Cerebral Cortex March 2012;22:527-536
doi:10.1093/cercor/bhr117
Advance Access publication June 21, 2011

Mechanisms of Hierarchical Reinforcement Learning in Cortico–Striatal Circuits 2: Evidence from fMRI

David Badre and Michael J. Frank

Department of Cognitive, Linguistic, and Psychological Sciences, Brown Institute for Brain Sciences, Brown University, Providence, RI 02912-1978, USA

Address correspondence to David Badre. Email: david_badre@brown.edu.

The frontal lobes may be organized hierarchically such that more rostral frontal regions modulate cognitive control operations in caudal regions. In our companion paper (Frank MJ, Badre D. 2011. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits I: computational analysis. 22:509–526), we provide novel neural circuit and algorithmic models of hierarchical cognitive control in cortico–striatal circuits. Here, we test key model predictions using functional magnetic resonance imaging (fMRI). Our neural circuit model proposes that contextual representations in rostral frontal cortex influence the striatal gating of contextual representations in caudal frontal cortex. Reinforcement learning operates at each level, such that the system adaptively learns to gate higher order contextual information into rostral regions. Our algorithmic Bayesian "mixture of experts" model captures the key computations of this neural model and provides trial-by-trial estimates of the learner's latent hypothesis states. In the present paper, we used these quantitative estimates to reanalyze fMRI data from a hierarchical reinforcement learning task reported in Badre D, Kayser AS, D'Esposito M. 2010. Frontal cortex and the discovery of abstract action rules. Neuron. 66:315–326. Results validate key predictions of the models and provide evidence for an individual cortico–striatal circuit for reinforcement learning of hierarchical structure at a specific level of policy abstraction. These findings are initially consistent with the proposal that hierarchical control in frontal cortex may emerge from interactions among nested cortico–striatal circuits at different levels of abstraction.

Keywords: basal ganglia, cognitive control, fMRI, prefrontal cortex, reinforcement learning

Introduction

Human behavior is marked by its flexibility. Even in novel circumstances, we are capable of adaptively choosing the courses of action that are likely to achieve our goals. This behavioral flexibility partly arises from our capacity for cognitive control, which is the mechanism by which we use plans, goals, or features of our environment to constrain action selection (Miller and Cohen 2001; Badre and Wagner 2004; Bunge 2004; O'Reilly and Frank 2006). Cognitive control function is known to partly depend on interactions between lateral frontal cortex and the basal ganglia (Cools et al. 2006; Frank and O'Reilly 2006; Cohen and Frank 2009). However, the precise mechanisms by which these interactions produce flexible goal-directed behavior remain underspecified.

Recent work has provided evidence that cognitive control function may be organized systematically along the rostro–caudal axis of the frontal lobes, such that control based on progressively abstract representations is associated with progressively rostral frontal regions (Koechlin et al. 2003; Koechlin and Jubault 2006; Badre and D'Esposito 2007; Badre 2008). Abstraction has been defined differently across these studies (see Badre 2008). However, one definition receiving empirical support—and that is adopted in the current work—is in terms of "policy abstraction." Policy refers to the mapping between a particular state, an action, and an anticipated outcome. Simple policy relates a state to an overt action, as in an arbitrary stimulus–response mapping. Policy is defined as more abstract to the extent that a state comes to represent a class of lower order policy. When manipulated experimentally, increases in selection demands at systematically higher levels of policy abstraction have been associated with progressively rostral activation in lateral frontal cortex (Badre and D'Esposito 2007; Badre et al. 2009).

Importantly, evidence from effective connectivity analysis of functional magnetic resonance imaging (fMRI) data (Koechlin et al. 2003) and from patients with lateral frontal damage (Badre et al. 2009) further indicates that dynamics along the rostro–caudal frontal axis may be hierarchical, in that processing in rostral regions appears to differentially influence processing in caudal regions more than vice versa. Prior theorizing has generally assumed that these rostro-to-caudal interactions to arise from cortico–cortical connections within lateral frontal cortex itself (Fuster 2001; Koechlin and Summerfield 2007; Badre and D'Esposito 2009). However, in a companion paper (Frank and Badre 2011), we develop an explicit model of hierarchical cognitive control in which rostro–caudal influences occur via nested cortico–striatal circuits rather than (or in addition to) direct cortico–cortical connections. However, to date, few studies have directly investigated interactions between frontal cortex and striatum during hierarchical cognitive control tasks, leaving key predictions from the model untested. To provide initial evidence, this paper presents a reanalysis of a hierarchical reinforcement learning task (Badre et al. 2010) based on novel quantitative predictions derived from the computational model described in the companion paper (Frank and Badre 2011).

In the experiment, human participants were scanned with fMRI while they learned, through reinforcement, 18 mappings between the conjunction of 3 features of a presented stimulus (shape, orientation, and color) and 1 of 3 finger responses on a keypad. Critically, each participant learned 2 such sets of 18 rules, and for 1 of the sets, abstract policy was available that would permit generalization across multiple individual mappings of stimuli and responses.

The results from this experiment demonstrated that 1) fMRI activation was evident in both dorsal premotor cortex (PMd)

and more rostral premotor cortex (prePMd) early in learning but declined in the prePMd by the end of learning when no abstract rule was available, 2) Participants were capable of rapidly discovering and applying the abstract rule when it was available, 3) Individual differences in the activation early in learning in prePMd, but not in PMd, were correlated with behavioral markers of the successful discovery of an abstract rule when one was available, and 4) striatum (caudate and putamen) showed functional connectivity with prePMd and PMd that was consistent across learning at different levels of the hierarchy. Hence, these results suggest that from the outset of learning, the search for relationships between context and action may occur at multiple levels of abstraction simultaneously and that this process differentially relies on systematically more rostral portions of frontal cortex, along with striatum, for the discovery of more abstract relationships.

Frank and Badre (2011) developed a mechanistic account of these findings using 2 models of hierarchical reinforcement learning at different levels of analysis (neural circuit and algorithmic). The companion paper provides theoretical and implementation details of these models along with analysis of their relationship to each other and to hierarchical learning and control more generally. Here, we present a high-level summary of each model and our approach to testing these models in fMRI data.

The neural network model was adapted for hierarchical learning from an established neural model of action selection, working memory, and cognitive control (Frank et al. 2001; Brown et al. 2004; Frank and O'Reilly 2006; Gruber et al. 2006; O'Reilly and Frank 2006). Frank et al. (2001) proposed a neural network architecture in which interactions between the prefrontal cortex (PFC) and the basal ganglia support an adaptive gating mechanism that determines whether or not contextual information is allowed to enter and to be maintained in working memory. While in working memory, contextual information can provide a top–down influence on action selection. In the model, the striatum disinhibits thalamic units, permitting certain components of the sensory input to update PFC working memory states (i.e., input gating). Using similar mechanisms, other striatal modules determine which of the currently maintained PFC representations should influence action selection (i.e., output gating), and in turn, which motor response should be emitted given this PFC state (i.e., response gating). The selection of which representations to gate at all striatal levels is learned via a common dopaminergic reward prediction error (RPE) signal (Montague et al. 1996) that modulates activity and plasticity in "Go" and "NoGo" striatal neuronal populations (Frank 2005; O'Reilly and Frank 2006; Shen et al. 2008).

The companion paper presented a modification to this model to accommodate hierarchical control and learning. Specifically, we proposed that information maintained in rostral regions of PFC provides contextual input to the striatal units that determine which of the more caudal PFC representations should be output gated (Fig. 1). Thus, in this model, hierarchical control emerges from a series of nested cortico–striatal loops arrayed along the rostro–caudal axis of the frontal lobe. Applied to the Badre et al. (2010) learning task, contextual information maintained in prePMd influences striatal units that output gate information maintained in PMd, which in turn influences motor response selection. We showed that this model facilitates hierarchical learning relative to models without such structure, by learning a gating policy abstraction and reducing the dimensionality of stimulus–response mappings.

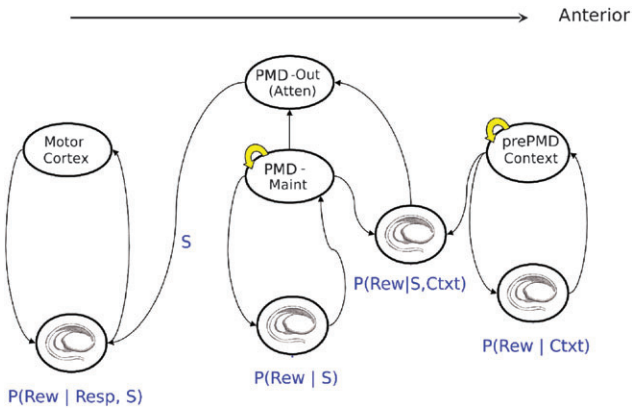

Figure 1. Schematic of hierarchical cortico–striatal circuit. In the standard response selection circuit, motor areas of the striatum interact with motor cortex to facilitate response selection based on the learned probability of reward given the current stimulus state. The PMd-maint layer represents possible stimuli to be actively maintained so as to constrain motor selection processes. Its corresponding striatal region learns which stimulus dimensions should be gated into PMd based on the learned probability that their maintenance is predictive of reward. The PMd-out layer represents the deep lamina (e.g., layers 5/6) of PMd in which only a subset of currently maintained PMd stimuli influences response selection, by projecting to the motor striatum. Its corresponding striatal area learns which of the maintained PMd stimuli should be output gated depending on context. The most anterior prePMd layer maintains stimulus features that act as context, by sending their axons to striatal output gating areas of PMd. Its corresponding striatal gating layer learns whether the maintenance of particular stimuli as higher order context in prePMd is predictive of reward.

In the companion paper, we also develop a more abstract Bayesian mixture of experts (MoEs) model of hierarchical rule learning, intended to correspond to key computational features of the neural model. This model can estimate latent states, that is, hypotheses about the relationships between context and action that are most likely being tested (including hierarchical hypotheses), in individual human learners given their trial-by-trial sequences of choices and rewards.

In the MoE, each expert represents a hypothesis that a participant could entertain about the task structure by focusing on a particular stimulus dimension or combination of dimensions and learning the probability of obtaining a reward for each motor response given the features present in their domain of expertise. For example, the orientation expert would learn "P(Rew|Response,Orient)" and so forth for other experts. Attentional weights assigned to the different experts correspond to the probability that expert contributes to response selection. These attentional weights are learned via a reinforcement learning credit assignment mechanism akin to that used in the neural model. In hierarchical experts, attentional weights to 1 of 2 lower dimensions (i.e., shape or orientation) are dynamically assigned conditional on the identity of the higher dimension (i.e., color). Finally, attentional weights to overall hierarchical relative to flat experts are learned as a function of their relative reward probabilities in the given task context (depending on whether the task structure is hierarchical).

In the companion paper, we showed that the MoE model provides a good quantitative fit to both human participant choices and to those of the cortico–striatal network model.

Moreover, by manipulating the ability of the cortico–striatal network to test hierarchical structure, we validated that the MoE can successfully infer, based on observed stimuli, choices, and rewards alone, the attentional weights to hierarchical rules.

Here, we use the MoE model-derived attentional weights for individual participants to test key aspects of the mechanistic hierarchical reinforcement learning in fMRI data. Importantly, the models make 3 central sets of predictions that will be the focus of the present paper.

First, the neural model predicts that prePMd activation during flat and hierarchical blocks reflects testing of higher order policies. Thus, the more activated that prePMd is, the more likely it is that hierarchical structure will be discovered. Consistent with this hypothesis, Badre et al. (2010) found that activation differences in prePMd early in the learning trial correlated with behavioral differences between the hierarchical versus flat learning conditions. However, the model permits this prediction to be tested even more directly. In particular, greater activation in prePMd during the hierarchical learning session should correlate with the extent to which attention is allocated to hierarchical structure, as estimated by the model. Thus, we test whether individual differences in prePMD activity during the hierarchical block are associated with individual differences in attention to hierarchical structure as estimated by the MoE.

Second, our model makes a specific prediction about the nature of the learning signals that reinforce the discovery of hierarchical structure. In particular, hierarchical learning in the model emerges from interactions among nested cortico–striatal loops operating at different levels of policy abstraction that get reinforced and punished as a function of dopaminergic RPE signals. Many human reinforcement learning studies report RPE correlates across the extent of the striatum (McClure et al. 2003; O'Doherty et al. 2003; Pessiglione et al. 2006; Rutledge et al. 2010), a finding that we replicate here. However, our hierarchical model makes the more specific prediction that blood oxygen level–dependent (BOLD) signal in a restricted striatal region—at the same rostro–caudal level as prePMd—should covary with RPE as a function of whether participants are attending to a hierarchical rule. We test this prediction by interrogating areas of the brain sensitive to RPE to determine to what extent these signals are modulated by attention to hierarchical rules, as estimated by the MoE.

Third, we test whether these striatal RPE signals are predictive of changes in frontal activation states as a function of learning. Specifically, to the extent that action selection is guided by a rule at a particular level of abstraction but does not produce positive outcomes (i.e., does not well describe the actual task contingencies), then negative RPE will punish the use of this rule, and the corresponding region of frontal cortex should decrease its participation in learning (see Frank and Badre 2011; Fig. 4). This mechanism provides an account of the decline in prePMd observed in Badre et al. (2010) during flat blocks. Recall that in the model, information represented in prePMd provides contextual input to output gating units determining which of the other dimensions should be output gated (attended) to guide motor response selection. When no hierarchical structure exists, there is no context that reliably predicts when the network should constrain attention to a particular stimulus feature. As such, the prePMd influence can actually hinder performance because it will force the model to focus on a subset of dimensions when it should instead learn about the conjunction of all stimulus features on each trial. Thus, any pattern of prePMd activity is predictive of poor performance and hence elicits a negative RPE which, in turn, drives NoGo learning in the associated BG circuit it so that it becomes less likely it to gate stimuli into (or out of) the prePMd. As a result, model prePMd activity levels decline with increasing trials in the flat condition. By contrast, in the hierarchical condition, BG gating units are positively reinforced when color is represented in prePMd, so that activity is maintained across trials. Thus, our model predicts that the decline in prePMd during the flat block should correlate with striatal negative prediction error signals deriving from reliance on the hierarchical rule.

To summarize, then, based on an individual's trial-to-trial sequence of responses and rewards, the MoE permits computation of 3 types of variables which can be correlated with the BOLD response: 1) the attentional weight for each expert (type of rule) that corresponds to the probability that its respective hypothesis is contributing to response selection on a given trial, 2) the RPE which corresponds to the difference between the actual and expected outcome on a given trial, and 3) the product of the RPE and the attentional weight which estimates to the extent to which RPE's act to reinforce or punish a particular rule type (including the hierarchical rule) on a given trial.

Using these MoE estimates to analyze fMRI data, we sought to assess 1) whether individual differences in activation in neural structures (i.e., prePMd) thought to support hierarchical control are predictive of model estimates of attention to hierarchical versus flat rule structure during learning, 2) whether model estimates of RPE, both generally and specifically associated with expected outcomes given a hierarchical rule, are systematically associated with regions of striatum and frontal cortex, and 3) whether negative prediction error associated with the hierarchical rule is correlated with individual differences in the decline in activation in prePMd during learning of the flat rule set, as predicted by the neural model.

Materials and Methods

Participants

We reanalyzed fMRI data collected from the experiment described in Badre et al. (2010) using estimates from the Bayesian MoE model. Details concerning participant characteristics are published in Badre et al. (2010). We note that 16 of the 20 participants included in Badre et al. (2010) were analyzed in the present project. This is due to the fact that 4 participants in the original experiment had asymmetric proportions of individual response mappings in one of the learning conditions. Though this had no bearing on the logic, results, or conclusions of the Badre et al. (2010) study, we chose not to include these participants in the current modeling analysis in order to avoid any potential bias of the model fits.

Hierarchical Learning Task

Participants were required to learn through reinforcement the mapping between a stimulus consisting of a shape at a particular orientation surrounded by a colored box and 1 of 3 responses on a keypad (Fig. 2). Each participant learned 2 rule sets: one that only could be learned as 18 individual mappings between the conjunction of color, shape, and orientation and a response (flat rule set; Fig. 2*b*) and one that offered the opportunity to learn an abstract rule (hierarchical rule set; Fig. 2*c*). Participants were not given an indication through an

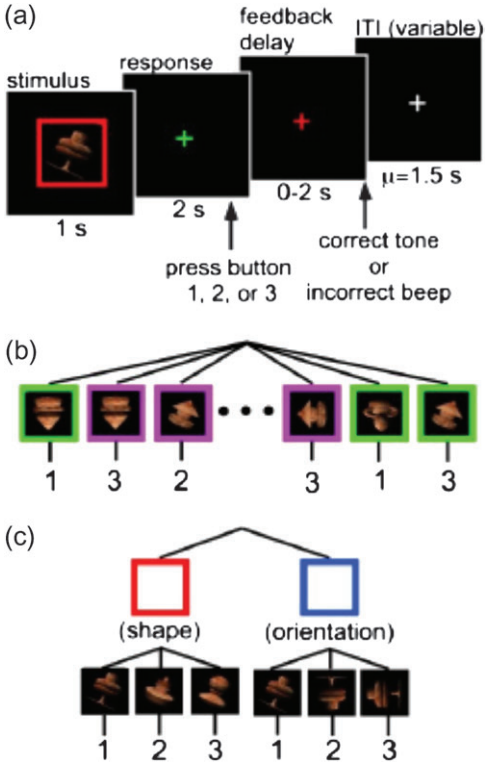

Figure 2. Schematics of the hierarchical learning task from Badre et al. (2010). (*a*) Depiction of trial events during both learning epochs. On each trial, the participant is presented with a shape, at a particular orientation, surrounded by a colored box. They then choose 1 of 3 responses on the keypad depending on these stimulus features. This is followed by feedback indicating whether the response was correct or not. Feedback was separated from stimulus onset to permit separate event-related analysis of these 2 phases. (*b*) Policy structure for the flat condition. In the flat condition, 18 unique mappings had to be learned between each conjunction of shape, orientation, and color and a response, yielding a wide flat first-order structure. (*c*) Policy structure for the hierarchical condition. If they learned the contingent relationship between color and orientation versus shape (second-order policy), participants could select a subset of shape- or orientation-based rules depending on color.

instruction or any other cue that a higher order structure existed in one of the rule sets. The order of trials and duration of intertrial intervals within a block were determined by optimizing the efficiency of the design matrix so as to permit estimation of the event-related response (Dale 1999). Efficiency was equated across rule sets, and the order of rule set learning (i.e., whether hierarchical or flat was learned first) was counterbalanced across participants. Additional details concerning presentation parameters, individual trial events, and stimulus features are described in Badre et al. (2010).

fMRI Procedures

Whole-brain imaging was performed on a Siemens 3T TIM Trio MRI system using a standard 12-channel head coil. Functional data were acquired using a gradient-echo echo-planar pulse sequence (time repetition = 2 s, time echo = 28 ms, flip angle = 90°; 29 axial slices, matrix = 128 × 128, field of view = 230 × 230 mm, slice thickness = 3 mm, 203 volume acquisitions per run). High-resolution T_1-weighted [magnetization prepared rapid gradient echo (MP-RAGE)] anatomical images were collected for anatomical visualization. Head motion was restricted using firm padding that surrounded the head. Visual stimuli projected onto the screen were viewed through a mirror attached to the head coil. Auditory feedback was presented through Siemens headphones provided as a stock component with the Trio scanner. All experimental scripts were programmed and run on a Macintosh computer using the Psychophysics Toolbox in MATLAB (http://psychtoolbox.org/).

Functional imaging data were processed using SPM2 (Wellcome Department of Cognitive Neurology, London). Following quality assurance procedures to assess outliers or artifacts in volume and slice-to-slice variance in the global signal, functional images were corrected for differences in slice acquisition timing by resampling all slices in time to match the first slice, followed by motion correction using sinc interpolation across all runs. The mean functional image was then coregistered with the high-resolution MP-RAGE anatomical image. After normalizing the MP-RAGE to Montreal Neurological Institute stereotaxic space, we applied the same normalization parameters (determined by a 12 parameter affine transformation along with a nonlinear transformation using cosine basis functions) to each of the realigned functional images. Images were resampled into 2 × 2 × 2 mm voxels and then spatially smoothed with an 8-mm full-width at half-maximum isotropic Gaussian kernel.

fMRI Analysis

Functional imaging data were analyzed using SPM2 (Wellcome Department of Cognitive Neurology, London). Statistical models were constructed under the assumptions of the general linear model (GLM). The fMRI analyses used quantitative estimates derived from the model fits to each subject of the MoE model. For details on model fitting and implementation, see the companion paper (Frank and Badre 2011). Here, we will provide summary details concerning the parametric regressors and how they were included in the GLM used to analyze the fMRI data.

Two statistical models were used to analyze the fMRI data. First, we constructed a model that included 4 event-related regressors corresponding to stimulus and feedback onsets for the hierarchical and flat learning sets. Two parametric regressors were included in association with each stimulus onset: 1) model estimates of trial-to-trial changes in attention to the correct hierarchical rule (i.e., shape or orientation given color; $w_{OS|C}$) and 2) model estimates of attention to the flat rule (i.e., the conjunction of shape, color, and orientation; w_{OSC}). As noted in the Introduction, these attentional weight parameters reflect the probability that a given hypothesis about the relationship between state and response is determining a response on a participant trial. Thus, $w_{OS|C}$ indexes the probability that the participants is testing the hypothesis that color determines whether orientation or shape are relevant to a response, a hierarchical rule. And, w_{OSC} indexes the probability that the participant is testing the hypothesis that the conjunction of shape, color, and orientation determine the response, a flat rule.

Three parametric regressors were associated with each feedback onset: 1) model estimates of RPE on every trial. RPE is defined as the actual reward outcome minus probability that the current response would be rewarding given the model output (using a weighted average of each expert's estimated reward probability for that response in the given state, scaled by the attentional weight to that expert, as estimated from the subject's choices); 2) RPE related to the hierarchical rule specifically, that is scaled by attentional weight to the that rule ($RPE_{Hmod} = RPE \times w_{OS|C}$); 3) RPE related to the flat rule ($RPE_{Fmod} = RPE \times w_{OSC}$).

Though not collinear, the parametric regressor for RPE was correlated in some subjects with RPE_{Hmod} and RPE_{Fmod}, potentially making it difficult to estimate the effects for the latter two conditions in this first model. Thus, we constructed a second statistical model that also estimated the effects of RPE_{Hmod} and RPE_{Fmod}. However, in this model, we modeled out variance due to positive or negative feedback rather than RPE as a parametric regressor. Specifically, event-related regressors were included for stimulus onset, positive feedback onset, and negative feedback onset for hierarchical and flat sets. Each stimulus onset was associated with parametric regressors for attentional weights to the hierarchical and flat rules ($w_{OS|C}$ and w_{OSC}). Each feedback onset was associated with parametric regressors for RPE_{Hmod} and RPE_{Fmod}. Thus, analyses reported in the results that are based on RPE_{Hmod} and RPE_{Fmod} are estimated from this GLM.

It is notable that credit assignment in the MoE model is not actually derived from the product of the individual attentional weight for each rule and the RPE (i.e., RPE_{Hmod} or RPE_{Fmod}). Rather, the MoE assigns credit though a system of filtered rewards, wherein experts are only

reinforced in the same direction as the actual RPE if the selected response was the same as that predicted by the expert (see Frank and Badre 2011, for detailed discussion). However, the use of binary or categorical filtered rewards as a regressor for fMRI analysis was not feasible because to do so requires accurately knowing whether a given expert contributed to a response on each trial (i.e., when an expert's assigned reward probability for the selected action is less than other actions, the filtered reward value is inverted). With this practical limitation in mind, RPE_{Hmod} and RPE_{Fmod} are good proxies for filtered rewards because they provide a probabilistic estimate of the likelihood that a given expert contributed to the response and therefore of the likelihood that the filtered reward is the same as the RPE. In other words, multiplying RPE by attention to the hierarchical or flat rules effectively filters the RPEs to be in the same direction as that expected by the filtered rewards used in the MoE. And, by contrast, when RPE_{Hmod} is low, RPEs are more likely to be uncorrelated with filtered rewards. Of course, as already noted, the RPE_{Hmod} and RPE_{Fmod} regressors are also consistent with the mechanistic neural circuit model, in that learning occurs particularly when representations are gated into frontal cortex (and hence "attended"). Hence, the use of RPE_{Hmod} and RPE_{Fmod} does not distinguish whether people only reinforce hypotheses that are currently attended (as in the neural circuit model) or whether they are also using counterfactual prediction errors to appropriately reinforce experts that did not gate responses (as in the MoE). Future study is required to investigate this specific question.

Statistical effects were estimated using a subject-specific fixed-effects model, with run and session-specific effects and low-frequency signal components (<0.01 Hz) treated as confounds. Linear contrasts were used to obtain subject-specific estimates for each effect. These estimates were entered into a second-level analysis treating subject as a random effect, using a 1-sample *t*-test against a contrast value of zero at each voxel. Whole-brain voxel-based group effects reported in the results were considered reliable to the extent that they consisted of voxels that exceeded a family-wise error (FWE) corrected threshold of $P < 0.05$.

Whole-brain voxel-wise event-related analysis was supplemented by region-of-interest (ROI) analysis. ROIs were defined in 2 ways: 1) the ROIs for PMd, prePMd, inferior frontal sulcus (IFS), and frontal polar cortex (FPC) were taken from a prior fMRI study of hierarchical cognitive control (Badre and D'Esposito 2007) based on their association with first-, second-, third-, and fourth-order rule execution, respectively and 2) all other ROIs were defined as all significant voxels within 8 mm of a maximum chosen from the contrast of RPE versus baseline in the current experiment. We note that though RPE shares a nonlinear relationship with RPE_{Hmod} and RPE_{Fmod}, defining ROIs based on RPE may introduce some bias in the assessment of the simple of effects of RPE_{Hmod} and RPE_{Fmod} against baseline. However, neither RPE_{Hmod} and RPE_{Fmod} were differentially correlated with RPE ($P = 0.99$) and so differences between RPE_{Hmod} and RPE_{Fmod} or region by effect interactions from these ROIs are unbiased.

Selective averaging with respect to peristimulus time was conducted using the Marsbars toolbox , permitting assessment of the signal change associated with each condition. ROI analysis of parametric effects was conducted by averaging beta estimates from voxels contained in the a priori defined ROI.

Median split analyses were conducted between subjects based on the neutrally defined peak stimulus-related percent signal change from an ROI. Specifically, participants were split into 2 groups based on whether they were above or below the median on a measure of interest, such as overall attention to hierarchy (w_H). Peak stimulus-related activation was then compared between these groups of participants using a *t*-test. Median split analyses were also followed with between-subject linear regression analyses that included peak stimulus-related percent signal change from an ROI as the dependent variable and a measure of interest, like w_H, as the independent variable.

Results

As described in the Introduction, we sought to assess 3 key predictions in reanalysis of fMRI data, using estimates from the MoE model: 1) Greater activation in prePMd will be associated with greater attention to hierarchical versus flat rule structure across individuals, 2) RPE associated with testing a hierarchical rule will be specifically associated with a local circuit between prePMd and striatum, and 3) RPE associated with hierarchical rule will correlate with individual differences in the decline in activation in prePMd during learning of the flat rule set.

Individual Differences in Attention to Hierarchy

Badre et al. (2010) found that activation differences in prePMd early in the learning trial correlated with behavioral differences between the hierarchical versus flat learning conditions, implicating this region in successful search for hierarchical rules. Importantly, Frank and Badre (2011) further posited that, even within the hierarchical block individually, participants who devote more attention to hierarchical rule hypotheses over the course of learning will be those who have greater activation in prePMd. However, this prediction requires an estimate of the extent to which individuals attend to hierarchical rules. [And, illustrative of the enhanced sensitivity afforded by using the model estimates, the median split within the hierarchical block based on behavioral measures such as terminal or mean accuracy across the session, does not yield within a session difference in prePMd that reaches significance ($ts < 1.9$)].

Given its assumptions, the MoE model provides an estimate of the extent to which each individual is attending to hierarchical structure over the course of learning (even if particular motor response mappings are not yet learned) and so may provide greater sensitivity than terminal or mean accuracy to test this hypothesis. Indeed, the MoE model can differentiate attention to hierarchy versus other hypotheses, even when accuracy is equivalent (as demonstrated by MoE fits to

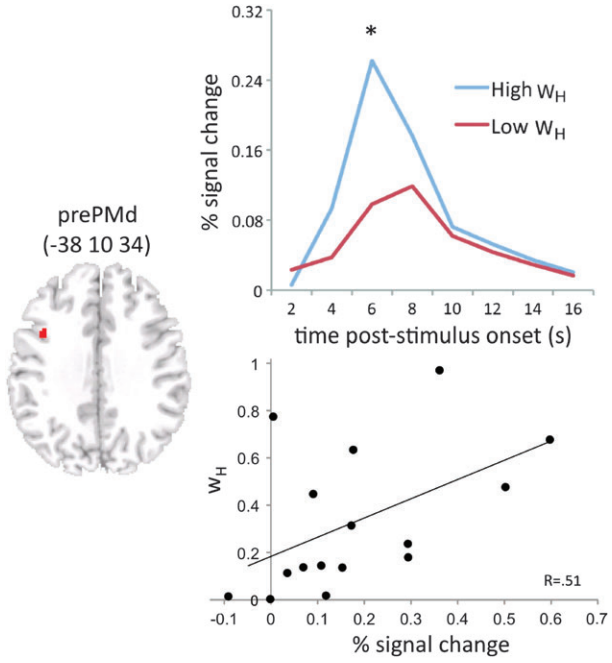

Figure 3. Individual differences analyses based on attentional weight to hierarchical (vs. flat) expert (w_H) shows differential activation in prePMd ROI (shown at left). (Top) The time course of BOLD response in prePMd ROI is plotted, showing significantly greater response in subjects with high attentional weight to hierarchical expert (w_H). (Bottom) Scatter plot of peak activation in prePMd (*x*-axis) against w_H (*y*-axis) with best fit trendline. Activation in prePMd and w_H are reliably correlated.

simulated data generated by cortico–striatal circuit models in which the ability to attend to hierarchical structure was manipulated; Frank and Badre 2011). When we fit the MoE to individual participants, those with greater attention to hierarchy (as indexed by a median split on w_H during the hierarchical learning session) exhibited greater prePMd activation during this session (t_{14} = 2.2, P < 0.05; Fig. 3). Note that this same attentional weight w_H differentiated between cortico–striatal network models depending on whether their architecture supported hierarchical structure or not (Frank and Badre 2011). Beyond this prePMd focus, no other ROI tested in frontal cortex (PMd, IFS, or FPC) or striatum showed a reliable median split effect.

We complemented the above median split analysis with a between-subjects correlation of w_H with prePMd activation at stimulus onset during the hierarchical block (Fig. 3). This analysis yielded a reliable positive correlation during hierarchical learning blocks (R = 0.51, P < 0.05) but not during flat blocks (R = 0.14, P = 0.4). Hence, consistent with the model, activation in prePMd only predicts an increase in attention to hierarchy when there is a hierarchical rule to discover.

RPE and Hierarchical Learning

Along with providing an estimate of attention to hierarchical and flat experts during learning, the MoE model also provides a trial-by-trial estimate of RPE (RPE defined as the reward

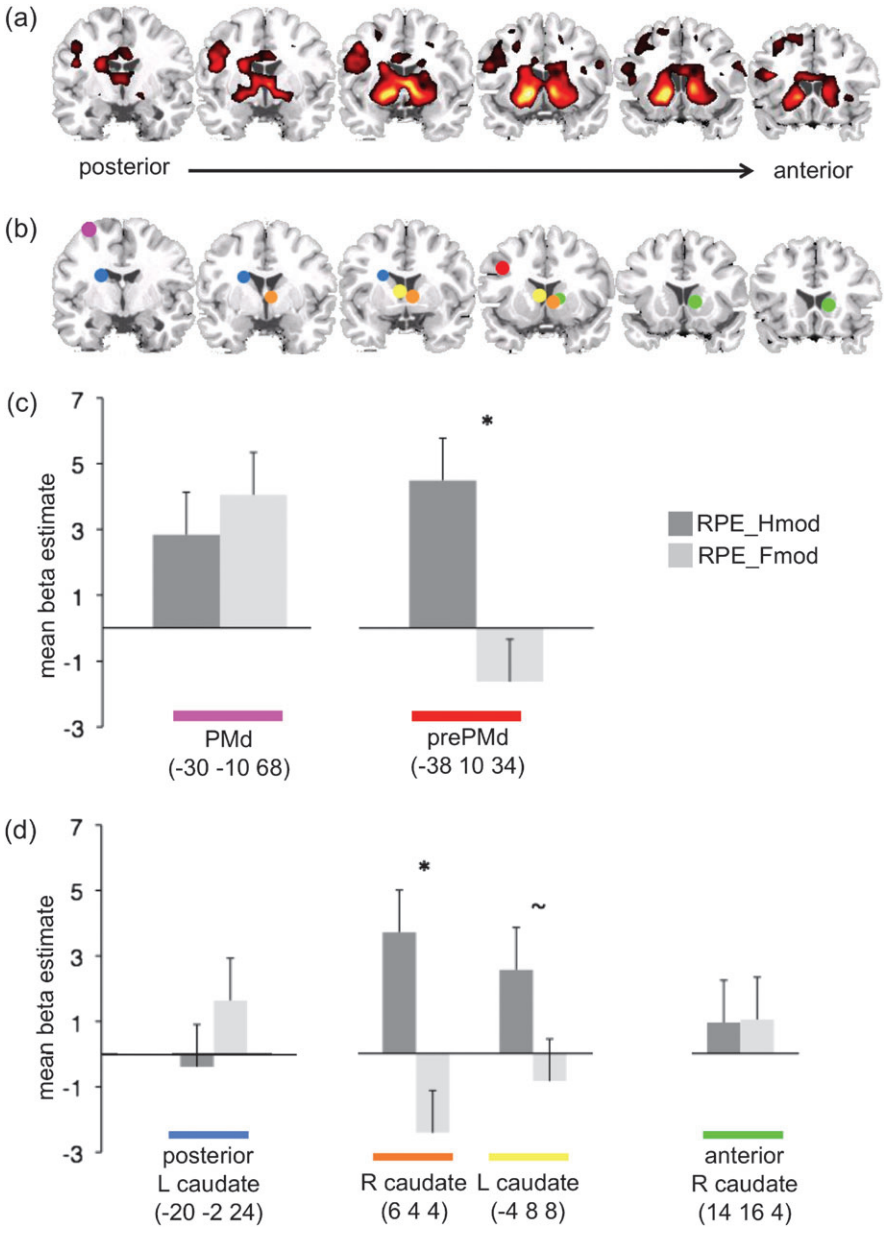

Figure 4. Model-based RPEs and frontostriatal activity. (*a*) BOLD response to brain areas that track RPE. Activations are observed in striatum and lateral frontal cortex. Note that for illustrative purposes, activations are plotted at an uncorrected threshold of $P < 0.001$. (*b*) Functionally defined ROI's for PMd, prePMd, and areas within caudate posterior to, at the same level as, and anterior to, prePMd. (*c*) Within cortical ROI's, prePMd tracks RPE specifically when model-derived attentional weight to hierarchical rule (RPE_{Hmod}), but not flat rule (RPE_{Fmod}), is high. PMd doesn't distinguish between hierarchical and flat rules in its sensitivity to RPE. (*d*) Within caudate, areas at the same anterior-to-posterior level as prePMd track RPE modulated by attention to hierarchical relative to flat rule. Caudate areas more posterior and more anterior to prePMd are not sensitive to this distinction.

delivered relative to the model output's predicted probability of reward for the chosen response; see Methods). We used this continuous estimate of RPE as a regressor during the feedback phase of the trial in a whole-brain voxel-wise analysis ($P < 0.05$, FWE corrected). This analysis yielded activation in bilateral caudate (x, y, z = −12 12 −2; 14 16 4; 6 4 4), left lateral frontal cortex (x, y, z = −40 24 20; −42 42 8; −50 34 12), and lateral occipital sulcus (−54 −54 −6) (Fig. 4). Considerable prior evidence from reinforcement learning has routinely located activation in medial and orbital frontal regions along with striatal regions in association with RPE during reinforcement learning tasks (McClure et al. 2003; O'Doherty et al. 2004; Gershman et al. 2009; Rutledge et al. 2010). Notably, when the whole-brain map was thresholded less conservatively ($P < 0.1$, FWE corrected), activation was also observed in orbitofrontal cortex (x, y, z = 0 42 −4) and an additional region of left caudate (−4 8 8).

Next, we sought to test the more specific model prediction that a subset of cortico–striatal circuitry would coactivate with RPE when participants were testing hierarchical structure. We thus created regressors that modulated RPE by the estimates of attention to the hierarchical rules (i.e., shape or orientation contingent on color; RPE_{Hmod}) to determine which areas of the brain track RPE's preferentially related to the specific hierarchical rules. In essence, RPE_{Hmod} tests the extent to which RPE-related activity is large when the model estimates that the participant is attending to the hierarchical rule to govern their response and not otherwise. Importantly, putative hierarchical rules could be tested and rewarded or punished across both hierarchical and flat blocks. So, even in the flat condition when participants start off with attention to hierarchy, we would expect RPE signals to activate this related circuitry. Hence, we tested the RPE_{Hmod} regressor on the fMRI data along with a corresponding RPE_{Fmod} regressor that tests the modulation of RPE as a function of attention to the flat rule (ie., fully conjunctive color–shape–orientation).

Notably, the whole-brain estimate of RPE_{Hmod} produced activation in left prePMd (x, y, z = −56 14 34), along with posterior parietal cortex, at an uncorrected threshold ($P < 0.001$). ROI analysis using an unbiased definitions from Badre and D'Esposito (2007) confirmed the effect of RPE_{Hmod} in prePMd, in addition to finding a reliable effect in IFS ($t_{15} = 2.4$, $P < 0.05$) and a trend in PMd ($t_{15} = 2.1$, $P = 0.06$). Next, we contrasted the difference in RPE_{Hmod} and RPE_{Fmod} in order to test which regions showed more sensitivity to RPE modulated by relatively greater attention to hierarchical versus flat rules. In frontal ROIs, IFS and prePMd showed a reliable difference ($ts_{15} > 2.7$, $Ps < 0.05$). Thus, in contrast to the more caudal PMd, prePMd was differentially sensitive to RPE related to hierarchical rather than flat rules (Fig. 4*c*). This difference between PMd and prePMd was supported by a reliable region by effect interaction ($F_{1,15} = 6.6$, $P < 0.05$).

In striatum, we sought to test whether the sensitivity to overall RPE is similarly modulated by attention to hierarchical rule in restricted regions. ROIs defined from the RPE contrast in bilateral caudate (x, y, z = 6 4 4; −4 8 8) revealed a reliable effect of RPE_{Hmod} in right caudate ($t_{15} = 2.7$, $P < 0.05$; Fig. 4*d*). Though, this simple effect could be biased to the extent that RPE_{Hmod} shares variance with RPE (see Methods). Thus, it is important to directly test the difference between RPE_{Hmod} and RPE_{Fmod}, which is unbiased by ROI selection. And, indeed, the focus in the right caudate ROI (x, y, z = 6 4 4) revealed a reliable difference between RPE_{Hmod} and RPE_{Fmod} ($t_{15} = 3$, $P < 0.01$). The homologous left caudate (x, y, z = −4 8 8) ROI showed a similar trending effect ($t_{15} = 2.0$, $P = 0.07$). It is notable that the effect of RPE_{Hmod} was selective to these bilateral subregions of the caudate that are at approximately the same point along the rostro–caudal axis as prePMd (Fig. 4*b*). To further test this selectivity, we tested a more anterior right caudate ROI (x, y, z = 14 16 4) and a more posterior left caudate ROI (x, y, z = −20 −2 24) that was at a similar caudal extent as PMd. Neither of these striatal ROIs showed an effect of RPE_{Hmod} over RPE_{Fmod} ($ts < 1.3$, $Ps > 0.2$). Indeed, if anything, the posterior caudate ROI showed a quantitative trend in the opposite direction, with RPE_{Fmod} trending greater than RPE_{Hmod}, a pattern more similar to the nearby PMd than prePMd (Fig. 4*c*,*d*). Thus, to summarize, the specific subregions of caudate that were sensitive to the difference in RPE for hierarchical versus flat rules were those closest in the rostro–caudal dimension to prePMd, the frontal region that was also differentially sensitive to these rules. We consider the implications of this finding further in the Discussion as it relates to the neural model. Importantly, these distinctions in sensitivity to RPE_{Hmod} and RPE_{Fmod} among rostro-to-caudal striatal ROIs were supported by region by effect interactions (right caudate × left posterior caudate: $F_{1,15} = 10.6$, $P < 0.01$; right caudate × right anterior caudate: $F_{1,15} = 11.6$, $P < 0.005$; left caudate × left posterior caudate: $F_{1,15} = 6.4$, $P < 0.05$; left caudate × right anterior caudate: $F_{1,15} = 4.5$, $P = 0.05$.

Learning Dynamics in Striatum and prePMd

The selectivity of RPE_{Hmod} is initially consistent with a key prediction of the cortico–striatal circuit model. In the model, the representation of contextual information in prePMd is not adaptive during flat blocks in which no hierarchical structure is present (because prePMd representations constrains attention to one of the dimensions in PMd). As such, prePMd layer activity comes to be associated with negative value and generates a negative RPE, which in turn allows the BG to learn not to gate (NoGo) contextual representations into this layer. Hence, this model predicts that RPE activity in the striatum when one is attending to the hierarchical rule (i.e., RPE_{Hmod}) during the flat block should be the basis of the decline in activation in prePMd during these blocks reported by Badre et al. (2010). Consistent with this prediction, individual differences in the effect of RPE_{Hmod} in the very same striatal ROI tested above (left caudate x, y, z = −4 8 8) during the flat block correlated with individual differences in the decline in activation in left prePMd (activation during the first third minus the last two-thirds of the learning trial) as estimated by Badre et al. ($R = 0.55$, $P < 0.05$; Fig. 5). A similar marginal correlation was evident with right caudate (x, y, z = 6 4 4; $R = 0.49$, $P = 0.057$; Fig. 5).

Importantly, this effect was highly specific to RPE_{Hmod} during the flat block only. No such correlation with the prePMd activation decline was evident using 1) overall RPE during the flat block unmodified by attention to hierarchy ($Ps > 0.75$), 2) attentional weight to hierarchy independent of RPE ($Ps > 0.5$), 3) RPE_{Hmod} from the hierarchical block ($Ps > 0.09$) or the modulation of RPE by attention to the flat rule during the flat block ($Ps > 0.85$). Thus, this correlation was selective to the modulation of RPE by attention to the hierarchical rule during the flat block, as predicted by the neural model—in which RPE's punish striatal gating of prePMd—and estimated by the MoE model.

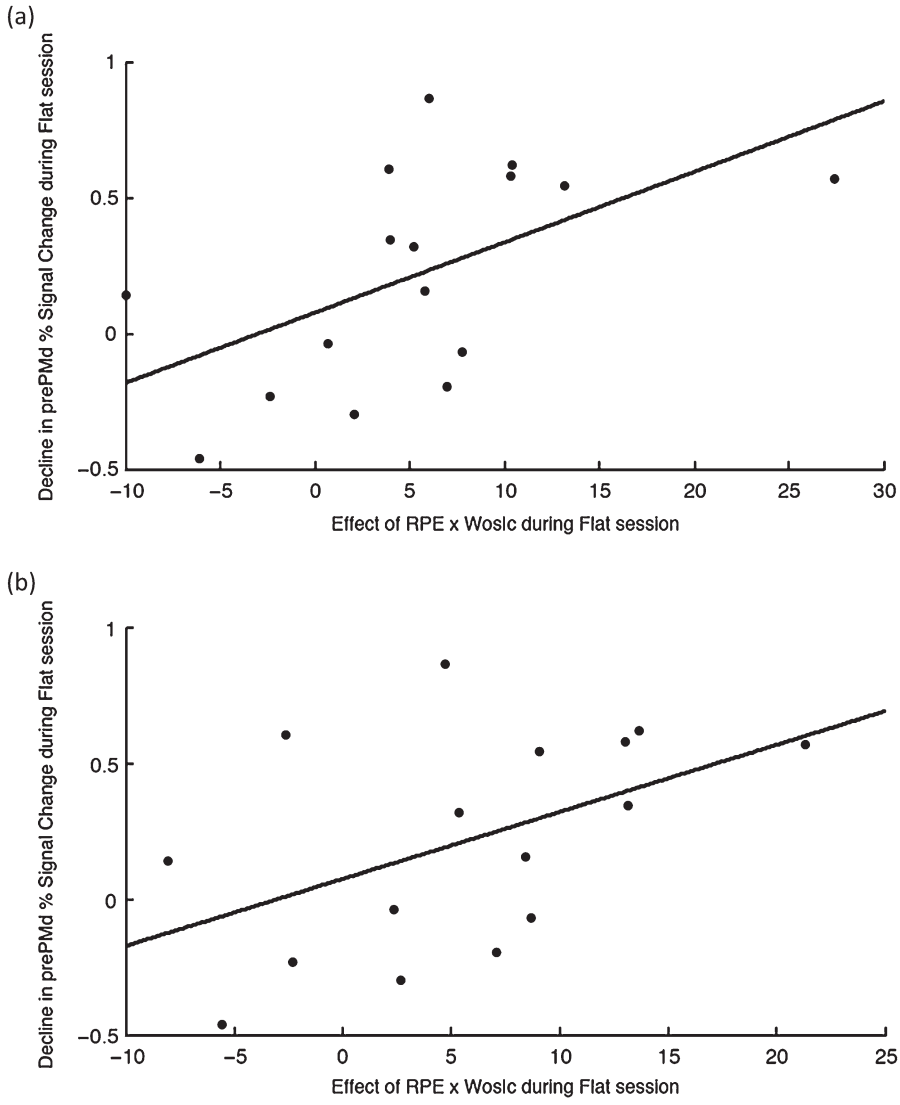

Figure 5. Between subjects, the BOLD response to RPE modulated by attention to the hierarchical rule (RPE_{Hmod} = RPE × $w_{OS|C}$) is predictive of the decline of prePMd activity in the flat condition in (*a*) left and marginally in (*b*) right caudate.

Discussion

The putative hierarchical organization of the frontal lobes has been a focus of increasing investigation in recent years. Much of this work has assumed that hierarchical interactions within frontal cortex arise from cortico-cortical connections within lateral frontal cortex itself. In our companion paper, we provide a novel model of hierarchical control in cortico–striatal circuits. This neural circuit model proposes that maintenance of contextual representations in rostral regions of frontal cortex influences striato-frontal gating policy in more caudal frontal cortex. Reinforcement learning operates at each level, such that the system adaptively learns to gate, represent and maintain higher order contextual information in rostral regions (e.g., prePMd), which serve to conditionalize gating (in this case by attentional selection) in more caudal regions (e.g., PMd) and ultimately influence response selection in motor cortex. An algorithmic Bayesian MoE model captures the key computations of this neural model and provides trial-by-trial estimates of the latent hypothesis state of each subject (Frank and Badre 2011). We used these estimates to reanalyze fMRI data from a hierarchical reinforcement learning tasks reported in Badre et al. (2010). Results from this model-based fMRI analysis provide evidence in initial support of the models.

First, we demonstrate that individual differences in prePMd activation during hierarchical learning blocks is positively correlated with the MoE estimates of attention to hierarchical structure. Previously, Badre et al. (2010) had reported that individual differences in activation early in learning in prePMd across both hierarchical and flat blocks was predictive of behavioral differences between the hierarchical and flat learning blocks. This provided initial evidence that prePMd may be critical in search and subsequent discovery of the hierarchical structure, which is presumed to be the source of the behavioral differences between the learning blocks. However, without a model of the latent states of the learner that contribute to a response, it was not possible to relate activation in prePMd to search and acquisition of hierarchical structure more directly.

The MoE model developed in the companion paper does provide estimates of these latent states, specifically indexed by the attentional weights. Thus, the attentional weight to the hierarchical expert (u_H) is an estimate of the probability that

a hierarchical is contributing to response selection on a given trial. When this weight increases, it indicates that participants have learned to rely on hierarchical structure and that this provides a better account of their sequence of choices, even compared with other models that could lead to good performance (i.e., the full conjunctive flat expert). Hence, the correlation between prePMd activation during hierarchical blocks and the attentional weight to hierarchy provides evidence that prePMd activation relates to discovery of hierarchical structure specifically, as opposed to other sources of behavioral variation.

It is, however, important to emphasize that prePMd activity is proposed to be necessary but not sufficient for attention to hierarchy. Indeed, when the MoE is fit to simulated behavior generated by the cortico–striatal circuit model, the estimated attentional weights to hierarchy are related to the development of an abstract striatal gating policy based on prePMd representations (Frank and Badre 2011), not to raw prePMd activity, per se. Thus, increases in activation in prePMd are not proposed to mirror changes in attention to hierarchy in a trial-by-trial manner, and the between-subject correlation should not be interpreted as indicating such to be the case. Indeed, for most subjects, w_H increases over the course of a block (see companion paper; Fig. 7) as the participant relies more on the hierarchical rule to make a response. In contrast, the brain activation in prePMd stays stable over the course of a block in which a hierarchical rule gets rewarded, declining during blocks when it is not rewarded (Badre et al. 2010). According to the neural model, this is because activation in prePMd reflects maintenance of contextual information potentially relevant to hierarchical rule choices. Its maintenance and subsequent association with reward and punishment to reinforce striatal gating are central to the search for hierarchical structure. Thus, participants who engage prePMd are more likely to discover the hierarchical rule and so their attention to hierarchy should be higher overall than those who do not engage prePMd (and hence cannot test for hierarchical structure). This would not necessarily be the prediction for flat blocks where search-related activation in prePMd cannot yield increases in attention to hierarchy.

Second, model-based regressors provided evidence that similar—putatively dopaminergic—RPE signals modulate learning of hierarchical structure in restricted prePMd and striatal regions in a manner analogous to that observed in more basic stimulus–response reinforcement tasks (McClure et al. 2003; O'Doherty et al. 2004; Pessiglione et al. 2006; Schonberg et al. 2010; Voon et al. 2010). Notably, the locus of these restricted striatal regions was at the same rostro–caudal extent subcortically as the prePMd on the lateral surface. This is an intriguing observation and is consistent with evidence from monkey tracing studies and probabilistic connectivity in humans showing a rostro–caudal organization of inputs from premotor/prefrontal cortex to corresponding regions of striatum (Inase et al. 1999; Lehericy, Ducros, Krainik et al. 2004; Lehericy, Ducros, Van de Moortele et al. 2004; Postuma and Dagher 2006; Draganski et al. 2008) and with the general principle that inputs to striatum are strongest from cortical areas of closest proximity (Kemp and Powell 1970).

Thus, these results provide novel evidence that a specific cortico–striatal circuit is involved in learning second-order hierarchical structure. The existence of such a local circuit devoted to one level of hierarchy is consistent with the broader organizational hypothesis, set forth by the neural circuit model, that hierarchical control emerges from nesting of these individual cortico–striatal controllers. However, to fully test the neural circuit hypothesis, future work will need to extend the present finding by locating evidence of additional controllers corresponding to other levels of the hierarchy and demonstrate that they interact in the hierarchically nested manner proposed by the model. The discovery of this individual cortico–striatal circuit that is sensitive to reinforcement learning at one level of policy abstraction is of particular interest because it also provides support for a second key assumption of the model. Specifically, that learning at different levels of the hierarchy can nevertheless arise from a single RPE signal, computed based on the expectation of the entire agent, comprising mixed hierarchical and flat experts. The specificity of this learning signal comes from its modulation by the attentional weight or the degree to which a rule at a particular level contributed to a response. The rationale for this dynamic is that in the neural model, there is a single "critic" which evaluates the reward value of the current state, comprising all input and frontal representations. This dopaminergic prediction error is then communicated to all striatal parts of the network, but its effect on neural activity depends on the strength of Go and NoGo unit activations in each "stripe" (subcircuit), which in turn reflect the action values associated with gating of the corresponding frontal hypothesis. Thus, if the prePMD represents hierarchical structure, the corresponding striatal area is predicted to reflect RPE activity modulated by the degree to which the prePMD currently represents such structure. Hence, the selectivity of the prePMd-striatal circuit is only observed when testing the interaction between RPE and attention to the hierarchical rule ($w_{OS|C}$).

Finally, the hierarchical RPE signal in this same striatal subregion was correlated with the decline in prePMd activity when no hierarchical structure existed (i.e., in the flat condition). This observation is consistent with the mechanistic explanation for this decline suggested by the neural model, in which this RPE signal reinforces or punishes the gating of prePMd hierarchical representations. In particular, the neural model predicts that during the flat block (when no hierarchical rule structure is present to be learned), choosing responses based on hypothetical hierarchical rules will lead to punishment, which will selectively punish the prePMd gating circuit. Thus, over time, prePMd activation will decline. The correlation of the observed decline in prePMd with the negative RPE related to hierarchy (RPE_{Hmod}) during the flat block provides initial evidence consistent with this hypothesis. It is worth noting the specificity of this finding. Tests on a number of controls indicated that this effect was specific to RPE_{Hmod} and not a general effect of RPE, a general effect of RPE_{Hmod} outside of the context of the flat block or a general effect of attention to the hierarchical rule ($w_{OS|C}$) independent of RPE. Moreover, the effect was specifically located in the ROI in caudate that was also selective for RPE_{Hmod}, again implicating this circuit in the learning of second-order policy.

Future work may continue to validate the attentional weight estimates from the MoE model. In the present study, the individual differences analysis highlighted prePMd activation in association with a higher likelihood of discovering second-order policy. This region has been highlighted in separate studies in relation to cognitive control at a second order of policy abstraction (Badre and D'Esposito 2007; Badre et al.

2009), and the exact same ROI from one of these prior experiments was used in this study to ensure a precise overlap. As such, this finding is highly consistent with past work. However, future studies may seek to further validate the attentional weight estimates from the MoE. For example, multivoxel pattern classification approaches may provide a key means of probing the content of representations being attended or maintained. Comparison of this type of index with the attentional weight estimates from the MoE will provide further validation of the model's ability to predict latent states during cognitive control and hierarchical reinforcement learning tasks.

Funding

National Institute of Neurological Disease and Stroke (R01 NS065046 to D.B.); National Institute of Mental Health (R01 MH080066 to M.J.F.).

Notes

The authors wish to thank A. S. Kayser and M. D'Esposito who were coauthors along with D.B. on the publication of the original hierarchical learning task reported in Badre et al. (2010) and who consented to our reanalysis of the fMRI data from that experiment (supported by National Institute of Health awards MH63901 and NS40813) for the present study. *Conflict of Interest*: None declared.

References

Badre D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn Sci. 12:193–200.

Badre D, D'Esposito M. 2007. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. J Cogn Neurosci. 19:2082–2099.

Badre D, D'Esposito M. 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? Nat Rev Neurosci. 10:659–669.

Badre D, Hoffman J, Cooney JW, D'Esposito M. 2009. Hierarchical cognitive control deficits following damage to the human frontal lobe. Nat Neurosci. 12:515–522.

Badre D, Kayser AS, D'Esposito M. 2010. Frontal cortex and the discovery of abstract action rules. Neuron. 66:315–326.

Badre D, Wagner AD. 2004. Selection, integration, and conflict monitoring; assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron. 41:473–487.

Brown JW, Bullock D, Grossberg S. 2004. How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. Neural Netw. 17:471–510.

Bunge SA. 2004. How we use rules to select actions: a review of evidence from cognitive neuroscience. Cogn Affect Behav Neurosci. 4:564–579.

Cohen MX, Frank MJ. 2009. Neurocomputational models of basal ganglia function in learning, memory and choice. Behav Brain Res. 199:141–156.

Cools R, Ivry RB, D'Esposito M. 2006. The human striatum is necessary for responding to changes in stimulus relevance. J Cogn Neurosci. 18:1973–1983.

Dale AM. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8:109–114.

Draganski B, Kherif F, Kloppel S, Cook PA, Alexander DC, Parker GJ, Deichmann R, Ashburner J, Frackowiak RS. 2008. Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. J Neurosci. 28:7143–7152.

Frank MJ. 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci. 17:51–72.

Frank MJ, Badre D. 2011. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits I: computational analysis 22:509–526.

Frank MJ, Loughry B, O'Reilly RC. 2001. Interactions between frontal cortex and basal ganglia in working memory: a computational model. Cogn Affect Behav Neurosci. 1:137–160.

Frank MJ, O'Reilly RC. 2006. A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav Neurosci. 120:497–517.

Fuster JM. 2001. The prefrontal cortex—an update: time is of the essence. Neuron. 30:319–333.

Gershman SJ, Pesaran B, Daw ND. 2009. Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J Neurosci. 29:13524–13531.

Gruber AJ, Dayan P, Gutkin BS, Solla SA. 2006. Dopamine modulation in the basal ganglia locks the gate to working memory. J Comput Neurosci. 20:153–166.

Inase M, Tokuno H, Nambu A, Akazawa T, Takada M. 1999. Corticostriatal and corticosubthalamic input zones from the presupplementary motor area in the macaque monkey: comparison with the input zones from the supplementary motor area. Brain Res. 833:191–201.

Kemp JM, Powell TP. 1970. The cortico-striate projection in the monkey. Brain. 93:525–546.

Koechlin E, Jubault T. 2006. Broca's area and the hierarchical organization of human behavior. Neuron. 50:963–974.

Koechlin E, Ody C, Kouneiher F. 2003. The architecture of cognitive control in the human prefrontal cortex. Science. 302:1181–1185.

Koechlin E, Summerfield C. 2007. An information theoretical approach to prefrontal executive function. Trends Cogn Sci. 11:229–235.

Lehericy S, Ducros M, Krainik A, Francois C, Van de Moortele PF, Ugurbil K, Kim DS. 2004. 3-D diffusion tensor axonal tracking shows distinct SMA and pre-SMA projections to the human striatum. Cereb Cortex. 14:1302–1309.

Lehericy S, Ducros M, Van de Moortele PF, Francois C, Thivard L, Poupon C, Swindale N, Ugurbil K, Kim DS. 2004. Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. Ann Neurol. 55:522–529.

McClure SM, Berns GS, Montague PR. 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron. 38:339–346.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 24:167–202.

Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 16:1936–1947.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. Neuron. 38:329–337.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 304:452–454.

O'Reilly RC, Frank MJ. 2006. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. 18:283–328.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 442:1042–1045.

Postuma RB, Dagher A. 2006. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. Cereb Cortex. 16:1508–1521.

Rutledge RB, Dean M, Caplin A, Glimcher PW. 2010. Testing the reward prediction error hypothesis with an axiomatic model. J Neurosci. 30:13525–13536.

Schonberg T, O'Doherty JP, Joel D, Inzelberg R, Segev Y, Daw ND. 2010. Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. Neuroimage. 49:772–781.

Shen W, Flajolet M, Greengard P, Surmeier DJ. 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. Science. 321:848–851.

Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, Dolan RJ, Hallett M. 2010. Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. Neuron. 65:135–142.