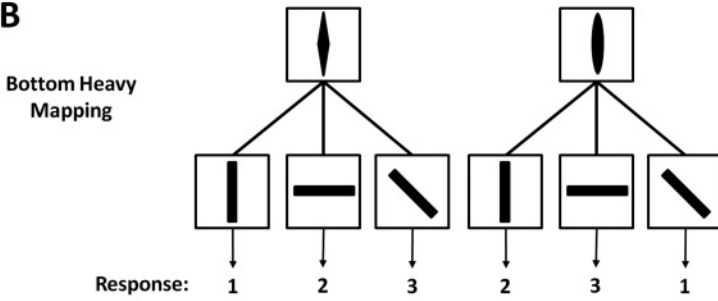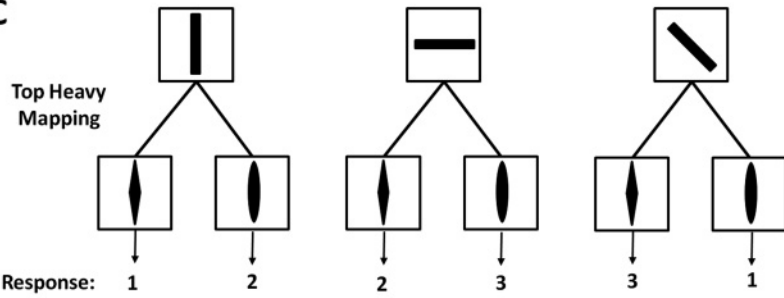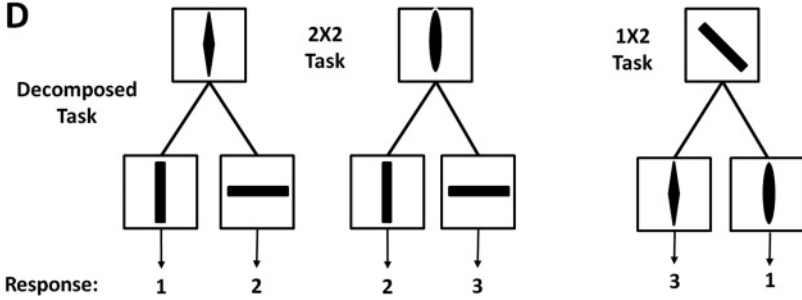
and three, respectively. This results in a task with three possible responses and two response mappings. Additional values could be included for each dimension of the stimulus (i.e., additional shapes, additional orientations) to generate new hierarchical tasks. It should be noted that structured tasks of this sort can be solved using a generic backpropagation network with a single hidden layer (functionally similar to a single layer of the HER model). Although such networks are not hierarchically structured, they are able to learn hierarchical tasks through learning conjunctions of features, represented by activity in hidden units, which can then be used to generate appropriate responses. In order to ensure that the model learned the task using the hierarchical model architecture, and not through representing conjunctions of features in the WM representation (similar to a hidden layer), weights from external stimuli (**s**) to internal representations (**r**) used a one-to-one mapping rather than the fully connected mapping as in the previous simulation. This ensures that the activity of each unit in the WM module is unable to represent feature conjunctions and instead represents only a single feature.

We conducted 10 "experiments" in which the model was simulated performing each of these tasks for 10,000 trials, and for each experiment, 100 simulations (each corresponding to a single subject) were performed. Each of the 10 experiments was conducted under two separate conditions: one in which the model learned the mapping of stimulus to hierarchical model layer autonomously and another in which each of the two stimuli were forced into a specific mapping (e.g., dimension 1 was always stored at the lowest hierarchical layer and dimension 2 at the second layer; see Figures 3B and 3C). On each trial, the value of elements in the binary feature vector

---

Figure 3: Generic hierarchically structured task. (A) A notional structured task in which subjects are presented with a series of oriented shapes and required to learn appropriate responses (see Badre et al., 2010). A task is structured if information about one dimension of the stimulus can be used to inform how a subject ought to respond to the second dimension. (B) A bottom-heavy hierarchical task structure. The stimulus dimension having two possible values (diamond or ellipse) serves as a context cue informing responses to the second stimulus dimension (orientation), which has three values. (C) A top-heavy task structure, in which the stimulus dimension with a greater number of possible values serves as the higher-level context cue. Note that in both structures, the responses for each unique combination of stimulus dimension (e.g., vertical and diamond) are identical. Hierarchically structured tasks of this sort can be extended to an arbitrary number of dimensions. (D) A mapping in which the task has been decomposed into two simpler tasks, a 2×2 and 1×2 task, and in which the stimulus dimension used as the higher-level context cue in one component is different from the dimension used as the context cue in the other component.

that serves as input to the model (**s**; see equation 2.4) corresponding to each of the stimulus dimensions was set to 1; that is, on each trial, the model was presented with two possible items eligible to be stored in WM, though it was still restricted to storing only a single item in the WM module of each hierarchical layer. To assess performance, the model was considered to have completed learning of the structured task on the first of 1000 consecutive trials in which no errors were committed.

2.3.3 Parameter Influence. Model parameters for both simulations were hand-tuned to yield acceptable performance. In order to provide insight into how the parameterization of the model contributes to its ability to learn the tasks described above, additional simulations were carried out to explore the performance of the model in different regions of the parameter space. The model was simulated on the 1–2AX task while the values of the λ and bias parameters were varied. These parameters are related to the temporal representation of a stimulus and the decision to update WM or maintain an item during ongoing performance, and thus the discrete nature of the atemporal structured task is not suitable for investigating these parameters. The values for parameters α and β, governing learning rate and WM update gain, were manipulated for simulations in which the model performed three variations of the structured task ($2 \times 2, 2 \times 3, 3 \times 3$). Simulations were conducted according to the previous descriptions and, with the exception of the manipulated parameters, parameter values were identical to those reported in Tables 1 and 2. Only one set of parameters (e.g., the three values for alpha at each hierarchical layer) was manipulated at a time.

3 Results

3.1 Simulation 1: 1-2AX CPT. In our first simulation, we demonstrate the ability of the HER model to successfully solve the 1-2AX CPT. Of the 1000 simulations conducted, the model successfully met criterion all 1000 times. On average, the number of cue presentations to criterion was 4491.6 (standard deviation 2756.2; median = 3691.5; interquartile range = 2606.5), equivalent to approximately 30 epochs in other reports (Krueger & Dayan, 2009; O'Reilly & Frank, 2006). Notably, the model significantly outperforms other computational models of WM, including the PBWM and LSTM (Hochreiter & Schmidhuber, 1997) models, as well as an LSTM model trained using a manual shaping procedure (Krueger & Dayan, 2009). If instead we use Krueger and Dayan's more stringent criterion, we find the HER model fails to meet criterion 1.2% of the time, while successful runs took an average of 4940.7 (median = 4107) cue presentations to reach criterion (standard deviation 3465.6, interquartile range = 3274). Using the more lenient criterion of two consecutive epochs with no errors as described in O'Reilly and Frank (2006) yields an average time to criterion of 2902.6

trials (approximately 19 epochs; standard deviation = 1340 trials, median = 2503 trials, interquartile range = 1426), compared to the approximately 300 epochs required by the PBWM and LSTM networks in O'Reilly and Frank.

Although the performance of the model on the 1-2AX task exceeds that of other models capable of learning the task, in comparison to human performance, the HER model requires many more trials to learn the task. In thesis work reported by Krueger (2011), human subjects were able to learn the 1-2AX task in approximately five epochs. One potential reason for this is that humans may use the information provided by the stimulus identity—letters and numbers—to aid in learning. Letter stimuli that fill similar functional roles, such as A and B, in the task are similar to one another in the sense of being sequentially adjacent in most alphabets. However, Krueger additionally reported the results of an experiment with the same overall task structure of the 1-2AX task, but in which stimuli were selected without such a preexisting relationship. This arrangement is similar to the binary feature vector used for representation in the HER model, in which representations corresponding with, for example, 1 and 2, are orthogonal, and thus is a more accurate comparison of model and human performance. The HER model was simulated on the adjusted 1-2AX task as reported in Krueger (2011) in the same fashion as described for simulations of the more typical 1-2AX task. Using the criterion described in that thesis of approximately 90% accuracy on potential target trials (X or Y) over two epochs, we find that the HER model meets criterion in a similar number of stimulus presentations as human subjects: 13.01 epochs to sustain 90% accuracy over two epochs compared to the 12 epochs needed by human subjects. Moreover, while only 25% of the human subjects met criterion, the HER model met criterion on this more difficult version of the 1-2AX task 100% of the time over 1000 simulations. Thus, the HER model matches human-level performance with regard to the speed with which it learns and outperforms naive subjects in being able to consistently solve the task.

A likely reason for this increase in performance is the organization of the model into hierarchical layers, which maps well to the innate hierarchy of the 1-2AX task. At the lowest layer of the hierarchy, the model learns base response rates associated with each cue and disregarding any additional information that may be supplied by higher-order cues. At the highest layer of the hierarchy, the WM component of the model learns to store and maintain context cues (1 and 2) over multiple inner loop sequences and updates only when a new context cue is presented. At the middle layer, WM updates following the presentation of an inner loop cue (A or B) and maintains the representation only until after the following cue has been presented. In fMRI studies using the 1-2AX task, this pattern of WM updating has been observed in human dlPFC (Nee & Brown, 2013).

Additionally, the model displays behavior similar to that of humans learning complex hierarchical tasks in that rather than following a smooth

trajectory, learning curves are observed to plateau at points before increasing again (see Figures 4A and 4B). Error rates in the model are observed to plateau as discrete components of the task are acquired. Following learning of the base response rates by the lowest layer (see Figure 4C), a second interval occurs during which the model learns the relationship of inner loop cues and target cues (see Figure 4D). Following acquisition of this relationship, another interval follows during which the model learns the relationship of context cues to inner loops sequences (see Figure 4E). This pattern of task acquisition is similar to behavioral shaping paradigms in which concrete stimulus-response relationships are trained before more abstract relationships, facilitating acquisition of a task relative to attempting to learn the complete task from the beginning. In this sense, the HER model can be considered to shape its own behavior. It is important to note that although the model learned to store more abstract cues at the highest hierarchical layer and more concrete cues at the lowest level of the model, this is an emergent rather than a prespecified property of the model. Task cues were eligible to be stored in any layer of the hierarchy when presented to the model.

The distinct learning periods observed in the model's behavior correspond to the development of weights in the model (see Figure 5). After 500 trials (see Figure 5, left panels), weights related to predicting action-outcome conjunctions in the first hierarchical layer begin to converge on the base response rates, at which point the model is incapable of improving its performance without the involvement of additional layers. Following this period, distinct patterns of errors generated by the first layer are observed by the second layer (see Figure 5, center panels). Weights in this layer reflect the association of low-level context cues A and B with errors observed during presentation of the X and Y cues, indicating that a potential target sequence has been observed. Finally, the third layer of the model learns the association between high-level context cues 1 and 2 with the errors generated by the second layer, after which the model has solved the task and performance is perfect (see Figure 5, right panels).

Figure 6 (see also the supplementary material for an extended example) depicts the top-down modulation of predictions in layer 1 by predictions in layer 2. When a low-level context cue (e.g., A in Figure 6, upper left panels) is stored in WM at layer 2, predictions associated with that cue are elicited. In the example depicted, the model predicts a positive error for a Target/Correct response/outcome and a negative error for NonTarget/Correct if an X is subsequently observed, while also predicting a negative error for a Target/Correct response and a positive error for a NonTarget/Correct response/outcome if a Y is subsequently observed. Note that since the model has solved the task, predictions regarding Error outcomes are near 0. If an X is subsequently observed (see Figure 6, lower left), layer 1 of the model weakly predicts both Target/Correct and NonTarget/Correct outcomes. Without additional contextualization, these predictions reflect
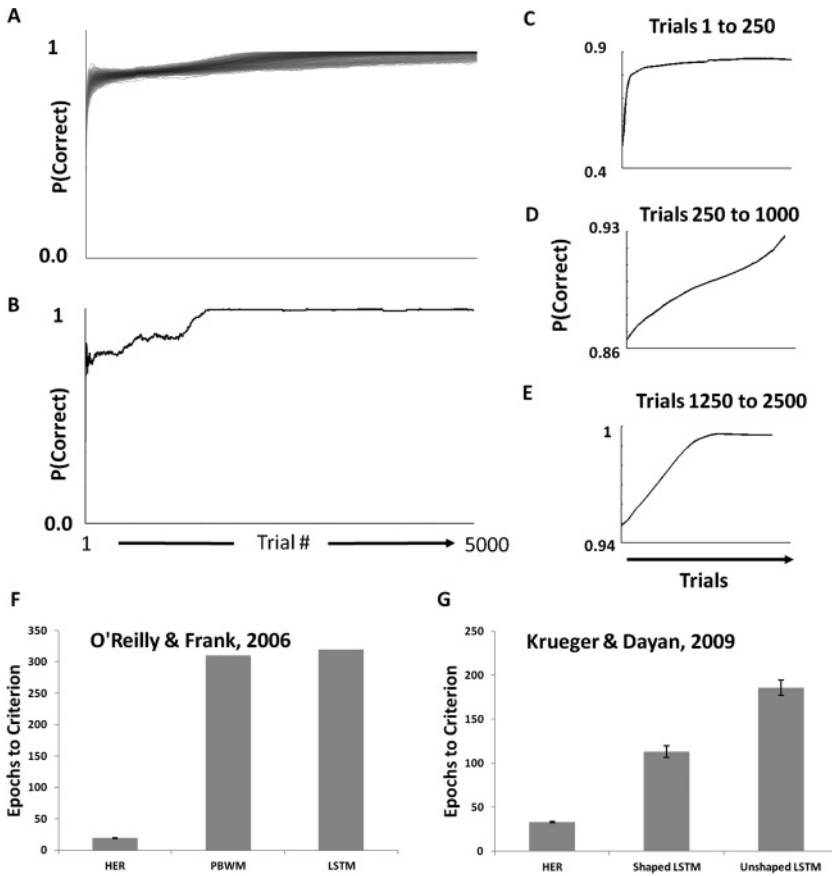
Figure 4: Behavioral performance. The HER model learns to perform the 1-2AX CPT quickly and consistently. (A) Percent correct trials for 1000 simulations of the HER model on the 1-2AX task. Values reflect the running average for a moving 500 trial window. (B) Percent correct trials for a single simulation of the HER model, again using a moving 500 trial window. Behavioral markers of hierarchical learning are apparent in the performance plateaus observable at different learning stages. Panels C–E display the same performance data as panel B but restricted to (C) early (1–250 trials), (D) intermediate (250–1000), and (E) late (1250–2500) learning periods. (F) Performance of the HER model in relation to simulations of the PBWM and LSTM models (reported in O'Reilly & Frank, 2006; values are approximate) using the criterion of two consecutive epochs without error. (G) Performance of the HER model in relation to simulations of a behaviorally shaped and unshaped LSTM model (Krueger & Dayan, 2009) using the criterion of no more than five errors in 30 epochs.
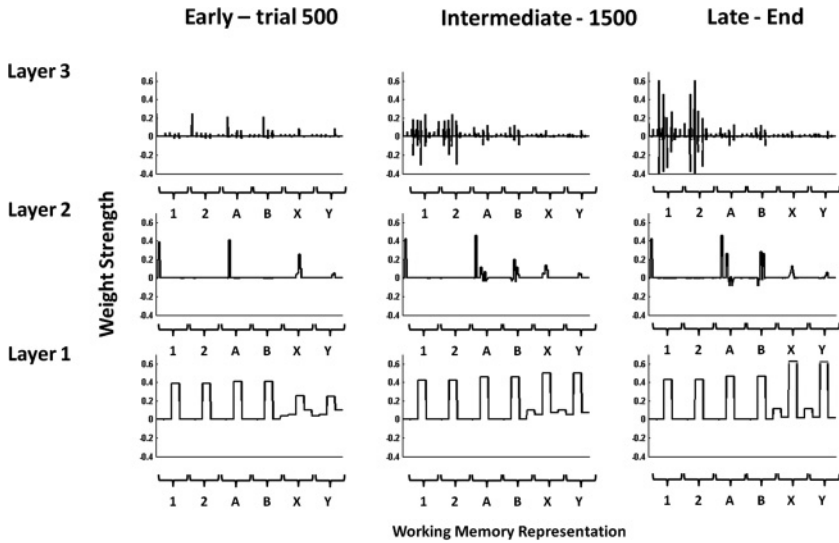
Figure 5: Associative weight development. The development of weights be-
tween working memory representations and outcome prediction units (see
equation 2.1) in the HER model during different learning stages. Brackets along
the x-axis indicate weights associated with working memory representations.
Note that at each layer of the hierarchy, the number of weights associated
with each working memory representation increases geometrically (see equa-
tion 2.7). In the first 500 trials (left column), weights in layer 1 (bottom row)
reflect base response rates for items in WM—observations of nontarget stimuli
in the task are strongly associated with nontarget/correct response-outcome
conjunctions, while potential target stimuli (X and Y) are partially associated
with nontarget/error RO conjunctions. During intermediate stages, weights in
layer 2 (middle row) associated with WM representations of A and B develop,
reflecting the increased probability of making target responses to X and Y, re-
spectively. Finally, during the final stage of learning, weights in layer 3 (top
row) associated with the high-level stimuli 1 and 2 develop, indicating the
model should respond only to target sequences (AX and BY) in the appropriate
contexts.

the base rates of those response/outcome conjunctions conditioned on X.
The error predictions in layer 2 are used to modulate associative weights in
layer 1 (see Figure 6, center), increasing the prediction of a Target/Correct
response/outcome and decreasing the prediction of a NonTarget/Correct
response/outcome. Following modulation (see Figure 6, lower right), the
prediction of a Target/Correct response/outcome is near unity. Note that
the above description assumes the appropriate high-level context variable
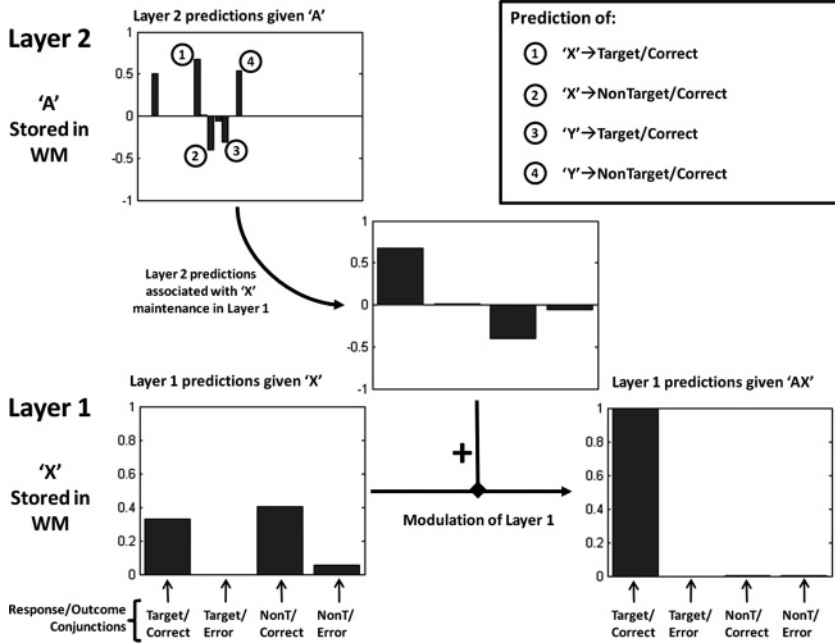(1) is stored in layer 3 of the hierarchy.

Figure 6: Modulation of prediction weights. Top-down modulation of lower layers is illustrated using a simplified version of the 1-2AX task (the AX CPT). Following learning of the task, the model reliably stores context variables A and B in the second model layer, while the less abstract variables X and Y are in the lowest hierarchical layer. The case of the sequence AX is represented in the figure. Upon the presentation of an X, WM representations in layer 1 are updated, resulting in the predictions indicated in the bottom left panel. Since X is not always associated with target responses (as in the case of a BX sequence), the model predicts both Target/Correct and NonTarget/Correct outcomes. In layer 2, the context variable A is maintained in WM following presentation of the X stimulus. Associated with the WM representation of A are predictions regarding possible sequences: if A is followed by an X, a Target/Correct outcome is more likely (indicated in the figure by the circled 1), while a NonTarget/Correct outcome is less likely (2). Similarly, if A is followed by a Y, the Target/Correct response (3) is less likely while the NonTarget/Correct response is more likely. Similar patterns of predictions exist, but are not depicted, for instances in which the model maintains B in the second layer. Since X was presented following A in this case, only the layer 2 predictions involving X (middle panel) have any effect on the lower layer; specifically, predictions at layer 2 additively modulate associative weights between WM representations and predictions at layer 2. Following modulation of layer 1 by layer 2, the final prediction indicated by layer 1 is a near certain estimation of observing a Target/Correct outcome.
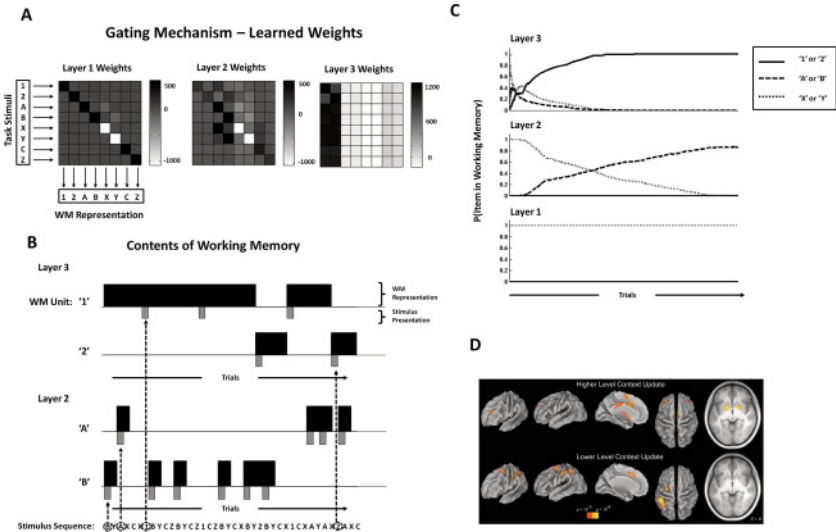
Weights in the WM gating mechanism (see Figure 7A) follow a similar time course. Because of the bias term in equation 2.5, the lowest layer of the hierarchy updates the contents of WM following the presentation of each stimulus. Although the gating mechanism is identical to higher layers, model performance depends on updating this layer on every trial. The BG weights at the second layer of the hierarchy learn to preferentially store the low-level context cues. It does this by learning when one of the potential target cues is presented; the value of maintaining a low-level context cue (i.e., how consistently it is associated with a particular error pattern) is greater than for maintaining the target cue in WM. Finally, BG weights in the third hierarchical layer learn to preferentially store high-level context cues. This pattern of weights supports the adaptive updating of WM in the model (see Figure 7B): when a high-level context cue is presented (indicated by the gray bar), it is selectively gated into WM at layer 3 and maintained until a different high-level context cue is presented (maintenance activation indicated by black boxes). Similarly, low-level context cues (A and B) are preferentially stored in layer 2. In contrast to the third hierarchical layer, however, low-level context cues are maintained for only a single stimulus presentation after they are stored in WM—long enough to contextualize the model's response to the potential target cues (X and Y), but not long enough to interfere with future responses. This adaptive gating develops over the course of training (see Figure 7C). Initially, when presented with a potential target cue, all model layers tend to represent the cue in WM. As the model develops, layer 2 learns to preferentially store low-level context cues in WM, while layer 3 learns to store high-level context cues. This pattern is consistent fMRI data from humans performing the 1-2AX CPT

(see Figure 7D; Nee & Brown, 2013), in which it is observed that activity in rostral regions of dlPFC is associated with high-level context updates, while more caudal regions are associated with low-level context updates.

3.2 Simulation 2: Atemporal Task Structure. As noted above, temporal structure may provide a valuable cue to learning hierarchically organized tasks. However, even without this cue, humans are able to learn structured

---

Figure 7: WM gating. (A) Weights (averaged over 1000 simulations) in the WM gating mechanism reflect the utility of maintaining versus updating WM representations following learning. At the lowest hierarchical layer, weights in the gating mechanism are confined primarily to the diagonal, indicating that the mechanism tends to update WM on each trial. Weights associated with the presentation of X and Y stimuli are strongly negative, indicating the mechanism's tendency to immediately "forget" those stimuli when any other stimulus is present. However, the model still stores the two variables in WM at layer 1 despite the strong negative weights due to the high bias for updating at layer 1. Weights in layer 2 reflect the tendency for the model to maintain representations of low-level context variables A and B rather than gating in X and Y representations. This is indicated by the positive off-diagonal weights associating X and Y inputs with A and B outputs. Finally, weights in layer 3 strongly favor maintenance of high-level context variables. The presentation of any stimulus other than a 1 or 2 is positively associated with maintenance of the WM representations for 1 and 2. (B) WM contents at layer 3 (top two rows) and two (bottom 2 rows). Gray vertical lines indicate the presentation of high- and low-level context cues, while solid black lines indicate the activity of corresponding WM units in the model. Following the presentation of high-level cues 1 and 2, WM units in layer 3 representing those cues become active and maintain their activity over extended durations until a new high-level context cue is presented. WM units in layer 2 preferentially maintain low-level context cues A and B and tend to maintain those items in WM for 1 model iteration—long enough to modulate responses to the immediately following X or Y cues. (C) Contents of WM over the course of learning in the first 1500 potential target trials (i.e., trials in which an X or Y is presented). The likelihood of storing items in WM at layer 2 and 3 changes as components of the task are learned. Initially all three layers represent potential target cues X and Y. After base response rates are learned, layers 2 and 3 represent A and B. Finally, at late stages of learning, layer 3 maintains only high-level context cues 1 and 2 in WM. (D) Dissociable regions of dlPFC are associated with updating in the 1-2AX task. Activity in rostral dlPFC (upper row) is observed following presentations of high-level context variables (1 and 2), while activity in caudal dlPFC (lower row) is observed following presentation of low-level context variables A and B. This organization is reflected in the mapping learned by the HER model performing the 1-2AX task. (Figure 7D reprinted with permission of Oxford University Press: *Cerebral Cortex*, Nee & Brown, 2013.)
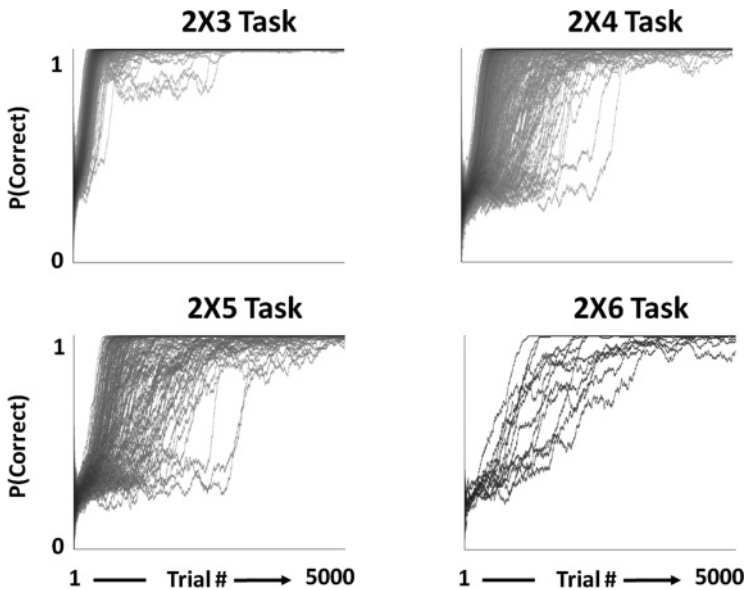
Figure 8: Behavioral performance. Plots show the probability of a correct response on the first 5000 trials (averaged over a moving 200 trial window) for various versions of a generic hierarchically structured task in which the number of possible values of 1 stimulus dimension is manipulated. Each line represents a single simulated subject, and for each version of the task, 1000 subjects were simulated. Simulations that failed to meet criterion are omitted. As the number of possible values for the manipulated stimulus dimension increases, so too does learning time.

tasks (Badre et al., 2010). Similarly, the HER model is able to learn to perform structured tasks in which all information is presented simultaneously (see Figure 8) and displays behavioral markers associated with human learning for the structured task, as discussed above. Specifically, in the course of learning, response accuracy in humans is observed to follow a pattern of rapid increase in performance followed by a plateau as components of the task are acquired.

Of particular interest is how the HER model learns to solve the structured tasks: How are the features used in structured tasks mapped to hierarchical layers in the model, and how might this mapping influence learning and behavior? Although in discussing structured tasks, we refer to two different stimulus dimensions, each having multiple values, this information is not hardwired in the model. Rather, values for each stimulus dimension presented on each trial are arbitrarily mapped to a binary feature vector, which constitutes the only information (disregarding feedback) available
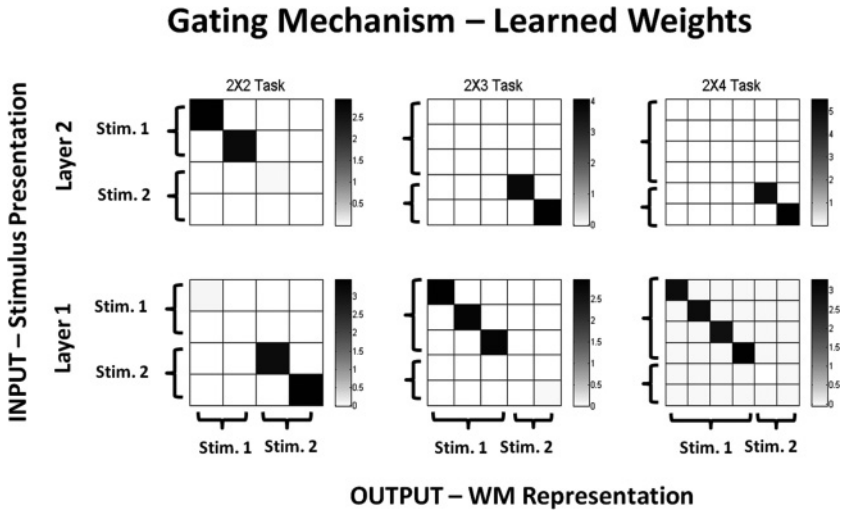
Figure 9: Emergent categorization in the WM gating mechanism. Learned weights in the WM gating mechanism reflect task structure. Although input to the model consists only of a binary feature vector containing no information about how low-level feature values relate to one another, values belonging to the same stimulus dimension are represented in the same layer of the HER model.

to the model. Given this, there may be multiple ways in which a hierarchical model may solve a structured task. For example, a 2×3 task (e.g., see Figure 3) may be decomposed into two component tasks: a 2×2 task, and a 1×2 task (see Figure 3D). Some features nominally assigned to one stimulus dimension may map to the lowest hierarchical layer, while other features belonging to the same stimulus dimension may map to higher layers. Analysis of learned weights in the WM gating mechanism (see Figure 9), however, reveals that when the model successfully learns a structured task, all features for a given stimulus dimension are consistently mapped to the same hierarchical layer. This mapping constitutes a categorization of unlabeled features into higher-order classes that emerges through the interaction of the model with a structured task, suggesting a mechanism by which rules and task sets, generally thought to be represented within dlPFC, may be acquired through experience.

Moreover, depending on the structure of the task performed by the model, the mapping of stimulus dimensions to hierarchical model layers follows a characteristic pattern. For cases in which the dimensionality of each stimulus dimension is equal, mapping is random, each stimulus dimension is equally likely to be mapped to either the first or second layer of the hierarchy. However, as the dimensionality of one stimulus
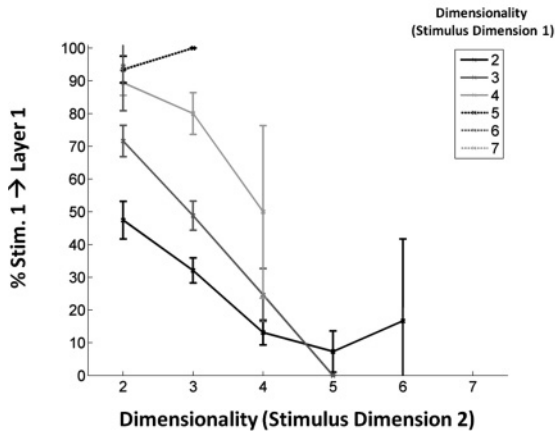
Mapping of Stimuli to Hierarchical Layers

Figure 10: Bottom-heavy mapping. The HER model's solutions to generic hierarchically structured tasks tend to favor a mapping in which the stimulus dimension with greater mutual information with the responses is represented at the lowest layer of the model. There is generally a strong positive correlation between the mutual information and a greater number of possible values in a given stimulus dimension (i.e., dimensionality), so here we plot the stimulus dimensionality as a proxy for mutual information. As the asymmetry between the dimensionality of the two stimuli increases, so too does the tendency of the model to solve tasks in a bottom-heavy fashion (i.e., higher mutual information dimensions map to lower layers). Vertical lines reflect standard deviation. Data are omitted for conditions in which the model failed to meet criterion at all over 1000 simulations.

dimension increases relative to that of the other, the model preferentially maps the higher-dimensional stimulus to the lowest hierarchical layer (see Figure 10). This is not due to the increased dimensionality of the stimulus dimension per se. As we will see, the major determinant is that stimulus dimensions that are better able to predict and reduce uncertainty about the response (as quantified by mutual information) will be mapped to lower layers (see Figure 13).

A perhaps unsurprising result is that the ability of the model to learn a structured task decreases as the dimensionality of each stimulus dimension increases, and for those models that are able to learn the task, the number of trials to criterion was higher. The model was most reliable and learned most rapidly for the simplest structured task (2×2), and performance in both metrics decreased for increases in both dimensions (see Figures 11A and 11B). One rather obvious explanation for the difference in learning time

is that as the dimensionality of each stimulus dimension increases, so too does the number of conjunctions the model must learn about. In a 2×2 task, a particular combination of stimuli dimensions will be seen much more frequently than in a 7×7 task. However, the pattern holds when the frequency with which each conjunction is observed is taken into account (see Figure 11B): as the dimensionality of each feature increases, the number of times the model must observe each conjunction before reaching criterion increases.

Moreover, the number of unique conjunctions for a given task structure is not the only factor influencing learning rate. Although the model tends to solve tasks by mapping the highest-dimensional stimulus to the lowest hierarchical layer, it is also possible for structured tasks to be accomplished by mapping the stimulus with the lowest dimensionality to the lowest layer (see Figure 3C). For simulations in which the model learns to map the stimulus with the higher dimensionality to the lowest model layer (a bottom-heavy representation), the number of trials to criterion is lower than when the model learns the task by mapping the stimulus with lower dimensionality to the lowest model layer (top-heavy). A two-way ANOVA was conducted using the number of trials to criterion in structured tasks in which the dimensionality of one of the stimulus dimensions was 2 (as these tasks were most reliably learned by the model) and reveals a significant effect of the dimensionality of the second dimension ($F(4,10) = 827.55$, $p < 0.001$) as well as the preferred mapping ($F(1,10) = 169.76$, $p < 0.001$).

One possible explanation for this difference is that for simulations in which a top-heavy representation is learned, the model may initially map task stimuli in the preferred bottom-heavy fashion, and subsequent adoption of the top-heavy mapping may be driven by random variation between simulations. In order to assess whether differences in trials to criterion are influenced by such variation, we conducted a second simulation in which the mapping of task stimuli to hierarchical layer was fixed: one stimulus dimension was always stored in WM at the same layer regardless of its dimensionality relative to the other stimulus. By fixing mappings in this manner, effects on learning related to the noisy, biased selection of stimulus mappings can be eliminated.

Even when controlling for this possible cause, however, the number of trials to criterion for top-heavy mappings remains higher than for bottom-heavy mappings (see Figure 12). For a given structured task in which the number of unique conjunctions was identical, the number of trials to criterion was higher when the stimulus with higher dimensionality was fixed at a superior hierarchical layer. Inspection of learned model weights (see Figure 13A) linking stimulus representations to outcome predictions suggests that prolonged learning periods in the top-heavy condition are due to increased ambiguity in the required responses. As noted in the 1-2AX simulations, the lowest hierarchical layer learns baseline response-outcome likelihoods. Since the total number of responses in the generic structured
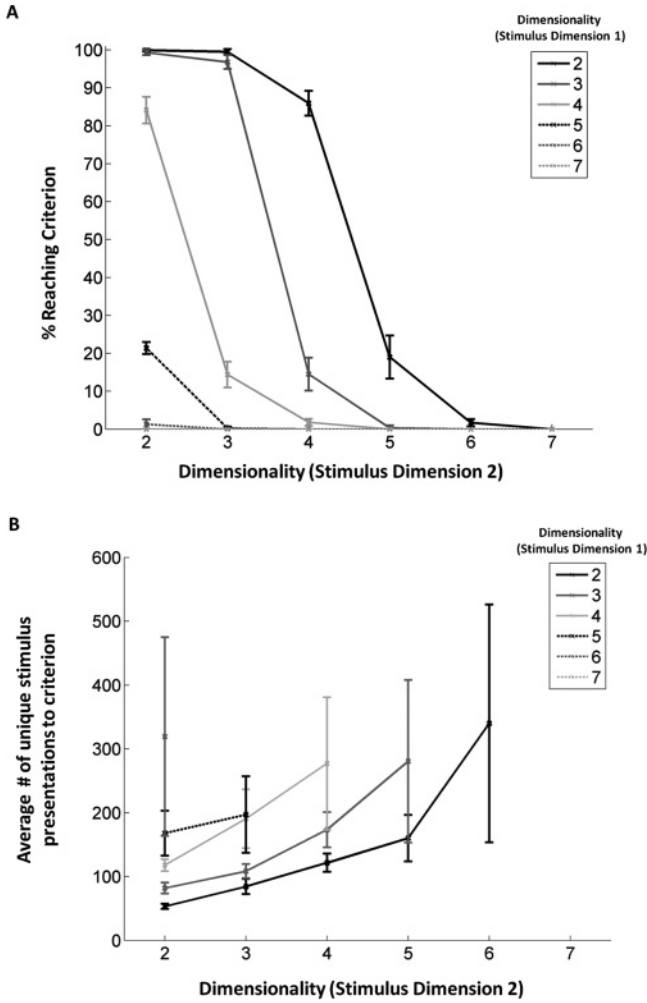
Figure 11: Dimensionality and performance. (A) Proportion of simulations (out of 1000 total) reaching criterion (vertical lines reflect standard deviation) in each condition. The HER model reliably learns hierarchically structured tasks provided the dimensionality of either stimulus is relatively low. The probability of reaching criterion drops precipitously when the dimensionality of any stimulus is five or greater. (B) Concurrently, the number of presentations of each unique stimulus conjunction required for the model to reach criterion increases with the overall dimensionality of both features (data omitted for conditions in which the model never met criterion).
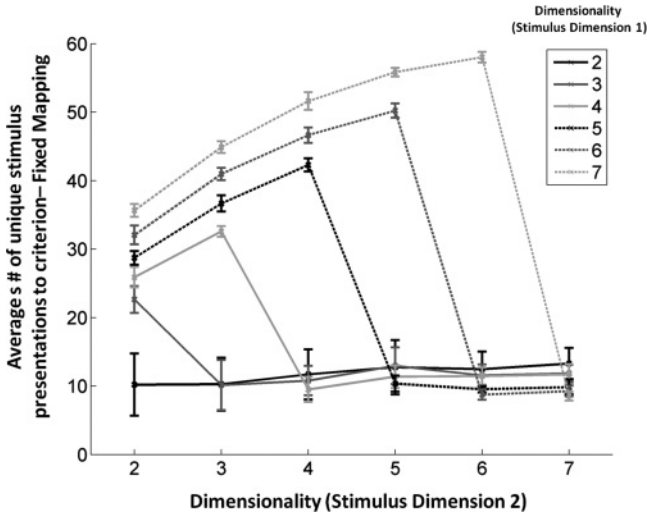
Figure 12: Fixed mapping. Number of presentations for each unique stimulus conjunction needed to reach criterion when the mapping of stimuli to hierarchical layers is fixed (i.e., model inputs associated with stimulus 1 are automatically gated into WM at layer 2, while inputs associated with stimulus 2 are gated into WM at layer 1). Although the model reliably learns the hierarchically structured task under fixed mapping conditions, the number of trials needed to meet criterion increases drastically when a top-heavy mapping is enforced.

task depends on the stimulus with the highest dimensionality, each of the stimulus dimensions for the lower-dimensional stimulus will be associated with all possible responses when it is mapped to the lowest hierarchical layer. Conversely, when the higher-dimensional stimulus is mapped to the lowest hierarchical layer, the number of responses for each dimension of that stimulus will be associated only with a number of responses equal to the dimensionality of the lower-dimensional stimulus. In the top-heavy condition, then, learning is protracted due to the less certain association of stimulus with response at the lowest hierarchical layer, reflected by larger average negative error components, indicating the failure of predicted outcomes to occur.

This analysis suggests a more general principle underlying the mapping of stimulus dimensions to hierarchical layers. Specifically, stimulus dimensions that reduce uncertainty about responses to a greater degree tend to be mapped to the lowest hierarchical layer, akin to the results of a principal component analysis, where the lowest layer in the hierarchy captures features analogous to the first principal component. Also, the degree to which a particular stimulus dimension reduces response uncertainty need not
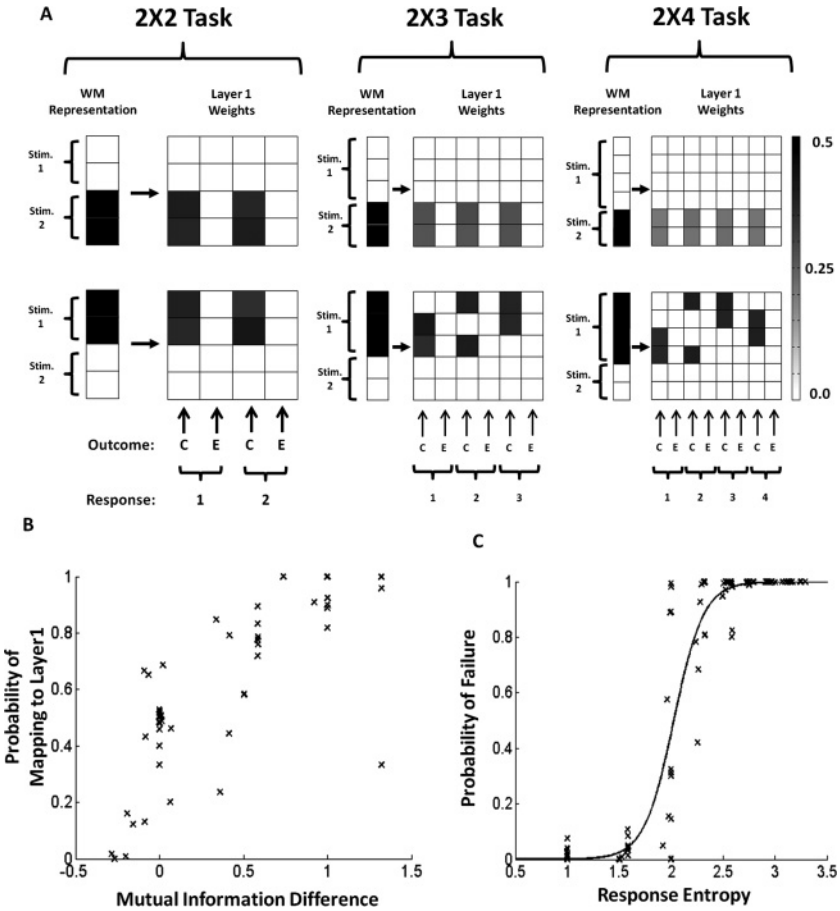
Figure 13: Weight development and uncertainty. (A) Under fixed mapping conditions, associative weights at layer 1 from WM representations to outcome predictions show qualitative differences in strength, with weights in the top-heavy condition (upper panels) reaching a lower asymptotic strength, while weights in the bottom-heavy condition (lower panels) reach a higher asymptotic strength. (B) The type of mapping preferred by the HER model correlates with the difference in mutual information between feature dimensions and responses. When one feature dimension reduces uncertainty regarding the response to a greater degree than the other, it is more likely to be mapped to the lowest hierarchical layer. (C) The probability of failing to solve a given structured task decreases with increases in response entropy. As the number of possible responses increases, the likelihood that the model will be able to solve the task decreases.

necessarily correspond to the number of values that stimulus dimension may take. For example, one may imagine that in the notional structured task depicted in Figure 3B, the response mappings may be arranged such that all stimuli in the shape of a diamond are associated with a response of 1, while ellipses are associated with a response of 2 regardless of the orientation. In this case, the stimulus dimension that most reduces uncertainty is the shape and not the orientation (which, in this example, contains no information regarding the appropriate response). If reduction of response uncertainty is indeed the principle determining the order in which stimulus dimensions are mapped to hierarchical layers in the model, one would expect that the shape dimension would be mapped to the lowest hierarchical layer.

To investigate whether uncertainty reduction could explain how the model maps stimulus dimensions to model layers, we undertook an additional round of simulations using the generic structured task. Simulations were run using structured tasks with two values for each stimulus dimension (2×2) up to tasks with five values (5×5) for each stimulus dimension as in the previous simulations. Additionally, the number of possible responses was varied for each structured task, from a minimum of 2 responses to a maximum of 10 responses. Response mappings for each response condition were selected to maximize the conditional entropy of the response with each of the stimulus dimensions.

The degree to which uncertainty regarding responses is reduced due to knowledge about a particular stimulus dimension can be quantified by the mutual information of the two, which is calculated as the entropy of the response minus the conditional entropy of the response given the stimulus dimension. For each condition, the mutual information between the correct response and each of the feature dimensions was calculated, and 1000 simulations of the model were conducted. Simulations of the model demonstrate that the difference in mutual information between the two stimulus dimensions predicts the mapping of stimulus dimensions to hierarchical layers in the model ($r = 0.75$, $p < 0.001$; see Figure 13B). Thus the tendency of the model to prefer a bottom-heavy mapping in our initial simulations is not a direct consequence of the number of dimensions of the stimulus but is due to the ability of higher-dimensional features to reduce uncertainty regarding the correct response to a greater degree than lower-dimensional features. Although dimensionality is not directly responsible for the mapping learned by the model, in general the mutual information of a stimulus dimension with the response for a given structured task will be greater for stimulus dimensions that can take a larger number of values than another dimension. For the simulations conducted, the difference in dimensionality between stimulus dimensions (larger minus smaller) is correlated with the difference in mutual information ($r = 0.81$, $p < 0.001$).

Overall, the dimensionality of each of the stimulus dimensions does not appear to influence the ability of the model to solve the structured

task per se, despite the strong correlation. Rather, the amount of response uncertainty (entropy) predicts whether the model will be able to solve a given structured task. As the number of responses increases, regardless of increases in the dimensionality of features, the probability that the model will fail to solve the task increases (see Figure 13C). Visual inspection of the data suggests this relationship may be captured by a logistic function; consequently, a two-parameter logistic function fit to the probability of failure data, with response entropy as the independent variable, yields an R^2 value of 0.85, while logistic fits in which the independent variable is the entropy of each of the feature dimensions provided substantially worse fits ($R^2 = 0.093$ and 0.16).

3.2.1 Simulation: Flat Model. Additional simulations were planned to investigate the importance of the hierarchical organization of the model on learning by using a flattened version of the model. In this version, the modules independently contributed to the overall outcome prediction and response generation. Initial simulations of the flat model showed that for the simplest structured task of 2×2 stimulus dimensions, the model failed to perform above chance after 10,000 trials. Simulations on the 2×3, 2×4, and 2×5 tasks showed a slight improvement above chance during training, and this improvement corresponded to learning the association of the higher-dimensional feature with the correct responses unique to that feature. In order to rule out the possibility that the failure of the flat model to learn the task was caused by dysfunction of the WM gating mechanism, each stimulus dimension was manually mapped to one of the modules. Although learning was more rapid in this case, the model still failed to completely solve the structured task after 10,000 trials, and performance asymptotes at the base response rates of the higher-dimension feature. The inability of the flat model to solve the structured task highlights the necessity of learning about feature conjunctions in order to perform the structured task. In the HER model, feature conjunctions are learned by exploiting the hierarchical organization of the model. However, this is not the only manner in which feature conjunctions can be formed. As noted above, a generic backpropagation network with at least one hidden layer is able to solve the structured tasks described in this manuscript, as well as the more complex task reported in Badre et al. (2010, unpublished simulations), where the activity of hidden units reflects learned conjunctions of features.

3.2.2 Simulation: Parameter Influence. Additional simulations of the HER model were undertaken to investigate the behavior of the model under alternative parameterizations. In order to investigate the influence of parameters related to temporal aspects of the model—eligibility traces (see equation 2.6) and WM update during ongoing performance (see equation 2.5), the λ and bias parameters were manipulated while the model performed the 1-2AX CPT. Changes to λ suggest a dual role in supporting

Table 3: Influence of Trace Parameter λ.

Task	.1 / .1 / .1	.5 / .5 / .5	.99 / .99 / .99	.1 / .5 / .99	.99 / .5 / .1
1-2AX	NA	NA	5935.2 / 6088 / 0.99	4227.3 / 2808 / 1	NA

Notes: Top row: Parameter values for layer 1/2/3. Cells: Mean trials to criterion/ interquartile range/proportion of simulations meeting criterion.

Table 4: Influence of Bias Parameter.

Task	1 / 1 / 1	.1 / .1 / .1	.01 / .01 / .01	.01 / .1 / 1
1-2AX	NA	10926 / 11270 / 0.88	9159.9 / 12060 / 0.72	NA

Notes: Top row: Parameter values for layer 1/2/3. Cells: Mean trials to criterion/ interquartile range/proportion of simulations meeting criterion.

learning in the model (see Table 3). First, simulations in which the value of λ was reduced specifically for the third layer of the hierarchy failed to reach criterion (0 successes out of 100 simulations), while simulations for which the value of lambda was maintained at the hand-tuned value at the third layer met criterion 99 and 100 times out of 100 simulations. This suggests that the ability to solve the 1-2AX task critically depends on the model's ability to maintain sufficiently strong trace activity in order to support learning in the WM gating mechanism for temporally distant stimuli. Second, learning of the task was impaired when the value of λ was increased for the lowest hierarchical layer. Although the model successfully acquired the task for the condition in which λ was set to 0.99 for all layers, the time required to do so was significantly prolonged ($t(198) = 2.88$, $p = 0.0043$), suggesting that learning appropriate responses to immediate stimuli at the first hierarchical layer may be impaired by the presence of information irrelevant to current task demands.

Similarly, manipulation of the bias parameter for WM updating (see equation 2.5) suggests that maintaining an item in WM for extended durations, especially at the top layer of the hierarchy, is critical to the model's ability to solve the 1-2AX (see Table 4). For simulations in which the bias was set to 1 (indicating a tendency to update WM with the current stimulus over maintaining the current contents of WM) at the top hierarchical layer, the model failed to meet criterion on any simulation. Conversely, the model met criterion 88 and 72 times out of 100 when the bias parameter for all layers was lowered to 0.1 and 0.01, respectively. Again, however, the time needed for these simulations to reach criterion suggests a dual role for the parameter. For these simulations, the model required on average 10,926 (bias = 0.1) and 9159.9 (bias = 0.01) trials to reach criterion, significantly different ($t(186) = 9.56$, $p < 0.001$; $t(170) = 6.99$, $p < 0.001$, respectively) from

Table 5: Influence of Learning Parameter.

Task	.01 / .05 / .1	.1 / .05 / .01	.01 / .01 / .01	.1 / .1 / .1
2 × 2	251.44 / 88.5 / 1	722.55 / 252 / 0.84	337.66 / 136 / 1	1351.2 / 0 / 0.2
2 × 3	407.49 / 194.5 / 1	1816.4 / 0 / 0.05	869.79 / 638 / 0.96	NA
3 × 3	580.84 / 248.5 / 0.99	5739 / 0 / 0.01	932.74 / 328.5 / 1	NA

Notes: Top row: Parameter values for layer 1/2/3. Cells: Mean trials to criterion/ interquartile range/proportion of simulations meeting criterion.

the trials required by the hand-tuned model in which the lowest layer of the hierarchy tended to update WM frequently (bias = 1) while the top layer of the hierarchy was less biased to do so (bias = 0.01). These results suggest that the hand-tuned parameterization in Table 1 balances the need to maintain information over extended periods of time in order versus updating information in order to meet rapidly changing immediate task demands.

The influence of the learning rate parameter α and the gain parameter γ for working memory updating were explored using three versions of the structured task: 2 × 2, 2 × 3, and 3 × 3. Manipulation of the learning rate suggests that the ability of the model to learn the task depends on its ability to develop stable representations at the lowest hierarchical level. For all versions of the task, the model was able to meet criterion only for simulations in which the learning rate for the lowest level was low. Simulations in which the learning rate was low across all hierarchical levels ($\alpha = 0.01$) produced results comparable to those from the hand-tuned model parameterization (summarized in Table 5). Interestingly, increases in α at successively higher hierarchical layers produced overall better performance than the hand-tuned parameterization (see Table 5, first column), especially as the number of possible conjunctions increases (e.g., in the 3 × 3 task, the mean number of trials to criterion is significantly lower than for the hand-tuned parameterization ($t(1065) = -5.83$, $p < 0.001$)). In contrast, as the learning rate in lower hierarchical layers increased, performance became qualitatively worse, suggesting that learning in the model depends partially on the stability of learned weights in lower hierarchical levels.

Finally, manipulations of the gain parameter, which governs the decisiveness of the WM gating mechanism in choosing between two potential items to store in WM, showed no obvious differences in performance on the structured task (see Table 6). The number of trials needed for the model to meet criterion was comparable across all manipulations to the hand-tuned parameterization. One possible reason no differences were observed is that the structured task simulations did not involve a substantial temporal component: the model had to choose between storing each of the two features, but each trial was independent of the last (i.e., the model did not have to maintain information over trials in order to successfully perform the task).

Table 6: Influence of Gain Parameter γ.

Task	5 / 5 / 5	10 / 10 / 10	20 / 20 / 20	5 / 10 / 20	20 / 10 / 5
2 × 2	214.8 /	212.77 /	242.26 /	224.9 /	214.2 /
	66 /	76 /	88 /	82.5 /	66.5 /
	1	1	1	1	0.99
2 × 3	519.9 /	520.71 /	462.4 /	430.37 /	429.99 /
	180 /	218.5 /	121.5 /	130 /	154.5 /
	1	1	0.99	1	0.99
3 × 3	1015.2 /	846.7 /	984.8 /	814.98 /	1065.8 /
	376.5 /	336 /	399.5 /	310 /	476.5 /
	0.98	0.97	0.97	1	0.98
1-2AX	18,204 /	5763.9 /	4124 /	4616.2 /	12,560 /
	18,064/	2705.5/	3358.5 /	2717 /	5717 /
	0.45	1	1	0.99	0.94

Notes: Top row: Parameter values for layer 1/2/3. Cells: Mean trials to criterion/ interquartile range/proportion of simulations meeting criterion.

In order to explore the influence of temporal contingencies in conjunction with manipulation of the bias parameter, the model was simulated on the 1-2AX task using the parameters given in Table 6. Here we see that the value of the gain parameter has a substantial effect on model performance specifically at the top layer of the hierarchy. In simulations in which the gain parameter was high, model performance was comparable to the hand-tuned parameters, while performance decreased as the gain parameter decreased. These results are similar to those observed in the bias parameter inasmuch as a low-gain parameter corresponds with a less deterministic choice to store (or maintain) an item in WM, while a high-bias parameter indicates a preference for storing new items in WM over potentially more appropriate items already represented.

While exploration of the parameter space indicates that especially in the case of the structured task simulations, the HER model is fairly robust to parameter changes, questions may remain regarding the ability of the model to solve both the 1-2AX task and the structured tasks with a single parameter set. As noted above, parameters for each task were chosen based on how quickly and consistently the model was able to learn each task, and so the parameterization for one task may not be suitable for performing the other task. In particular, bias values used for the structured task simulations are set to 0, reflecting our intuition that for the structured task, WM would be updated at each layer on each trial, since information from previous trials has no bearing on a current trial in that task. In order to demonstrate the ability of the model to perform multiple tasks with a single parameterization, the initial round of structured task simulations was rerun using the parameters from Table 1. Detailed results are included in

supplementary material; generally performance of the model is comparable to the hand-tuned parameterization given in Table 2.

4 Discussion

In this letter, we have described a new computational model of ACC/mPFC and dlPFC building on a previous model of mPFC, and specifying how a multidimensional error signal generated by mPFC may be used to train representations of expected error in order to support behavior. Simulations of the model demonstrate its ability to rapidly learn complex cognitive tasks that involve maintaining information over extended periods of time and to discover abstract rules that govern responses.

A central feature of the model is the ability to represent tasks as having hierarchical structure. In principle, though, our tasks could be represented as flat tasks, in which all possible combinations of conditions collectively form a (flat) basis for responding. In that case, the tasks could be solved in principle without a hierarchical representation. There are two reasons for representing the tasks as having hierarchical structure. First, although animals can treat compound stimuli of the sort used in structured tasks as being single stimuli (Kehoe,1986), a condition necessary to form the basis of a flat task, it is not obvious that humans would do so for arbitrary compound stimuli, particularly when the dimensions of the compound stimuli are not related to one another except through their involvement in a structured task. Second, recent work suggests that humans spontaneously attempt to represent tasks with a hierarchical rather than flat structure, even when a hierarchical structure is unnecessary (Collins & Frank, 2013).

4.1 Mapping the Model to Brain Regions. The HER model was developed with the intent of accounting for the complementary roles of mPFC and dlPFC in cognitive control. As noted in section 1, the frequently observed coactivation of these regions implies some degree of functional overlap, and this is reflected in the organization of the model, in which task-related stimuli produce predictions, which then modulate predictions generated at lower hierarchical layers, following which errors are calculated and used to drive learning at higher hierarchical layers. Thus, both regions can be regarded as engaging in both prediction and error processing, making it somewhat unclear what unique role the HER model ascribes to mPFC and dlPFC.

The distinction between the two regions suggested by the model is most apparent at the highest hierarchical layer. In the 1-2AX task, the third layer learns to preferentially store high-level context variables 1 and 2. These variables are associated with errors generated by the immediately inferior layer (see Figure 5, top panels). Critically, the error predictions learned by the top layer inform the predictions of lower levels for events that may or may not happen. When the model stores the high-level context variable 1

at the top layer, the predictions generated at that layer increase the model's estimate of the probability of a target response should the sequence A-X be observed, while decreasing the estimate of target response probability for the sequence B-Y. When these predictions are elicited following the storage in WM of the variable 1, either of these events is possible (as well as B-X and A-Y sequences), but neither has occurred. Following the presentation of the low-level context variable (A, as an example), which is stored in the second hierarchical layer, only predictions generated at the third layer relating to inner-loop sequences beginning with A are relevant. However, since it is possible in the 1-2AX task to observe multiple inner-loop sequences following the presentation of a high-level context variable, predictions regarding sequences beginning with B need to persist until the next presentation of a high-level context variable, even though they may not be needed given the current low-level context variable.

In the HER model, the division of labor between maintaining predictions of all possible events that may be observed versus applying a currently valid set of predictions is assigned to dlPFC and mPFC, respectively (see Figure 1C and 1D). Thus, when the high-level context variable 1 is presented, predictions regarding possible inner-loop sequences are presumed to be elicited and maintained in rostral dlPFC, while at the presentation of the first stimulus in an inner-loop sequence, rostral mPFC regions apply those predictions that are currently valid to more caudal PFC regions. Similarly, at the presentation of the lower-level context cue, A, predictions regarding potential target presentations are elicited and maintained in caudal dlPFC, modulated by predictions selected by rostral mPFC (but maintained in rostral dlPFC).

This distinction is somewhat obfuscated by the manner in which top-down modulation is implemented in the model. According to equations 2.9 and 2.10, modulation of predictions at a given layer by the predictions generated by the immediately superior layer is tonically applied. All weights in the weight matrix **W** are adjusted regardless of whether the stimulus associated with those weights is currently stored in WM. Although weights associated with stimuli not stored in WM are modulated, they have no influence on the predictions generated by the layer being modulated. Only the subset of top-down predictions corresponding to weights connected to the currently active WM representation has an impact on predictions; thus, the selection carried out by mPFC in the model is implicit rather than explicit.

4.2 Relationship with Known Anatomy. The HER model proposes a number of mechanisms regarding the interaction of regions of the brain which. While promising in that they combine to produce a model capable of performing complicated tasks, the exact mapping of model components to brain structures remains open to question.

One such question is the role of mPFC in driving learning of working memory updating. In the model, error signals generated by the mPFC are used to train weights that govern how and when a stimulus may be stored in working memory. Although a role for BG in WM updating has been proposed previously (O'Reilly & Frank, 2006), the BG is selectively active in higher-level working memory updating, while lower-level working memory updating involves mainly cortico-cortical interactions (Nee et al., 2013). The mechanism we propose differs from others in crediting mPFC with a functionally important role in training the working memory updating process. In our model, mPFC error signals train working memory updates via backpropagation, a mechanism that is biologically implausible. However, it is conceivable that a more biologically realistic account of mPFC interactions with working memory updating may be developed. In particular, activation in mPFC elicited either through electrical (Gariano & Groves, 1988; Taber & Fibiger, 1993) or chemical stimulation (Murase, Grenhoff, Chouvet, Gonon, & Svensson, 1993) has been observed to drive burst firing in the ventral tegmental area (VTA), a major source of dopamine (DA) projections to cortex, while inactivation of mPFC by means of lidocaine (Murase et al., 1993) reduces burst firing. In previous models (O'Reilly & Frank, 2006), dopaminergic activity was exploited to allow BG to learn the value of storing a stimulus in WM. Thus, the current computationally convenient but implausible backpropagation method may be replaced by a more plausible reinforcement learning mechanism (see the supplementary material). The HER model suggests that mPFC may be a key contributor to working memory updating, particularly for stimuli that are not innately valuable.

Additionally, we include a bias parameter in our gating mechanism governing how likely a currently presented stimulus is to be gated in to WM versus maintaining the current WM contents. In the HER model, behavior is observed to degrade when bias values for the top layers of the hierarchy are higher than they are for the bottom layer; indeed, behavior appears to be optimal when the bias value is highest at layer 1 of the model and lowest at layer 3. This finding suggests a correspondence with DA innervation of PFC, which has been observed to follow an anterior-posterior gradient (Williams & Goldman-Rakic, 1993), wherein posterior medial and lateral PFC are more densely innervated relative to more anterior regions. This pattern of innervation may correspond to the hypothesized abstraction gradient posited for dlPFC, with more anterior regions of dlPFC representing information that is more abstract compared to posterior regions. DA is known to be modulate WM updating (Brunel & Wang, 2001; Muly, Szigeti, & Goldman-Rakic, 1998; Murphy, Arnsten, Goldman-Rakic, & Roth, 1996), and the higher degree of DA innervation of posterior PFC may support frequent updates of information in WM, while the relative sparseness of DA innervation of anterior PFC may allow information in WM to be maintained over longer durations, thus allowing the formation of abstract relationships between temporally

distant stimuli. Overall, though, DA levels must be maintained in an optimal range for successful working memory maintenance (Brunel & Wang, 2001; Muly et al., 1998). Dysfunction of the DA system is implicated in a range of behavioral pathologies, including obsessive-compulsive disorder and schizophrenia (Abi-Dargham et al., 2002; Cohen & Servan-Schreiber, 1993; Denys, Zohar, & Westenberg, 2003; Goodman et al., 1990).

The HER model suggests a hierarchical organization of mPFC, with discrete regions predicting outcomes at various levels of abstraction and signaling discrepancies. Although mPFC, and especially ACC, is often referred to unitarily, a wealth of evidence suggests that cingulate is segregated into multiple regions based on functional, connectivity, and meta-analyses (Jahn et al., 2014; Nee, Kastner, & Brown, 2011). Recent evidence from neuroanatomical parcellation (Yu et al., 2011), functional connectivity (Taren et al., 2011), and neuroimaging studies (Amiez & Petrides, 2014) suggests that connections between dACC and dlPFC follow a posterior to anterior pattern, with posterior and anterior regions of ACC connecting to posterior and anterior regions of dlPFC, respectively. Previous functional imaging results have suggested a correspondence between medial and lateral PFC hierarchies, arguing that medial PFC provides motivational signals, while lateral PFC provides selection, that is, cognitive signals (Kouneiher, Charron, & Koechlin, 2009). Our model differs from this earlier proposal in that rather than providing motivational signals to lateral PFC, the mPFC provides prediction error signals that train the lateral PFC regarding what information must be maintained for successful task performance. While these views are not directly contradictory, the HER model, like the PRO model on which it is based, emphasizes predictive processes that are not explicitly tied to value.

There is some degree of ambiguity regarding the precise nature of the contents of working memory. The HER model suggests that dlPFC specifically represents potential errors that are associated with task stimuli. Thus, although within the model, it is the stimuli themselves that are stored in WM, the model does not claim that dlPFC represents stimulus identity, but rather that it stores specifically the information associated with a particular stimulus that is needed to perform a task. Recent neuroimaging studies have failed to find stimulus-specific properties represented in dlPFC (Riggall & Postle, 2012), although increased activity was observed in the region during a delay period during which it is presumably engaging in some aspect of WM maintenance. Other recent work using MVPA suggests that information is stored (and decodable) in dlPFC specifically if it is necessary to guide task performance (Nee et al., 2011). This helps to explain the inconsistency in that sometimes the dlPFC appears to encode stimuli in working memory (Hussar & Pasternak, 2013; Nee & Brown, 2013) and sometimes not (Riggall & Postle, 2012). Essentially, dlPFC representations are prospective in nature, reflecting anticipated task contingencies. Error representations in the HER model may then be thought of as probabilistic

if-then rules—biases on responses to be made in the event a particular stimulus is observed. Thus, single neurons in dlPFC may encode stimuli in working memory not as sensory information per se, but rather as selected information that is anticipated to be necessary for guiding task performance (Rainer, Rao, & Miller, 1999).

Additionally, it is remains ambiguous as to whether activity in dlPFC, especially in rostral and caudal dlPFC, reflects maintenance of ongoing task demands or if activity in dlPFC is primarily involved in updating proactive control plans in response to shifting task contingencies. On one hand, evidence suggests that activity in these regions corresponds to the presentation of high-level context cues that need to be maintained over extended periods of time in order to provide information necessary for successful task performance (Koechlin et al., 2003). On the other hand, studies that have looked at the role of caudal and rostral dlPFC during task performance suggest that transient activity, associated with updating WM and instantiating control processes which may then be maintained elsewhere, contributes more to dissociable signals within dlPFC (Reynolds et al., 2012). The HER model reflects this ambiguity inasmuch as the predictions generated at each layer of the hierarchy are instantiated and maintained using the same units and weights. While the model assigns the role of maintaining task-relevant predictions over extended periods of time to dlPFC, it may be possible that only transient dlPFC activity is required to instantiate relevant task rules, which are then maintained by mPFC. Recent evidence suggests distinct roles of mPFC in signaling prediction error and maintaining task-related predictions (Jahn et al., 2014).

Although the activity of single neurons involved in WM has been associated with coding specific task rules (Asaad, Rainer, & Miller, 2000; Wallis & Miller, 2003; Yamada, Pita, Iijima, & Tsutsui, 2010), it is not clear that such coding is devoted to representing unitary rules exclusively. Activity in dlPFC neurons is quite heterogeneous, and neurons have been reported whose activity appears to code for various stimulus properties as well as response plans (Genovesio, Brasted, Mitz, & Wise, 2005; Jun et al., 2010; Miller, Erickson, & Desimone, 1996). The level of heterogeneity observed within dlPFC neurons would seem to militate against a simple link between the instantiation of a rule and a corollary activation of a neuron coding for that rule. An alternative interpretation offered by the HER model is that rules are represented in dlPFC through a distributed set of neurons whose activity corresponds to primitive behavioral elements that constitute a rule. Such representations occur by virtue of the fact that each working memory cell encodes the potential for a particular kind of error (i.e., if the appropriate information is not used to guide the action choice), and the errors may occur in more than one context.

Efferent connections to dlPFC from mPFC are distributed through both deep and superficial layers of dlPFC (Barbas & Pandya, 1989; Medalla & Barbas, 2009), indicating that interactions between the two regions may

support multiple forms of processing, with a slight predominance of net inhibition from ACC to dlPFC (Medalla & Barbas, 2009). This is generally consistent with the HER model connections from error signals (presumably from mPFC; see Figure 1) to higher-level outcome signals (presumably in dlPFC). Similarly, prediction signals may be represented in mPFC and modulate (and train) lower-level working memory representations in dlPFC. This training function may be separate from a more general inhibitory function, by which mPFC may suppress or reset the current task set (Donoso et al., 2014; Kolling et al., 2012).

There may also be substantial cortico-cortical interactions within dlPFC, as higher (more anterior) regions provide direct modulation of lower (more posterior) regions. This is consistent with anatomical studies showing that more anterior regions originate projections in more superficial layers, and these terminate in deeper layers of more posterior regions (Barbas & Rempel-Clower, 1997).

4.3 Relationship to Existing Models. The HER model synthesizes a number of ideas that have previously appeared in the neuroscientific and modeling literature. First, hierarchical organization of PFC has been previously proposed, primarily with regard to dlPFC (Badre & D'Esposito, 2009; Koechlin et al., 2003; Nee et al., 2013), but additional evidence suggests the mPFC follows a similar organization (Amiez & Petrides, 2014; Dixon, Fox, & Christoff, 2014; Taren et al., 2011). Second, the idea of hierarchically organized regions of the brain learning expected or residual errors of inferior layers has previously been applied to explaining extraclassical receptive fields in visual cortex (Rao & Ballard, 1999). Third, the HER model incorporates a mechanism that allows for task stimuli to be gated into working memory, inspired by the function of basal ganglia proposed by O'Reilly and Frank (2006). Fourth, the concept of expected error as applied in this model is similar to entropy (or free energy) in information-theoretic accounts of brain function (Friston, 2009, 2010), in that the HER model is continually learning the environment in such a way as to minimize prediction error. Finally, we incorporate our own previously published model of mPFC function, which characterizes the role of mPFC as predicting the likely outcomes of actions and signaling unexpected deviations from predictions (Alexander & Brown, 2011).

The HER model simulates tasks at the event level, wherein a number of discrete processing steps intervene between the presentation of a stimulus and the generation of a response and delivery of feedback to the model. Processing related to stimulus presentation (i.e., WM updates, prediction modulation) occurs prior to responses and outcomes, which are treated as a single conjoined event. The use of response-outcome conjunctions in the model is derived from their utility in the PRO model in accounting for a range of results observed in mPFC. However, this does not rule out the possibility that additional processing steps may intervene between the

generation of a response and its associated outcome, particularly in tasks in which a response is not necessarily diagnostic of eventual feedback (e.g., probabilistic tasks in which the outcome of a choice is not known until feedback versus typical cognitive control tasks in which success or failure can be assessed based only on the identity of the response). More generally, stimuli, responses, and outcomes may be more broadly construed as generic events: salient stimuli (whether exogenous or endogenous) that could conceivably form the basis for additional processing (Alexander & Brown, 2014). As such, the grouping of responses and outcomes used in the HER model was chosen based on the types of tasks the model was expected to simulate and does not rule out the possibility that mPFC and dlPFC may also use alternative associative pairings (e.g., stimulus-response, stimulus-outcome, response-outcome).

The HER model provides a novel, general solution to the problem of how hierarchical task stimuli may be functionally self-organized into corresponding hierarchical task representations. One reason for the tidiness of the HER model's learned associations may be that it is the only way for task performance to be optimal given the smallest possible number of layers. For each level of the task hierarchy, exactly one cue will be valid. In the 12-AX task for example, either the 1 or the 2 cue is valid at any given time. Conveniently, only one learned cue representation will be stored at each layer at a given time. This means that storage is most efficient if each level of the task hierarchy has its own corresponding model layer. If instead the 1 were stored in one layer and the 2 were stored in another, then it would not be possible to store every combination 1A, 1B, 2A, and 2B across only two layers. The HER model will not settle on such a solution as it entails greater prediction errors. Overall, this insight captures the essence of a stimulus dimension: that instances of properties on a given dimension are mutually exclusive. In the example here, 1 versus 2 collectively form a feature dimension. By implication, feature dimensions may be abducted in general by grouping feature instances that never co-occur. This approach to abductive inference about feature dimensions may be more general than earlier proposals that allow feature dimensions to be abducted only when the same feature is presented repeatedly over a long series of trials (Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005).

In our previous work, we suggested that multidimensional error signals generated by mPFC provide a plausible learning signal that may be used by a model-based RL algorithm, while other studies have associated activity in dlPFC with such a signal (Barto et al., 1995; Gläscher et al., 2010; Sutton, 1990). The HER model, based on the PRO model of mPFC, suggests that signals in dlPFC corresponding to a model-based learning signal may be driven in part by mPFC. To the extent that the HER model uses reinforcement learning formulations to learn tasks that require the maintenance of an environmental model in order to complete, it may be considered a model-based RL algorithm. However, while typical model-based RL algorithms

often render models of the environment as a state-transition matrix enumerating the probabilities of transition between all states (Gläscher et al., 2010; Jong & Stone, 2008), no one component of the HER model corresponds with such a unitary model. Rather, the model of the environment learned by the HER model is distributed across hierarchical layers, each of which relies on relatively simple model-free RL formulations. Model-based behavior exhibited by the HER model then emerges from the interaction of multiple model-free RL learners.

To the extent that the HER model follows a hierarchical organization, it may also be considered a hierarchical RL (HRL) algorithm. In general, HRL models attempt to learn to accomplish tasks composed of abstract goals, rules, and concrete elementary behaviors organized into behavioral schemata (Barto & Mahadevan, 2003; Botvinick, 2008; Botvinick, Niv, & Barto, 2009; Jong & Stone, 2008). For example, the abstract goal of "making a cup of coffee" may include rules such as grinding beans, heating water, and so on, and the rules themselves may be composed of behavioral primitives such as "opening" or "filling" (Cooper & Shallice, 2000). While organizing these elements into a coherent goal-based behavior is a significant challenge, typical HRL models take the existence of such elements as given. In contrast, the HER model assembles the schemata needed to accomplish an abstract task starting from concrete responses and learning more abstract "rules" in progressive stages.

Recent theoretical work has also proposed an interaction of ACC and dlPFC in the context of hierarchical behavior (Holroyd & Yeung, 2012). In this work, ACC is thought to constitute the highest layer of a hierarchy and is involved in maintaining goals and policies necessary for realizing these goals, while lower layers represent task sets (dlPFC), subgoals (OFC), action primitives (dorsal striatum), and value (ventral striatum) required to implement such policies and evaluate success or failure. In this scheme, ACC constitutes the top level of a hierarchy, and thus is substantially different from the HER model in that the HER model proposes that distinct regions of dlPFC and mPFC are involved at all layers of a hierarchy.

In the HER model, hierarchical layers learn successively more abstract error predictions wherein the prediction to be learned at particular layer in the hierarchy is the error term of the immediately inferior layer, provided that the error can reliably be associated with the prior presentation of a stimulus. Previous work has implicated mPFC, as well as neuromodulatory systems innervated by mPFC, especially locus coeruleus (Aston-Jones & Cohen, 2005), in the processing of unexpected errors, as opposed to expected errors, related to shifts in environmental contingencies (Behrens et al., 2007; Holroyd et al., 2004; Yu & Dayan, 2005). Error-related signals in the HER model, associated with mPFC activity, are modulated by error predictions at superior hierarchical layers: when an error occurring at a lower layer can be predicted, error signals for that layer decrease, while errors that cannot be predicted continue to produce robust signals. By indirectly decreasing the

effective learning rate of the model (through modulation of predictions and, consequently, error signals needed to drive learning) in conditions when errors are predictable, the HER model suggests that dlPFC may play a role in resolving the stability-plasticity dilemma (Grossberg, 1980) by decreasing the effect of errors in predictable, though error-prone, environments, while permitting rapid learning in response to unexpected errors.

Error signals in the HER model are calculated at each hierarchical layer, suggesting that, corresponding to the hierarchical organization of mPFC, distinct subregions of mPFC, and particularly ACC, may produce error-related activity related to more or less abstract errors (e.g., response errors, rule errors). Previous models of PFC have also posited error signals and updating at multiple levels of abstraction (Collins & Frank, 2013; Collins & Koechlin, 2012), and fMRI studies have observed distinct regions within cingulate that appear to signal error related to the degree of abstraction (Donoso et al., 2014; Zarr & Brown, 2015). The HER model differs from these previous models in that the same mechanism is used to compute error, and drive learning, at each level of abstraction, while the PROBE (Collins & Koechlin, 2012) and C-TS model (Collins & Frank, 2013) use RL formulations for learning concrete actions in conjunction with Bayesian approaches for updating beliefs about task sets.

4.4 Limitations. Although the HER model was developed in order to investigate the function of dlPFC and its interaction with ACC, a number of assumptions are incorporated into the model that may limit the scope of its utility in this regard.

In the HER model, prediction units at each layer reflect a combinatorial enumeration of all possible sequences that may be observed in inferior layers. At the lowest layer, all possible conjunctions of responses and outcomes are represented by prediction units, and are associated with task-related stimuli, while at the next higher layer, prediction units represent all possible stimulus-response-outcome conjunctions, and so on, resulting in a geometric growth of the dimensionality of the network with each additional layer. In terms of neural plausibility, this representation scheme is clearly deficient. If such a scheme were employed by a behaving animal, it would require the existence of specialized neurons coding a specific sequence of events that may never be observed, as well as an absolutely enormous rostral dlPFC. However, we note that this problem is not unique to the HER model; practically all models of working memory include hard-wired representations in some form, as well as a degree of specificity in the representational scheme, for the sake of simplicity. In practice, there is likely to be a sparse and adaptive coding such that stimuli that are experienced more frequently will be represented by more neurons (Clark, Allard, Jenkins, & Merzenich, 1988). Generally it is assumed that such representations reflect the dynamic mapping of an external stimulus to neurons involved in WM; that is, simply because the activity of a unit within a

computational model reflects the presence or absence of an item in WM, it is not presumed that the such a unit corresponds with a real neuron that represents only a very specific type of stimulus. Rather, the unit in the model is a proxy for a real neuron that in the context of a specific task, happens to code for a specific stimulus. Similarly, the problem of representing multiple high-dimensional stimulus-outcome conjunctions may be solved by such dynamically remapped neurons rather than the allocation of additional cortical real estate to such representations.

The HER model as simulated may be unable to generalize rules across tasks. Because of the specificity of representation at each hierarchical layer, the HER model is unable to generalize associations learned in one context to a novel one. For example, if the identity of stimuli in the 1-2AX task were altered, an instance of the HER model that had learned the task with the original stimulus set would have to relearn the task *ab initio*, even though the task structure remains the same. Nevertheless, this limitation may be overcome if the stimulus representations respond to more general categories of stimuli rather than specific stimulus instances.

In much the same way, it may not be required that a subject have extensive experience with a task in order to acquire error representations appropriate to that task. In previous work (Alexander & Brown, 2011), we described the role of mPFC as learning to predict the outcomes of actions, or, more generally, generic events (Alexander & Brown, 2014), and signaling deviations from predicted outcomes. In the model we present in this letter, we specifically focus on the role of mPFC in error signaling as a mechanism to train representations in dlPFC. However, once such representations are formed, prediction-related activity in mPFC may be sufficient to elicit associated activity in dlPFC that may then be dynamically mapped to external stimuli. Conceptually, this is consistent with models of cognitive control in which mPFC is thought to signal the need for control, which is then implemented by dlPFC (e.g., the conflict model (Botvinick et al., 2001; Yeung et al., 2004). However, the error representation hypothesis suggests that the signal generated by mPFC, rather than reflecting a nonspecific need for increased control, may in fact contain detailed information regarding how top-down control of behavior should be implemented (i.e., specific predictions about likely outcomes given a particular stimulus history).

The HER model incorporates a gating mechanism used to determine whether a task stimulus should be gated into WM. The gating mechanism was inspired by previous computational models of WM in which biologically detailed models of basal ganglia were used to learn the value of storing stimuli in working memory. As our goal in developing the HER model was not to investigate the function of basal ganglia, we opted to use the less biologically plausible mechanism of backpropagation to train the gating mechanism in the HER model. While ACC is known to interact with basal ganglia, it is unlikely that the detailed information carried by the error signal in the HER model is used to train a gating mechanism as implemented in real brains. Rather, it is more likely that such training

involves a scalar signal, typically associated with dopaminergic activity (Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997). While there is evidence that activity in ACC is able to drive dopaminergic activity, it is unclear if this interaction is necessary or sufficient to train a WM gate. Although the specific role assigned to ACC by the HER model remains open to question, additional simulations of the HER model in which the WM gating mechanism is trained by a scalar error signal derived from RL demonstrate the model's ability to solve the 1-2AX task even without the benefit of error backpropagation (see the supplementary material).

Throughout this letter, we refer to WM primarily in the context of maintenance of stimuli associated with information needed to successfully perform a task (i.e., task-related WM). This type of WM may be distinct from other forms of memory involved in, for example, maintaining sensory information over extended periods of time (Braver et al., 2001; D'Esposito, Postle, Ballard, & Lease, 1999; Riggall & Postle, 2012). This means that the HER model's WM capacity of three should not be understood as an absolute limit on WM capacity. Instead, WM includes both the task WM of the HER model specifically and also other kinds of sensory WM. The HER model may be considered to have a rudimentary form of sensory information in the form of eligibility traces (see equation 2.6) that maintain temporally extended representations of recently observed stimuli, allowing the model to learn relationships between temporally distant events. Under this interpretation, the WM mechanism in the HER model, which allows for only a single item to be stored per hierarchical level, may be thought of as an attentional spotlight (Awh & Jonides, 2001; Cowan, 1988; Garavan, 1998) that allows additional processing of individual cues at the expense of others. However, while it is presumably the case that items encoded in a sensory memory store could be attended to in a dynamic, flexible manner, in the HER model, items are eligible to be stored in WM only at the time they are presented. One possible extension to the model, then, would be to more fully account for the interaction of sensory and task-related working memory (e.g., in the form of a visuospatial sketchpad or phonological loop; Baddeley & Hitch, 1974).

The types of tasks the HER model was developed to perform generally involve integrating structured information presented previously in order to generate an appropriate response in a current state. This type of task is fundamentally different from other sorts of tasks thought to involve dlPFC in which a behaving agent is required to generate responses that will lead from a beginning state to a desired end state (i.e., goal-directed behavior). In its present form, the HER model is unable to perform such tasks and therefore is an incomplete account of dlPFC function. It remains an open question as to whether the basic structure of the model is capable of supporting goal-directed behaviors, and to the extent that it could in principle do so, it will likely require augmentation reflecting the function of additional brain regions also known to be involved in goal representation, such as OFC or vmPFC (O'Doherty, 2011).

5 Conclusion

Despite these potential limitations, the HER model provides a promising theoretical framework for interpreting the function and interaction of dlPFC and mPFC, as well as their contribution to behavior. One goal of computational neural modeling is to show how sophisticated cognitive processes can be supported given constraints imposed by the structure and function of the brain. With the exception of the backpropagated error term in equation 2.6 (but see the supplementary material), the computations and architecture of the HER model are relatively straightforward; nevertheless, the model is capable of autonomously learning complex, structured tasks. In doing so, the model displays emergent effects that result from the interaction of the model with a task. These effects include autonomous behavioral shaping in which the model learns concrete stimulus-response mappings prior to higher-level contingencies, category formation, and the mapping of external stimuli to hierarchical layers of the model. These effects were not hardwired into the model and provide a plausible account of how humans may learn such tasks.

A second goal of computational neural modeling is to provide evidence not only that a particular mechanism is plausible given neurobiological constraints, but also that such a mechanism can account for existing data and provide predictions regarding future observations. In this letter, we have focused on describing the function of the HER model without regard for how well it may (or may not) satisfy this goal. Future work with the HER model will aim to demonstrate the ability of the model to account for a range of data from fMRI, EEG, single-unit neurophysiological, lesion, and behavioral studies involving dlPFC. More generally, the HER model, and the error representation hypothesis outlined in section 1, suggests how rules and task sets may be represented in dlPFC. Specifically, the activity of a single prediction unit in the model reflects a prediction of future error signals, with rules being represented in a distributed fashion over several units, each of which signifies some component of that rule. Additional work is needed to identify whether such a representation scheme is implemented in lateral PFC or whether individual "rule neurons" exist for a given task.

References

Abi-Dargham, A., Mawlawi, O., Lombardo, I., Gil, R., Martinez, D., Huang, Y., . . . Laruelle, M. (2002). Prefrontal dopamine D1 receptors and working memory in schizophrenia. *Journal of Neuroscience*, *22*(9), 3708–3719.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.*, *14*, 1338–1344. http://doi.org/10.1038/nn.2921

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience*, *8*, 69. http://doi.org/10.3389/fncom.2014.00069

Amador, N., Schlag-Rey, M., & Schlag, J. (2000). Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J. Neurophysiol.*, *84*, 2166–2170.

Amiez, C., & Petrides, M. (2014). Neuroimaging evidence of the anatomo-functional organization of the human cingulate motor areas. *Cerebral Cortex*, *24*(3), 563–578. http://doi.org/10.1093/cercor/bhs329

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, *230*(4724), 456–458. http://doi.org/10.1126/science.4048942

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, *84*(1), 451–459.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403–450. http://doi.org/10.1146/annurev.neuro.28.061604.135709

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. http://doi.org/10.1016/S1364-6613(00)01593-X

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Orlando, FL: Academic Press. http://www.sciencedirect.com/science/article/pii/S0079742108604521

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200. http://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*(12), 2082–2099. http://doi.org/10.1162/jocn.2007.19.12.2082

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*(9), 659–669. http://doi.org/10.1038/nrn2667

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico–striatal circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. http://doi.org/10.1093/cercor/bhr117

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315–326. http://doi.org/10.1016/j.neuron.2010.03.025

Barbas, H., & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *286*(3), 353–375. http://doi.org/10.1002/cne.902860306

Barbas, H., & Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, *7*(7), 635–646. http://doi.org/10.1093/cercor/7.7.635

Barto, A., Bradtke, S. J., & Singh, S. P. (1995). Learning to act using real-time dynamic programming. *Artificial Intelligence*, *72*(1–2), 81–138. http://doi.org/10.1016/0004-3702(94)00011-O

Barto, A., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, *13*(4), 341–379. http://doi.org/10.1023/A:1025696116075

Barto, A., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-13*(5), 834–846. http://doi.org/10.1109/TSMC.1983.6313077

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.*, *10*, 1214–1221.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. http://doi.org/10.1016/j.tics.2008.02.009

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. C. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Botvinick, M. M., Niv, Y., & Barto, A. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. http://doi.org/10.1016/j.cognition.2008.08.011

Braver, T. S., Barch, D. M., Kelley, W. M., Buckner, R. L., Cohen, N. J., Miezin, F. M.,. . . Petersen, S. E. (2001). Direct comparison of prefrontal cortex regions engaged by working and long-term memory tasks. *NeuroImage*, *14*(1), 48–59. http://doi.org/10.1006/nimg.2001.0791

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the Anterior cingulate cortex. *Science*, *307*, 1118–1121.

Brunel, N., & Wang, X.-J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience*, *11*(1), 63–85. http://doi.org/10.1023/A:1011204814320

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, *4*, 215–222.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749. http://doi.org/10.1126/science.280.5364.747

Clark, S. A., Allard, T., Jenkins, W. M., & Merzenich, M. M. (1988). Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature*, *332*(6163), 444–445. http://doi.org/10.1038/332444a0

Cohen, J., & Servan-Schreiber, D. (1993). A theory of dopamine function and its role in cognitive deficits in schizophrenia. *Schizophrenia Bulletin-PsycArticles 1993*, *19*(1), 85–104.

Collins, A., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. http://doi.org/10.1037/a0030852

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biol*, *10*(3), e1001293. http://doi.org/10.1371/journal.pbio.1001293

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*, 297–338.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. http://doi.org/10.1037/0033-2909.104.2.163

Denys, D., Zohar, J., & Westenberg, H. (2003). The role of dopamine in obsessive-compulsive disorder: Preclinical and clinical evidence. *Journal of Clinical Psychiatry*, *65*(Suppl. 14), 11–17.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fmri study. *Brain and Cognition*, *41*(1), 66–86. http://doi.org/10.1006/brcg.1999.1096

Dixon, M. L., Fox, K.C.R., & Christoff, K. (2014). Evidence for rostro-caudal functional organization in multiple brain areas related to goal-directed behavior. *Brain Research*, *1572*, 26–39. http://doi.org/10.1016/j.brainres.2014.05.012

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–1486. http://doi.org/10.1126/science.1252254

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. http://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. http://doi.org/10.1038/nrn2787

Garavan, H. (1998). Serial attention within working memory. *Memory and Cognition*, *26*(2), 263–276.

Gariano, R. F., & Groves, P. M. (1988). Burst firing induced in midbrain dopamine neurons by stimulation of the medial prefrontal and anterior cingulate cortices. *Brain Research*, *462*(1), 194–198. http://doi.org/10.1016/0006-8993(88)90606-3

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: An event-related potential accompanying errors. *Psychophysiology*, *27*, S34.

Genovesio, A., Brasted, P. J., Mitz, A. R., & Wise, S. P. (2005). Prefrontal cortex activity related to abstract response strategies. *Neuron*, *47*(2), 307–320. http://doi.org/10.1016/j.neuron.2005.06.006

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. http://doi.org/10.1016/j.neuron.2010.04.016

Goodman, W. K., McDougle, C. J., Price, L. H., Riddle, M. A., Pauls, D. L., & Leckman, J. F. (1990). Beyond the serotonin hypothesis: A role for dopamine in some forms of obsessive compulsive disorder? *Journal of Clinical Psychiatry*, *51*(Suppl.), 36–43.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*(1), 1–51. http://doi.org/10.1037/0033-295X.87.1.1

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. http://doi.org/10.1162/neco.1997.9.8.1735

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psych. Rev.*, *109*, 679–709.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nat. Neurosci.*, *7*, 497–498.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. http://doi.org/10.1016/j.tics.2011.12.008

Hussar, C. R., & Pasternak, T. (2013). Common rules guide comparisons of speed and direction of motion in the dorsolateral prefrontal cortex. *Journal of Neuroscience*, *33*(3), 972–986. http://doi.org/10.1523/JNEUROSCI.4075-12.2013

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, *17*(10), 2443–2452. http://doi.org/10.1093/cercor/bhl152

Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *NeuroImage*, *95*, 80–89. http://doi.org/10.1016/j.neuroimage.2014.03.050

Jong, N. K., & Stone, P. (2008). Hierarchical model-based reinforcement learning: R-max + MAXQ. In *Proceedings of the 25th International Conference on Machine Learning* (pp. 432–439). New York: ACM. http://doi.org/10.1145/1390156.1390211

Jun, J. K., Miller, P., Hernández, A., Zainos, A., Lemus, L., Brody, C. D., & Romo, R. (2010). Heterogenous population coding of a short-term memory and decision task. *Journal of Neuroscience*, *30*(3), 916–929. http://doi.org/10.1523/JNEUROSCI.2062-09.2010

Kehoe, E. J. (1986). Summation and configuration in conditioning of the rabbit's nictitating membrane response to compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*(2), 186–195. http://doi.org/10.1037/0097-7403.12.2.186

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.*, *9*, 940–947.

Klopf, A. H. (1972). *Brain function and adaptive systems: A heterostatic theory*. Hanscom Air Force Base, MA: Air Force Cambridge Research Labs.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181–1185. http://doi.org/10.1126/science.1088545

Kolling, N., Behrens, T.E.J., Mars, R. B., & Rushworth, M.F.S. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–98. http://doi.org/10.1126/science.1216930

Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, *12*(7), 939–945. http://doi.org/10.1038/nn.2321

Krueger, K. A. (2011). *Sequential learning in the form of shaping as a source of cognitive flexibility*. London: UCL (University College London).

Krueger, K. A., & Dayan, P. (2009). Flexible shaping: How learning in small steps helps. *Cognition*, *110*(3), 380–394. http://doi.org/10.1016/j.cognition.2008.11.014

MacDonald, A. W. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–1838.

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. http://doi.org/10.1037/h0076778

Medalla, M., & Barbas, H. (2009). Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron*, *61*(4), 609–620. http://doi.org/10.1016/j.neuron.2009.01.006

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. http://doi.org/10.1146/annurev.neuro.24.1.167

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154–5167.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.

Muly, E. C., Szigeti, K., & Goldman-Rakic, P. S. (1998). D1 receptor in interneurons of macaque prefrontal cortex: Distribution and subcellular localization. *Journal of Neuroscience*, *18*(24), 10553–10565.

Murase, S., Grenhoff, J., Chouvet, G., Gonon, F. G., & Svensson, T. H. (1993). Prefrontal cortex regulates burst firing and transmitter release in rat mesolimbic dopamine neurons studied in vivo. *Neuroscience Letters*, *157*(1), 53–56.

Murphy, B. L., Arnsten, A. F., Goldman-Rakic, P. S., & Roth, R. H. (1996). Increased dopamine turnover in the prefrontal cortex impairs spatial working memory performance in rats and monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1325–1329.

Nee, D. E., & Brown, J. W. (2012). Rostral-caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *NeuroImage*, *63*(3), 1285–1294. http://doi.org/10.1016/j.neuroimage.2012.08.034

Nee, D. E., & Brown, J. W. (2013). Dissociable frontal-striatal and frontal-parietal networks involved in updating hierarchical contexts in working memory. *Cerebral Cortex*, *23*(9), 2146–2158. http://doi.org/10.1093/cercor/bhs194

Nee, D. E., Jahn, A., & Brown, J. W. (2013). Prefrontal cortex organization: Dissociating effects of temporal abstraction, relational abstraction, and integration with fMRI. *Cerebral Cortex*, *24*, 2377–2387. bht091. http://doi.org/10.1093/cercor/bht091

Nee, D. E., Kastner, S., & Brown, J. W. (2011). Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. *NeuroImage*, *54*(1), 528–540. http://doi.org/10.1016/j.neuroimage.2010.08.027

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research*, *171*(2), 213–224.

O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, *1239*, 118–129. http://doi.org/10.1111/j.1749-6632.2011.06290.x

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.*, *18*, 283–328.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

Rainer, G., Rao, S. C., & Miller, E. K. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience*, *19*(13), 5493–5505.

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

Rescorla, R. A. (1971). Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, *2*(2), 113–123. http://doi.org/10.1016/0023-9690(71)90002-6

Reynolds, J. R., O'Reilly, R. C., Cohen, J. D., & Braver, T. S. (2012). The function and organization of lateral prefrontal cortex: A test of competing hypotheses. *PLoS One*, *7*, e30284. http://doi.org/10.1371/journal.pone.0030284

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, *32*(38), 12990–12998. http://doi.org/10.1523/JNEUROSCI.1892-12.2012

Rougier, N. P., Noelle, D. C., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7338–7343. http://doi.org/10.1073/pnas.0502455102

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci*, *28*, 13775–13775. http://doi.org/10.1523/JNEUROSCI.3541-08.2008

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. http://doi.org/10.1126/science.275.5306.1593

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*(1), 203–219. http://doi.org/10.1146/annurev.neuro.051508.135546

Seo, H., Barraclough, D. J., & Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cerebral Cortex*, *17*(Suppl. 1), i110–117. http://doi.org/10.1093/cercor/bhm064

Servan-Schreiber, D., Cohen, J. D., & Steingard, S. (1996). Schizophrenic deficits in the processing of context. A test of a theoretical model. *Archives of General Psychiatry*, *53*(12), 1105–1112.

Shima, K., & Tanji, J. (1998). Role of cingulate motor area cells in voluntary movement selection based on reward. *Science*, *282*, 1335–1338.

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. *Proceedings of the Seventh International Conference on Machine Learning* (pp. 216–224). San Mateo, CA: Morgan Kaufmann.

Sutton, R. S., & Barto, A. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel& J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). Cambridge, MA: MIT Press.

Taber, M. T., & Fibiger, H. C. (1993). Electrical stimulation of the medial prefrontal cortex increases dopamine release in the striatum. *Neuropsychopharmacology*, *9*(4), 271–275. http://doi.org/10.1038/npp.1993.63

Taren, A. A., Venkatraman, V., & Huettel, S. A. (2011). A parallel functional topography between medial and lateral prefrontal cortex: Evidence and implications for cognitive control. *Journal of Neuroscience*, *31*(13), 5026–5031. http://doi.org/10.1523/JNEUROSCI.5762-10.2011

Wallis, J. D., & Miller, E. K. (2003). From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology*, *90*(3), 1790–1806. http://doi.org/10.1152/jn.00086.2003

Walton, M. E., Devlin, J. T., & Rushworth, M.F.S. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nat. Neurosci.*, *7*, 1259–1266. http://doi.org/10.1038/nn1339

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *1960 Wescon Convention Record* (pp. 96–104).

Williams, S. M., & Goldman-Rakic, P. S. (1993). Characterization of the dopaminergic innervation of the primate frontal cortex using a dopamine-specific antibody. *Cerebral Cortex*, *3*(3), 199–222.

Yamada, M., Pita, M. del C. R., Iijima, T., & Tsutsui, K.-I. (2010). Rule-dependent anticipatory activity in prefrontal neurons. *Neuroscience Research*, *67*(2), 162–171. http://doi.org/10.1016/j.neures.2010.02.011

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol. Rev.*, *111*, 931–959.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. http://doi.org/10.1016/j.neuron.2005.04.026

Yu, C., Zhou, Y., Liu, Y., Jiang, T., Dong, H., Zhang, Y., & Walter, M. (2011). Functional segregation of the human cingulate cortex is confirmed by functional connectivity based neuroanatomical parcellation. *NeuroImage*, *54*(4), 2571–2581. http://doi.org/10.1016/j.neuroimage.2010.11.018

Zarr, N., & Brown, J. (2015, March 6). *Hierarchical error representations in medial prefrontal cortex*. Manuscript submitted for publication.