Adaptive learning via BG-thalamo-cortical circuitry

Qin He[1,2], Daniel N. Scott[1,2], Michael J. Frank[1,2,3]
Cristian B. Calderon[*4] and Matthew R. Nassar[*1,2]

[1]Department of Neuroscience, Brown University, Providence, United States
[2]Carney Institute for Brain Science, Brown University, Providence, United States
[3]Department of Cognitive and Psychological Sciences, Brown University, Providence, United States

[4]Centro Nacional de Inteligencia Artificial (CENIA) Macul, Chile

March 2025

[*]Co-senior authors.

1 Abstract

People adjust their use of feedback over time through a process referred to as adaptive learning. We have recently proposed that the underlying mechanisms of adaptive learning are rooted in how the brain organizes time into similarly credited units, which we refer to as latent states. Here we develop a BG-thalamo-cortical circuit model of this process and show that it captures both the commonalities and heterogeneity in human adaptive learning behavior. Our model learns incrementally through synaptic plasticity in PFC-BG connections, but upon observing discordant information, produces thalamocortical reset signals that alter PFC connectivity, driving attractor state transitions that facilitate rapid updating of behavioral policy. We demonstrate that this mechanism can give rise to optimized learning dynamics in the context of either changepoints or reversals, and that under reasonable biological assumptions the model is able to generalize efficiently across these conditions, adjusting behavior in a context-appropriate manner. Taken together, our results provide a biologically plausible mechanistic model for adaptive learning that explains existing behavioral data and makes testable predictions about the computational roles of different brain regions in complex learning behaviors.

2 Introduction

People and animals exploit the statistics of their environments to optimize behavior through a process that we refer to as adaptive learning (Behrens et al., 2007; Donahue & Lee, 2015). For example, in environments that demand we occasionally learn completely new state-action associations humans and animals prioritize information collected after "changepoints" in the environment to quickly learn policies corresponding to the new contingency (Li et al., 2019; Nassar et al., 2010). In environments where optimal policy alternates occasionally between two or more alternatives, people and animals can quickly recognize the changes and apply previously learned policies (which we call "reversals" here) (Collins & Koechlin, 2012; Wilson et al., 2014). Bayesian models that reflect knowledge about the structure of the environment have been used to demonstrate the utility of adaptive learning behaviors observed in specific environments (i.e. changepoints and reversals) (Heilbron & Meyniel, 2019; Piray & Daw, 2021; Wilson et al., 2010). However, how adaptive learning is implemented in the brain, and how a fixed neural circuit might implement different forms of adaptive learning, depending on the underlying structure of the environment, remain open and important questions.

One mechanism with potential to achieve adaptive learning across different temporal structures is latent state inference (Collins & Frank, 2013; Gershman & Niv, 2010; Yu et al., 2021). Adaptive learning across a wide range of statistical environments can be cast as a predictive inference problem, in which the solution requires inferring and dynamically updating inferences about the "latent state" – an unobservable variable that gives rise to observable experiences (e.g., task feedback). In this view, partitioning state-action mappings by recruiting distinct neuronal populations for novel states would avoid interference between learning episodes. Moreover, revisiting previously experienced states and reusing established mappings enables faster inference compared to learning from scratch when new experiences closely resemble past ones. While optimal adaptive learning is computationally intractable for complex environments, it can be reasonably well approximated through simplifying assumptions that eliminate the need for attribution of all possible experiences to all possible latent states of the world (Collins & Koechlin, 2012; Fearnhead & Liu, 2007; Lloyd & Leslie, 2013; Nassar et al., 2010; Wilson et al., 2013; Yu & Dayan, 2005).

Neural networks incorporating latent state inference have shed light on a number of behavioral phenomena, even without completely explaining adaptive learning. For example, biologically motivated neural networks have demonstrated that hierarchies of cortico-striatal processing can be used to recruit latent state representations, which in turn constrain downstream action selection (Collins & Frank, 2013; Frank & Badre, 2012). Such models propose on how the brain might transfer learning across stimuli and reuse previously learned policies, and explain human behavior across a range of stimulus-response paradigms (Collins & Frank, 2013; Frank & Badre, 2012). In principle, a hierarchy through which prefrontal representations constrain lower-level action selection could also give rise to the full range of adaptive learning phenomena. While this has yet to be tested in a fully elaborated biological network, one recent study used a two-layer feed-forward network to show that representational drift in input layers occurring in response to unexpected feedback at latent state transitions can reproduce normative and human-like adaptive learning behaviors in such tasks (Razmi & Nassar, 2022). This model provided a proof-of-principle, but it abstracted over details that would be critical for implementing the core computations, building them into a real-time biological network, and affording generality across environments

(the original model used an "oracle" to dictate how drift should occur in each environment). Thus, there is currently no model that can simultaneously: (i) account for the wide range of adaptive learning behaviors observed in humans, (ii) implement it in a real-time biologically supported circuit, (iii) and generalize across structures (ie. changepoint/reversals).

To solve these problems, and produce a biologically realistic neural network which accords with human data across tasks, we draw inspiration from empirical observations from systems neuroscience. First, task-relevant representations are often coded in neural population attractors (Ebitz & Hayden, 2021), which can emerge in recurrent neural networks, and could potentially be used to store latent state information that persist across time (Brunel, 2003; Hopfield, 1982; Litwin-Kumar & Doiron, 2014; Maes et al., 2020; Recanatesi et al., 2022). Second, context information in hierarchically structured tasks is known to be available within cortico-basal ganglia loops, where corticothalamic synaptic plasticity facilitates learning that links concrete or abstract states to actions that are rewarded within them (Chatham et al., 2014; Collins & Frank, 2013; Frank & Badre, 2012). Third, thalamocortical projections enable switches between neural representation in the cortex, as shown by experimental studies (Lam et al., 2024; Remington et al., 2018; Wang et al., 2018) and supported by theoretical work (Calderon et al., 2022; Recanatesi et al., 2022). Finally, the timing of neural and behavioral state transitions in rodents can be accounted for through dynamic changes in random connectivity that drive activity patterns in recurrent neural networks (Mazzucato, 2022; Recanatesi et al., 2022). These findings collectively imply that attractors storing latent state information in prefrontal regions might be linked to actions through learned associations with the striatum and flexibly recruited by the thalamus altering activity dynamics and facilitating human-like adaptive learning.

Based on the biology just reviewed, we developed a neural model of adaptive learning. Our model can capture human behavior across various tasks and reproduce the emergent behaviors described by Bayesian normative models. It combines active memory representations in prefrontal cortex (Euston et al., 2012; Funahashi & Kubota, 1994; Preston & Eichenbaum, 2013; Rugg et al., 1996; Stokes, 2015; Wall & Messier, 2001) with synaptic plasticity in cortico-striatal connections (Calabresi et al., 2007; Charpier & Deniau, 1997; Fino et al., 2005; Kreitzer & Malenka, 2008; Reynolds & Wickens, 2002). Prefrontal attractors store state representations which become associated with policies through corticostriatal Hebbian learning. Mismatches between expectations and observations trigger thalamocortical transients that drive prefrontal attractor switches (as seen empirically; Lam et al., 2024; Rikhye, Gilra, and Halassa, 2018; Scott et al., 2024). This model captures a wide range of human behaviors, including both trends and individual differences, and can account for normative aspects of adaptive learning behavior across different conditions and temporal environments. Our results provide insight into how learning and inference, implemented through plasticity and activity dynamics, can be combined in biologically plausible circuits to afford adaptive learning across different temporal contexts.

3 Results

We tested our neural model on changepoint and reversal tasks, and we fit it to human data collected previously (McGuire et al., 2014). We found that the canonical cortico-basal-ganglia-thalamo-cortical architecture we implemented was capable of capturing this participant data, and furthermore, that model thresholds for differentiating states and the structured connectivity controlling state transitions were key parameters impacting behavior. More specifically, we found that the model required an adaptive threshold for state separation in order to effectively cope with different levels of observation noise, and that this adaptive threshold explained human adjustment to noise as well. Furthermore, we discovered that the model required connectivity between motor and prefrontal cortex, beyond that we initially hypothesized, in order to adapt to environments exhibiting mixes of changepoints and reversals. Below, we outline the specific tasks we used, and demonstrate both the model fits and the impacts of different parameters on model function, then we conclude by discussing further connections between our model and other work, along with shortcomings and hypotheses future work should address.

3.1 Simulated tasks

We tested our model in a predictive inference task environment that has played a primary role in uncovering the dynamics of adaptive learning in humans (Bruckner et al., 2025; Nassar et al., 2010, 2016). The task required participants to predict the next in a series of trial outcomes, making behavior minimally complex and maximally informative (McGuire et al., 2014; Nassar & Gold, 2013). In a popular gamification of the task, a helicopter (occluded by clouds and thus not visible to participants) drops a

bag of coins on each trial. Participants are required to catch these bags by positioning a bucket, which serves as a prediction for where the next bag will fall (Figure 1A). After each bag is dropped, participants reposition the bucket according to their updated beliefs about the position of the helicopter, providing an experimental readout of how the most recent bag location affected underlying beliefs about the helicopter location.

In the "changepoint" version of the task, bag locations are sampled from a Gaussian distribution centered on a helicopter location that changes occasionally to an unpredictable new location (Nassar, Bruckner, & Frank, 2019; Nassar et al., 2010). The helicopter location serves as a latent state which must be inferred, and abrupt changes in helicopter location, at so-called changepoints, demand rapid adjustments of behavior. These changepoints render past bag locations irrelevant to the problem of predicting future ones, so that the association of actions with latent states requires that these latent states be partitioned at changepoints. Figure 1B shows an example of the behavior of a human participant performing the helicopter task. Note that the example participant shows the hallmark of adaptive learning in the presence of changepoints, quickly updating predictions after changepoints in the helicopter location but adjusting predictions during periods of stability much more incrementally. Beyond addressing changepoints, predictive inference tasks can be used to reveal adaptive learning behaviors across a much broader range of statistical environments, including reversals and tasks with uninformative latent state noise, thereby providing us with a testbed for generalizing latent state inference mechanisms over environments (D'Acremont & Bossaerts, 2016; Nassar, Bruckner, & Frank, 2019; Piray & Daw, 2024). As we discuss further below, we therefore also run our model on so-called "reversal" tasks, showing that it can also flexibly reuse information in environments containing recurrent contingencies.

3.2 Learning as entropy-induced-attractor-switch model

Our goal was to build a biologically constrained model capable of achieving human-like adaptive learning in predictive inference tasks. Inspired by empirical observations involving the basal ganglia-thalamo-cortical circuitry we propose the **L**earning as **E**ntropy-**I**nduced **A**ttractor switch (LEIA) model (Figure 2). The model has a PFC, a BG, and a motor output, with the PFC representing latent states of the environment that the BG transforms into an action selection in the motor layer. Latent states in the PFC are encoded as attractors, which the thalamus controls switches between. LEIA performs our predictive inference tasks sequentially over trials, by first making decisions based on internal activity then adjusting internal weights and activity based on feedback. We describe the model elements in more detail using an idealized noise-free changepoint task; later, we discuss realistic noisy tasks, and the reader is referred to the Methods section 5 for mathematical and implementation details.

To be more specific, the model implements a detailed set of biological circuits, making use of several specific hypotheses about the representations and dynamics in these elements. In the LEIA model, the PFC encodes latent state representations as attractor states within an RNN, which projects to the basal ganglia (BG; cyan area in Figure 2) and subsequently to motor outputs, thereby performing learning and action selection (Frank & Claus, 2006; Karlsson et al., 2012; Nassar, McGuire, et al., 2019). These attractor states are characterized by stable activations of a specific subset of excitatory neurons (e.g., blue nodes in the yellow shaded area of Figure 2 A). Attractor states represent specific contexts (i.e., moments when the helicopter at a certain position, although, in principle, they can encode any type of latent state). The network then proposes an action associated with each active RNN state to the striatal "Go" cells (Frank & Claus, 2006; O'Reilly & Frank, 2006), which is subsequently relayed through the Globus Pallidus (GP) and motor thalamus. The thalamus projects to a motor output that generates actions based on BG input. Following action selection, the motor cortex receives supervised feedback from the environment—encoding the observed bag location in the helicopter task—and projects back to the motor thalamus and striatal "Go" cells. This feedback, combined with cortico-striatal plasticity, enables PFC-striatal synapses to learn associations between specific latent states and the observed bag locations.

The core activity and learning dynamics LEIA can be clearly visualized in an idealized noise-free changepoint task that consists of three contexts (or hidden states) with two changepoints (Figure 2C; see below for realistic noisy tasks). To demonstrate this, we analyzed LEIA dynamics in four crucial task transitions, focusing on shifts between stable periods and changepoints (color coded, respectively, by yellow, cyan, purple, and green colors in Figure 2C). The LEIA model quickly adapts to the environment statistics (purple line and grey circles in Figure 2C); in particular, the model performs strong updates after changepoints and quickly reduces update magnitudes while responses to consistent latent states are stable, as one would expect from human adaptive learning (Nassar, Bruckner, & Frank, 2019; Nassar,

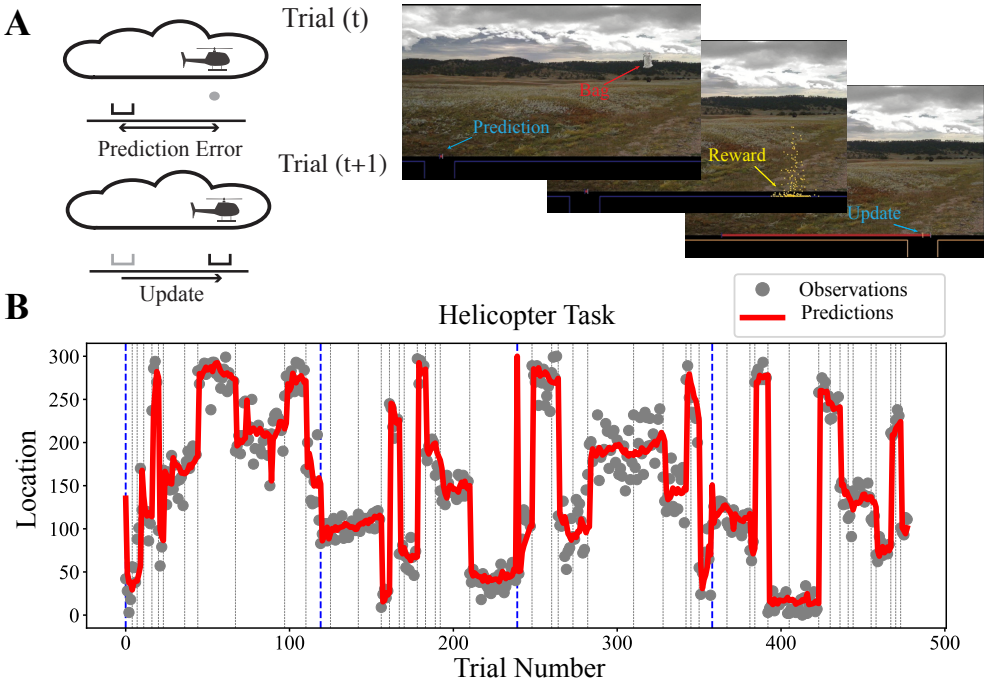

Figure 1: An illustrative example of the helicopter task. **A**, Cartoon illustration on the left and a screenshot on the right of helicopter task. In the task, human participants place a bucket at the predicted location of a hidden helicopter (prediction) in order to catch a bag of coins that would be dropped from it (outcome). After observing each outcome, participants updated their predictions for the subsequent trial based on their prediction error (outcome(t) – prediction(t)). Participants were incentivized to make good predictions since they earned rewards proportional to the number of coins caught. **B**, Example predictions from an example human participant (red) in response to bag locations (gray) are plotted across trials (abscissa). Bag locations were sampled from a Gaussian distribution, with a mean that was reset at changepoints (gray vertical lines) and a standard deviation ('noise') that was experimentally manipulated across blocks of 120 trials (demarcated by vertical blue dashed lines). Note that predictions are updated dramatically after changepoints, but much more slowly during periods when the distribution of bags is relatively stable.

McGuire, et al., 2019; Nassar et al., 2010).

The way the LEIA model accomplishes the human-like learning just described involves two mechanisms for learning: weight-based plasticity and activation based dynamics (Frank & Claus, 2006; Russin et al., 2024; Wang et al., 2018). First, after receiving feedback (motor dynamics in Figure 2F), cortico-striatal plasticity (Figure 2E, i.e., dashed red lines in Figure 2A) allows LEIA to progressively adjust state-action associations based on the magnitude of the prediction error (PE), which allows the network to learn incrementally from PEs. Second, a more pronounced change in learning occurs by altering activity dynamics within the RNN (Figure 2F2 and F4 for changepoints colored in cyan and green, respectively). Entropy over thalamic population firing rates, which reflects prediction errors, controls these state transitions. When entropy is higher than a threshold in the model (Figure 2D2 and D4), asymmetric attractor connectivity (Figure 2B) switches PFC states, implementing one possible mechanism for dynamics documented in recent experimental work (Calderon et al., 2022; Halassa & Sherman, 2019; Rikhye, Gilra, & Halassa, 2018; Rikhye, Wimmer, & Halassa, 2018). Therefore, weight and activation based dynamics allows LEIA to either integrate new observations into an existing state, and learn incrementally via Hebbian learning (see methods 5.1.3), or, to partition learning into a new PFC latent state and its associated BG connections. Learning can thus proceed incrementally "in-weights" or rapidly via PFC activation dynamics that recruit new PFC populations and cortico-striatal synapses (Frank & Claus, 2006; Russin et al., 2024; Wang et al., 2018).

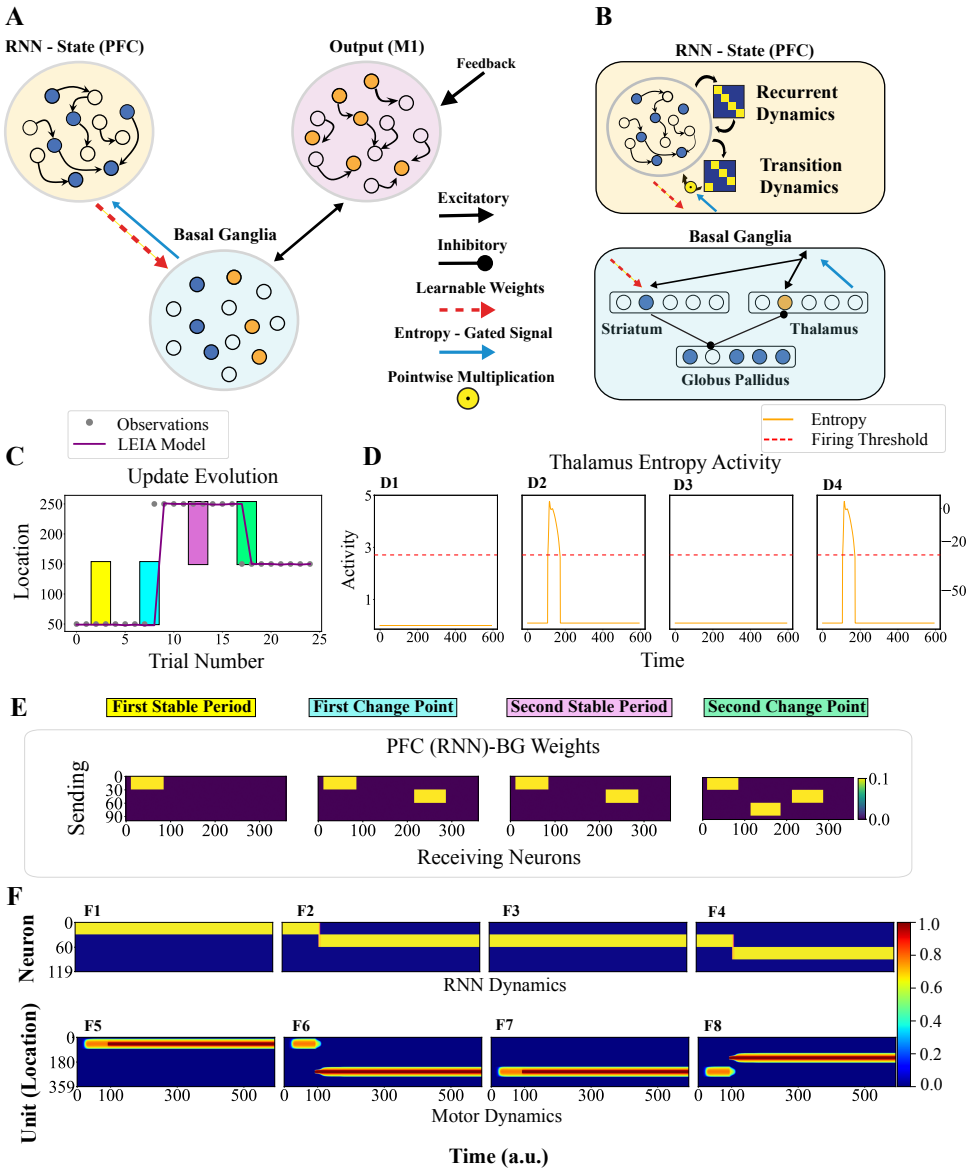

Figure 2: LEIA model architecture and dynamics in noise-free environment. **A**, LEIA model consists of three modules. The RNN-state module (yellow shaded area) is composed of a Hopfield-like recurrent neural network with a recurrent connection (inhibitory neurons are encoded in off-diagonal blocks) and entropy-gated asymmetric connectivity. The RNN - state module projects toward the basal ganglia (BG) module (cyan-shaded area). Each attractor state learns (via Hebbian learning; dashed red line) to couple with the appropriate action node (i.e., Go Cells) within the striatum. Through the double inhibition from the striatum to the globus pallidus (GP) and then to the thalamus, the network proposes an action to the motor output module. The motor output (purple-shaded area) displays bump attractor dynamics over potential actions and represents the action being executed by the model. Finally, the cortico-basal ganglia loop is completed by a projection from the thalamus to the RNN, which transmits a dynamic measure of the entropy (surprise) that serves as a signal, activating the transition dynamics in RNN when it above a certain threshold. **B**, detailed architecture of the RNN and BG. **C**, Representative moments of a noise-free environment for changepoints tasks. **D**, **E** and **F** show the LEIA model dynamics evolution against time and each column across all panels represents the LEIA model dynamics in different modules at the same moment.

3.3 LEIA mirrors human adaptive learning

People performing predictive inference tasks tend to rely heavily on information immediately following changepoints, but differ in the overall rate of learning, the degree of adjustment at changepoints, and the use of relative uncertainty (which occurs during stable periods). We hypothesized that parameters controlling attractor transitions and Hebbian learning in the LEIA model could recapitulate both trends and differences in human learning behavior. In particular, we expected our threshold parameter for the entropy (which may represent balance between stabilizing, inhibitory and destabilizing, disinhibitory thalamocortical inputs (Lam et al., 2024; Mukherjee & Halassa, 2024; Scott et al., 2024)) to account for individual differences in a combination of overall learning and the degree of adjustment to changepoints, while Hebbian learning would impact the degree of learning within context.

Indeed, we found that the model recapitulated group trends in normative learning and that different entropy thresholds in LEIA captured an important source of variation in human learning behavior. In agreement with normative predictions from Bayesian models, network simulations revealed learning rates, measured as the degree to which predictions are updated for a given prediction error, that were maximal on changepoint trials and decreased as a function of trials thereafter (Figure 3A). However, both the degree of overall learning and its subsequent adjustment depended critically on the entropy threshold, with higher values leading to faster and more stable overall learning rates (Figure 3A; colors refer to different entropy thresholds). Fitting LEIA to subject behavior by adjusting the threshold parameter yielded average adaptive learning dynamics well matched to those in human participants (Figure 3B).

Closer examination of model behavior revealed that LEIA adjusted its reliance on recent prediction errors according to normative prescriptions. Normative models of adaptive learning rely on two key factors to control sensitivity to prediction errors: 1) changepoint probability (CPP), accounting for the probability that the helicopter has changed locations since the last observation, and 2) relative uncertainty (RU), accounting for the degree of uncertainty about the current underlying helicopter position. To better understand how LEIA responded to these factors, we fit a linear regression, which predicted updates according to prediction errors along with their interactions with relative uncertainty and changepoint probability (McGuire et al., 2014; Nassar et al., 2010, 2012; Razmi & Nassar, 2022). LEIA effectively captured the average fixed rate of learning (β_{PE}) as well as the degree of learning adjustment according to CPP and RU (Figure 3C)(McGuire et al., 2014). Thus, LEIA adjusted its use of prediction errors dynamically as prescribed by normative models and in a similar manner to human study participants.

LEIA not only accounts for overall trends across participants, but also individual differences, including a strong negative relationship between the overall rate of learning and its adjustment according to normative factors (Figure 3C; compare the pattern of gray lines across human and model behavior; when the PE coefficient is high, the CPP and RU coefficients are low, and visa versa). In the case of LEIA, these individual differences are driven by a single factor: the entropy threshold. The PE (β_{PE}) coefficient captures the overall learning rate, and participants with higher values of β_{PE} are best fit using lower entropy thresholds (aggressive learner) 3D. CPP (β_{CPP}) and RU (β_{RU}) coefficients are positively correlated with entropy threshold, indicating that participants with higher entropy thresholds (structural learner) more effectively respond to the structure of the task. Therefore, those with higher entropy thresholds make more accurate and conservative use of latent state representations, whereas those with lower thresholds are expected to segregate learning across a greater number of distinct inferred contexts.

Figures 4A and 4B give examples of the LEIA model's fitting to participants with high and low average learning rates, which are captured using low and high entropy thresholds, respectively. The more aggressive learner makes predictions based almost completely on the most recent observation even during stable periods of the task 4A, whereas the more structured learner makes predictions based on previous trials and only adapts to use high learning rates following changepoints 4B. For the aggressive learner, learning rates remain consistently at high level regardless of changepoints, whereas for the structured learners, the learning rates start high and gradually decrease to lower levels over time after changepoints (Figure 4C and 4D). LEIA captures behavior on both of these extremes, accounting for not only changes in overall learning rate, but also its dynamics (Figure 4C and 4D), through adjustments of a single threshold parameter.

While the entropy threshold in the model can be thought of as setting a tolerance for errors beyond which a new latent state is recruited and captures overall learning variability, it imposes a tradeoff between appropriately responding to CPP and being overly sensitive to noise. As higher entropy threshold reduces noise sensitivity and delays learning in response to CPP simultaneously. This raises a question of how a single person might adjust their own threshold based on the amount of noise in the current environment(Nassar et al., 2010; Piray & Daw, 2021, 2024), which we address below.

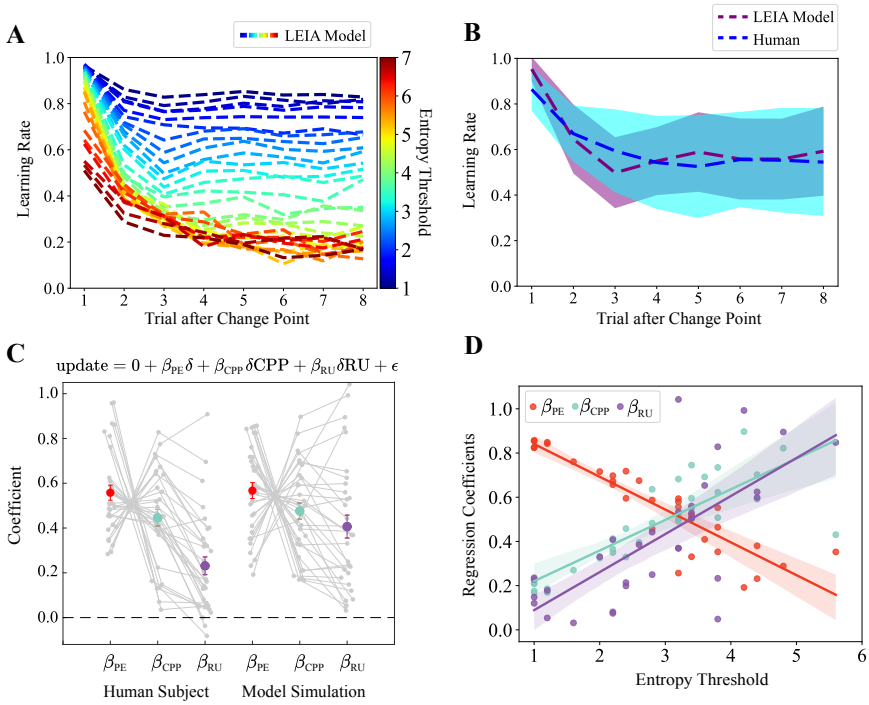

Figure 3: Parameter exploration for entropy threshold and human data regression analysis. **A**, Parameter search for entropy firing threshold against the learning rate after changepoints. We apply different entropy thresholds (ranging from 1 to 7, step size 0.2) to the model and calculate the average learning rate as a function of trials after changepoints across human data. The results show that the learning rate decreases after changepoints and the LEIA model captures a wide range of human learning patterns. **B**, The best-fitting LEIA model and human average learning patterns. Shaded areas represent the standard deviation. The best-fitting model is determined by minimizing the mean square error between human and model average learning rate post changepoints. **C**, coefficients derived from a linear regression model (equation above the panel C) fit to individual trials to characterize the overall learning patterns [β_{PE}: Fixed LR (learning rate)], adjustments in learning with respect to change point probability (CPP) [β_{CPP}: CPP Driven Learning] and relative uncertainty (RU) [β_{PE}: RU Driven Learning]. Gray circles represent fits to individual human subjects (left) and simulated data from the best models from panel B (right), whereas colored circles/bars reflect mean/SEM of each coefficient across participants. **D**, regression analysis between individual best-fitting entropy firing thresholds and coefficients β_i. High β_{PE} indicates a aggressive learning patterns, while low β_{PE} indicates structural learning patterns. The best-fitting model thresholds inversely correlated with β_{PE} suggesting that a high learning rate is associated with a low entropy firing threshold, resulting in increased response and belief updates when presented with noisy contexts and vice versa. For structural learners, the CPP (β_{CPP}) and RU (β_{CPP}) coefficients exhibit a positive relationship with the entropy threshold, implying that higher CPP and RU values enhance learning whereas aggressive learners do the opposite. The shaded area represents a 95% confidence interval.

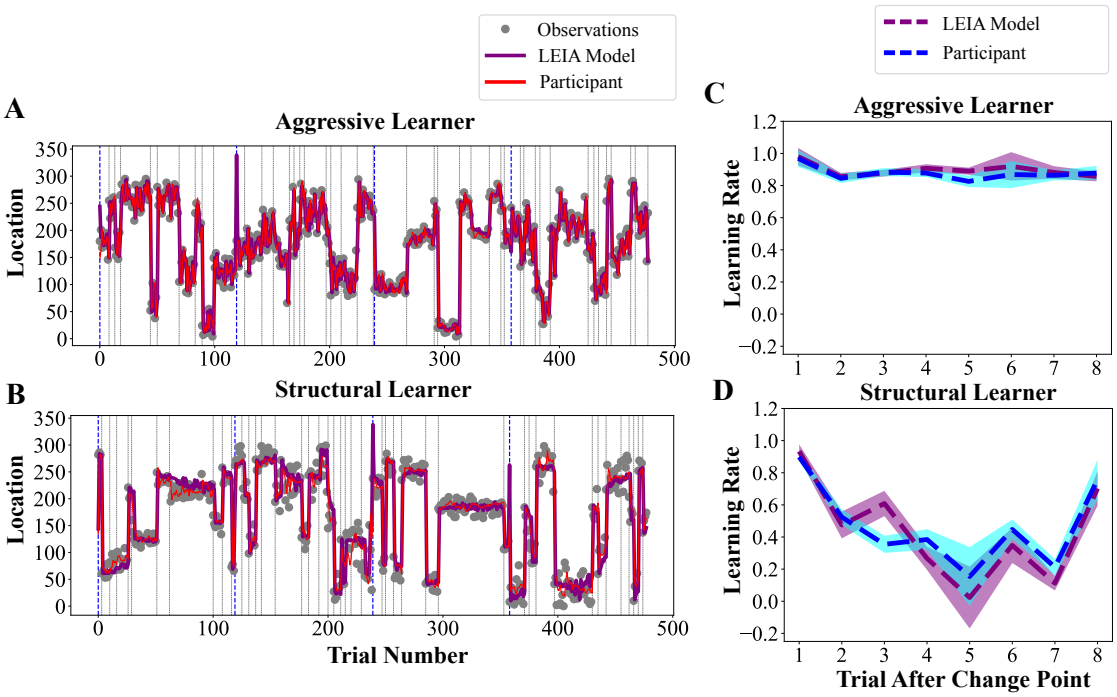

Figure 4: Examples of LEIA models adapt to aggressive learner and structural learner. **A** and **B** show participants with high and low learning rates, respectively. Red and purple lines show participant's and the LEIA model predictions, respectively. Blue dashed lines indicate the beginning of a new block (the experiment consists of four blocks in total), while gray dashed lines mark different contexts. **C** and **D** show the respective learning rates as functions of trials following changepoints for the aggressive learner and the structural learner. Blue and purple lines demonstrate human and the LEIA model's learning rates, respectively. The shaded areas represent the standard deviation across all trials at the same position after the changepoints. The results indicate that the LEIA model captures not only human average learning patterns but also individual variability.

3.4 LEIA captures learning adjustments in response to noise

To mitigate the tradeoff between accurately identifying new states and avoiding excessive sensitivity to noise, we extended the LEIA model to incorporate an adaptive threshold. The fixed threshold simulations above were able to capture individual differences in learning observed in a single noise condition, however, previous work has shown that human participants adjust their learning according to the level of environmental noise. By including an adaptive threshold, the model can capture differential learning dynamics across conditions with different levels of noise, as we demonstrate below (see methods 5.2). Specifically we tested the model in conditions differing in their level of noise and compared the performance of LEIA to that of human participants, as well as to a (reduced, approximate) normative Bayesian inference model (Nassar et al., 2010) with full knowledge of the level of noise in each condition 5A. This model maintained a belief state at each time based on the computation of changepoint probabilities and outcome expectations conditioned on changepoint occurrance and non-occurance.

LEIA showed qualitative adjustments to noise prescribed by normative learning to a degree that roughly matched that observed in human participants. Qualitatively, the LEIA model shows a similar dependence on noise as the optimal model in series of generative data (Figure 5B). We pre-trained the model on a sequence of data with a fixed noise level, and then tested on a single prediction error to examine the degree to which it updated in response (defining a learning rate). The adaptive LEIA model required a larger error to increase its learning rate in the face of higher levels of noise, whereas the model without an adaptive threshold presented the same amount of learning behavior for different errors. This adjustment was qualitatively similar to that of a normative learning agent that had full knowledge of the noise conditions, albeit smaller. The adaptive threshold LEIA model shows a shift comparable to that seen in humans across noise levels tested in human experiments (Figure 5C). Taken together, these results suggest that equipping LEIA with an adaptive threshold facilitates normative noise-related adjustments in learning dynamics consistent with human learning.

3.5 LEIA learns new policies and revisits old ones

Humans employ different forms of adaptive learning in different environments. As described above, in situations where changepoints are prevalent, people prioritize experiences at and after such changepoints. However, in environments where contexts are repeated, people often make use of previous learning, for example by recognizing that a previously learned action-outcome contingency would work in a new situation (Yu et al., 2021). This sort of behavior can be measured in tasks that involve contingency reversals, where the appropriate action-outcome mapping occasionally reverts to one that was previously experienced (Collins & Koechlin, 2012; Wilson et al., 2014). The LEIA model described above used a hard-coded gated-connectivity structure appropriate for changepoint environments, and therefore is unable to reuse information after reversals. However, we show in this section that by replacing the hard-coded thalamus-gated PFC connectivity with two biologically supported forms of gated connectivity, LEIA becomes a flexible adaptive learner, learning new policies or reverting to old ones as appropriate given the environmental demands.

We enable LEIA to reuse information, in cases where it would be beneficial, by modifying the machinery governing RNN attractor state transitions. Specifically, we allow the RNN attractor state usage to be informed by both new and previously encountered experiences. The version of LEIA that we described in previous sections is not capable of reusing previously learned policies since the thalamus-gated connectivity was designed to always "push" the RNN into a new attractor state. In order to enable LEIA to reuse previously learned attractor states (and hence behavioral policies), we made two changes to the model that were motivated by both known biology and computational considerations. The first change was to the thalamus-gated RNN connectivity; we replaced the connectivity enforcing transitions to the next state with correlated random connectivity gated by the thalamocortical inputs, consistent with that previously proposed to explain state transitions in cortical population activity and behavior (Recanatesi et al., 2022). The second change was to incorporate feedback projections from the motor output layer directly to the RNN, with weights that are learned slowly through Hebbian mechanisms. These two mechanisms combine to enable attractor switches that favor novel states when encountering observations unlike those seen previously, but previously used ones when encountering observations that closely match previously observed ones.

To assess the flexibility and generality of the augmented model, we compared the previous and the new LEIA models across varying environments. These environments differed in the number of changepoints and reversals (as illustrated in Figure 6B). As the ratio of changepoints to reversals increases, the task gradually shifts from having only reversals to mixed scenarios, and ultimately to containing only

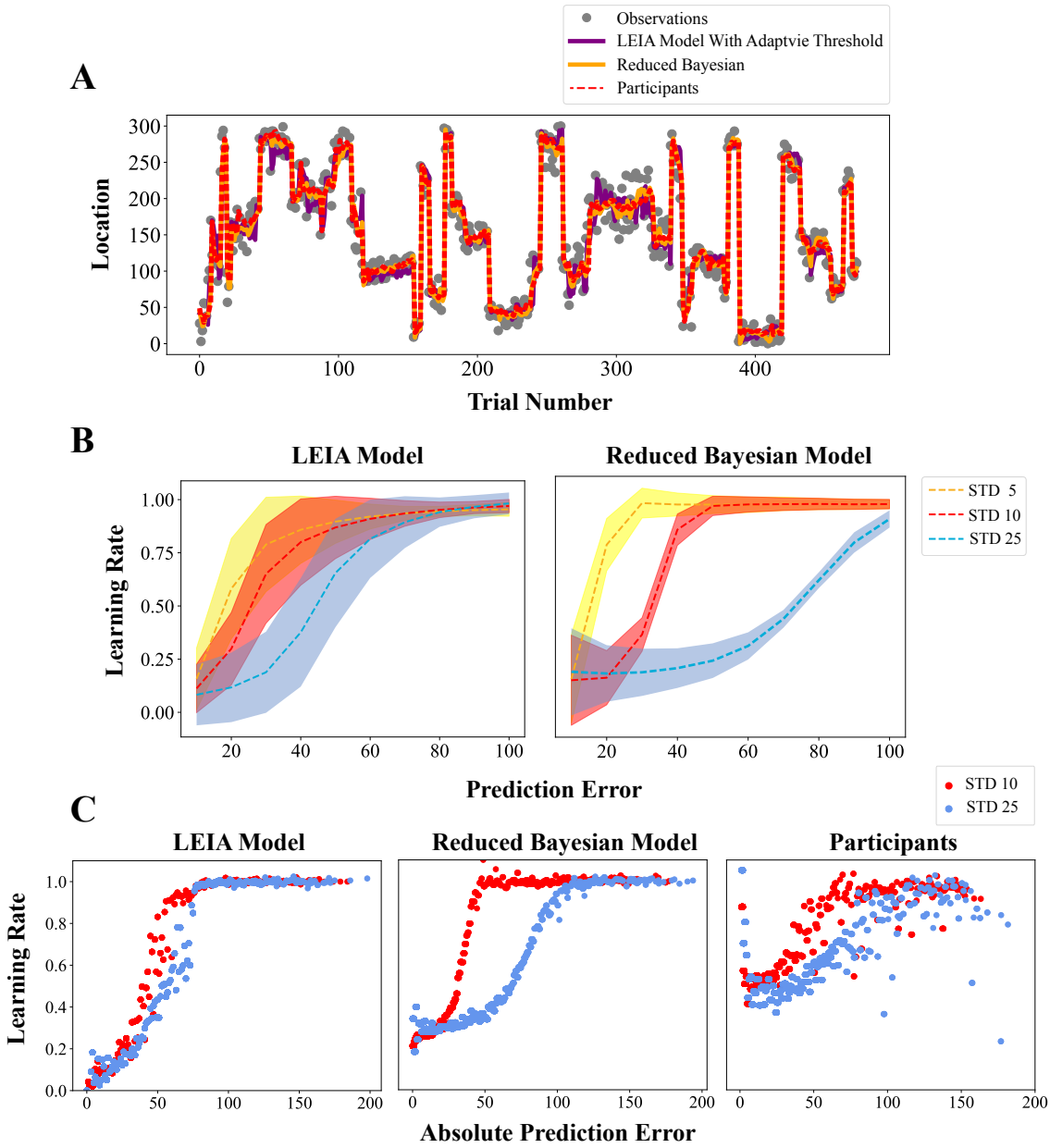

Figure 5: Comparative analysis of models and human performance across varying noise levels. **A**, predictions of two models and a participant in the same experiment. The LEIA model with adaptive threshold (see methods 5.2) performs comparably to the reduced Bayesian model in capturing structural information under different noise levels and captures human behavior in general. **B**, learning rate against prediction error dynamics under different noise levels across various methods in a generative model. In each measurement series, we first produce 10 trials drawn from a context with a mean of 50 and a given noise level (i.e., 5, 10, or 25 standard deviations). Based on the 10th prediction, we create 2 more trials, modulating the prediction error from 10 to 100. For each level of prediction error, we make 1000 different measurement series and calculate the average learning rate for the last trial. Shaded areas represent the standard deviation. The results demonstrate that the learning rates of both models are affected by noise and lead to obvious shifts. **C**, learning rate against prediction error dynamics under different noise levels across various methods in human experiments. We derive the average learning rate from average updates and absolute prediction errors within bins of step 1. Red and blue dots represent different noise levels. Both the LEIA model and the reduced Bayesian model, as well as human participants, show obvious shifts in learning with different observation noise.

changepoints. Simulation results indicate that our previous, changepoint-only model consistently infers too many changepoints in mixed scenarios, regardless of the number of changepoints and reversals that actually occurred (6D). We counted the total number of states inferred by the two LEIA models as a function of trials observed (see methods 5.5), finding that the LEIA model considers each context as a new state and the number of inferred states across different environmental settings remain the same. In contrast, the augmented LEIA model maintains a low count in reversal-only scenarios, since it can revert to a previous attractor (relating to this experience) and increases the latent state count linearly as more changepoint contexts are introduced (Figure 6D). This allows it to reuse previous knowledge whenever it encounters previously experienced contexts. In individual simulations, this can be seen in the model's rapid adjustments to the mean of a previously experienced distribution, as opposed to updating to the newest data point(Figure 6B – compare blue and red curves). As a result of this, the flexible model considerably out-performs the changepoint-only model in mixed environments, where reusing information is more beneficial (Figure 6C).

Together with the results from the previous sections, our findings show how BG-thalamo-cortical circuitry can afford human-like adaptive learning across environments by combining incremental learning in cortico-striatal synapses with fast inference facilitated by thalamo-cortical projections that drive activity dynamics resulting in selection of context-appropriate cortical populations.

4 Discussion

We developed a neural network model constrained by biology capable of operating on neuronal timescales to execute adaptive learning that generalizes across temporal structures. Our LEIA model builds on cognitive models of latent state inference that require two operations: 1) the ability to infer the latent state corresponding to the active action-outcome contingency and 2) the ability to learn appropriate policies for the active latent state (Collins & Koechlin, 2012; Eckstein & Collins, 2020; Gershman & Niv, 2010; Yu et al., 2021). The LEIA model accomplishes these two operations in a corticostriatothalamic architecture whereby inference of the active latent state is accomplished through attractor switches in a prefrontal reccurrent neural network and incremental policy learning is accomplished through synaptic plasticity at coritico-striatal synapses. A base version of the LEIA model captures the full range of individual differences in human adaptive learning through adjustment of a single threshold parameter and adjusting this threshold dynamically facilitates human-like adaptive adjustments to environmental noise. Finally, we show that structural assumptions that are typically necessary to implement latent state inference, such as asymmetric connectivity between attractor states, can be relaxed in LEIA and replaced with two biologically motivated elements: 1) connections between motor and prefrontal cortex tuned through slow hebbian learning (Gabbott et al., 2005), and 2) random connectivity in the prefrontal recurrent neural network that is gated by suprathreshold levels of response entropy signaled through thalamo-cortical projections. These elements allow the model to achieve near-optimal inference across environments spanning a range from those that demand reuse of previously learned policies to those that demand rapid relearning in the face of completely new outcome contingencies.

Our model has several implications related to adaptive learning behavior. First, it achieves many of these behaviors, which have traditionally been considered to potentially rely on synaptic learning rate adjustments (Behrens et al., 2007; Browning et al., 2015; Nassar et al., 2010, 2012), without changing synaptic learning rates. Dynamics in the learning behavior of the LEIA model are driven by transitions in a recurrent neural network that provides input to a cortico-striatal learning system. Such transitions allow LEIA to reset policies when entering a new context or to reuse an old policy when entering a familiar one (see Figure 6A). This distinction is important because recent work has highlighted that neural signatures of adaptive learning previously thought to reflect dynamic learning rates (Behrens et al., 2007; Fischer et al., 2015; Nassar et al., 2012) have relationships to behavior that are highly context dependent (D'Acremont & Bossaerts, 2016; Nassar, Bruckner, & Frank, 2019; O'reilly, 2013). Such context dependence is not easy to reconcile with the idea of learning rate adjustment, but is instead better explained by the idea that such signals reflect subjective assessments of latent state transitions (Razmi & Nassar, 2022; Yu et al., 2021). Here we show how such signals, used by LEIA to update prefrontal attractor states, can be harnessed to achieve a wide range of adaptive learning behaviors including some interpretable through the lens of learning rate adjustment (See Figure 3) as well as some that cannot (See Figure 6B).

A second behavioral implication of LEIA is that PFC-BG-Thalamic networks using the well-attested representations and learning rules incorporated in our model should be able to learn efficiently in environments with more complex latent state structure than that assessed here. Latent states learning,

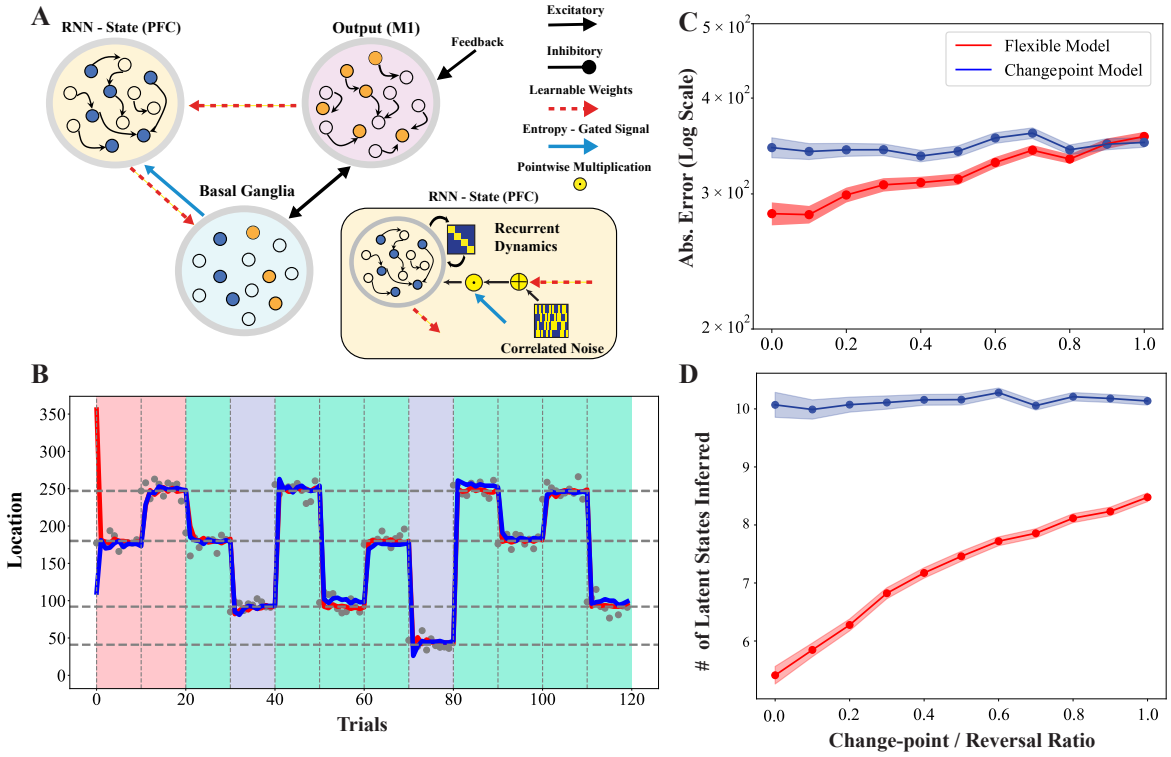

Figure 6: Flexible model structure and comparison of two LEIA models in complex environments. **A**, flexible model structure. Based on the previous model, we incorporate connection from the motor output layer to the RNN to address reversal learning tasks. To release model attractor switches constraints in complex environments, we replace the asymmetric transition dynamics with correlated noise. Upon reaching a changepoint, the model will randomly switch to an unused attractor rather than forcing a transition to the next adjacent one. **B**, example of a complex environment setting. We established a generative model for a complex environment that includes multiple changepoints and reversals, with their ratio set as a hyperparameter. Red shaded areas indicate "warm up" contexts (see methods 5.5), green shaded areas represent reversal contexts, and purple shaded areas denote changepoint contexts. The changepoint model considers each context as novel and learns anew. The flexible model, which adaptively reverts to the previous attractor or jumps to a random new attractor depending on the context, demonstrates the best learning strategy. **C**, performance comparison between two models. The Y-axis represents the absolute error of predictions against observations, while the X-axis indicates the ratio of changepoints to reversals. The absolute error of changepoint model maintains the same regardless of the changepoints/reversal ratio. The flexible model can better captures structural information in reversals and show comparable ability to the changepoint model in changepoints tasks. Shaded areas represent the standard error of mean across 300 measurements. **D**, the number of latent states inferred in complex environments. We counted the latent states inferred by the models. The latent states inferred by the changepoint model remain the same with respect to the change point/reversal ratio. In contrast, the flexible model infers few number of latent states during reversals, and scales linearly as the number of changepoints increases.

when implemented through probabilistic reasoning, requires making assumptions about the underlying structure of the environment, and prior implementations of latent state learning in cognitive and network models have "baked in" the transition structure of the environment in order to achieve behaviors that are efficient for a given task environment. However, people are capable of behaving appropriately across a wide range of different environments, including ones that require rapidly relearning after environmental changes, as well as those that require re-using information learned from an experience in the distant past (Collins & Koechlin, 2012; Nassar et al., 2021). Here we provide a first clue to how people might achieve such flexibility, in that we show how a specific architecture making use of learned feedback projections and gated random connectivity, can support latent state transitions that are appropriate for the current environment, without explicit learning of the environmental transition structure. Here we focus on one specific dimension of transition structure, whether changes in the environment reflect new contexts or repetitions of old ones, and we hope that our work inspires future studies to characterize the breadth of transition structures over which humans can learn, as well as the broader characteristics of network architectures that affect whether and how networks manage learning when confronted with them.

Beyond behavior, LEIA makes specific predictions about its biological basis. A core element of the model is a recurrent neural network that maintains active latent state representations and sends projections to the striatum. A key signature of this sort of representation is that it should change rapidly at transitions in the environment, thereby enabling new subpopulations of neurons to control behavior after these transitions, which minimizes interference between different contexts. While we broadly hypothesize that the attractor network exists in frontal cortex, there is evidence for latent-state like representations in a number of different frontal regions, including dorsolateral prefrontal cortex (Vaidya & Badre, 2022; Vaidya et al., 2021), orbitofrontal cortex (Nassar, McGuire, et al., 2019; Schuck et al., 2016; Wilson et al., 2014) , and anterior cingulate cortex (Karlsson et al., 2012; Powell & Redish, 2016). It is noteworthy that all of these regions project to the striatum, albeit to different subregions thereof. Furthermore, "network reset" phenomena subsequent to environmental transitions and tightly linked to behavioral change have been observed in both orbitofrontal cortex in humans (Muller et al., 2019; Nassar, McGuire, et al., 2019) and medial frontal regions of rodents thought to correspond to ACC in primates and humans (Karlsson et al., 2012; Powell & Redish, 2016). Our model proposes that these changes in the cortical representations are provided to the striatum and play a causal role in increasing the rate of behavioral adjustment at environmental changes. This prediction is distinct from other theories based on learning rate adjustment, which suggest instead that periods of rapid learning are accomplished through enhancement of dopaminergic signaling in the striatum (Diederen et al., 2016, 2017; Haarsma et al., 2021), and is broadly in line with the view that an integrated hierarchy of activity and plasticity mechanisms across scales generates metaplasticity that improves behavior Scott and Frank, 2023. Recent findings that rodent adaptive learning behaviors are not accompanied by scaling of dopamine learning signals (Mah et al., 2024), along with our alternative model of how such adaptive learning behaviors might emerge, suggest the need for new methods that would allow perturbations at the population level to test the causal link between reset of cortical networks and their downstream consequences for learning. Such tests might allow for better distinction between the different cortical regions that have been shown to represent latent states in different tasks and conditions, and potentially reveal their differential contributions to behavior.

While our model provides a substantial advance in understanding how adaptive learning could be implemented in cortico-striato-thalamic circuitry, it has a number of limitations. First, while our extended architecture is capable of generalization across environments, this generalization is limited to environments with persisting state transitions, such as changepoints and reversals. Thus, while in principle latent state inference could be used to improve learning in environments with sequences or transient outliers (Bakst & McGuire, 2023), accomplishing this in our model would require additional mechanisms for explicit learning of the environmental structure, in particular the transition function defining how latent states are likely to transition to one another. In principle this might be accomplished by extending LEIA through the introduction of plasticity into the RNN, such that gated attractor transitions are biased not only by learned associations to motor outputs, but also through learned associations between latent states that have occurred sequentially in the past.

A second limitation of our model is that it incorporates simplified RNN dynamics and action policies. These simplifications emerged because the goal of our modeling efforts were to explain behavioral data from predictive inference tasks where optimal policies are somewhat simple (report the mean of the underlying distribution) and conditional only on the active generative mean in the absence of trial-specific cues. Here we have attempted to illustrate exactly how combining RNN transitions with a coritico-striatal learning system can be used to facilitate adaptive learning under these restrictive conditions, however

the principles that we demonstrate here could be applied to more complex problems such as those faced in other tasks or in real world learning problems. For example, cue information could also be provided to the RNN to yield more complex state representation including both stimulus information as well as its temporal context. Similarly, more complex dynamics could be incorporated into the RNN to yield policies that change over time, or change according to actions initiated. Thus, while we have focused our exploration of the LEIA model on tasks that have been the subject of extensive investigation of behavior and its underlying neural mechanisms (McGuire et al., 2014; Nassar, Bruckner, & Frank, 2019; Nassar, McGuire, et al., 2019; Nassar & Troiani, 2021; Nassar et al., 2010, 2012, 2021), the basic architecture is ripe for extension.

In summary, the LEIA model provides a biologically motivated account of adaptive learning through latent state inference. It explains human behavior, recapitulates normative adjustments in learning, and can rapidly re-learn or reuse information as appropriate for a given situation. These findings provide a principled explanation for how cortical activity dynamics and striatal plasticity mechanisms can be combined to produce flexible and efficient learning. Future work should test the key biological predictions of our model and build on our results to address more complex statistical environments, and to integrate our model's account of plasticity with the growing body of related work which suggests that a hierarchy of different mechanisms, distributed within and between areas, supports behavioral plasticity.

5 Methods

In this section, we begin by describing the mathematical formulation of the LEIA model. In the model, the PFC encodes latent states as RNN states, which are supported by a recurrent connection matrix. Then, the PFC projects to the BG via canonical disinhibitory connections—first from striatal "Go" cells to the GP, and subsequently from the GP to the motor thalamus. The BG module in our model is characterized by a topographical organization of actions, as implemented in many previous cortico-BG models (e.g., Frank and Claus, 2006; Gurney et al., 2001) and motivated by neuroanatomical data (Alexander et al., 1986). After an action is selected and executed, supervised feedback enters the network, reinforcing existing state-action associations or establishing new ones through plasticity at the PFC-striatal synapses. First, we discuss layer dynamics in the LEIA changepoint model, followed by the flexible model. Subsequently, we describe the model's connectivity and learning rules. Next, we discuss how the model makes a decision, and how it functions in the helicopter task. Finally, we address other details in the human experiments. The code and parameter values for all the simulations reported here can be found at **xxxxx** (All code will be publicly available on after acceptance).

5.1 The basic LEIA model

5.1.1 The PFC and latent state representations

The dynamics within the LEIA model illustrate the sequential unfolding of the RNN-BG-Motor (cortico-basal ganglia) loop, as shown in Figure 2. This loop starts with the activation of a cluster of excitatory RNN neurons, whose dynamics is governed by the following equations:

$$\tau_{\text{rnn}} \frac{dx_i}{dt} = -r_i + \sum_{j=1}^{N} W_{ij} r_j + \delta \left(\sum_{k=1}^{N} T_{ik} r_k \right) \tag{1}$$

$$r_i(t) = \max(0, \tanh(x_i(t))) \tag{2}$$

Here, x_i and r_i represent membrane potentials of neurons in PFC (RNN) and firing rates, respectively. W is the recurrent connection matrix, RNN has $N = 600$ neurons and T is an asymmetric state transition matrix. An entropy-based gating signal δ, computed in the thalamus controls state transitions.

More specifically, the stable recurrent connectivity W, which implements attractor dynamics in the PFC, is implemented via a symmetric matrix that encodes P patterns in the following way:

$$W_{ij} = \frac{F}{N} \sum_{i=1}^{P} e_i^T e_i \tag{3}$$

where e_i is a pattern encoding vector in $\mathbb{R}^{1\times N}$ which is defined as a subset of excitatory neurons ($N_{patt} = 30$) that are surrounded by inhibitory neurons ($N - N_{patt}$), and $\frac{F}{N}$ is a scaling factor. The symmetric

term W_{ij} is structured as blocks of diagonal excitatory neurons surrounded by off-diagonal inhibitory neurons. The asymmetric transition weights T control the RNN's transition dynamics by shifting activity from one attractor to another. These weights are essentially a circular shift of those in W by N_{patt} units, as shown in Figure 2B.

5.1.2 The thalamus and thalamic entropy

The thalamus in our model is a feed-forward layer of $M = 360$ neurons with topographic input from the GP and outputs to the motor layer. Entropy is computed in the thalamus and controls the PFC gating signal δ. It is a dynamical measure of the conflict between proposed output and supervised feedback within the thalamus layer. The (scaled) entropy is computed, over a normalized vector T of thalamic activities:

$$S = -g\mathbb{E}[TlogT] \tag{4}$$

Here, g is a scaling factor we introduced to adjust the entropy to match the magnitude of the RNN dynamics. We applied a step function with a threshold S_{th} to define the sensitivity of conflict measurements to entropy, thereby enabling effective switching within the model.

The PFC gating signal is computed from the entropy:

$$\begin{cases} \tau_{en}\frac{d\delta}{dt} = -m_1\delta + S & \text{if } t_{ref} = 0 \text{ and } \delta < \delta_{th} \\ \tau_{en}\frac{d\delta}{dt} = -m_2\delta + \delta_{rest} & \text{Otherwise} \end{cases} \tag{5}$$

$$t_{ref} := \begin{cases} \text{Refractory time} & \text{if } t_{ref} = 0 \text{ and } \delta \geq \delta_{th} \\ t_{ref} - 1 & \text{Otherwise} \end{cases} \tag{6}$$

That is, the entropy-based gating signal integrates inputs with a time constant τ_{en}, decays to a resting state δ_{rest}, and is driven by entropy S. It is subject to thresholding by δ_{th} and refractoriness. Specifically, inspired by the membrane potential model, the state transition signal takes entropy as input and fires upon reaching the threshold δ_{th} for state transition when surprise is elevated.

We introduced the transition signal δ, rather than using entropy S directly, to prevent excessive fluctuations in the RNN's asymmetric component, and causing waves of network activity. Following a surprise event, the entropy value remains above the threshold for a certain duration until a new state-action pair is established. Using the entropy signal S directly can lead to continuous switching of attractors due to sustained high entropy levels. In contrast, the entropy-gated transition signal δ generates a spiking signal such that the RNN only switch to the next adjacent attractor in response to a surprise event.

5.1.3 The basal-ganglia and action selection

The RNN excitatory units project to the second, basal-ganglia, action selection module (cyan shaded area in Figure 2), and in particular to the striatum. This module is composed of 3 layers of M neurons each, the striatum, globus pallidus and thalamus. These are governed by the follwing equations, respectively:

$$\tau_{\text{msn}}\frac{dx_s}{dt} = -r_s + \alpha\sum_{i=1}^{N} W_{si}r_i + \sum_{o=1}^{M} W_{so}r_o + b_{msn} \tag{7}$$

$$\tau_g\frac{dx_g}{dt} = r_g - \sum_{s=1}^{M} W_{gs}r_s + b_{gp} \tag{8}$$

$$\tau_v\frac{dx_v}{dt} = -r_v + \sum_{o=1}^{M} W_{vo}r_o - \sum_{g=1}^{M} W_{vg}r_g + b_{th} + \epsilon \tag{9}$$

$$r_q(t) = \max(0, \tanh(x_q(t))), \quad q = s, g, v \tag{10}$$

Vectors x_s, x_g, x_v are the membrane potentials of neurons in the striatum, global pallidus and thalamus, r_s, r_g, r_v are their firing rates. τ_{msn}, τ_g and τ_v respectively represent the time encoding constant to these neurons. In equation (7), W_{si} and W_{so} are the weight matrices representing the RNN to the striatum connectivity and the output to striatum connectivity. r_i and r_o are the firing rates of RNN and output layer neurons, and we use a rectified hyperbolic tangent function to guarantee the non-negativity of

excitation (the same applies for equations (1), (8), (7), (9), (13), (18)). In equation (8), W_{gs} is the weight matrix representing the connectivity between the striatum and the globus pallidus. In equation (9), W_{vo}, W_{vg} are the weights matrices coding for the motor output to thalamus connectivity, and the globus pallidus to thalamus connectivity. $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ is private gaussian noise with $\mu = 0$ and $\sigma = 0.05$. The connectivity between the layers in the basal ganglia, namely W_{gs} (Eq. (8)) and W_{vg} (Eq. (9)), are identity matrices in $\mathbb{R}^{M \times M}$. Thus, connectivity between the layers in the basal ganglia is represented by one-to-one topographical connections which is supported by biological evidence.

Synaptic plasticity in the LEIA changepoint model takes place within the projection from the PFC to the BG (dashed red line from yellow-shaded area to cyan-shaded area in Figure 2). Hebbian learning in these connections is governed by the following equations:

$$\Delta W_{si} = \beta \left((r_i - \gamma_1) \cdot (r_s - \gamma_2) \right) \tag{11}$$

$$W_{si} := W_{si} + \Delta W_{si} \tag{12}$$

where β is the learning rate, γ_1 and γ_2 are learning thresholds, and r_i and r_s are RNN layer firing rate and striatum layer firing rate. We clipped the learning increments to the range $[W_{min}, W_{max}]$, which represent the minimum and maximum learning thresholds.

5.1.4 The motor cortex and action execution

The motor output module (purple shaded area in Figure 2) is composed of a M neuron layer, receiving inputs from the thalamus layer, and whose dynamics are defined as follows:

$$\tau_o \frac{dx_o}{dt} = -r_o + \sum_{v=1}^{M} W_{ov} r_v + \sum_{o=1}^{M} W_{oo} r_o + \eta J^{In} \tag{13}$$

where x_o is the membrane potentials of neurons in the motor output and r_o is firing rate. τ_o is time encoding constant, W_{ov} and W_{oo} are the weight matrices representing the thalamus to output connectivity and within output layer connectivity, respectively. η is the supervisory input gain and J^{In} supervision input vector.

Within-layer connectivity is set to produce bump attractor dynamics, with self-local excitation and surround inhibition. These matrices in $\mathbb{R}^{M \times M}$ are built by filling each row with a vector in $\mathbb{R}^{1 \times M}$ representing a normal probability distribution function that is scaled by h, w, and b to control the height, width, and baseline values of the bell curve, respectively.

The LEIA model makes a decision when the internal representation within the output (motor) layer is stable and converges to a bump attractor. At every timestep during the decision phase, we compute the hopfield energy of the motor layer, update a history vector containing 5 time-points of these values, and we evaluate whether it is increasing by taking the slope over these 5 data points. If the slope is lower than 0.1, we assume that dynamics at the output layer are stable, and select the action that is associated to the highest activation node. At this point the supervised signal is given to the output layer, and the dynamics continue for 500 iterations (i.e., learning phase), in order for an attractor switch in the RNN to potentially occur. The state of the RNN and striatum activations during this last iteration are used in our Hebbian update equations, such that W_{si} is updated between trials.

5.2 LEIA with an adaptive entropy threshold

To improve the ability of the model of capturing with varying noise levels in both generative model data and human experimental data (Figure 5 B and C), we implement an adaptive entropy threshold mechanism. This mechanism is a form of exponential decay which dynamically adjusts the entropy firing threshold based on the maximum entropy value associated with the current inferred state as given by the following equation:

$$S_{th} := (1 - \lambda_{en}) S_{th} + \lambda_{en} (S_{max} + \theta_{en}) \tag{14}$$

where λ_{en} is a decay constant, S_{max} is the maximum entropy value the model has received so far in the current inferred state (i.e, the current attractor state), and θ_{en} is a constant. Whenever inferred state changes, S_{th} will be reset to the initial value and S_{max} will be reset to zero. Notably, S_{th} is evaluated between trials. We adapted this mechanism in the LEIA changepoint model and performed learning rate

analysis against prediction error in a genrative model (Figure 5B) and use the following sigmoid function with 3 DOF to fit original data.

$$\text{sigmoid}(x) = \frac{d}{1 + e^{-(ax+b)}} \tag{15}$$

5.3 LEIA with correlated noise

The LEIA changepoint model quickly adapts to new contexts after detecting a surprise, that is, a changepoint. However, if the new context is similar to a previously experienced one, the basic model tends to learn state - action mapping from anew rather than reusing previous information, which conflicts with real-life experiences. We further extend the LEIA changepoint model to deal with complex environments, i.e., a mix of changepoints and reversals. The ideal scenario is the attractor switching to any unused states and building a new state-action pair whenever a new context is encountered, while reverting to previously used attractors during reversals.

To meet our desiderata, we use correlated noise (correlated with attractor position) governed by the following equations to replace asymmetric transition connectivity for attractor switch.

$$\xi \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C}) \tag{16}$$

$$C_{i,j} = \begin{cases} 1 & \text{if } \left\lfloor \frac{i}{N_{\text{patt}}} \right\rfloor = \left\lfloor \frac{j}{N_{\text{patt}}} \right\rfloor \\ e^{-\lambda \left| \left\lfloor \frac{i}{N_{\text{patt}}} \right\rfloor - \left\lfloor \frac{j}{N_{\text{patt}}} \right\rfloor \right|} & \text{otherwise} \end{cases} \tag{17}$$

where $i, j \in [0, N]$, $\lfloor x \rfloor$ denotes the floor function, λ is decay constant, ξ is a noise drawn from a multivariate standard Gaussian distribution with means of $\mathbf{0}$ and covariance matrix $\boldsymbol{C}$. Unlike an asymmetric matrix, which always switches to the next adjacent attractor, noise can cause the attractor to shift to any unused attractor randomly. This enables more flexible learning by allowing the model to either build a new state-action mapping or revert to a previously experienced mapping upon encountering changepoints or reversals.

For the model with correlated noise to perform reversal learning, we discovered that it was crucial to augment the basic Hebbian learning by thresholding it. Learning in the flexible model occurs in two connections: RNN - striatum and motor output - RNN projections with the same learning and updates rules in equations (11) and (12). In the flexible model, we set the learning threshold γ in equation (11) higher to enable the learning of negative weights in connection from the motor output to the RNN. Negative weights are able to inhibit the currently active attractor during state switches. The RNN dynamics in flexible model read as:

$$\tau_{\text{rnn}} \frac{dx_i}{dt} = -r_i + \sum_{j=1}^{N} W_{ij} r_j + S \sum_{j=1}^{M} W_{io} r_o + \kappa \xi \tag{18}$$

where κ is the noise gain and ξ is the correlated noise drawn from equation (16).

5.4 Helicopter task and performance analysis

We analyze human behavior, with a cohort of 32 participants, and model performance in a previously described predictive inference task (McGuire et al., 2014). In the task, there is an animated helicopter occluded by clouds and drops bags of coin in an open field. Participants were pre-trained to move a bucket horizontally to catch falling bags. During the experiment, the helicopter occasionally changed locations, and positions of bags were sampled from Gaussian distributions with a standard deviation of 10 or 25 centered around the helicopter's locations. The objective of the task is to catch as many bags as possible.

We fit the LEIA changepoint model to the data from all 32 participants, using an entropy firing threshold ranging from 1 to 7 with a step size of 0.2 (Figure 3A). We then calculated the learning rate as a function of post-changepoint trials. The learning rates for both the model and the human participants were derived from linear regression of all trials' updates versus prediction errors at specific positions (i.e., 1, 2, ... ,8) in the data of all 32 participants.

We fit linear models, following earlier work on changepoint tasks (Nassar et al., 2010, 2012), to both subject and network-based behavior. These models had the form

$$update = \beta_{PE}\delta + \beta_{CPP}\delta CPP + \beta_{RU}\delta RU + \epsilon \tag{19}$$

Changepoint probability (CPP) and relative uncertainty(RU) are statistics that can be defined during the derivation of normative models for our tasks (Figure 5). In approximations to these models, updates are depend on these terms linearly (Figure 3C). We give a brief introduction of CPP and RU here, more details are discussed in previous work. CPP is a reflection of the probability of a change in helicopter location given the most recent observation. RU is a measurement of the uncertainty about the location of the helicopter in the next prediction. The normative model takes additional environment information (i.e., hazard rate, noise level) and two latent variables described above to capture human learning behaviors.

When fitting the linear model to human data, we first extract the human prediction errors and updates. Next, we incorporate environmental information to compute the CPP and RU. Finally, we fit the update to the corresponding data using equation (19). Similarly, for the model data, we first determine the optimal thresholds for each individual by minimizing the MSE of the average learning rates after changepoints. Then, we run the model using the best threshold on the same experiment and extract the relevant data to fit the linear model (Figure 3C).

For the LEIA model with an adaptive threshold, the entropy firing threshold follows equation (14). Whenever the model switch to a new attractor, S_{th} is initialized as 5 and then adaptively adjusts to the current noise level. In Figure 5B, we fit the model to an experiment consisting of 11 trials. We cached the model dynamics for previous 10 trials, and controlled the prediction error for the 11th within the range of 10 to 100 under varying noise levels. The original data was then fitted using equation (15) in Figure 5B. The reduced Bayesian model was processed in the same manner.

After confirming that the LEIA model with an adaptive threshold exhibits clear separation across different noise levels, we applied the LEIA model with apative threshold to human data. The data points were averaged using a sliding window with a step size (absolute prediction error) of 1, as shown in Figure 5C. The results indicate that the LEIA model qualitatively approximates human behavior more closely than the normative model.

5.5 Applying the model to mixed changepoint and reversal environments

We established a generative model to evaluate performance in complex environments for the changepoint model and flexible model. For each set of generative data, the experiment consists of two initial contexts followed by ten subsequent contexts, which include a mix of changepoint tasks and reversal tasks, with changepoint-to-reversal ratio set as a hyper-parameter. Each context includes 10 trials sampled from a Gaussian distribution with a mean ranging from 30 to 300 and a standard deviation of 10. The first two contexts are designated as "warm-up" contexts (red regions, Figure 6B), as they must be changepoint tasks. The following 10 contexts consist of either changepoint tasks (purple, 6B) or reversal tasks (green regions) controlled by the ratio.

We applied the LEIA model to these environments by generating 300 random initial environments across the two models and performed analyses based on the predictions. The number of inferred states in Figure 6D refers to the total number of states inferred by the models at the last trial and is measured by counting the state - action associations (yellow blocks) in the RNN-striatum connection (Figure 2C). In mixed environments, the LEIA flexible model has lower absolute error (measured as the prediction deviation from the mean of the Gaussian distribution) and fewer inferred states, indicating that it reverts to previous states and reuses learned knowledge. In contrast, the LEIA changepoint model treats each incoming context as a completely new environment, learning from scratch each time, which results in a higher absolute error and more inferred states.

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, *9*(1), 357–381.

Bakst, L., & McGuire, J. T. (2023). Experience-driven recalibration of learning from surprising events. *Cognition*, *232*, 105343.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214–1221.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature neuroscience*, *18*(4), 590–596.

Bruckner, R., Nassar, M. R., Li, S.-C., & Eppinger, B. (2025). Differences in learning across the lifespan emerge via resource-rational computations. *Psychological Review*.

Brunel, N. (2003). Dynamics and plasticity of stimulus-selective persistent activity in cortical network models. *Cerebral Cortex*, *13*(11), 1151–1161.

Calabresi, P., Picconi, B., Tozzi, A., & Di Filippo, M. (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends in neurosciences*, *30*(5), 211–219.

Calderon, C. B., Verguts, T., & Frank, M. J. (2022). Thunderstruck: The acdc model of flexible sequences and rhythms in recurrent neural circuits. *PLoS computational biology*, *18*(2), e1009854.

Charpier, S., & Deniau, J. (1997). In vivo activity-dependent plasticity at cortico-striatal connections: Evidence for physiological long-term potentiation. *Proceedings of the National Academy of Sciences*, *94*(13), 7036–7040.

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

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological review*, *120*(1), 190.

Collins, A. G., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS biology*, *10*(3), e1001293.

D'Acremont, M., & Bossaerts, P. (2016). Neural mechanisms behind identification of leptokurtic noise and adaptive behavioral response. *Cerebral Cortex*, *26*(4), 1818–1830.

Diederen, K. M., Spencer, T., Vestergaard, M. D., Fletcher, P. C., & Schultz, W. (2016). Adaptive prediction error coding in the human midbrain and striatum facilitates behavioral adaptation and learning efficiency. *Neuron*, *90*(5), 1127–1138.

Diederen, K. M., Ziauddeen, H., Vestergaard, M. D., Spencer, T., Schultz, W., & Fletcher, P. C. (2017). Dopamine modulates adaptive prediction error coding in the human midbrain and striatum. *Journal of Neuroscience*, *37*(7), 1708–1720.

Donahue, C. H., & Lee, D. (2015). Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nature neuroscience*, *18*(2), 295–301.

Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. *Neuron*, *109*(19), 3055–3068.

Eckstein, M. K., & Collins, A. G. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, *117*(47), 29381–29389.

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*(6), 1057–1070.

Fearnhead, P., & Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(4), 589–605.

Fino, E., Glowinski, J., & Venance, L. (2005). Bidirectional activity-dependent plasticity at corticostriatal synapses. *Journal of Neuroscience*, *25*(49), 11279–11287.

Fischer, A. G., Jocham, G., & Ullsperger, M. (2015). Dual serotonergic signals: A key to understanding paradoxical effects? *Trends in cognitive sciences*, *19*(1), 21–26.

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral cortex*, *22*(3), 509–526.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, *113*. https://doi.org/10.1037/0033-295X.113.2.300

Funahashi, S., & Kubota, K. (1994). Working memory and prefrontal cortex. *Neuroscience research*, *21*(1), 1–11.

Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centers. *Journal of Comparative Neurology*, *492*(2), 145–177.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current opinion in neurobiology*, *20*(2), 251–256.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, *84*, 401–410.

Haarsma, J., Fletcher, P., Griffin, J., Taverne, H., Ziauddeen, H., Spencer, T., Miller, C., Katthagen, T., Goodyer, I., Diederen, K., et al. (2021). Precision weighting of cortical unsigned prediction error

signals benefits learning, is mediated by dopamine, and is impaired in psychosis. *Molecular psychiatry*, *26*(9), 5320–5333.

Halassa, M. M., & Sherman, S. M. (2019). Thalamocortical circuit motifs: A general framework. *Neuron*, *103*(5), 762–770.

Heilbron, M., & Meyniel, F. (2019). Confidence resets reveal hierarchical adaptive learning in humans. *PLoS computational biology*, *15*(4), e1006972.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554–2558.

Karlsson, M. P., Tervo, D. G., & Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*, *338*(6103), 135–139.

Kreitzer, A. C., & Malenka, R. C. (2008). Striatal plasticity and basal ganglia circuit function. *Neuron*, *60*(4), 543–554.

Lam, N. H., Mukherjee, A., Wimmer, R. D., Nassar, M. R., Chen, Z. S., & Halassa, M. M. (2024). Prefrontal transthalamic uncertainty processing drives flexible switching. *Nature*, 1–10.

Li, Y. S., Nassar, M. R., Kable, J. W., & Gold, J. I. (2019). Individual neurons in the cingulate cortex encode action monitoring, not selection, during adaptive decision-making. *Journal of Neuroscience*, *39*(34), 6668–6683.

Litwin-Kumar, A., & Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature communications*, *5*(1), 5319.

Lloyd, K., & Leslie, D. S. (2013). Context-dependent decision-making: A simple bayesian model. *Journal of The Royal Society Interface*, *10*(82), 20130069.

Maes, A., Barahona, M., & Clopath, C. (2020). Learning spatiotemporal signals using a recurrent spiking network that discretizes time. *PLoS computational biology*, *16*(1), e1007606.

Mah, A., Golden, C. E., & Constantinople, C. M. (2024). Dopamine transients encode reward prediction errors independent of learning rates. *Cell reports*, *43*(10).

Mazzucato, L. (2022). Neural mechanisms underlying the temporal organization of naturalistic animal behavior. *Elife*, *11*, e76577.

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881.

Mukherjee, A., & Halassa, M. M. (2024). The associative thalamus: A switchboard for cortical operations and a promising target for schizophrenia. *The Neuroscientist*, *30*(1), 132–147.

Muller, T. H., Mars, R. B., Behrens, T. E., & O'Reilly, J. X. (2019). Control of entropy in neural models of environmental state. *elife*, *8*, e39404.

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related eeg signals and learning. *elife*, *8*, e46975.

Nassar, M. R., Bruckner, R., Gold, J. I., Li, S.-C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nature communications*, *7*(1), 11609.

Nassar, M. R., & Gold, J. I. (2013). A healthy fear of the unknown: Perspectives on the interpretation of parameter fits from computational models in neuroscience. *PLoS Computational Biology*, *9*(4), e1003015.

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable forms of uncertainty-driven representational change across the human brain. *Journal of Neuroscience*, *39*(9), 1688–1698.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, *15*(7), 1040–1046.

Nassar, M. R., & Troiani, V. (2021). The stability flexibility tradeoff and the dark side of detail. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 607–623.

Nassar, M. R., Waltz, J. A., Albrecht, M. A., Gold, J. M., & Frank, M. J. (2021). All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain*, *144*(3), 1013–1029.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378.

O'reilly, J. X. (2013). Making predictions in a changing world—inference, uncertainty, and learning. *Frontiers in neuroscience*, *7*, 105.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural computation*, *18*(2), 283–328.

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature communications*, *12*(1), 6587.
Piray, P., & Daw, N. D. (2024). Computational processes of simultaneous learning of stochasticity and volatility in humans. *Nature Communications*, *15*(1), 9073.
Powell, N. J., & Redish, A. D. (2016). Representational changes of latent strategies in rat medial prefrontal cortex precede changes in behaviour. *Nature communications*, *7*(1), 12830.
Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current biology*, *23*(17), R764–R773.
Razmi, N., & Nassar, M. R. (2022). Adaptive learning through temporal dynamics of state representation. *Journal of Neuroscience*, *42*(12), 2524–2538.
Recanatesi, S., Pereira-Obilinovic, U., Murakami, M., Mainen, Z., & Mazzucato, L. (2022). Metastable attractors explain the variable timing of stable behavioral action sequences. *Neuron*, *110*(1), 139–153.
Remington, E. D., Egger, S. W., Narain, D., Wang, J., & Jazayeri, M. (2018). A dynamical systems perspective on flexible motor timing. *Trends in cognitive sciences*, *22*(10), 938–952.
Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural networks*, *15*(4-6), 507–521.
Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature neuroscience*, *21*(12), 1753–1763.
Rikhye, R. V., Wimmer, R. D., & Halassa, M. M. (2018). Toward an integrative theory of thalamic function. *Annual review of neuroscience*, *41*(1), 163–183.
Rugg, M., Fletcher, P., Frith, C., Frackowiak, R., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*(6), 2073–2083.
Russin, J., Pavlick, E., & Frank, M. J. (2024). Human curriculum effects emerge with in-context learning in neural networks.
Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, *91*(6), 1402–1412.
Scott, D. N., & Frank, M. J. (2023). Adaptive control of synaptic plasticity integrates micro-and macroscopic network function. *Neuropsychopharmacology*, *48*(1), 121–144.
Scott, D. N., Mukherjee, A., Nassar, M. R., & Halassa, M. M. (2024). Thalamocortical architectures for flexible cognition and efficient learning. *Trends in Cognitive Sciences*.
Stokes, M. G. (2015). 'activity-silent'working memory in prefrontal cortex: A dynamic coding framework. *Trends in cognitive sciences*, *19*(7), 394–405.
Vaidya, A. R., & Badre, D. (2022). Abstract task representations for inference and control. *Trends in cognitive sciences*, *26*(6), 484–498.
Vaidya, A. R., Jones, H. M., Castillo, J., & Badre, D. (2021). Neural representation of abstract task structure during generalization. *ELife*, *10*, e63226.
Wall, P. M., & Messier, C. (2001). The hippocampal formation—orbitomedial prefrontal cortex circuit in the attentional control of active memory. *Behavioural brain research*, *127*(1-2), 99–117.
Wang, J., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nature neuroscience*, *21*(1), 102–110.
Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural computation*, *22*(9), 2452–2476.
Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS computational biology*, *9*(7), e1003150.
Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279.
Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692.
Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience and Biobehavioral Reviews*, *128*, 270–281.