

RESEARCH ARTICLE

Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: An Active Inference model

Anna C. Sales^{1*}, Karl J. Friston², Matthew W. Jones¹, Anthony E. Pickering^{1,3}, Rosalyn J. Moran^{1,4}

1 School of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, United Kingdom, **2** Wellcome Trust Centre for Neuroimaging, UCL, London, United Kingdom, **3** Anaesthesia, Pain and Critical Care Sciences, Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom, **4** Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom

* anna.sales@bristol.ac.uk



OPEN ACCESS

Citation: Sales AC, Friston KJ, Jones MW, Pickering AE, Moran RJ (2019) Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: An Active Inference model. *PLoS Comput Biol* 15(1): e1006267. <https://doi.org/10.1371/journal.pcbi.1006267>

Editor: Samuel J. Gershman, Harvard University, UNITED STATES

Received: June 4, 2018

Accepted: October 24, 2018

Published: January 4, 2019

Copyright: © 2019 Sales et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data was generated using MATLAB code which is freely available via the public GitHub repository <https://github.com/AnnaCSales/ActiveInference>. This code is a modified version of general Active Inference implementations which are also freely available at the public repository: <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>

Funding: ACS is supported by the Wellcome Trust PhD programme in Neural Dynamics: from synapses to systems in health and disease at the

Abstract

The locus coeruleus (LC) in the pons is the major source of noradrenaline (NA) in the brain. Two modes of LC firing have been associated with distinct cognitive states: changes in tonic rates of firing are correlated with global levels of arousal and behavioural flexibility, whilst phasic LC responses are evoked by salient stimuli. Here, we unify these two modes of firing by modelling the response of the LC as a correlate of a prediction error when inferring states for action planning under Active Inference (AI). We simulate a classic Go/No-go reward learning task and a three-arm ‘explore/exploit’ task and show that, if LC activity is considered to reflect the magnitude of high level ‘state-action’ prediction errors, then both tonic and phasic modes of firing are emergent features of belief updating. We also demonstrate that when contingencies change, AI agents can update their internal models more quickly by feeding back this state-action prediction error—reflected in LC firing and noradrenaline release—to optimise learning rate, enabling large adjustments over short timescales. We propose that such prediction errors are mediated by cortico-LC connections, whilst ascending input from LC to cortex modulates belief updating in anterior cingulate cortex (ACC). In short, we characterise the LC/ NA system within a general theory of brain function. In doing so, we show that contrasting, behaviour-dependent firing patterns are an emergent property of the LC that translates state-action prediction errors into an optimal balance between plasticity and stability.

Author summary

The brain uses sensory information to build internal models and make predictions about the world. When errors of prediction occur, models must be updated to ensure desired outcomes are still achieved. Neuromodulator chemicals provide a possible pathway for triggering such changes in brain state. One such neuromodulator, noradrenaline,

University of Bristol (ref. 108899/Z/15/Z). KJF is funded by a Wellcome Trust Principal Research Fellowship (ref: 088130/Z/09/Z). AEP is funded by a Wellcome Trust Senior Clinical Fellowship (ref. 088373/Z/09/A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

originates predominantly from a cluster of neurons in the brainstem—the locus coeruleus (LC)—and plays a key role in behaviour, for instance, in determining the balance between exploiting or exploring the environment. Here we use Active Inference (AI), a mathematical model of perception and action, to formally describe LC function. We propose that LC activity is triggered by errors in prediction and that the subsequent release of noradrenaline alters the rate of learning about the environment. Biologically, this describes an LC-cortex feedback loop promoting behavioural flexibility in times of uncertainty. We model LC output as a simulated animal performs two tasks known to elicit archetypal responses. We find that experimentally observed ‘phasic’ and ‘tonic’ patterns of LC activity emerge naturally, and that modulation of learning rates improves task performance. This provides a simple, unified computational account of noradrenergic computational function within a general model of behaviour.

Introduction

The locus coeruleus (LC) is the major source of noradrenaline (NA) in the brain, projecting to most territories from the frontal cortex to the distal spinal cord. Changes in LC firing have been associated with behavioural changes, most notably the switch from ‘exploiting’ to ‘exploring’ the environment, and the facilitation of appropriate responses to salient stimuli [1,2].

Tonic LC activity is correlated with global levels of arousal and behavioural flexibility, where firing rates increase with rising levels of alertness [1]. At the extreme, high rates of tonic firing have been causally related to behavioural variability and stochastic decision making [3]. This ‘tonic mode’ has previously been modelled as a response to factors such as declining utility in a task [4] or ‘unexpected uncertainties’ [5], triggering behavioural variability and a switch from ‘exploiting’ a known resource to ‘exploring’ for a new resource.

The LC also fires in short, high frequency bursts. Such phasic activity occurs in animals in response to behaviourally relevant salient stimuli [1,6–8]. This phasic response has been described as a ‘network interrupt’ or ‘reset’, which facilitates a shift to shorter-term behavioural planning [9,10]. Activating stimuli are those which have an established behavioural significance; for instance, signalling the location of food or the presence of a predator. They may also include stimuli that are highly unexpected [1,11]—although the phasic response will habituate rapidly to novelty alone in the absence of behavioural salience [12].

A series of studies has provided evidence of further nuance to phasic LC responses. Similar to the well-known dopaminergic response, as an animal learns a cue-reward relationship, phasic LC responses will transfer from temporal alignment with an unconditioned stimuli (US) to a predictive, conditioned stimuli (CS+)[13]. Additionally, rarer stimuli, or those predicting a large reward, elicit a stronger LC response [6,8]. In contrast if predictive cues are delivered consecutively, the size of the response appears to decrease [6]. The rich array of factors affecting the nature of the phasic response suggests that LC activation is linked to both facilitation of behavioural response and to internal representations of uncertainties and probabilities.

Despite the increasing body of knowledge about the impact of the LC on behaviour, a comprehensive computational account remains elusive—in contrast to the more developed account of other neuromodulators; most notably dopamine, which has been interpreted as a signal of reward prediction error. In particular, existing modelling approaches have generally tackled the tonic and phasic firing responses of the LC as separate modes with distinct functional significance, triggered by different circumstances [4,5,9,10].

Here, we propose that a critical computational role of the LC-NA system is to react to high level ‘state-action’ prediction errors upstream of the LC and cause appropriate flexibility in belief updating via feedback projections to cortex. In brief, our account of noradrenergic activity is based on the fact that the degree of belief updating reflects volatility in the environment and can therefore inform the optimal rate of evidence accumulation and plasticity. The ‘state-action’ prediction error considered in this work is the ‘Bayesian surprise’ or change in probabilistic beliefs before and after observing some outcome. We develop these ideas as neural correlates of discrete updates and action planning under the formalism of Active Inference (AI). AI offers an effective mathematical framework for such modelling, unifying inferences on states and action planning and providing a detailed description of beliefs at each step of a behavioural task [14–17]. In taking this formal approach, our description of the LC is integrated into a general theory of the brain function and uses constructs that underwrite the normal cycle of perceptual inference and action selection. This contrasts with previous LC modelling approaches, which have invoked the separate monitoring of statistical quantities (such as unexpected uncertainty) outside of the action selection cycle [4,5].

In the following we apply AI to simulate the updating of beliefs about states of the world—and actions—as a synthetic agent engages with two scenarios (a Go/No-go task with reversal and an ‘explore/exploit’ task) that elicit archetypal LC responses. Using this approach, we show that the ‘state-action prediction error’ offers an effective predictor of LC firing over both long (tonic) and short (phasic) timescales, without the need to invoke switches between distinct modes. Furthermore, we described how the signal may be broadcast back to cortex to affect appropriate updates to internal models of the environment. This links the error via the LC to model flexibility—bringing two key concepts of the LC together: ‘explore-exploit’ and ‘network reset’. It also produces behavioural changes that agree with experimental knowledge of animal behaviours under noradrenergic manipulation. Finally, the simulations produce realistic LC firing patterns that could, in principle, be used to model empirical responses.

Models

Brief overview of Active Inference

Active Inference is a theory of behaviour that has previously been mapped to putative neural implementations [14]. The basic premise of AI is that to stay in states compatible with survival, an agent must create and update a generative model of the world [14,18,19]. To do this effectively the agent represents the true structure of the world with an internal model that is a good approximation of how its sensations are generated. (Note that in this paper, we often use the term ‘model’ to refer to the agent’s beliefs about states and actions in the world. Technically, these beliefs are posterior probability distributions, which require a generative model to exist).

The generative model encompasses a set of discrete states and transition patterns that probabilistically capture all the agent’s beliefs about the world and likely outcomes under different actions. The model is formulated as a Partially Observable Markov Decision Process (POMDP), under which the agent must infer its current state, make predictions about the outcome of actions in the future and make postdictions about the landscape it has just traversed. In this context the word ‘state’ refers to a combination of features relevant to the agent, including its location and the cognitive context of that location; i.e., states of the world that matter for its behaviour.

To optimise this model, the agent constantly seeks to minimise variational free energy. This free energy is a mathematical proxy for the difference between the agent’s generative model and a ‘perfect’ or ‘true’ model of the world, and thus must be continually updated for the agent

to survive. Estimates of the free energy can be obtained over time by comparing predictions from the generative model with the results of actions in the real world, for instance, by checking whether an action produces the expected sensory feedback. Using this information from the real world, the agent minimises free energy in two ways: by adjusting the parameters of the generative model itself, and by picking actions that it believes will be associated with the lowest free energy. This allows the agent to both optimise the model and change its action plans. Updating proceeds in cycles, with each round of model updates accompanied by predictions that are then checked by selecting and executing an action—in turn allowing a new round of updates (Fig 1).

This framework means that each round of updates combines perceptual inference with action selection. Mathematically, this takes the form of a series of iterative updates to parameters that are repeated until convergence. Specifically, each new observation from the environment enables posterior beliefs about states to be updated via iterating expressions that minimise free energy. Similar iterative updates are then applied to posterior beliefs about competing policies and precision parameters (step 4 of Fig 2). Finally, the updated beliefs are used to select an action which in turn generates a new observation (step 5 of Fig 2). It is this machinery that we will map to LC/NA firing and function. A derivation of the Active Inference framework is provided in Appendix 1; S1 Fig shows hierarchical dependencies within the model. Appendix 2 also gives an overview of the implementation in code.

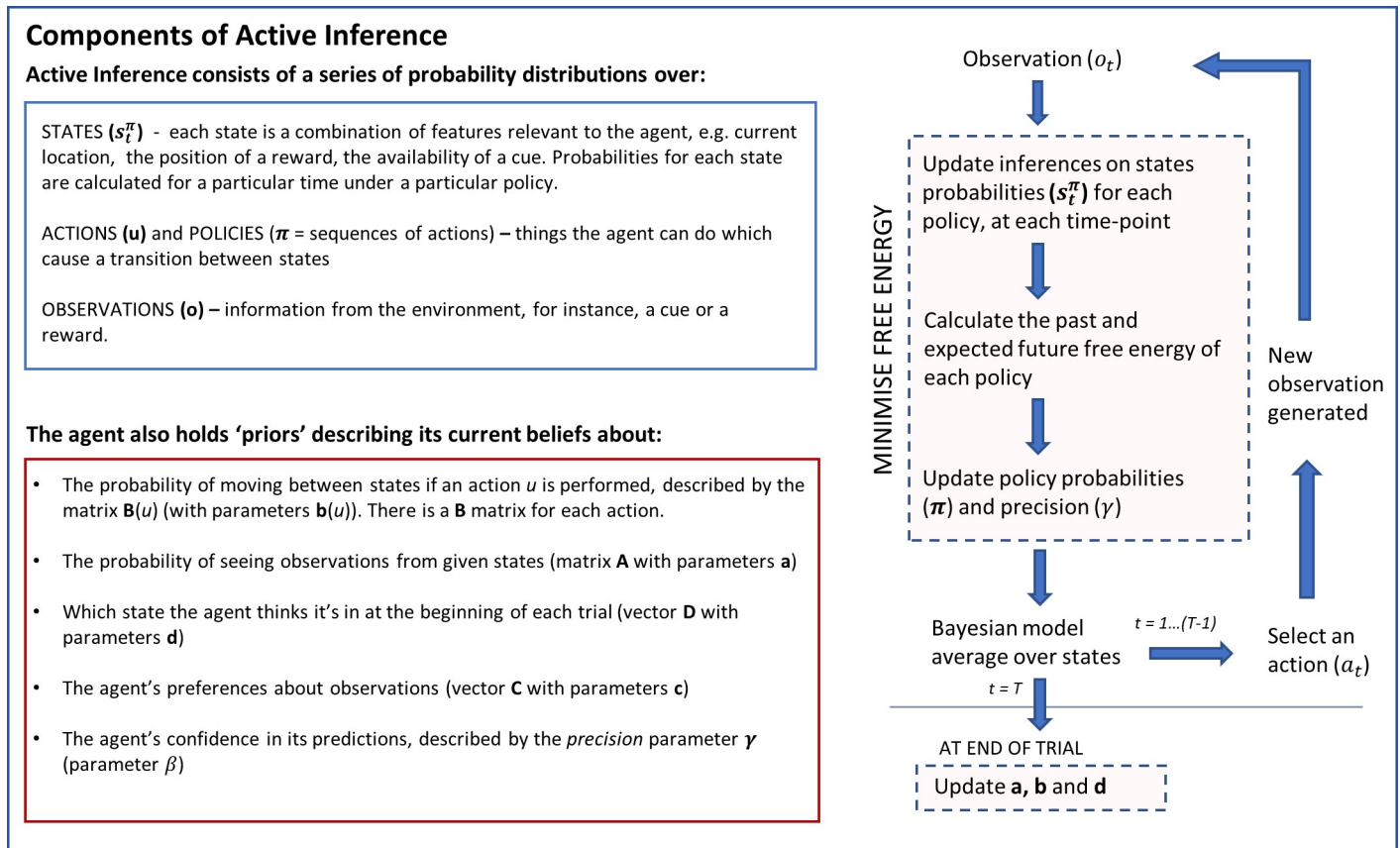


Fig 1. A quasi-mathematical description of the framework of Active Inference (based on [14]). The flow chart on the right shows the sequence of updates occurring over a single trial consisting of time-steps $t = 1 \dots T$.

<https://doi.org/10.1371/journal.pcbi.1006267.g001>

1. 'True' posterior distribution from the agent's generative model:

The probability of a particular combination of states, outcomes, actions and model parameters, defined as the product of all of the individual component probabilities.

For the POMDP over action-dependent time:

$$P(\tilde{o}, \tilde{s}, \pi, \alpha, b, d, \beta) = P(\pi)P(\alpha)P(b)P(d)P(\beta) \prod_{t=1}^T P(o_t | s_t)P(s_t | s_{t-1}, \pi)$$

2. The agent's approximate posterior:

Probabilities for states, actions and model parameters which are optimised by the agent via minimising free energy.

$$Q = Q(s_1 | \pi) \dots Q(s_T | \pi) Q(\pi) Q(A) Q(B) Q(D) Q(\gamma)$$

3. Expressions for free energy (FE) :

Minimising free energy is a proxy for reducing the KL divergence between P and Q, (or equivalently, maximising model evidence $\ln P(\tilde{o} | x)$). Here, x represents all parameters and variables in the model.

$$FE = D_{KL}[Q(x) \parallel P(x|\tilde{o})] - \ln P(\tilde{o} | x)$$

$$Q(x) = \arg \min_{Q(x)} FE$$

The free energy is calculated for the entire posterior (right) and used to optimise all parameters.

Past and estimated future free energies are also required for each policy individually:

Beliefs about the past free energy of policies:

$$F_{\pi} = \sum_{\tau} F(\pi, \tau) \quad F(\pi, \tau) = D_{KL}[Q(s_{\tau} | \pi) \parallel P(s_{\tau} | s_{\tau-1}, \pi)] - E_Q[\ln P(o_{\tau} | s_{\tau})]$$

Beliefs about the future free energy of policies:

$$G_{\pi} = \sum_{\tau} G(\pi, \tau) \quad G(\pi, \tau) = D_{KL}[Q(o_{\tau} | \pi) \parallel P(o_{\tau})] + E_Q[H[P(o_{\tau} | s_{\tau})]]$$

4. To minimize free energy, carry out the following updates:

The expressions below are derived by differentiating the full expression for free energy with respect to each individual parameter/hyperparameter and setting the resulting expression to zero. This yields a set of update equations which can be iterated until convergence at each time within the trial.

Update state probabilities, using latest observed outcome (o) and current model of the environment (A and B):

$$s_{\tau}^{\pi} = \sigma(\hat{A} \cdot o_{\tau} + \hat{B}'_{\tau-1} s_{\tau-1}^{\pi} + \hat{B}'_{\tau} \pi s_{\tau+1}^{\pi})$$

Where B'_{τ} is the matrix for the action prescribed by policy π at time τ , and σ represents the softmax function. F_{π} and G_{π} are also updated alongside s_{τ}^{π} (expressions not shown, see Appendix 1 for full details).

Update policy probabilities based on their free energies in the past and future, so that policies with a lower free energy are more probable:

$$\pi = \sigma(-F - \gamma \cdot G)$$

Update estimate of precision based the change in policy estimates given new observations:

$$\beta = \beta + (\pi - \pi_0) \cdot G, \quad \pi_0 = \sigma(-\gamma \cdot G)$$

5. Bayesian Model Average (BMA) and action selection:

The BMA is an overall estimate of state probabilities in the past and future, encompassing all updated variables:

$$S_{\tau} = \sum_p \pi_p \cdot s_{\tau}^{\pi_p} \quad \text{Here, } \pi_p \text{ is the probability of policy } p, \text{ and } s_{\tau}^{\pi_p} \text{ are the state probabilities under policy } p \text{ at time } \tau.$$

The BMA provides the roadmap for action selection: at each time-step t an action is selected which the agent believes will bring about the observations associated with S_{t+1} :

$$u = D_{KL}(AS_{t+1} \parallel AB(u)S_t)$$

An action generates a new observation, which is used in the next round of updates.

6. Update a, b and d

At the end of each trial, updates are applied to a , b and d parameters, using the history of state estimations and observations seen during the trial. For d - the agent's belief about the state it occupies at the start of each trial - this is

$$d = d + S_1$$

Where s_1 is the final BMA for the state occupied at $t = 1$, d is the updated value and d is the old value. Updates for a and b are:

$$a = a + \sum_t o_t \otimes S_t \quad b(u) = b(u) + \sum_n \pi_n \cdot s_{\tau}^n \otimes s_{\tau-1}^n \quad \text{For full derivation and further details, see Appendix 1}$$

Fig 2. Mathematical outline of the framework of Active Inference.

<https://doi.org/10.1371/journal.pcbi.1006267.g002>

There are two more subtleties that should be noted in this brief description. Firstly, the requirement to minimise free energy in action selection means that actions are driven by twin goals—the future attainment of states that the agent holds valuable (utility), as well as the attainment of information when performing an action (epistemic value). Formally, these describe the path integral of free energy expected under competing policies (see [14] and Appendix 1). Thus, agents that act to minimise free energy will end up where they hoped to, while resolving uncertainty about their environment. If policies do not differ in their ability to resolve uncertainty (i.e. no policy will harvest more information) then utility will drive policy selection. It has already been established that this particular cost function explores and exploits in a predictable and mathematically well-defined manner, depending on the relative utility of outcomes and on the uncertainty with which the agent views its environment [15–17,20].

The second important component is the timespan covered by inferences. The agent continually updates its understanding of the past, the present and the future. This means that observations in the present can be used to update inferences on states that occurred in the past—in this way, past events continue to be useful for belief updating long after they occurred. This is just a formalisation of our ability to postdict (e.g., “I started in this context, even if I didn’t know at the time”). Equally, the agent’s knowledge of the world is used to form predictions at future times (e.g., “These are the outcomes I expect under this policy”). The agent not only attempts to use events that have already happened to minimise free energy, but also tries to select actions and inferences which it believes will minimise free energy of future observations.

A Bayesian model average drives action selection

As outlined in Figs 1 and 2, the generative model comprises probability distributions over states, sequences of actions, precision (confidence in predictions) and observations. At each time step, the agent updates its beliefs about these probability distributions over states, actions and precision by minimising free energy.

Once all updates have been completed the agent combines all of its inferences to produce a Bayesian Model Average (BMA) of states under possible actions. This can be considered as a summary of everything the agent knows about its place in the world—an overall ‘map’ of the states it believes it occupied in the past, the state it occupies now and the states it believes it will occupy in the future. The distribution implicitly includes action planning that is informed by inferences about events in the past. The Bayesian Model Average is then used by the agent to select an action, as described in Fig 2. The action causes a transition to a new state, which generates an observation from the environment.

State-action prediction errors as a driver of LC activity

Any large change in the state-action heatmap between time steps represents a *state-action prediction error*. These errors indicate that the agent’s beliefs about its past and future states have changed substantially after receiving a fresh observation. Such prediction errors indicate that the agent’s model of the world—including its plan for actions—must change. This may either be because an unexpected stimulus has occurred, requiring an abrupt change in behaviour, or because observations over longer timescales are consistently demonstrating that key components of the model (for example, the observation likelihood (*A*) and state transition (*B*) matrices) are no longer fit for purpose. Crucially, errors originating from both situations are reflected in the state-action prediction error. We propose that they are a driver of LC activity.

The BMA is estimated for each time point within the task (indexed by τ) and takes the form of a weighted sum over state probabilities (states are weighted by the probability of each policy predicting that state at the given time). To estimate the state-action prediction error during a

task, we take the Kullback-Leibler divergence between Bayesian Model Average (BMA) distributions at successive time steps. Mathematically, this reflects the degree of belief updating induced by each new observation. It is often known as a relative entropy, information gain or Bayesian surprise. The following expressions describe the BMA (upper equation) and the state-action prediction error (lower):

$$S_{\tau} = \sum_p \pi_p \cdot s_{\tau}^p$$

$$SAPE(t) = \sum_{\tau} D_{KL}[(S_{\tau}^t) || (S_{\tau}^{t-1})]$$

Here, S_{τ} is the BMA over states for time τ within the task, while s_{τ}^p is the vector of probabilities for states at time τ under policy p , which has a probability π_p . In the expression for state-action prediction error, superscripts (either t or $t - 1$) refer to the time at which the estimate is calculated.

Prediction errors over shorter timescales (i.e. between actions, during the iterative cycle of belief updating) are an integral feature of AI. The state-action prediction error, in contrast, represents a global error: it is expressed over the timescale of a behavioural epoch as a *response* to the outcome of belief updating that facilitates action selection.

In the implementation, the state-action prediction error is calculated immediately after the BMA over states, as shown in Fig 3.

LC feedback: Flexible model learning promoted by state-action prediction errors

Why might it be useful for the LC to respond to state-action prediction errors? We suggest that one important function is that such errors require a specific modulation of distributed cortical activity encoding representations of the structure of the environment, particularly in frontal cortex. This modulation would boost the flexibility of internal representations (where our matrices would be formed by connected cell assemblies in frontal cortex) and increase their responsiveness to recent observations. In vivo, this may be mediated by the release of noradrenaline from LC projections to the frontal cortex occurring in response to state-action prediction errors.

The need for flexible model updating is directly relevant to a related challenge for Active Inference models; namely, the rate at which the agent's experience is assimilated into its model. Addressing this issue provides a pathway for modelling the effect of LC activation and closes the feedback loop between brainstem and cortex. So what computational role does NA have in facilitating adaptive flexibility?

Under AI, the agent's model of the world is encoded by a set of probability distributions that keep track of the mappings between states and outcomes, and between states occupied at sequential time points. These mappings are encoded by Dirichlet distributions, the parameters of which are incremented with each instance of a particular mapping the agent experiences (as shown in step 6, Fig 2 and Appendix 1) [14,20]. However, difficulties arise when environmental contingencies change, because the gradual accumulation of concentration parameters is essentially unlimited. Accumulated experience can come to dominate the agent's model, with new information having little effect on the agent's decisions. This occurs because the generative model does not allow for fluctuations in probability transitions, i.e. environmental volatility. This issue can be finessed by adding a volatility or decay factor (α), which effectively endows the generative model with the capacity to 'forget' experiences in the past that are not relevant if environmental contingencies change.

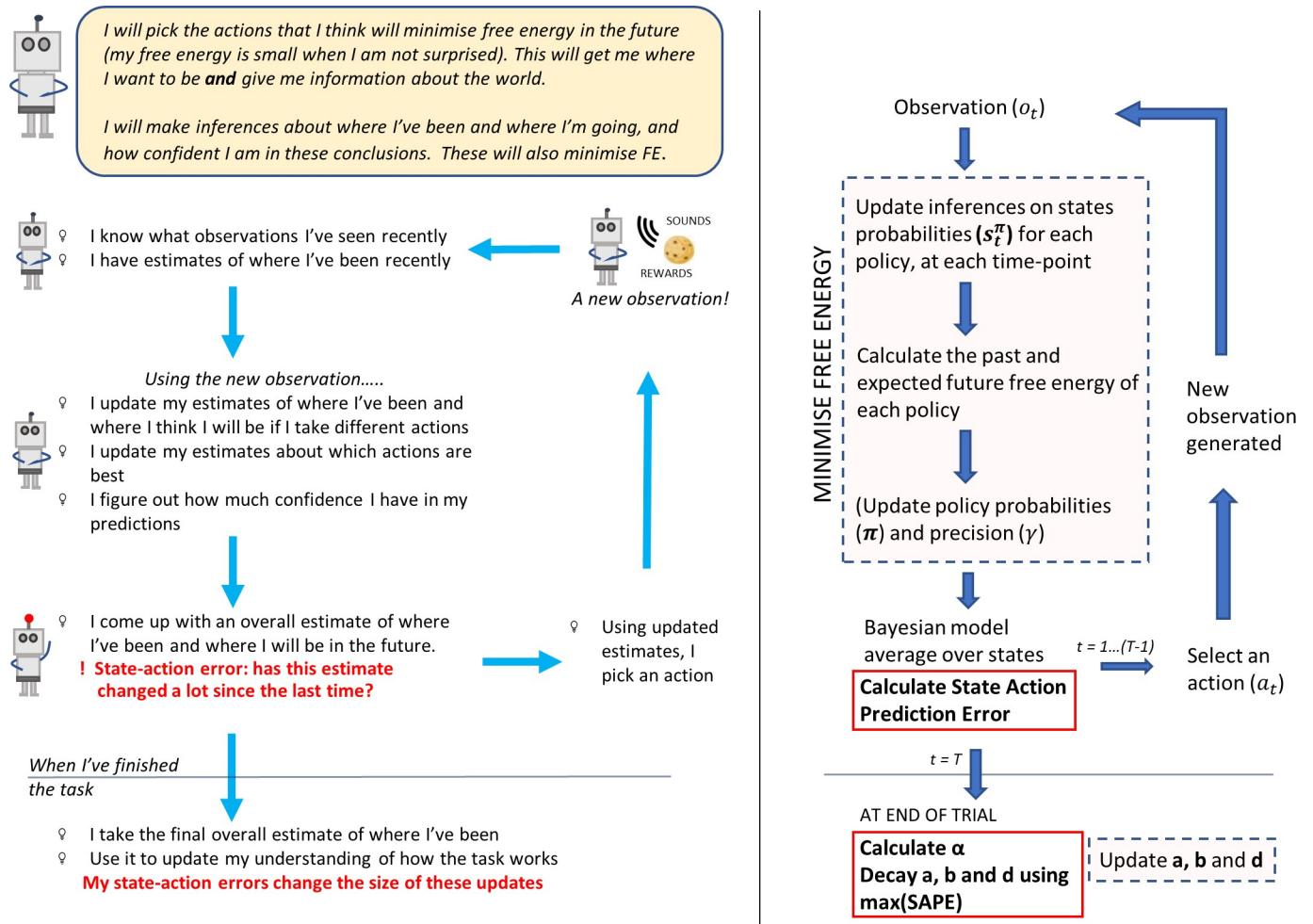


Fig 3. The cycle of updates under Active Inference, expanded to show the calculation of the state- action prediction error and the application of model decay. Left: non-mathematical 'cartoon' explanation of the cycle of updates. Right: more detailed update cycle, to be compared with the version shown in Fig 1.

<https://doi.org/10.1371/journal.pcbi.1006267.g003>

This introduces a modification to the update equations shown in Fig 2, of the form:

$$\text{old update : } \mathbf{d} = \mathbf{d} + S_1$$

$$\text{new update : } \mathbf{d} = \mathbf{d} + S_1 - \frac{(\mathbf{d} - 1)}{\alpha}$$

Where \mathbf{d} and d are the updated and existing beliefs respectively. In the original update, the \mathbf{d} vector (which describes the agent's prior about its state at $t = 1$) is simply incremented by adding the agent's beliefs about the state it occupied at $t = 1$. This update is then applied after each trial. In the new version, the same increment occurs, but with a 'decay' of the values in \mathbf{d} that is controlled by α . The same modification is made to the updates for \mathbf{a} and \mathbf{b} (the updated forms are given explicitly in Appendix 1).

In the context of reversal learning, this is not a trivial adjustment but a crucial addition to the generative model which enables AI agents to adapt flexibly. However, the level at which to set the decay term poses a further challenge: if the decay is too big, the model is too flexible and will be dominated by its most recent experiences (as all the other terms will have decayed).

If the decay is too small concentration parameters may accumulate too slowly, rendering the model too stable.

There are several ways one can optimise this ‘forgetting’ in volatility models. One could equip the Markov decision process with a further hierarchical level modelling fluctuations from trial to trial—as in the hierarchical Gaussian filter [21]. A simpler (and biologically parsimonious) solution is to link the decay factor to recent values of state-action prediction error via the LC. In other words, equip the agent with the prior that if belief updating is greater than expected, environmental contingencies have become more volatile.

This produces flexibility in model learning when state-action prediction error is high (low α) but maintains model stability when state-action prediction error is low (high α). We have modelled this feedback using a simple logistic function to convert the error into a value for α :

$$\alpha = \alpha_{min} + \frac{\alpha_{max}}{1 + e^{k(SAPE-m)}}$$

where *SAPE* is the state-action prediction error seen during the trial (in tasks with more than one prediction error per trial, the maximum error is used), *k* is a gradient and *m* is a mean (i.e., expected) value. In all simulations presented below, $\alpha_{min} = 2$, $\alpha_{max} = 32$, $k = 8$, and *m* was set one standard deviation above the mean error value encountered in 100 trials of each task with $\alpha = 16$.

Under this scheme, a brief but large state-action prediction error ‘boosts’ the impact of a recent experience upon the agent’s model of the world. This occurs by temporarily increasing the attrition of existing, experience dependent parameters encoding environmental contingencies. Crucially, this causes recent actions and observations to have a greater effect on the Dirichlet distributions than they would otherwise. If errors then decrease, the model stabilises again. However, if actions consistently produce large state-action prediction errors then the underlying model parameters will gradually lose their structure—equivalent to the flattening of probability distributions that form the agent’s model—leading to greater variability in action selection. This ‘flat’ model does not need to track volatility separately: we instead incorporate LC/NA directly into the decision-making loop.

General MATLAB code implementing Active Inference can be found at <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/> (in the folder toolbox/DEM). Code from this toolbox was modified to perform the simulations described in this paper; examples are available at <https://github.com/AnnaCSales/ActiveInference>.

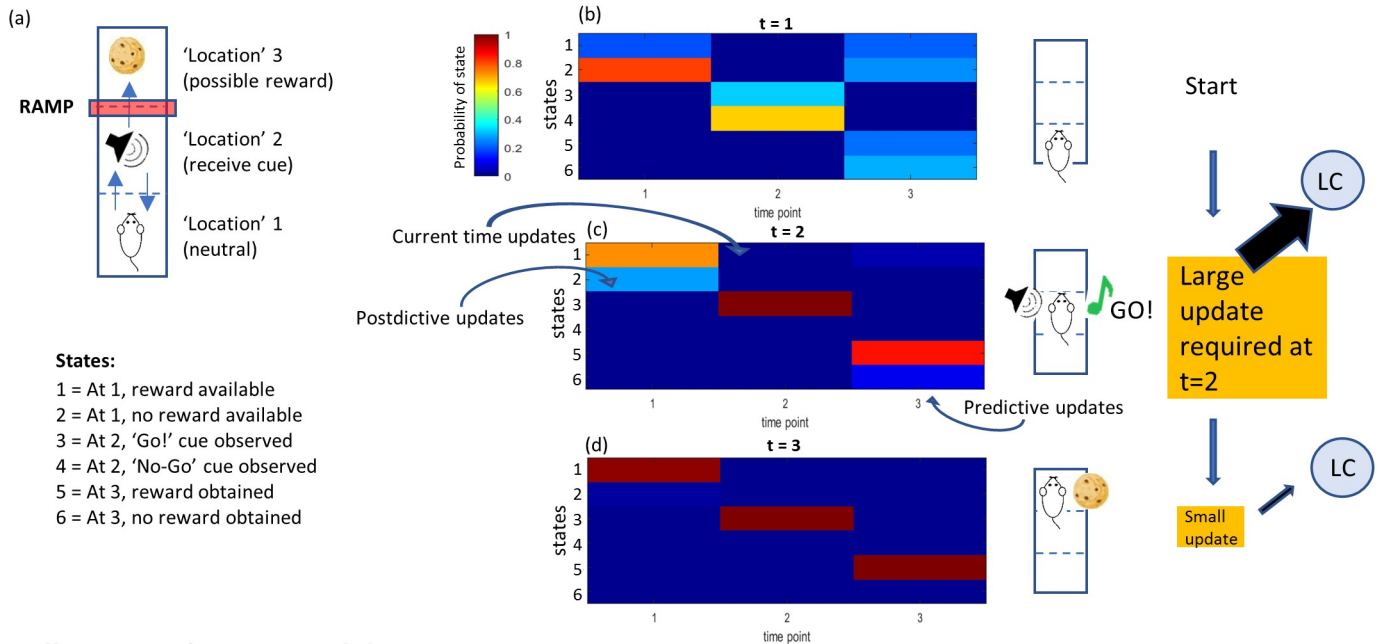
Results

The simulations reported in this paper suggest that behavioural contexts that produce large state-action prediction errors are also those that produce archetypal LC responses in experimental environments. Below, we describe the emergence of phasic and tonic activity in two tasks as a response to changes in state-action prediction error. We initially present results without the LC feedback (model decay) in place before showing how both simulations are improved by modelling the LC as a link between state-action prediction errors and model decay / volatility.

State-action prediction errors display tonic and phasic patterns of response

Go/no-go task. A simple go/no-go task modelled under AI is shown in Fig 4. In this task, the agent (depicted as a rat) starts in a ‘ready’ state—location 1—and must move to location 2 to receive a cue. When the cue is received the agent may either move back to location 1 or seek a reward at location 3. The agent has a strong preference for receiving the reward but an

Go-No-Go task



Full Active Inference model

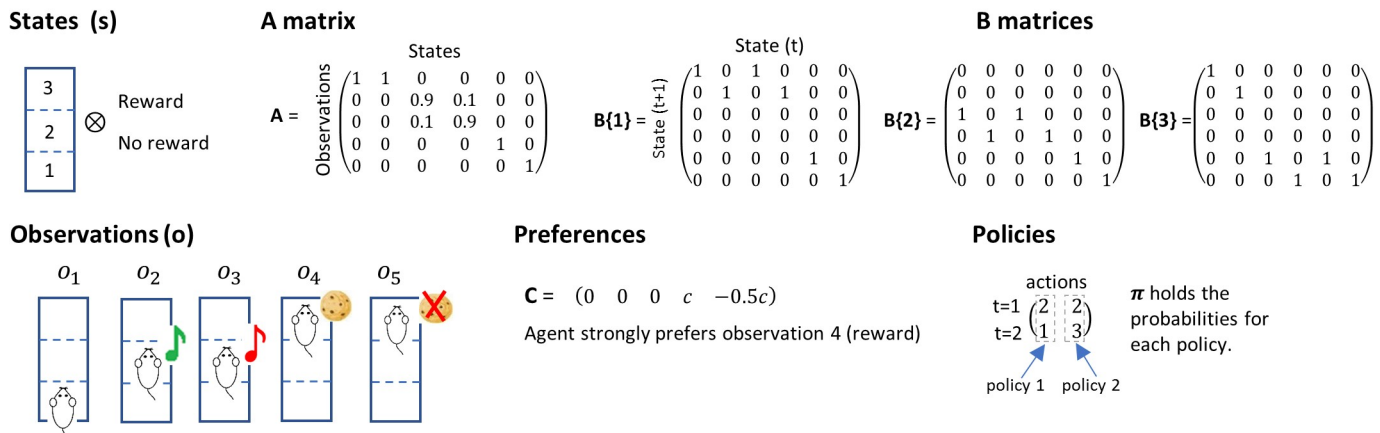


Fig 4. Simple go/no-go task modelled under AI. (a) Structure of the task (see main text) (b)-(d) The state-action heatmap showing inferences on the agent's state over a rare 'Go!' trial. Large updates are required at $t = 2$, after the animal receives the 'go' cue which forces it to update its action plans and state inferences. This update is proposed to cause a large, time specific input into LC, which causes a sudden phasic burst of LC activity. The lower part of the Fig shows the full modelling of the go/no-go task, with components as described in Fig 1.

<https://doi.org/10.1371/journal.pcbi.1006267.g004>

aversion to moving to location 3 and remaining unrewarded. This is represented in the task by a notional ramp which forces the agent to expend physical effort in seeking the reward. There are six available states, which between them describe the different combinations of features relevant to the agent during the task. Learning is mediated through updates to the **A** and **D** matrices, which encode likelihood mappings between hidden states of the world and outcomes—and prior beliefs about initial states.

At each time point, the agent's beliefs are summarised in the Bayesian Model Average, represented graphically as a state-action heatmap. The heatmap shows how the likelihood of different states evolves over time as evidence accumulates and beliefs are updated. Fig 4(B) shows

a representation of the agent's beliefs about states at the beginning of a new trial in which the 'go' cue is heard. The agent is 'well trained'; that is, it has an accurate understanding of the relationship between the cue and the availability of the reward, and of the fact that the 'go' cue is rare (here, the cue probability is 10%). In our modelling, we trained the synthetic rat by running the simulation for 750 trials. We then used the learnt priors as the starting point for the 'well trained' case.

Given its knowledge of the task, the agent begins with a strong belief that it is beginning the trial in state 2 (in which a reward will not be available). It also makes predictions for the states it believes it will occupy later in the trial: at $t = 2$, it believes it is likely to occupy state 4 –corresponding to the occurrence of the 'no-go' cue, but also entertains a slight possibility that the 'go' cue might still appear. The agent is much less certain in its predictions for $t = 3$, but still holds a higher probability that it will end up in one of the unrewarded end states.

At the next time point (at $t = 2$, Fig 4(C)), the agent updates its state-action heatmap, making new inferences on the probabilities of different states in the past, present and future, based on its most recent observations. If it has received the rare 'go' cue, it will have to update its predictions for its state at the end of the task, in addition to altering its inferences about the state in which it started at $t = 1$ (a process of postdiction about past states based on new information). The agent therefore has to make a large, sudden update to its BMA heatmap at $t = 2$. By $t = 3$ (Fig 4(D)), the agent has received the reward as predicted, and knows with certainty where it is and where it has been. Only small updates are required to its estimates at this point.

Simulated state-action prediction errors during this task are shown in Fig 5. In this simulation the state-action prediction error does not modulate learning and the decay parameter α has been set to a fixed value ($\alpha = 16$). During the task, an agent who is well trained shows large peaks of state-action prediction error when the reward-predicting cue is presented, resulting in phasic activity in the LC as seen experimentally [6,22]. The underlying reason for this error is a large, quick shift in action planning, from the (more likely) 'No-go' outcome to the rare 'Go' situation.

'Explore / Exploit' task. We also modelled a task designed to offer the agent a choice between exploiting a known resource or exploring for new sources of reward (depicted in Fig 6). On every trial in this task the agent searches for a reward in one of three arms. In one arm, the probability of finding a reward is high (90%), whilst in the others the probability is low (10%). The probabilities are held constant for a set number of trials, during which time the agent accumulates beliefs about the likelihood of finding a reward in each location. Typically, once the agent has been rewarded in one location numerous times it will build a strong prior probability on the availability of a reward in that location (reflected in updates to elements of the **B** matrix). In the example shown in Fig 6 the agent begins by exploring the arms until it has seen a reward in arm 1, after which it continues to visit this location. After a set number of trials, the location of the high probability arm is shifted. When this happens, the agent's established model of the world no longer provides an accurate explanation of its experiences. As expected rewards fail to materialise, state-action prediction errors arise. Under our model, this causes an increasing tonic rate of LC activity whilst new priors are learnt and behaviour changes.

Flexibility in model learning improves task performance and enables reversal learning

We now turn to simulations in which the state-action prediction error is linked—via LC activity—to the model decay parameter. When this link is introduced there are improvements in performance in the simulations of both the Go/no-go and explore/exploit tasks (Figs 7 and 8).

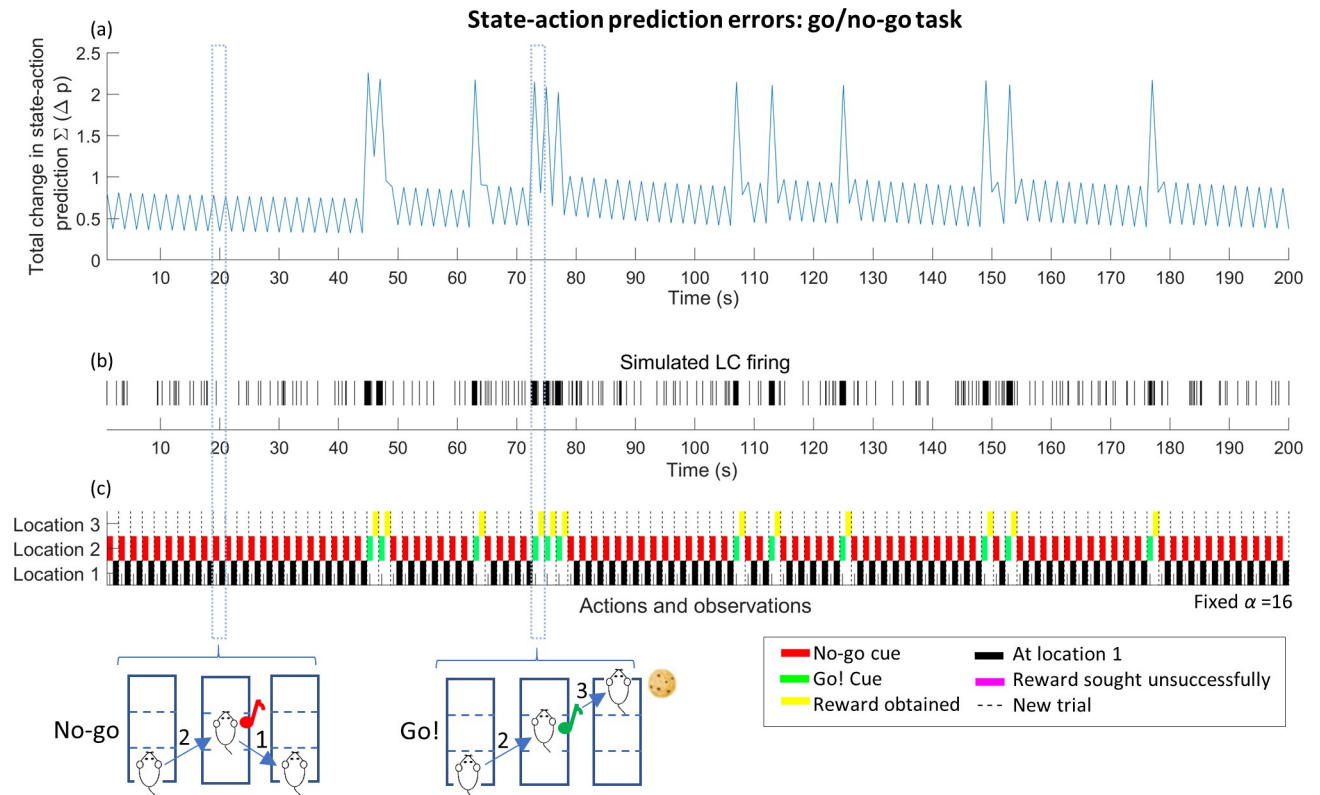


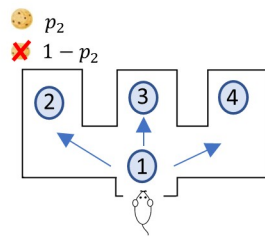
Fig 5. Plot of state-action prediction error, simulated LC spiking and behaviour during 100 trials of the go/no-go task, (for agents with a fixed value of model decay parameter α not linked to state-action prediction error). Each point within the task is assumed to last 1s and is associated with a single state-action prediction error. In (a) the raw prediction error is extracted for $t = 2$, when the animal receives a cue (this is the error between $t = 1$ and $t = 2$) and $t = 3$ when the animal receives feedback on its response to the cue (the error between $t = 2$ and $t = 3$). Because the prediction error explicitly evaluates differences between update cycles, there is no error available for the first time point. Each trial has therefore been collapsed to two time points, each lasting 1 second. In (a) the occurrence of the ‘go’ cue causes strong peaks in prediction error. This is converted into a simulated LC firing rate in (b). To visualise LC firing, a firing probability p is calculated for each second using the state-action prediction error (SAPE) as the input into a logistic function, so that $p = \frac{1}{1 + e^{-k(SAPE - m)}}$ where $k = 8$, and m is as above. Each second was then further split into 0.1s bins, during which the unit generated a single spike with probability p . This gives a physiologically reasonable [1,22,23] maximum firing rate of 10hz if $p = 1$. This is converted into a simulated LC firing rate in (b), showing phasic LC activation when the ‘go’ cue is heard. Plot (c) is a graphical representation of behaviour during the task at times $t = 2$ and $t = 3$ for each trial, in which the position of the coloured block describes the agent’s location and the colour shows the agent’s observation after moving.

<https://doi.org/10.1371/journal.pcbi.1006267.g005>

In the explore/exploit task, the dynamic modulation of model building allows state-action prediction errors to reduce more quickly when the rat is settled into the ‘exploit’ mode of harvesting a reward in a reliable location, promoting model stability. When the reward is no longer available, errors mount and the increase in model decay causes the agent to make more explorative choices. This contrasts with the same task simulated with fixed values of α (Fig 7 (A) and 7(B)): when the model is hyper-flexible ($\alpha = 2$), the agent often switches behavioural strategy after a single failed trial; when the model is inflexible, the agent takes a large number of trials to visit a new location ($\alpha = 2$). The highly flexible agent performs well when the structure of the environment is volatile (Fig 7C). In this context, even small errors may reasonably indicate that the underlying rules of the task have changed, and the agent’s rapid shifts in strategy yields rewards. Over multiple simulations the flexible agent obtains a significantly higher total reward than the inflexible agent (averages over 50 simulations of 150 trials each with rules as shown in Fig 7). Conversely, the inflexible agent obtains significantly more overall reward when the environment is more stable (5(d)). Fig 7(C) and 7(D) also show the performance of

Explore/exploit task

(a) Active Inference model



A matrix

$$A = \begin{matrix} & \text{States} \\ \text{Observations} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

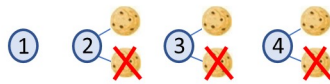
Policies

Actions
(1 2 3 4)

Preferences

$$U = (0 \quad c \quad -0.5c \quad c \quad -0.5c \quad c \quad -0.5c)$$

7 States and 7 observations



B matrix example

$$B\{2\} = \begin{matrix} & \text{State (t)} \\ \text{State (t+1)} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & p_2 & p_2 & p_2 & p_2 & p_2 & p_2 \\ 1-p_2 & 1-p_2 & 1-p_2 & 1-p_2 & 1-p_2 & 1-p_2 & 1-p_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

(b) Modelling of responses

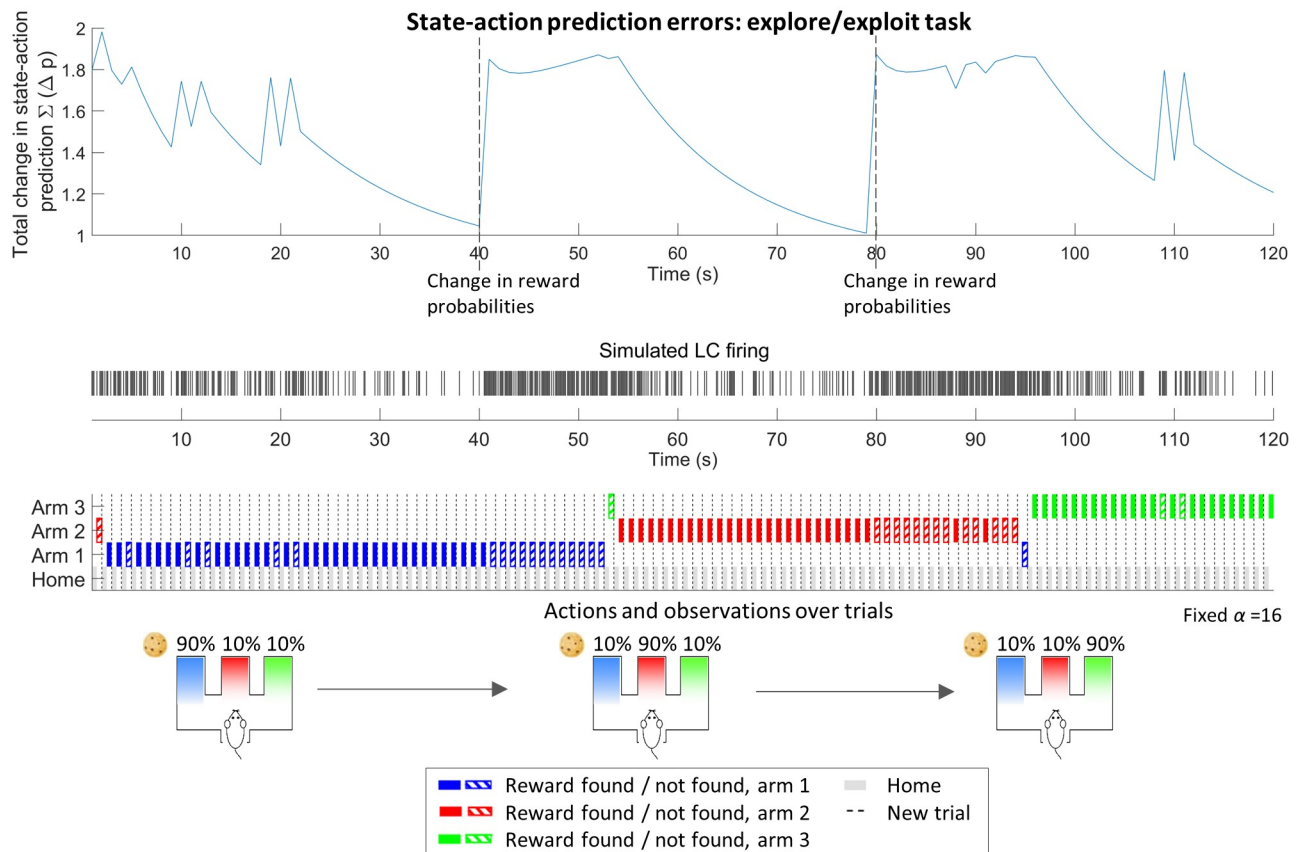
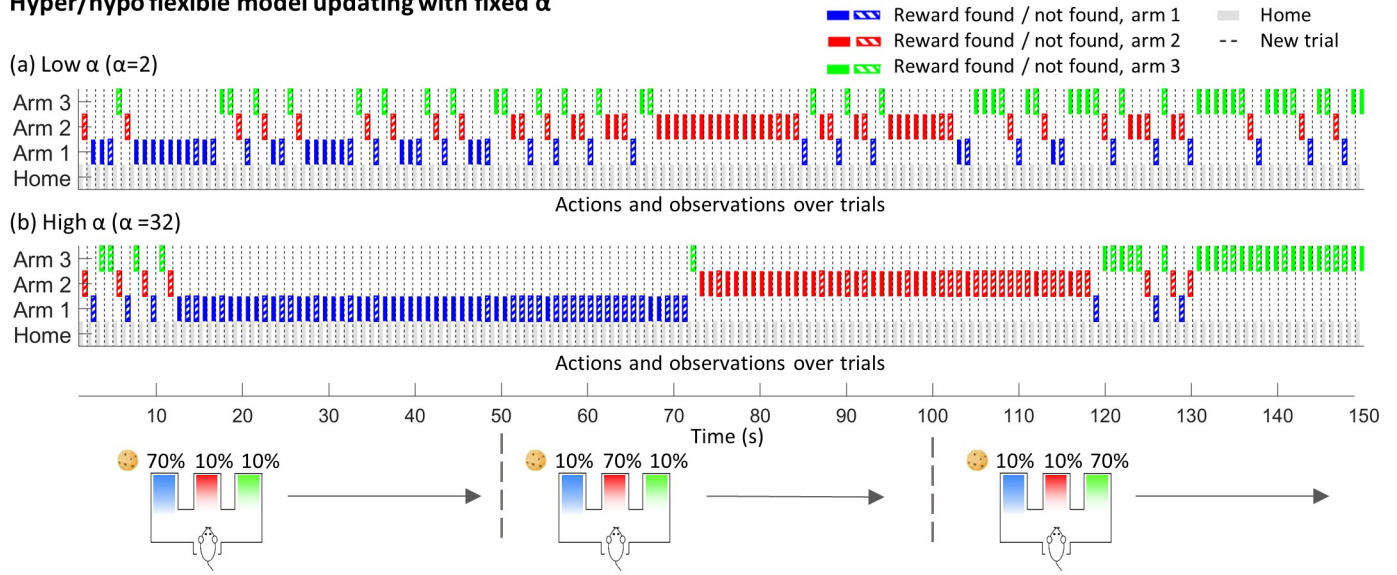


Fig 6. Modelling a 3-arm explore/exploit task under Active Inference. (a) shows the mathematical structure of the task. There are seven states, including one neutral starting point and 3 arm locations which can be combined with either a reward / no reward. There are 7 observations; here these have a 1-to-1 mapping to states (A matrix). Actions 1–4 simply move the agent to locations 1–4 respectively. The probability of obtaining a reward in a given arm (p_2 for action 2, above) is held static for a fixed number of trials, with one arm granting a reward with a 90% probability and the others with 10% probability. This is then switched, so that the agent must adjust its priors and its behaviour. (b) shows the state-action prediction errors and simulated LC responses over a typical run of 100 trials for an agent with a fixed ($\alpha = 16$) value of the model decay parameter.

<https://doi.org/10.1371/journal.pcbi.1006267.g006>

Hyper/hypo flexible model updating with fixed α



Results over multiple trials: explore/exploit task

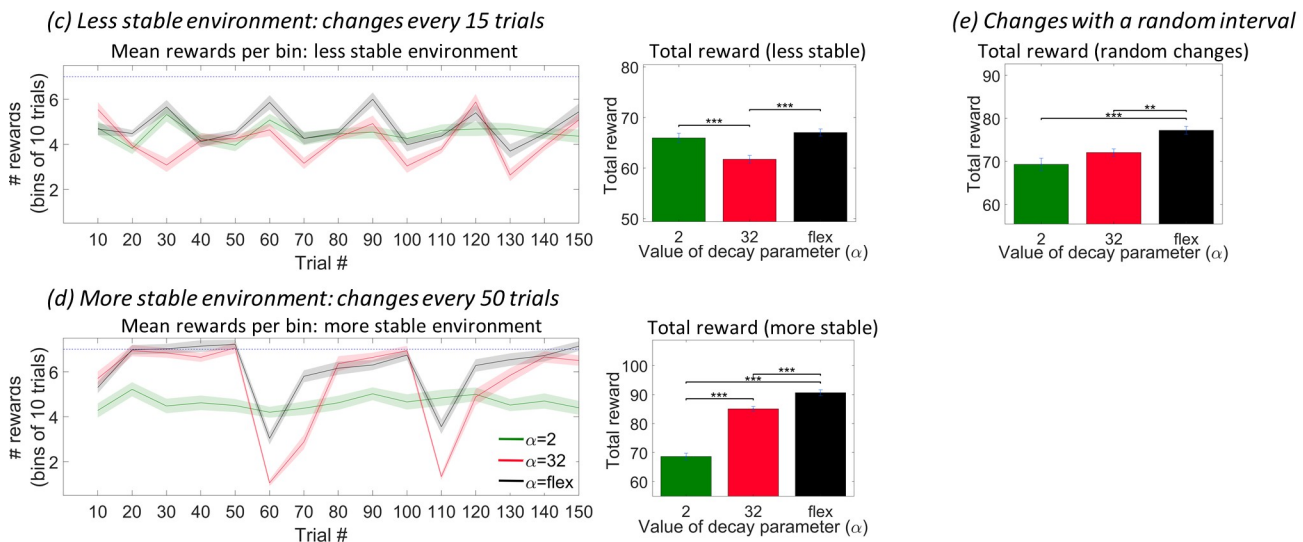


Fig 7. The explore/exploit task simulated with fixed and flexible values of model decay. (a) and (b) show the behavioural output from the explore/exploit task for agents with a fixed α parameter, specifically $\alpha = 32$ (slow model decay) or $\alpha = 2$ (fast model decay). The agent with $\alpha = 2$ is hyperflexible in its behaviour and changes its strategy after single failed trials. In contrast, the $\alpha = 32$ agent is inflexible and persists in seeking reward in the same location despite multiple failed trials. (c) and (d) show the outcome of simulations involving fixed α agents contrasted with the performance of an agent with a flexible value of α set by the state-action prediction error. Each simulation consisted of 150 trials in which the location of the high probability arm changed either every 15 or every 50 trials. The simulation was repeated 50 times. (c) and (d) show the average reward obtained in bins of 20 trials (shaded errors show standard error of the mean), alongside the mean total reward gained by each agent (error bars show S.E.M.; *** $P < 0.0001$, one way ANOVA followed by Tukey posthoc test). The less stable/more stable environments favour the $\alpha = 2$ / $\alpha = 32$ agent respectively; however, the flexible agent is able to perform as well (or better) in both scenarios. In (e) the location of the reward changes after random intervals, and the flexible agent clearly outperforms both of the fixed- α agents.

<https://doi.org/10.1371/journal.pcbi.1006267.g007>

an agent with a dynamic α with a value determined by the state-action prediction error, ranging between $\alpha = 2$ and $\alpha = 32$. This agent performs as well as, or better than, the fixed α agents in both contexts, responding with a rapid changes to its model (and resulting behaviour) when errors are high for a sustained period, but stabilising when errors decrease. This agent also outperforms both fixed α agents when the arm locations change after random intervals (Fig 7(D)).

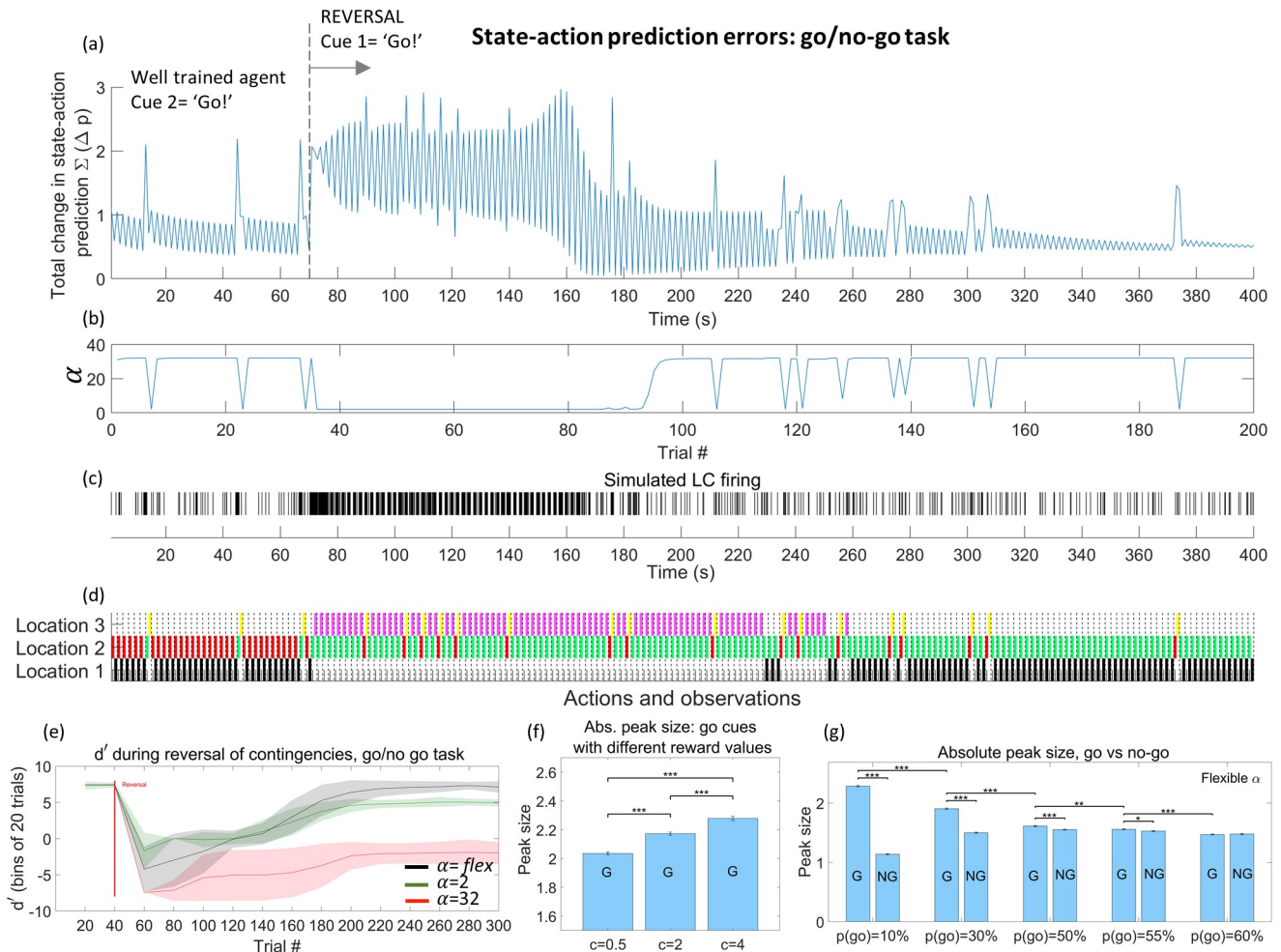


Fig 8. Reversal learning during the go/no-go task. (a)–(c) show the performance of an agent with a value of model decay determined by state-action prediction error during a reversal of cues in the go/no-go task. The agent begins with a well-trained understanding (via 750 trials of training) that cue 2 indicates that a reward is available. At trial 35 ($t = 70$) the cue/context relationship is reversed, and the agent must now learn that cue 1 indicates the ‘Go’ context. This initially causes numerous unsuccessful trials, violating the learnt model and producing high prediction errors (a). Note that prediction errors are initially elevated at both timepoints in each trial because both the previously rare cue and the subsequent lack of reward are unexpected. These prediction errors result in a lowering in the parameter decay factor (b), which in turn flattens the agent’s priors causing more variability in behaviour. Eventually the agent learns the new contingencies and the model stabilises, with the re-emergence of phasic bursts of LC activity on ‘Go’ trials (a, c). From trial 125 onwards, the peak of phasic activity begins to transition towards the presentation of the cue rather than the reward. Plot (d) is a graphical representation of behaviour during the task at times $t = 2$ and $t = 3$ for each trial, in which the position of the coloured block describes the agent’s location and the colour shows the agent’s observation after moving. (e) shows performance over 50 repeats of the reversal learning task shown in (a), for agents with a fixed or flexible value of α . All agents begin with a near optimal d' value (measured over bins of 20 trials). However, only the agent with α determined by the state-action prediction error is able to return to optimal levels of performance within the 300 trials shown. (f) and (g) show characteristics of the mean prediction error response to ‘go’ and ‘no-cue’ cues during the static (non-reversed) task as reward and probability parameters are varied, for agents with a flexible value of α (** $P < 0.0001$; one-way ANOVA between different c values, followed by Tukey posthoc test, (g) ** $P < 0.001$, *** $P < 0.0001$; two tailed Student’s t -test between go/no go contexts for fixed cue probabilities, one way ANOVA followed by Tukey posthoc test for ‘go’ peaks with different cue probabilities).

<https://doi.org/10.1371/journal.pcbi.1006267.g008>

A full plot of state-action prediction errors, simulated LC firing and behavioural output for the explore/exploit task with the flexible α agent is also shown in S2 Fig.

The Go/No-Go task was also simulated during a reversal of cue meanings (Fig 8). As expected, the well-trained agent begins the session by showing a phasic response in state-action prediction error / LC firing in response to the ‘Go’ cue (cue 1). At trial 35, the meaning of the two cues switches so that the ‘Go’ context is predicted by cue 2. At the reversal, state-

action prediction errors cannot be resolved and LC firing switches to a higher tonic level. During this period, model updating—and behaviour—becomes more flexible and the new rules of the task are learnt. Eventually the high levels of tonic activity fall away and phasic responses to the new ‘Go’ cue re-emerge; coupled with a lower level of tonic activity. This mirrors the pattern of LC firing recorded in monkeys during the same task [22]. Note that after the reversal, phasic responses emerge initially in response to the reward itself, then to both the cue and reward, and finally only in response to the cue. Over multiple trials of the reversal, only the agent with a flexible α linked to state-action prediction error is able to learn the new contingencies and return to optimum performance levels (Fig 8(E), for which the reversal was repeated 50 times).

The characteristics of the state-action prediction error for this agent were then examined in more detail (Fig 8(F) and 8(G)). 2000 trials were run of the go/no-go task in which the cue meanings were held constant (no reversal), and the agent started each trial with ‘well-trained’ priors obtained through 750 trials of training, as above. We find that the size of the state-action prediction error—the proposed input into the LC—changes in ways that are consistent with experimentally reported LC activation (see discussion below). As shown in Fig 8, the size of the phasic peak in state-action prediction error is larger for rarer ‘go’ stimuli (Fig 8(G), in which ‘G’ = go cues, ‘NG’ = no go cues). When the probability of the ‘go’ cue is held fixed, the resulting peak increases as the reward becomes more valuable to the agent (represented by the value held in the agent’s c matrix, Fig 8(F)). Finally, when consecutive ‘go’ trials occur, the second peak is reduced in size (mean reduction of $12.9\% \pm 1.4\%$, see S3 Fig).

As expected, when the ‘go’ cues are rare (e.g. $p(g) = 10\%$) the state-action prediction error response to the ‘go’ cue is significantly larger than the response to the ‘no-go’ cue. Interestingly, this is still true when the cues occur with equal probability ($p(\text{go}) = 50\%$), and when the ‘go’ cue is slightly more probable ($p(\text{go}) = 55\%$, Fig 8G).

Many of these response characteristics are also present in agents with a fixed α and are an inherent feature of the state-action prediction error (see S3 Fig). However only the agent with flexible α displays *both* the correct profile of prediction error responses and the ability to learn the reversal of contingencies shown above.

A full plot of state-action prediction errors, simulated LC firing and behavioural output for the static go/no-go task for the flexible α agent is shown in S2 Fig.

Discussion

We propose that the LC fulfils a crucial role, linking state-action prediction errors (or Bayesian surprise) during the planning of actions to model decay—a form of learning rate. Using this approach, we have reproduced the following experimentally observed LC characteristics:

- Phasic responses during a Go/No-Go paradigm as described experimentally in [6,13,22]. Here, cues predicting a reward (for which the animal must perform an action) elicit clear phasic LC responses, which stand out against a background of lower overall tonic activity.
- A more general link between the ‘exploration’ mode of behaviour and high tonic levels of LC activity. Whilst direct measurements of LC activity during explore-exploit paradigms are lacking, the link is strongly suggested by indirect experimental evidence. For instance, Tervo et al [3] demonstrated highly variable behavioural choices in rats when the activity of LC units projecting to ACC was held artificially high via optogenetic manipulation. Other studies have also demonstrated [24,25] that an increase in pupil size (a correlate of LC activity) occurs in parallel with behavioural flexibility and task disengagement.

We also reproduce more subtle characteristics of state-action prediction error responses. Simulations of the go/no-go task demonstrate a progression of phasic responses during the learning period that parallels the development of responses reported in a similar go/no-go task in rats [13], in which phasic responses occur initially for the reward alone, then for both the cue and reward, and finally, only for the cue itself. Additionally, during the go/no-go task the model reproduces the following empirical results:

- a reduction in the size of the phasic state-action prediction error when ‘go’ cues were presented consecutively, as reported for an identical go/no-go task [6];
- an increase in response size when target ‘go’ cues are rarer [6];
- a larger response to ‘go’ cues than to ‘no-go’ cues when the two are equally probable [6]. The model further predicts that the larger response to the ‘go’ cue will persist even when this cue is slightly *more* probable than the ‘no-go’ cue (up to 55%—see Fig 8 and S3 Fig) and that the responses will equalise or reverse as the probability of the ‘go’ cue increases further;
- a larger state-action prediction error response to the ‘go’ cue when the cue predicts a greater reward (when the probability of the cue is fixed; see Fig 8F), in line with results reported in monkeys in similar contexts [8].

The results above for both the go/no-go and explore/exploit tasks suggest that the LC should be phasically active when a strongly predicted reward is absent (for example, the unrewarded trials in Fig 6). This is in conflict with the results of a similar go/no-go task [13], which reported no LC activation when predicted rewards were omitted. However, another study—using pupil dilation as a measure of LC response—observed an increase in pupil size upon presentation of rewards that were either significantly higher or lower than expected [26], as predicted here.

We also note that Nassar et al. [27], have shown that pupil linked arousal systems are intimately linked to assessments of environmental instabilities and were predictive of the influence of new data on subsequent inferences. This is entirely complementary to the model presented above, in which the same principles are successfully applied to demonstrate realistic responses in both the go/no-go and explore/exploit tasks.

We do not address in detail the timing of phasic LC activation relative to cues and actions. However, under the model the LC is activated in response to the completed updating of beliefs. This updating is an iterative process that, in real life, may take a variable amount of time—in contrast to the selection of action which occurs automatically based on the probability distributions produced. This is consistent with empirical results showing that the activation of the LC is more tightly locked to action than to preceding cues (see, for instance [13]). A detailed examination of the interplay between LC activation, cued-actions and levels of motivation (including responses to ‘stop’ signals, as explored in [28]) are also outside the scope of the tasks described here but present important avenues for future modelling.

Neurobiology

In previous Active Inference literature the calculation of Bayesian Model Averages has been mapped to the dorsal prefrontal cortex [14]. This is one of the frontal regions known to send projections to LC [29,30] and is a candidate for the calculation of state-action prediction error (although we accept that without further experimental work such anatomical attributions are largely speculative). Experimental evidence for a neural representation of a distinct prediction error based on states, rather than rewards, has also been found in dorsal regions of the frontal cortex in a human MRI study [31].

Turning to the LC-prefrontal connections and the modulation of model updating, converging experimental evidence suggests that working models of the environment are reflected by ACC activity. Activity in the ACC has been shown to correlate to many factors relevant to the maintenance of a generative model, including reward magnitude and probability (for review see [32]), estimation of the value of action sequences and subsequent prediction errors [33,34] and the value of switching behavioural strategies [35]. Marked changes in activity in ACC have been observed at times thought to coincide with significant model updating and occur in parallel with explorative behaviour—an event that has been directly linked to increased input from locus coeruleus [3,36]. Similarly, a direct ACC/ LC connection has also been found in response to task conflicts [37]. ACC activity is also correlated with learning rate during times of volatility, such that when the statistics of the environment change, more recent observations are weighted more heavily in preference to historical information [38]. This evidence provides a solid foundation for the hypothesis that the LC modulates learning rate by governing model updating via ACC. Specifically, we propose that the release of noradrenaline would cause a temporary increase in the susceptibility of model-holding networks to new information. At a cellular level, this would lead to NA effectively breaking and reshaping connections amongst cell assemblies.

In vitro investigation of the cellular effects of noradrenaline provides support for this idea, indicating that noradrenaline may suppress intrinsic connectivity of cortical neurons, causing a relative enhancement of afferent input [1,39,40]. Sara [41] and Harley [42] also suggest that LC spiking synchronises oscillations at theta and gamma frequencies, allowing effective transfer of information between brain regions during periods of LC activity. This may allow enhanced updating of existing models with more recent observations. A role for the LC in prioritising recent observations during times of environmental volatility has been explicitly suggested experimentally [43] and is supported by evidence regarding the critical role of LC activation in reversal learning, e.g. [44].

We note that if the LC is indeed responding to state-action prediction errors, model updating is likely not the only functionality it has. For instance, LC activation has been experimentally linked to the potentiation of memory formation [41,45,46], analgesic effects [47,48] and changes to sensory perception for stimuli occurring at the time of LC activation [1,49,50]. These are all reasonable responses to a large state-action prediction error: the increase in gain on sensory input may ensure that salient stimuli are more easily detectable in the future, whilst enhanced formation of memory might ensure that mappings between salient stimuli and states are remembered over longer timeframes. Similarly, the temporary suppression of pain may facilitate urgent physical responses to important stimuli (for instance, allowing action in response to a stimulus indicating the presence of a predator). The possibility that the LC has the capacity to provide a differentiated response to state-action prediction error is supported by recent work indicating that existence of distinct subunits with preferred targets producing different functional effects [48,51–53].

Relationship to existing models of LC function

The ideas described above are not a radical departure from existing models of LC function—but use the theory of active inference to integrate similar concepts into a general theory of brain function, without invoking the need for monitoring of ad-hoc statistical quantities.

The adaptive gain theory proposed by Aston Jones and Cohen [4] proposes that the LC responds to ongoing assessments of utility in OFC and ACC by altering the global ‘gain’ of the brain (the responsiveness of individual units). Phasic activation produces a widespread increase in gain which enables a more efficient behavioural response following a task-related decision;

however, when the utility of a task decreases, the LC switches to a tonic mode which favours task disengagement and a switch from ‘exploit’ to ‘explore’.

The mechanism we have described reproduces many elements of the adaptive gain theory, with the important exception that different LC firing patterns promoting explorative or exploitative behaviour are an emergent property of the model rather than a dichotomy imposed by design. Since the probability assigned to individual policies is explicitly dependent on their utility (in combination with their epistemic value) a large state-action prediction error will ultimately reflect changes in the availability of policies which lead to high utility outcomes. This may be a positive change, as is the case when a cue indicates that a ‘Go’ policy will secure a reward, or a negative change, when rewards are no longer available in the explore/exploit task. This link is demonstrated in Fig 6 for the explore/exploit task, where increases in state-action prediction error / LC firing occur in tandem with abrupt changes in the agent’s assessment of a given policy’s utility. Both the LC response, and the underlying cause (state-action prediction error), show a shift between ‘phasic’ and ‘tonic’ modes (although it is entirely possible that coupling mechanisms within the LC also act to exaggerate the shift and cause the LC to fire in a more starkly bi-modal fashion, as suggested by computational modelling of the LC [4,54]). As described above, a short prediction error will act to heighten the response to a salient cue over the short term, whilst a large, sustained prediction error—occurring in parallel with declining utility in a task—will act to make behaviour more exploratory.

Yu and Dayan have proposed an alternative model where tonic noradrenaline is a signal of ‘unexpected uncertainty’, when large changes in environment produce observations which strongly violate expectations [5]. This is described as a ‘global model failure signal’ and leads to enhancement of bottom-up sensory information. We have focused on a similar ‘model failure’ signal which allows larger changes to learning about the structure of the model itself—but using the inferences of states within AI as our driver, avoiding explicit tracking of the statistics of ‘unexpected uncertainty’. Rather, we compute model failure in terms of ‘everyday’ errors in predicted actions and sensations. Our model is also in line with the ‘network reset’ theory proposed by Bouret et al, in which LC phasic activation promotes rapid re-organisation of neural networks to accomplish shifts in behavioural mode [10], see also [9]. Large changes in configuration of the state-action heatmap alongside the updates to internal models above would similarly constitute network re-organisations with the result of changing behaviour. Importantly, state-action updates precede action selection, placing LC activation after decision making / classification of stimuli, but before the behavioural response. This order of events is in keeping with experimental evidence showing that LC responses do indeed consistently precede behavioural responses [13,55]. This also parallels the ‘neural interrupt’ model of phasic noradrenaline proposed by Dayan and Yu [56], in which uncertainties over states within a task are signalled by phasic bursts of noradrenaline, causing an interrupt signal during which new states can be adopted.

More recently Parr et al have described an alternative active inference-based model of noradrenaline in decision making [57]. Under this model, noradrenaline and acetylcholine are related to the precision assigned to beliefs about outcomes and beliefs about state transitions. That is, the agent assigns a different weight to any inferences made using the **A** matrix (modulated by release of acetylcholine) or the **B** matrix (modulated by noradrenaline) in its updates. This approach captures some of the interplay between environmental uncertainty and release of noradrenaline. Our formulation also speaks to these uncertainties—without the need to introduce new volatility parameters, or to segregate cholinergic / noradrenergic response into separate modulators of likelihood and transition (i.e., **A** and **B** matrices). Both approaches target the coding of contingencies in terms of connectivity (i.e., probability matrices). Parr et al consider the optimisation of the precision of contingencies. Conversely, we consider the

optimisation of precision from the point of view of optimal learning rates. In other words, the confidence or precision of beliefs about outcomes likelihoods and state transitions can itself be optimised based on inference (about states) or learning (about parameters) in the generative model.

The key contribution of the current work is to link inference to the precision of beliefs about parameters via learning. This addresses the issue of how model parameters are learned and updated and allows an AI agent to make substantial changes to the architecture of its model in times when environmental rules have shifted. The ensuing behaviour produces the archetypal phasic-tonic shifts in LC dynamics, and links LC responses to the outcome of decision on stimuli, as suggested by in-vivo recordings; summaries of which can be found in [4,11].

The difference between these two applications of Active Inference illustrates a broader point about the way in which the theory is used to describe neuromodulation. Current versions of Active Inference have conceived of neuromodulatory systems as reflections of precision, altering the weights assigned to components of the agent's model *during* a continuous cycle of updates. This underlying modulation is capable of drastically altering the inferences the agent makes about likely states and actions. Here we have offered a different view, in which noradrenaline is proposed to respond to the *outcome* of an update cycle. This enables us to endow an active inference agent with a noradrenergic response which relates activity in the locus coeruleus to the outcome of decisions and to subsequent changes to action planning. These responses are then linked back to changes in the underlying structure of the agent's model—again outside of the cycle of inferences. Placing such responses above the update cycle moves them closer to the level of action selection and allows us to reproduce many aspects of LC dynamics observed empirically.

Once validated through experimental work, models of this type can provide insight into symptoms of disorders which have been linked to LC dysfunction. For example, attention deficit hyperactivity disorder (ADHD), which is characterised by inattention and hyperactivity, has been associated with elevated tonic LC activity [1]. Under our model, high tonic firing rates would cause a persistently high 'model decay'. This would cause similar outcomes to those demonstrated for the hyper-flexible explore/exploit agent (Fig 7), which cannot build a stable structured model of the environment and reacts to even minor violations of predictions by changing its behavioural strategy. Pharmacological interventions which lower tonic LC firing rates may ameliorate symptoms by allowing structured models to emerge, guiding appropriate phasic responses and producing more focused behavioural strategies.

Several lines of future work are available to test components of the state-action prediction error / LC theory. Firstly, a clearer understanding of the drivers of LC responses could be pursued through in-vivo recordings in PFC, ACC and LC. This would help to confirm if calculations of state-action prediction error (or utility/estimation uncertainty, under other theories) originate in frontal cortex, rather than being calculated in the LC itself or elsewhere. Simultaneous recordings with high temporal resolution in-vivo will also help to delineate cause and effect in frontal cortex/LC interactions and will complement the accumulating data from human fMRI / pupillometry. Specific details of the above theory could then be tested; for example, in comparing the predictions for an LC driven purely by consideration of utility or estimation uncertainty, rather than by a state-prediction error as prescribed by Free Energy-based estimates. In-vivo recordings during the two tasks described here could also be examined for the characteristic patterns. For instance, in the pattern of LC firing predicted for the explore/exploit task, the above modelling shows a sudden transition to a higher tonic level of activity during a change in the environmental statistics, and a much slower decay of activity occurring as rules stabilise. Triggering or blocking such patterns of firing during task performance would be particularly revealing regarding the proposed role of the LC.

Finally, we have not addressed the role of other neuromodulators that have related effects on behaviour. Whilst dopamine is explicitly included in Active Inference models as a precision parameter, other neuromodulators (e.g. serotonin) do not yet have a clear place within the model. Understanding the interplay between these systems will be crucial for placing LC activity in context—and will enable the explanatory power of Active Inference to be fully harnessed.

Supporting information

S1 Appendix. Derivations of Active Inference update equations and expressions for state-action prediction errors and model decay.

(PDF)

S2 Appendix. ‘Pseudocode’ overview of the implementation of the Active Inference model.

(DOCX)

S1 Fig. Bayesian dependency graph showing the overall generative model employed during decision making. First the free energy and expected free energy of policies is evaluated. Next these serve as priors on the probability of a policy, which in turn leads to a selected action, state transitions and outcomes. In Appendix 1 we provide a full description of the model.

(TIF)

S2 Fig. Full plots of state-action prediction error, behaviour and simulated LC firing in tasks performed by the agent with a flexible α . (a) shows the performance of the agent in the go/no-go task. The prediction error output is similar in form to the output from the task played by the agent with $\alpha = 16$ in Fig 5, except for the more pronounced reduction in the size of the state-action prediction error peaks in response to consecutive cues. (b) shows the performance during the explore/exploit task with the same parameters as in Fig 7(A) and 7(B), that is, a with changes in the location of the high probability arm ($p = 0.7$) every 50 trials.

(TIF)

S3 Fig. Characteristics of state-action prediction error peak responses during the no-go trial for different fixed α agents. All plots show averages over 2000 trials. (a) shows the changes in prediction error peak responses during either ‘go’ (marked ‘G’) cues or ‘no-go’ (marked ‘NG’) cues. All agents display a larger ‘go’ response for rarer cues. Additionally, responses for ‘go’ cues are consistently larger than those for ‘no-go’ cues when the ‘go’ is more probable than, or equally probable as, the ‘no-go’ cue. This effect persists up to a ‘go’ probability of 55%. When the probability of the cue is increased further, the peaks are equal in size or reversed. (b) shows the effect of changing the reward size. As in Fig 8, all agents display a larger state-action prediction error response when the reward is larger. (c) shows the reduction in peak size caused by presenting ‘go’ cues consecutively. This reduction is greater for the more flexible (lower α) agents, reflecting the larger changes to the agent’s model caused by the consecutive ‘go’ cues.

(TIF)

Author Contributions

Conceptualization: Anna C. Sales, Matthew W. Jones, Anthony E. Pickering, Rosalyn J. Moran.

Data curation: Anna C. Sales.

Formal analysis: Anna C. Sales, Karl J. Friston, Rosalyn J. Moran.

Funding acquisition: Matthew W. Jones, Anthony E. Pickering, Rosalyn J. Moran.

Investigation: Anna C. Sales, Anthony E. Pickering.

Methodology: Anna C. Sales, Anthony E. Pickering, Rosalyn J. Moran.

Resources: Anthony E. Pickering, Rosalyn J. Moran.

Software: Anna C. Sales, Karl J. Friston, Rosalyn J. Moran.

Supervision: Anthony E. Pickering, Rosalyn J. Moran.

Validation: Anna C. Sales, Anthony E. Pickering, Rosalyn J. Moran.

Visualization: Anna C. Sales, Anthony E. Pickering, Rosalyn J. Moran.

Writing – original draft: Anna C. Sales.

Writing – review & editing: Anna C. Sales, Karl J. Friston, Matthew W. Jones, Anthony E. Pickering, Rosalyn J. Moran.

References

1. Berridge C, Waterhouse BD. The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res Rev.* 2003; 42(1):33–84. PMID: [12668290](#)
2. Foote SL, Bloom FE, Aston-Jones G. Nucleus locus ceruleus: new evidence of anatomical and physiological specificity. *Physiol Rev.* 1983 Jul; 63(3):844–914. <https://doi.org/10.1152/physrev.1983.63.3.844> PMID: [6308694](#)
3. Tervo DGR, Proskurin M, Manakov M, Kabra M, Vollmer A, Branson K, et al. Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. *Cell.* 2014; 159(1):21–32. <https://doi.org/10.1016/j.cell.2014.08.037> PMID: [25259917](#)
4. Aston-Jones G, Cohen JD. AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annu Rev Neurosci.* 2005 Jul 21; 28(1):403–50.
5. Yu AJ, Dayan P. Expected and unexpected uncertainty: ACh and NE in the neocortex. *Adv neural Inf Process* 2003; 15:157–64.
6. Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T. Locus Coeruleus Neurons in Monkey Are Selectively Activated by Attended Cues in a Vigilance Task. *J Neurosci.* 1994; 14(7):4467–4460. PMID: [8027789](#)
7. Aston-Jones G, Bloom FE. Norepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *J Neurosci.* 1981/08/01. 1981; 1(8):887–900. PMID: [7346593](#)
8. Bouret S, Richmond BJ. Sensitivity of Locus Ceruleus Neurons to Reward Value for Goal-Directed Actions. *J Neurosci.* 2015; 35(9):4005–14. <https://doi.org/10.1523/JNEUROSCI.4553-14.2015> PMID: [25740528](#)
9. Dayan P, Yu AJ. Phasic norepinephrine: A neural interrupt signal for unexpected events. *Netw Comput Neural Syst* December. 2006; 17(4):335–50.
10. Bouret S, Sara SJ. Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. Vol. 28, *Trends in Neurosciences.* 2005. p. 574–82. <https://doi.org/10.1016/j.tins.2005.09.002> PMID: [16165227](#)
11. Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus—norepinephrine system. *Psychol Bull.* 2005; 131(4):510–32. <https://doi.org/10.1037/0033-2909.131.4.510> PMID: [16060800](#)
12. Aston-Jones G, Bloom FE. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep-waking cycle. *J Neurosci.* 1981 Aug; 1(8):876–86. PMID: [7346592](#)
13. Bouret S, Sara SJ. Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. *Eur J Neurosci.* 2004 Aug 1; 20(3):791–802. <https://doi.org/10.1111/j.1460-9568.2004.03526.x> PMID: [15255989](#)
14. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active Inference: A Process Theory. *Neural Comput.* 2017 Jan; 29(1):1–49. https://doi.org/10.1162/NECO_a_00912 PMID: [27870614](#)

15. Friston K, Schwartenbeck P, Fitzgerald T, Moutoussis M, Behrens T, Dolan RJ. The anatomy of choice: active inference and agency. *Front Hum Neurosci*. 2013; 7(September):598.
16. Schwartenbeck P, FitzGerald THB, Mathys C, Dolan R, Kronbichler M, Friston K. Evidence for surprise minimization over value maximization in choice behavior. *Sci Rep*. 2015 Nov 13; 5:16575. <https://doi.org/10.1038/srep16575> PMID: 26564686
17. Schwartenbeck P, FitzGerald T, Dolan RJ, Friston K. Exploration, novelty, surprise, and free energy minimization. *Front Psychol*. 2013; 4:710. <https://doi.org/10.3389/fpsyg.2013.00710> PMID: 24109469
18. Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G. Cognitive Neuroscience Active inference and epistemic value. *Cogn Neurosci*. 2015;
19. Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ. The anatomy of choice: dopamine and decision-making. *Phil Trans R Soc B*. 2014; 369(1655):20130481. <https://doi.org/10.1098/rstb.2013.0481> PMID: 25267823
20. Friston K, Fitzgerald T, Rigoli F, Schwartenbeck P, O’doherly J, Pezzulo G. Active inference and learning. *Neurosci Biobehav Rev*. 2016; 68:862–79. <https://doi.org/10.1016/j.neubiorev.2016.06.022> PMID: 27375276
21. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*. 2011; 5:39. <https://doi.org/10.3389/fnhum.2011.00039> PMID: 21629826
22. Aston-Jones G, Rajkowski J, Kubiak P. Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience*. 1997; 80(3):697–715. PMID: 9276487
23. Foote SL, Aston-Jones G, Bloom FE. Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal. *Proc Natl Acad Sci U S A*. 1980 May 1; 77(5):3033–7. PMID: 6771765
24. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci*. 2010 Jun; 10(2):252–69. <https://doi.org/10.3758/CABN.10.2.252> PMID: 20498349
25. Jepma M, Nieuwenhuis S. Pupil Diameter Predicts Changes in the Exploration?Exploitation Trade-off: Evidence for the Adaptive Gain Theory. *J Cogn Neurosci*. 2011 Jul 10; 23(7):1587–96. <https://doi.org/10.1162/jocn.2010.21548> PMID: 20666595
26. Preuschoff K, BM ’t Hart, Einhauser W. Pupil dilation signals surprise: evidence for noradrenaline’s role in decision making. *Front Neurosci*. 2011 Sep 30; 5:115. <https://doi.org/10.3389/fnins.2011.00115> PMID: 21994487
27. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasley B, Gold JI. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci*. 2012 Jul 3; 15(7):1040–6. <https://doi.org/10.1038/nn.3130> PMID: 22660479
28. Kalwani RM, Joshi S, Gold JI. Phasic activation of individual neurons in the locus coeruleus/subcoeruleus complex of monkeys reflects rewarded decisions to go but not stop. *J Neurosci*. 2014 Oct 8; 34(41):13656–69. <https://doi.org/10.1523/JNEUROSCI.2566-14.2014> PMID: 25297093
29. Arnsten AF, Goldman-Rakic PS. Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. *Brain Res*. 1984 Jul 23; 306(1–2):9–18. PMID: 6466989
30. Jodo E, Chiang C, Aston-Jones G. Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. *Neuroscience*. 1998 Mar; 83(1):63–79. PMID: 9466399
31. Gläscher J, Daw N, Dayan P, O’Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*. 2010 May 27; 66(4):585–95. <https://doi.org/10.1016/j.neuron.2010.04.016> PMID: 20510862
32. Rushworth MFS, Behrens TEJ. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci*. 2008 Apr 26; 11(4):389–97. <https://doi.org/10.1038/nn2066> PMID: 18368045
33. Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci*. 2007 May 22; 10(5):647–56. <https://doi.org/10.1038/nn1890> PMID: 17450137
34. Holroyd CB, Yeung N. Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn Sci*. 2011; 16:121–7.
35. Hayden BY, Pearson JM, Platt ML. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci*. 2011 Jul 5; 14(7):933–9. <https://doi.org/10.1038/nn.2856> PMID: 21642973
36. Karlsson MP, Tervo DGR, Karpova AY. Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty. *Science (80-)*. 2012; 338(6103):135–9.

37. Ebitz RB, Platt ML. Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*. 2015; 85(3):628–40. <https://doi.org/10.1016/j.neuron.2014.12.053> PMID: 25654259
38. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. *Nat Neurosci*. 2007 Sep 5; 10(9):1214–21. <https://doi.org/10.1038/nn1954> PMID: 17676057
39. Hasselmo ME, Linstor C, Patil M, Ma D, Cekic M. Noradrenergic Suppression of Synaptic Transmission May Influence Cortical Signal-to-Noise Ratio. *J Neurophysiol*. 1997 Jun; 77(6):3326–39. <https://doi.org/10.1152/jn.1997.77.6.3326> PMID: 9212278
40. Kobayashi M, Imamura K, Sugai T, Onoda N, Yamamoto M, Komai S, et al. Selective suppression of horizontal propagation in rat visual cortex by norepinephrine.
41. Sara SJ, J. S. Locus Coeruleus in time with the making of memories. *Curr Opin Neurobiol*. 2015 Dec; 35:87–94. <https://doi.org/10.1016/j.conb.2015.07.004> PMID: 26241632
42. Walling SG, Brown RAM, Milway JS, Earle AG, Harley CW. Selective tuning of hippocampal oscillations by phasic locus coeruleus activation in awake male rats. *Hippocampus*. 2011 Nov; 21(11):1250–62. <https://doi.org/10.1002/hipo.20816> PMID: 20865739
43. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasley B, Gold JI. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci*. 2012;
44. Rorabaugh JM, Chalermphanupap T, Botz-Zapp CA, Fu VM, Lembeck NA, Cohen RM, et al. Chemo-genetic locus coeruleus activation restores reversal learning in a rat model of Alzheimer's disease. *Brain*. 2017 Nov 1; 140(11):3023–38. <https://doi.org/10.1093/brain/awx232> PMID: 29053824
45. Takeuchi T, Duzskiewicz AJ, Sonneborn A, Spooner PA, Yamasaki M, Watanabe M, et al. Locus coeruleus and dopaminergic consolidation of everyday memory Studies of memory for over a century. *Nature*. 2016; 537:5–7.
46. Wagatsuma A, Okuyama T, Sun C, Smith LM, Abe K, Tonegawa S. Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proc Natl Acad Sci*. 2018 Jan 9; 115(2):E310–6. <https://doi.org/10.1073/pnas.1714082115> PMID: 29279390
47. Hickey L, Li Y, Fyson SJ, Watson TC, Perrins R, Hewinson J, et al. Optoactivation of locus coeruleus neurons evokes bidirectional changes in thermal nociception in rats. *J Neurosci*. 2014; 34(12):4148–60. <https://doi.org/10.1523/JNEUROSCI.4835-13.2014> PMID: 24647936
48. Hirschberg S, Li Y, Randall A, Kremer EJ, Pickering AE. Functional dichotomy in spinal- vs prefrontal-projecting locus coeruleus modules splits descending noradrenergic analgesia from ascending aversion and anxiety in rats. *Elife*. 2017 Oct 13; 6.
49. Raquel A, Martins O, Froemke RC. Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex.
50. Hurley L, Devilbiss D, Waterhouse B. A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. *Curr Opin Neurobiol*. 2004 Aug 1; 14(4):488–95. <https://doi.org/10.1016/j.conb.2004.06.007> PMID: 15321070
51. Kebschull JM, Garcia Da Silva P, Reid AP, Peikon ID, Albeanu DF, Zador Correspondence AM, et al. High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA NeuroResource High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron*. 2016; 91:975–87. <https://doi.org/10.1016/j.neuron.2016.07.036> PMID: 27545715
52. Chandler DJ, Gao W-J, Waterhouse BD. Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cortices. *Proc Natl Acad Sci*. 2014 May 6; 111(18):6816–21. <https://doi.org/10.1073/pnas.1320827111> PMID: 24753596
53. Uematsu A, Tan BZ, Ycu EA, Cuevas JS, Koivumaa J, Junyent F, et al. Modular organization of the brainstem noradrenergic system coordinates opposing learning states. *Nat Neurosci*. 2017 Sep 18;
54. Usher M, Cohen JD, Servan-Schreiber D, Rajkowski J, Aston-Jones G. The Role of Locus Coeruleus in the Regulation of Cognitive Performance. *Science* (80-). 1999; 283(5401).
55. Clayton EC, Rajkowski J, Cohen JD, Aston-Jones G. Phasic Activation of Monkey Locus Coeruleus Neurons by Simple Decisions in a Forced-Choice Task. *J Neurosci*. 2004 Nov 3; 24(44):9914–20. <https://doi.org/10.1523/JNEUROSCI.2446-04.2004> PMID: 15525776
56. Dayan P, Yu AJ. Phasic norepinephrine: A neural interrupt signal for unexpected events. *Netw Comput Neural Syst*. 2006 Jan 18; 17(4):335–50.
57. Parr T, Friston KJ. Uncertainty, epistemics and active inference. *J R Soc Interface*. 2017 Nov 1; 14(136):20170376. <https://doi.org/10.1098/rsif.2017.0376> PMID: 29167370