

Adaptive Learning through Temporal Dynamics of State Representation

Niloufar Razmi^{1,2} and  Matthew R. Nassar^{1,2}

¹Robert J. & Nancy D. Carney Institute for Brain Science, Brown University, Providence, Rhode Island 02912-1821, and ²Department of Neuroscience, Brown University, Providence, Rhode Island 02912-1821

People adjust their learning rate rationally according to local environmental statistics and calibrate such adjustments based on the broader statistical context. To date, no theory has captured the observed range of adaptive learning behaviors or the complexity of its neural correlates. Here, we attempt to do so using a neural network model that learns to map an internal context representation onto a behavioral response via supervised learning. The network shifts its internal context on receiving supervised signals that are mismatched to its output, thereby changing the “state” to which feedback is associated. A key feature of the model is that such state transitions can either increase learning or decrease learning depending on the duration over which the new state is maintained. Sustained state transitions that occur after changepoints facilitate faster learning and mimic network reset phenomena observed in the brain during rapid learning. In contrast, state transitions after one-off outlier events are short lived, thereby limiting the impact of outlying observations on future behavior. State transitions in our model provide the first mechanistic interpretation for bidirectional learning signals, such as the P300, that relate to learning differentially according to the source of surprising events and may also shed light on discrepant observations regarding the relationship between transient pupil dilations and learning. Together, our results demonstrate that dynamic latent state representations can afford normative inference and provide a coherent framework for understanding neural signatures of adaptive learning across different statistical environments.

Key words: adaptive learning; Bayesian; neural network; P300; pupillometry; representation

Significance Statement

How humans adjust their sensitivity to new information in a changing world has remained largely an open question. Bridging insights from normative accounts of adaptive learning and theories of latent state representation, here we propose a feedforward neural network model that adjusts its learning rate online by controlling the speed of transitioning its internal state representations. Our model proposes a mechanistic framework for explaining learning under different statistical contexts, explains previously observed behavior and brain signals, and makes testable predictions for future experimental studies.

Introduction

People and animals are often required to adjust their behavior in the face of new information, especially in environments that evolve over time (Behrens et al., 2007; Wilson et al., 2010; Donahue and Lee, 2015; Massi et al., 2018; Farashahi et al., 2019; Li et al., 2019). In recent years, Bayesian probability theory has been applied to develop learning algorithms for changing environments and to provide normative accounts for learning

behavior. Probabilistic models prescribe learning that is more rapid during periods of environmental change and slower during periods of stability (Adams and MacKay, 2007; Behrens et al., 2007; Wilson et al., 2010). These models have provided insight into why people adjust learning according to uncertainty (Browning et al., 2015; Muller et al., 2019) and the probability with which an observation reflects a changepoint (Adams and MacKay, 2007; Nassar et al., 2010). In this framework, the human brain is viewed as implementing an optimal learning algorithm that embodies the statistical properties of the world it operates in (O'Reilly, 2013; Meyniel and Dehaene, 2017).

While probabilistic modeling provides a normative account for many of the adjustments in learning rate observed in humans and animals (Behrens et al., 2007; Wilson et al., 2010; Nassar et al., 2019b), it has thus far failed to clarify the underlying neural mechanisms. One issue is that exact Bayesian inference can be closely approximated by many qualitatively different algorithms (Yu and Dayan, 2005; Nassar et al., 2010; Bernacchia et al., 2011;

Received Feb. 21, 2021; revised Jan. 24, 2022; accepted Jan. 26, 2022.

Author contributions: N.R. and M.R.N. designed research; N.R. performed research; N.R. analyzed data; N.R. and M.R.N. wrote the paper.

This work was supported by Department of Health and Human Services | National Institutes of Health | National Institute on Aging Grant R00-AG-054732 to M.R.N. We thank Linda Yu, Robert Wilson, Olga Lositsky, Romy Frömer, and Cristian Buc Calderon for helpful discussion and comments.

The authors declare no competing financial interests.

Correspondence should be addressed to Matthew R. Nassar at matthew_nassar@brown.edu.

<https://doi.org/10.1523/JNEUROSCI.0387-21.2022>

Copyright © 2022 the authors

Table 1. Summary of the parameters used for simulation of the probabilistic inference task and neural network training

Neural network parameter	Value	Description
Number of neurons in the input layer (N_{in})	63	Equally spaced points between $[-\pi, +\pi]$ incrementing by 0.1
Concentration (κ)	32	Concentration of the von Mises pdf used in the input layer
Number of neurons in the output layer (N_{out})	41	Equally spaced points between -50 and 350 , incrementing by 10
Synaptic learning rate (η)	0.1	
WD threshold	0.01	
WD rate	0.1	
Model hazard rate	0.6	The model uses a higher value compared with the actual hazard rate for optimal performance
Input layer threshold	0.0001	Neurons firing above this threshold constitute the active “context” on each trial
Task parameter		
Hazard rate (H)	0.1	Probability of a changepoint/oddball trial
Noise (σ_N)	25	SD of random process generating outcomes
SD of drift rate (σ_{drift})	10	SD of the random process generating drift rate in oddball condition

List of all parameters used in the current neural network modeling simulations as well as the original task parameters used in a previous behavioral study involving the same task (McGuire et al., 2014).

Mathys et al., 2011; Wilson et al., 2013; Igaya, 2016; Farashahi et al., 2017). One such approximation that relies on a single dynamic learning rate can capture behavior across a wide range of statistical environments (Nassar et al., 2021). Direct implementation of this model requires a dynamic learning rate signal that drives learning across statistical contexts; yet, such a context-invariant learning rate signal has yet to be observed in the brain, despite several attempts to do so (D’Acromont and Bossaerts, 2016; Li et al., 2019; Nassar et al., 2019b). In contrast, brain signals that predict more learning in discontinuously changing environments (Behrens et al., 2007; Nassar et al., 2012; O’Reilly et al., 2013; McGuire et al., 2014; Jepma et al., 2016) fail to do so across contexts (D’Acromont and Bossaerts, 2016), and in one case change the sign of their relationship to learning in environments with outliers or oddballs (Jepma et al., 2016, 2018; Nassar et al., 2019b). Other approximations to normative learning have been connected to neural signals, but typically in a limited range of environments and without a clear indication that they could capture the versatility of human learning behavior (Bernacchia et al., 2011; Mathys et al., 2011; Farashahi et al., 2017). In sum, while previous models have explored the potential neural mechanisms for adaptive learning, no algorithm has captured the range of human behavior and its neural correlates across environments.

In the present article, we propose such a generalized framework based on the idea that adaptive learning is accomplished by controlling internal representations according to environmental structure (Yu et al., 2021). We implement this idea with a feed-forward neural network in which learning rate dynamics emerge via an internal context that transitions rapidly after surprising events much like patterns of activity previously observed in prefrontal cortex (Karlsson et al., 2012; Nassar et al., 2019a). Within our framework, context transition signals provide a mechanistic explanation for brain signals, such as the EEG P300, that show context-dependent relationships to learning (Nassar et al., 2019b). Together, our results suggest that adaptive learning behavior could emerge through abrupt transitions in mental context, rather than through adjusting an explicit learning rate.

Materials and Methods

Experimental task. We examine human and model behavior in a predictive inference task that has been described previously (McGuire et al., 2014; Nassar and Troiani, 2021). The predictive inference task is a computerized task in which an animated helicopter drops bags in an open field. In the pretraining session, human subjects learned to move a bucket with a joystick beneath the helicopter to catch bags that could contain valuable contents. During the main phase of the experiment, the

helicopter was occluded by clouds and the participants were forced to infer its position based on the locations of bags it had previously dropped.

Our initial simulations focus on dynamic environments in which surprising events often signal a change in the underlying generative structure (changepoint condition; Figs. 1–5). In the changepoint condition, bag locations were drawn from a distribution centered on the helicopter with a fixed SD of 25 (unless otherwise specified in the analysis). The helicopter remained stationary on most trials, but occasionally and abruptly changed its position to a random uniform horizontal screen location. The probability of moving to a new location on a given trial is controlled by the hazard rate ($H = 0.1$). Unless otherwise noted, our modeling results are presented with 32 simulated subjects, to correspond to the sample size in the study by McGuire et al. (2014), and with the parameters listed in Table 1.

We also considered a complementary generative environment in which surprising events were unrelated to the underlying generative structure (oddball condition; Fig. 6; Nassar and Troiani, 2021). In the oddball condition, the helicopter would gradually move in the sky according to a Gaussian random walk (drift rate = 10). In the oddball condition, bags were typically drawn from a normal distribution centered on the helicopter, as described above, but on occasion a bag would be dropped in random location unrelated to the position of the helicopter. The location of an oddball bag was sampled from a uniform distribution that spanned the entire screen. The probability of an oddball event was controlled by a hazard rate ($H = 0.1$).

Normative learning model. A simple δ rule can perform the predictive inference by incrementally updating beliefs about the helicopter location according to prediction errors, as follows:

$$B_{t+1} = B_t + \alpha \delta \quad (1)$$

$$\delta = \text{Bag Position}(t) - B_t(t), \quad (2)$$

where B is the belief about the helicopter position on each trial, δ is the prediction error observed on that trial, and α is the learning rate. With a constant α , the model assigns the same weight to all predictions and outcomes. Previous work has shown that Bayesian optimal inference can be reduced to a δ rule learning under certain approximations, leading to normative prescriptions for learning rate that are adjusted dynamically (Nassar et al., 2010, 2019b). The resulting normative learning model takes information that human subjects would normally obtain during the pretraining sessions, including hazard rate and SD, but also computes two latent variables by using the trial-by-trial prediction error, as follows: (1) changepoint probability, which is computed after an outcome is observed and indicates the probability that the observed outcome has reflected a change in the helicopter location; and (2) relative uncertainty, which is computed before making the next prediction and indicates the uncertainty of the model about the location of the helicopter. Detailed information regarding how changepoint probability (CPP) and relative uncertainty (RU) are calculated can be found in previous work (Nassar et al., 2019b).

In the changepoint condition, the normative learning rate α_t is defined by the following:

$$\alpha_t = \text{CPP} + \text{RU} - \text{CPP} \times \text{RU}. \quad (3)$$

Using these two latent variables, both of which track the prediction error, but with different temporal dynamics (McGuire et al., 2014), the model computes a dynamic learning rate that increases after a changepoint and gradually decreases in the following stable period after a changepoint.

The same approximation to Bayesian inference can be applied in the oddball condition to produce a normative learning model that relies on oddball probability and relative uncertainty to guide learning. While the latent variables and the form of the model mimic that in the changepoint condition, the learning rate differs in that it is reduced, rather than enhanced, in response to outcomes that are inconsistent with prior expectations, as follows:

$$\alpha_t = \text{RU} - \text{OBP} \times \text{RU}, \quad (4)$$

where OBP is the posterior probability estimate of the model that an outcome was an oddball event, and RU reflects the uncertainty of the model about the current helicopter location. Thus, normative inference in the oddball condition requires decreasing learning according to the probability of an extreme event (oddball), whereas normative inference in the changepoint condition required increasing it.

Neural network models. To better understand how normative learning principles might be applied in a neural network we created a series of neural network models that use supervised learning rules to generate predictions in the predictive inference task. Specifically, we created a two-layer feedforward neural network that can perform the predictive inference task.

Network architecture includes two layers. The input layer is composed of N neurons with responses characterized by a von Mises (circular) distribution with mean (m) and fixed concentration equal to 32. We implemented several versions of this model depending on how m changes on a trial-by-trial basis.

The output layer contains neurons corresponding to spatial location of the bucket on the screen. The response of output layer neurons was computed by the weighted sum of the input layer, as follows:

$$r_j = \sum_{i=1}^{N_{in}} x_i w_{ij}, \quad (5)$$

where x_i is the activation of neuron i in the input layer, r_j is the response of neuron j in the output layer, and w_{ij} is the connection weight between neuron i and neuron j . The bucket position chosen by the model on each trial was computed as a linear readout of the output layer, as follows:

$$\text{Estimate} = \sum_{j=1}^{N_{out}} L_j r_j, \quad (6)$$

where L_j is the location encoded by each corresponding unit r_j in the output layer. The weight matrix is randomly initialized with a uniform distribution of a mean of 0 and an SD equal to 5×10^{-4} . The network is then trained on each trial by modifying the weight matrix according to the following:

$$w_{ij} = (1 - \eta)w_{ij} + \eta y_j x_i, \quad (7)$$

where y_j is the probability on a normal distribution centered on the observed outcome evaluated at L_j with an SD of 25 (equal to the SD of the outcome generative process), and η is a constant synaptic learning

rate controlling the weight changes of the neural network and was set to 0.1 for all model simulations. Although this value was chosen somewhat arbitrarily, additional simulations using network learning rates in the range of 0.01–0.6 did not change the qualitative predictions of the model.

Fixed-context shift models. In the first models we consider, fixed-context shift models, m reflects the position of the most active neuron in the input layer and is computed on each trial as follows:

$$m_{(t+1)} = m_{(t)} + \Delta m_f. \quad (8)$$

Here, Δm_f takes a fixed value for all trials throughout the simulation (Fig. 2B,C). We considered 50 different Δm_f values ranging from 0 to 2 to study the effect of context shifts on model performance. The word “context” refers to the subpopulation of input layer neurons that are firing above the threshold (here 0.0001, although the results are robust if using a range of values between 0.001 and 0.00001) on each trial. As m increases from one trial to the next, the activity bump in the input layer shifts clockwise (Fig. 2B,C), which can be thought of as the associative context being changed to some degree (Δm_f) on each trial. The architecture of the input layer is arranged in a circle so that hypothetically the context would be able to shift clockwise indefinitely, and thus context shifts are reported in radians. To minimize interference from previous visits to a set of context neurons, we implemented weight decay ($\text{WD} = 0.1$) on each time step according to the following rule:

$$W_{t+1}(x_i < \text{threshold}) = W_t(x_i < \text{threshold}) \times \text{WD} \quad (9)$$

Note that this weight decay is not intended as a biological assumption, but rather a convenient simplification to allow the model to represent a large number of contexts with a small pool of neurons.

Therefore, on each trial, first the model would make a prediction based on the weighted sum of the active input, observe an outcome, shift the context by the assigned context shift and store the supervised signal in the new context. This new context is in turn used at the beginning of the next trial to produce a response.

Ground truth context shift model. To leverage the benefits of different fixed context shifts, we designed a model that used dynamic context shifts that were optimized for each trial. The ground truth context shift model has the same design of a fixed-context shift model except instead of the constant term Δm , the model computes Δm in a manner that depends on whether the current trial is a changepoint:

$$\Delta m = \begin{cases} \max(\Delta m_f) & \text{if } t \text{ is changepoint} \\ 0 & \text{otherwise} \end{cases}. \quad (10)$$

Dynamic context shift models. The ground truth context shift model assumes full knowledge of changepoint locations, whereas humans and animals must infer changepoints based on the data. Here we build plausibility into the ground truth model by controlling context shifts according to subjective estimates of CPP that are based on the observed trial outcomes, as follows:

$$\Delta m = f(\text{CPP}). \quad (11)$$

The function f provides a fixed level of context shift according to the estimated changepoint probability by inverting the relationship between the context shift and the effective learning rate observed in the fixed-context shift models and are plotted in Figure 2D. Thus, on each trial, the model will choose a context shift belonging to a fixed-context shift model that has the closest effective learning rate to CPP. Thus, more surprising outcomes that yield higher values of CPP will consistently result in larger context shifts, with a changepoint probability of one resulting in the maximal context shift and a changepoint probability of zero resulting in no context shift at all.

CPP was computed either using the Bayesian normative model described above (Bayesian context shift) or from an approximation derived from the neural network itself (network-based context shift). In

the network-based version, the probability of a state transition is subjectively computed by the following equation:

$$\frac{H/41}{H/41 + r_{X_t}(1 - H)}, \quad (12)$$

which can be interpreted as a network-based approximation to Bayesian CPP estimation (for more details, see additional information at: <https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>) or in terms of a nonlinear activation over prediction errors such as has been proposed in various conflict models (Botvinick et al., 2001; Cockburn and Frank, 2013). H can be thought of in Bayesian terms as a hazard rate, or in neural network terms, as controlling the threshold of the activation function, and r_{X_t} is the firing rate of the output unit corresponding to the location X_t , which can be thought of as providing a readout of the outcome probability based on a Bayesian population code. The 41 reflects the total number of output units in our population, and, since outcomes could occur that were in between the tuning of these units, in practice we used linear interpolation to estimate r_{X_t} based on the two output units closest to the actual outcome location. The H value was set to 0.6 for the changepoint condition to optimize performance (i.e., minimize prediction errors). While the hazard rate in our model was selected to optimize predictions, it also is consistent with the behavior of human participants, whose m values, when fit with the model, yielded higher hazard rates than the generative hazard rate in our experiment (see additional Fig. 1 at: <https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>).

Mixture model. We also considered a model that uses context shifts intermediate between our fixed- and dynamic-context shift models. Specifically, this model shifted context according to a weighted mixture of the context shift from the best performing fixed-context shift model and the network-based context shift model, as follows:

$$\text{context shift} = \text{mixture} \times \text{fixed context shift} + (1 - \text{mixture}) \times \text{dynamic context shift}. \quad (13)$$

Quantitative model fitting. We fit the behavior of human subjects from a previous study (McGuire et al., 2014) with the following five different models: (1) the network-based context shift model with a single free parameter (hazard rate); (2) a fixed-context shift model with one parameter (i.e., fixed-context shift); (3) a mixture model in which context shifts are a weighted mixture of the best fixed-context shift model (fixed-context shift = 0.4) and a network-based dynamic context shift where the mixture weight is one free parameter (Eq. 13); (4) a mixture model as described above, but with the fixed-context shift fit as a second free parameter; and (5) a mixture model as described above, but with both the fixed-context shift and hazard rate fit as additional free parameters (in addition to mixture weight). Parameters that minimized the sum of squared errors were attained through a grid search over the parameter space ($[0, 1]$ for fixed-context shift, mixture weight, and hazard rate). The model comparison used Bayesian information criteria (BIC) to penalize for differential complexity across our model set, as BIC provided the most reliable model recovery for synthetic data.

Extension of network models to the oddball condition. To test our proposed models in a variation of the task where prediction errors are not indicative of a change in context (i.e., oddball condition), we use the same design of neural network but with a simple modification in temporal dynamics of context shifts.

The task involved the same paradigm described above, but with outcomes (i.e., bag locations) determined by a different generative structure. In particular, the helicopter location gradually changed its position in the sky with a constant drift rate, and bags were occasionally sampled from a uniform distribution spanning the range of possible outcomes, rather than being “dropped” from the helicopter itself (Nassar and Troiani, 2021; Nassar et al., 2021).

The ground truth neural network model was modified to incorporate the alternate generative structure of the oddball condition. In particular, on each trial, the input activity m value was changed by (1) maximally context shifting in response to oddballs at the time of feedback, (2) “returning” from the oddball-induced context shift at the end of the

feedback period before the subsequent trial, and (3) adding a constant value (0.05) proportional to the fixed drift rate of the random walk process before making the prediction (for choosing this constant drift rate, we ran simulations with different values of drift rate and chose one that produced optimal behavior). Thus, after a prediction is made on trial context, the mean changes according to the following:

$$\Delta m_1 = \begin{cases} \max(\Delta m_f), & \text{if } t \text{ is oddball} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

$$m_{t+1} = m_t + \Delta m_1. \quad (15)$$

But, after the model receives the supervised signal (represented by a normal distribution that is centered on the bag position with SD corresponding to the SD of bag drops) and stores it in a new context, the context transitions back to the following:

$$m_{t+1} = m_t - \Delta m_1 + \Delta m_2. \quad (16)$$

where Δm_2 is a constant (here 0.05) proportional to the drift rate of the random walk process. This leads to the information from the oddball trial being stored in a different context that will not influence the upcoming prediction of the model.

The dynamic context shift models were constructed to follow the same logic, but using subjective measures of oddball probability rather than perfect knowledge about whether a trial is an oddball. Specifically, we updated the context on observing feedback according to the probability that the feedback reflects an oddball (OP), as follows:

$$\Delta m_1 = f(\text{OP}) \quad (17)$$

$$m_{t+1} = m_t + \Delta m_1, \quad (18)$$

and before making a prediction for the subsequent trial returned to the previous context, except with a slight shift modeling to account for the drift in the helicopter position because of the following random walk:

$$m_{t+1} = m_t - \Delta m_1 + \Delta m_2. \quad (19)$$

This model captures the intuition that if an outcome is known to be an outlier, it should be partitioned from knowledge that pertains to the helicopter location, rather than combined with it. To accomplish this, the model changes the context first according to the oddball probability, or Δm_1 in the above equation, and after storing the supervised learning signal in the new context, the model transitions back to its previous context by subtracting the first context shift term Δm_1 . It then moves the context according to a constant shift proportional to the drift rate Δm_2 . The Δm_1 term causes significant shifts on oddball trials, but after that the model transitions back to the previous context and shifts according to the Δm_2 , which would not be influenced by oddball trials. Similar to the changepoint condition, here, we also made a version of the dynamic Bayesian context shift model, which used network output layer activity to compute subjective measures of oddball probability.

Representational similarity analysis. We computed a trial-by-trial dissimilarity matrix where each cell in the matrix represents the number corresponding to the dissimilarity between the input layer activity on two trials. The dissimilarity matrix (D) of the dynamic context shifts model uses Euclidean distance and is computed by the following:

$$D_{ij} = \sqrt{\sum_{q=1}^{N_m} (x_{(i,q)} - x_{(j,q)})^2}, \quad (20)$$

where $x_{(i,q)}$ reflects the firing of neuron q in the input layer on trial i .

Behavioral analysis. Behavioral analyses are aimed at understanding the degree to which we revise our behavior in response to new

observations. To quantify this, we define an “effective learning rate” as the slope of the relationship between trial-to-trial predictions errors (i.e., the difference between the bucket position and bag position) and trial-to-trial updates (i.e., the change in bucket position from one trial to the next). The adjective “effective” is chosen here so that this learning rate will not be mistaken by the reader for the following two other learning rates used in this article: (1) the fixed synaptic learning rate of the neural network; and (2) the normative learning rate prescribed by the reduced Bayesian model. To measure effective learning rate, we regressed updates (UPs) onto the prediction errors (PEs) that preceded them, as follows:

$$UP = \beta_0 + \beta_1 \times PE. \quad (21)$$

The resulting slope term β_1 captures the effective learning rate or the amount of update expected for a given prediction error. We also performed a more extensive regression analysis that included terms for (1) prediction error, (2) prediction error times change-point probability, and (3) prediction error times relative uncertainty (Fig. 4D).

Comparison with P300 analysis. To analyze the effect of trial-to-trial variability in context shifts from the dynamic context shift model on the effective learning rate produced by that model, we fit the regression model above to simulated predictions for the dynamic context shift models, but did so while splitting data into quartiles according to the size of the context shift size that the model underwent on a given trial. The corresponding figure (Fig. 6E) of P300 signal and learning rate are from the study by Nassar et al. (2019b).

Pupil response simulation. We modeled 480 trials of a predictive inference task for each of the two conditions (oddball, changepoint). We created synthetic pupil traces by defining time points for feedback-locked context shifts, which occurred 400 ms after oddball or changepoint, and preprediction context shifts at 900 ms after oddball events (Eqs. 10, 13). We used measurements of context shift for the respective changepoint and oddball trials (Eqs. 11, 16) at these time points and convolved these measurements with a γ distribution to create simulated time courses of a pupil response under the assumption that the pupil signal reflects the need for a context shift. We analyzed this signal with a regression model that was applied to all synthetic data in sliding windows of time. Explanatory variables in our model included surprise (changepoint/oddball probability computed from the normative model) and learning (trial-by-trial learning rate computed from the normative model).

Data availability. All analysis and modeling code (including code for generating the figures) has been made available on GitHub (<https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>).

Results

To test whether changes to latent state representations can facilitate adaptive learning behavior, we modeled a predictive inference task designed to measure adaptive learning in humans (Fig. 1; McGuire et al., 2014). In the task, a helicopter, which is hidden behind clouds, drops visible bags containing valuable contents from the sky (Fig. 1A, right). On each trial, the subject moves a bucket to the location where they believe the helicopter to be, such that they can catch potentially valuable bag contents. Subjects can move the bucket to a new position on each trial to update and improve their prediction (Fig. 1A, left, B, orange arrow). In the “changepoint” variant of the task, bag locations were sampled from a Gaussian distribution centered on the helicopter, which occasionally relocated to a new position on the screen. Such abrupt transitions in helicopter location led to changes in the statistical context defining the bag locations (context shifts), which could be inferred by monitoring the size of prediction errors (Fig. 1B, red arrow). Therefore, the helicopter position is a dynamic latent variable that must be inferred from noisy observations (i.e., dropped bags) on each trial to yield

optimal task performance. Previous work has shown that human behavior can be captured by a normative learning model that relies on a dynamic “learning rate” adjusted from trial to trial according to CPP and RU (Fig. 1B,C), but failures to identify neural signals that reflect this dynamic learning rate consistently across conditions cast doubt on its biological relevance (Nassar et al., 2012, 2019b; O’Reilly et al., 2013; D’Acemont and Bossaerts, 2016). Here we explore whether normative learning may instead be achieved in the brain by a neural network that undergoes dynamic transitions in the mental context to which associates are bound, thereby adjusting where information is stored, rather than the degree to which storage occurs.

A neural network test bed for exploring adaptive learning

To examine how normative updating could be implemented in a neural network, we devised a two-layer feedforward neural network in which internal representations of context are mapped onto bucket locations by learning weights using a supervised learning rule (Fig. 2B; see Materials and Methods). Units in the output layer of the network represent different possible bucket locations in the predictive inference task, and a linear readout of this layer is used to guide bucket placement, which serves as a prediction for the next trial. After each trial, a supervised learning signal corresponding to the bag location is provided to the output layer, and weights corresponding to connections between input and output units are updated accordingly.

The input layer of our model is designed to reflect the mental context to which learned associations are formed, and its activity is given by a Gaussian activity bump with a mean denoting the position of the neuron with the strongest activity and an SD denoting the width of the activity profile. The primary goal of this work is to understand how changes to the mean of the activity bump, across trials, affect learning within our model. Since the input layer of the network reflects mental context, it does not receive any explicit sensory information, and we can manipulate its activity across trials to provide a flexible test bed for how different task representations (i.e., mental context dynamics) might affect the performance of the model. In particular, we examine how displacing the mean of the activity bump in the input layer across trials affects the rate and dynamics of the networks learning behavior. In the simplest case, a nondynamic network, the mean of the activity bump in the input layer is constant across all trials, reflecting learning onto a fixed “state.” A slightly more complex mental context might be one that drifts slowly over time, such that the mean of the activity bump changes a fixed amount from one trial to the next leading trials occurring close in time to be represented more similarly. In this case, learning would occur in an evolving temporal state representation. In a more complex (but maybe more intuitive) case, the subset of active neurons in the input layer could correspond to the current “helicopter context” (Fig. 1B) or period of helicopter stability. In this case, the mean of the activity bump would only transition on trials where the helicopter changes position and thus could be thought of as representing the underlying latent state of the helicopter (e.g., this is the third unique helicopter position I have encountered)—albeit without any explicit encoding of its position.

Context shifts facilitate faster learning

We first examined the performance of models in which the mean of the input activity bump transitioned by some fixed amount on each trial. This set included 0 (fixed stimulus representation), small values in which nearby trials had more similar input activity profiles (timing representation) and extreme cases

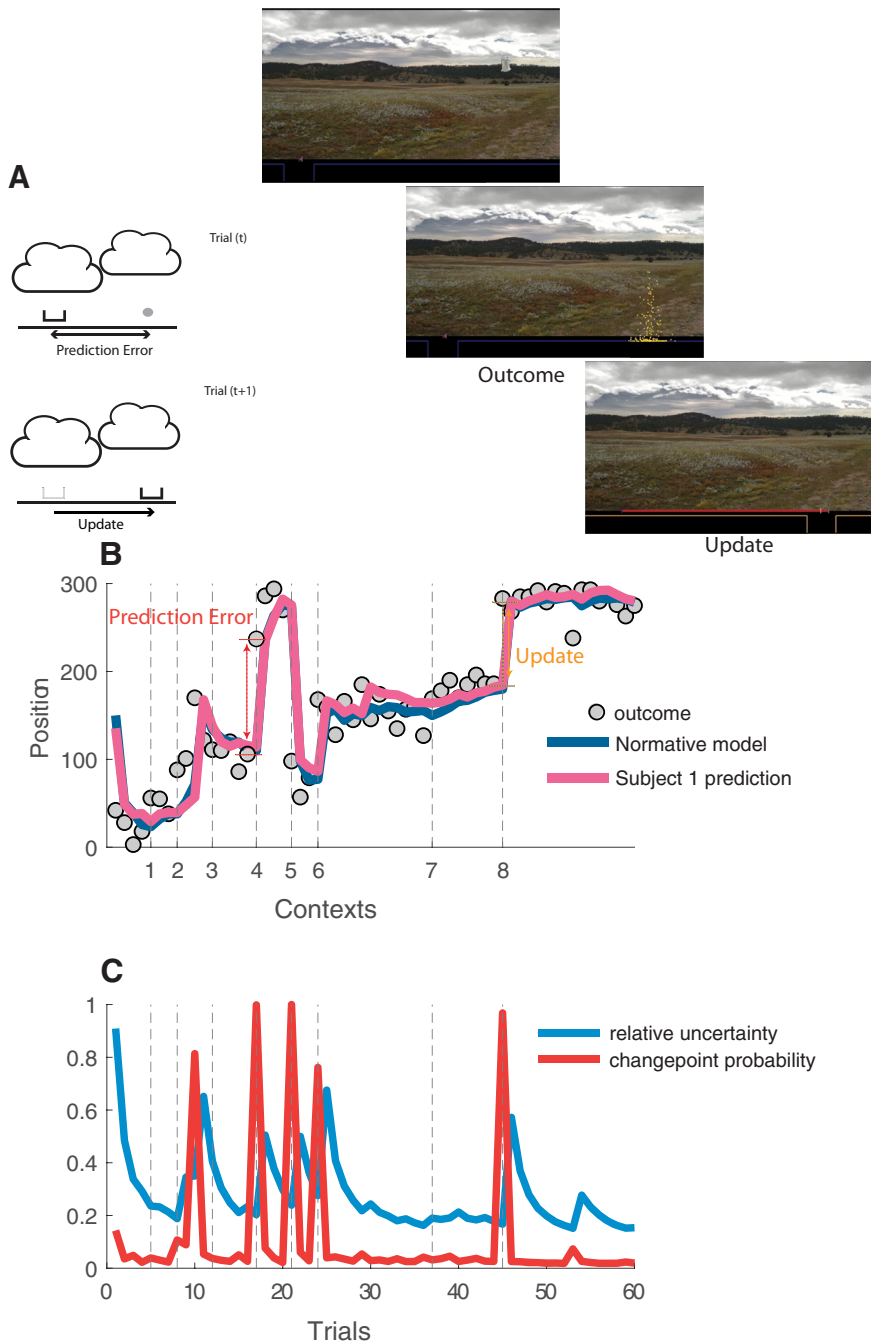


Figure 1. Predictive inference task to measure the dynamics of adaptive learning. **A**, Schematic illustration (left) and screenshots (right) of the predictive inference task. Human subjects place a bucket at a horizontal location on the bottom of the screen to catch a bag of coins that will be subsequently dropped from a hidden helicopter. After observing the bag location (outcome) at the end of each trial, along with their prediction error (distance between bucket and outcome), the subject could improve their response by adjusting their bucket position (update). In the changepoint condition, the helicopter typically remains stationary but occasionally moves to a completely new location. **B**, The sequence of bag locations (outcome; ordinate) is plotted across trials, which are segmented into discrete contexts reflecting periods with a stationary mean. Context transitions (dotted vertical lines) reflect changepoints in the position of the helicopter. Bucket placements made by a subject (pink) and normative model (navy) are shown with a representation of an example prediction error and outcome: [prediction error = outcome (t) – estimate (t) and update = estimate ($t + 1$) – estimate (t)]. **C**, The learning rate, which defines the degree to which the normative model updates the bucket in response to a given prediction error, depends on two factors, CPP (red) and RU (blue), which combine to prescribe learning that is highest at changepoints (CPP) and decays slowly thereafter (RU).

where there was no overlap between the input layer representations on successive trials (individuated trial representation). We defined the fixed shift in the mean of the activity profile as the “context shift” of our model. This shift is depicted in Figure 2C as the nodes shown in “hot colors” (i.e., active neurons) in the input layer of the neural network moving to the right; note how the size of rightward shift in the schematic neural network is constant in all four trials shown. We used increments starting from 0 (the same input layer population) to a number corresponding to a complete shift (completely new population) in each trial. Learning leads to fine-tuning of the weights by strengthening connections between active input neurons and the output neurons nearby the outcome location (bag position) on each trial. We observed that moderate shifts in the input layer (context shifts) led to the best performance in our task (Fig. 2E), and that the effective learning rate describing the behavior of the model monotonically scaled with context shift (Fig. 2D). We also compared the performance of these models to a δ rule equipped with learning rates matched to those empirically observed in each fixed-context shift model (Fig. 2D). The performance of fixed-context shift networks mirrored that of δ rule models, both in terms of overall performance and the advantage conferred to moderate context shifts in the network (Fig. 2E, black), or learning rates in the δ rule (Fig. 2E, red). Together, these results support the notion that context shifts could be used to enhance the sensitivity of behavior to new observations, analogous to adjusting the learning rate in a δ rule.

Dynamic context shifts can improve task performance

The higher performance of moderate context shift models (Fig. 2E) might be thought of intuitively as navigating the classic trade-off between flexibility and stability. A higher learning rate, which can be effectively produced by a larger context shift, promotes flexibility and leads to better performance in response to environmental changes that render past observations irrelevant to future ones (Fig. 3C). In contrast, smaller learning rates, which are effectively produced by smaller context shifts, yield stable predictions that facilitate a performance advantage in stable but noisy environments by averaging over the noise (Fig. 3D). More concretely, when the helicopter remains in the same location, small context shifts

improve performance by pooling learning over a greater number of bag locations to better approximate their mean, but large context shifts can improve performance after changes in helicopter location by reducing the interference between outcomes before and after the helicopter relocation. Inspired by the observed relationship between context shift and accuracy, we next modified the model to dynamically adjust context shifts to optimize performance. In principle, based on the intuitions above, we might improve on our fixed-context shift models by only shifting the activity profile of the input layer at a true context shift in the task (i.e., allow the input layer to represent the latent state). Since such a model requires pre-existing knowledge of changepoint timings, we refer to it as the ground truth model (Fig. 3, top). Indeed, we observed that the ground truth model performs as well as the best fixed-context shift model after changepoints (Fig. 3C), and better than the best fixed-context shift model during periods of stability (Fig. 3D), yielding overall performance better than any fixed-context shift model (Fig. 3E).

Needless to say, the brain does not have access to perfect information regarding whether a given trial is a changepoint or not. Is it possible to make a more realistic version of this optimal model, using information that the brain does have access to? To answer this question, we built models that infer changepoint probability based on experienced prediction errors. We built two versions of this model, one that computed CPP explicitly according to Bayes rule (Wilson et al., 2010), and one that approximated CPP according to the mismatch between output activity in the network and the observed outcome (i.e., supervised signal). In both cases, hazard rates necessary for computing CPP were optimized for performance, resulting in model hazard rates exceeding their experimental values (see additional Fig. 1 at: <https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>). Both models achieved good performance after changepoints by elevating context shifts (Fig. 3C) and during periods of stability by reducing context shifts (Fig. 3D), yielding overall performance better than any fixed-context shift model, and only slightly worse than the ground truth model (Fig. 3E,F). These results were consistent across different noise conditions (see additional Fig. 2 at: <https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>).

Dynamic context shifts capture key behavioral and neural signatures of adaptive learning in humans

Not only was the dynamic context shift model able to outperform fixed-context shift models, it did so by capturing behaviors that are observed in people. The model updated predictions

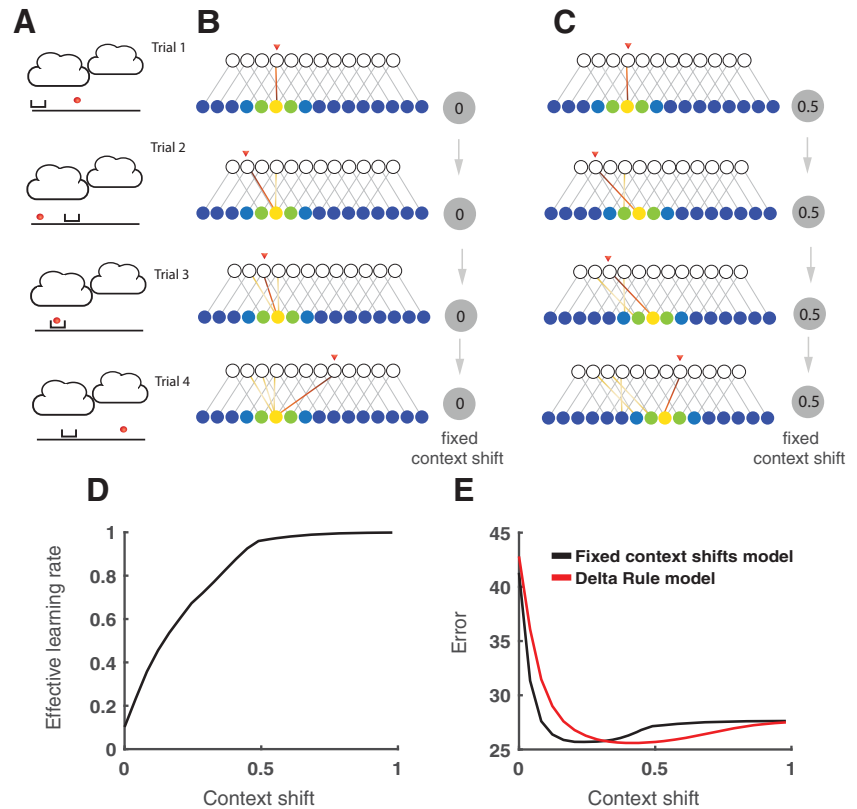


Figure 2. A neural network with fixed-context shifts can approximate any constant learning rate. **A–C**, Network structure and weight updates for two fixed-context shift models (**B**, **C**) are depicted across four example trials of a predictive inference task (**A**). For all networks, feedback was provided on each trial corresponding to the observed bag position (**A**, circle; **B**, **C**, red arrow), and the weights of the network were updated using supervised learning. Only a subset of neurons (circles) and connections between them (lines) are shown in the neural network schematic. Activation in the input layer was normally distributed around a mean value that was constant in **B** and shifted by a fixed amount on each trial in **C** (context shift). Learned weights (colored lines) were all assigned to the same input neuron when context shift was set to zero (**B**) but assigned to different neurons when the context shift was substantial (**C**). **D**, The effective learning rate (ordinate), characterizing the influence of an unpredicted bag position on the subsequent update in bucket position, increased when the model was endowed with faster internal context shifts (abscissa). **E**, Mean absolute prediction error (ordinate) was minimized by neural network models (black line) that incorporated a moderate level of context shift from one trial to the next (abscissa). The mean error of a simple δ rule model using various learning rates is shown in red (x -axis values indicate the context shift equivalent to the fixed δ rule learning rate derived from **D**). For each simulated δ rule model, we plotted the x position according to the amount of context shift that yielded the corresponding learning rate from that fixed-context shift model; thus, the position on the x -axis reflects the same amount of average learning of the two models, but the mechanics of how learning is generated differs across the two models. Note that neural networks with fixed-context shifts achieve similar task performance to more standard δ -rule models that use a constant learning rate.

according to prediction errors, but relied more heavily on prediction errors from certain trials (Fig. 4A). We can quantify the effective learning rate as the slope of the relationship between the bucket update of the model and its previously observed prediction error to compare the behavior of different models (Fig. 4A). To look more closely at the relationship between our proposed model and human behavior, we first created a mixture model that updated context as a linear mixture of those prescribed by the network-based dynamic model and those prescribed by the best fixed-context shift model (for a more detailed description, see Materials and Methods). We then fit the following five different models to data from human subjects: (1) the network-based context shift model with a single free parameter (hazard rate); (2) a fixed-context shift model with one parameter (i.e., fixed-context shift); (3) a mixture model in which context shifts are a weighted mixture of those used by the “best” fixed-context shift model and those used by the best network-based context

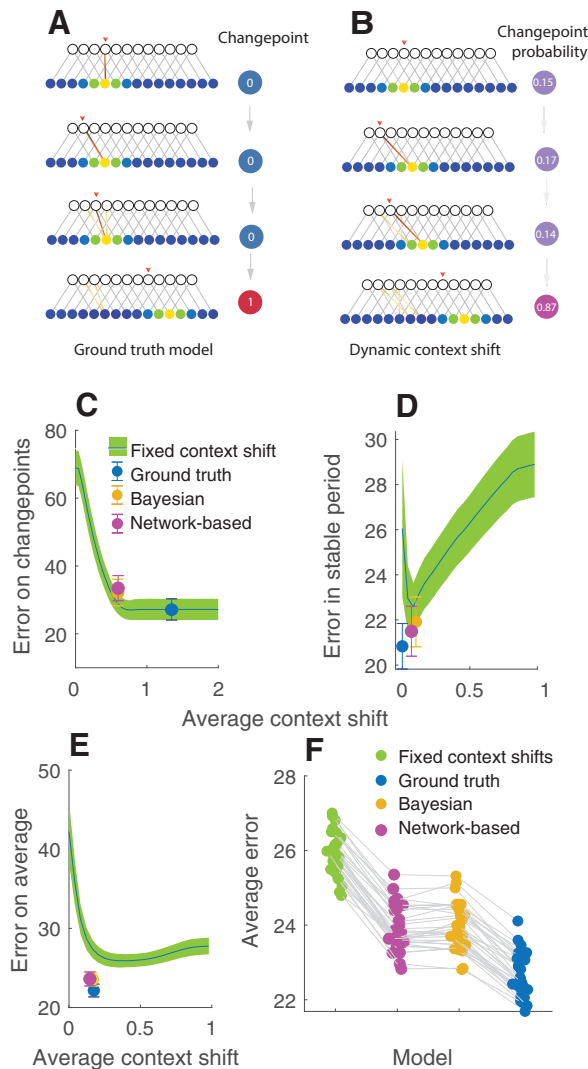


Figure 3. Dynamic context shifts facilitate better task performance. **A**, Schematic diagram of the ground truth model network (left), which is provided with objective information about whether a given trial is a changepoint or not (right) and uses that knowledge to shift the context only on changepoint trials. **B**, The dynamic context shift network uses a subjective estimate of changepoint probability based on a statistical model (Bayesian) or the network output (Network based) to adjust its context shift on each trial. All of these models shift context to a greater degree on changepoint trials (bottom row) than on non-changepoint trials (top three rows). **C**, Performance on trials immediately following a changepoint was best for models using the largest context shifts. Mean error on trials following a changepoint (ordinate) is plotted as a function of context shift (abscissa) for fixed (line/shading) and dynamic (points) context shift models. Shading and error bars indicate standard error of the mean. The ground truth model (blue point) minimized error after changepoints through large context shifts, and the dynamic context shift models, which made moderately large context shifts after changepoints, also approached this level of performance (yellow and pink). Note that since the optimal policy on changepoint trials is to use a learning rate of 1, any model with a large enough context shift would be able to achieve optimal performance on this subset of trials (note the performance of highest fixed-context shift models). **D**, The smallest errors on trials during periods of stability (>5 trials after changepoint; ordinate) were achieved by models that made smaller context shifts (abscissa). All dynamic context shift models (ground truth, Bayesian, network based) made relatively small context shifts for stable trials, yielding good performance. **E**, Across all trials, subjective dynamic context shift models yielded better average performance than the best fixed-context shift model and approached the performance of the ground truth model. **F**, Average error for individual simulations showing the Bayesian (yellow) and network-based (pink) context shift models beat the best fixed-context shift model (green) consistently across simulated task sessions. Note that on some occasions the ground truth model yields higher errors than the network-based context shift model. Exhaustive simulations with larger numbers of trials revealed that these apparent performance failures of the ground truth model

shift model with one free parameter (mixture weight); (4) the mixture model described above, but with the fixed-context shift fit as a second free parameter; and (5) the mixture model as above, but with the fixed-context shift and hazard rate fit as additional free parameters (in addition to the mixture weight).

Comparing the five models described above, we observed that human subject data were best described by the most complex model, which included aspects of both fixed and dynamic context shifts (Fig. 4B). The winning model assumes that context shifts as a mixture of network-based dynamics (with hazard rate as a free parameter) and a fixed level on each trial (controlled by a fixed-context shift parameter). The relative contribution of these two influences is controlled by the mixture weight, such that the model can model participants who adjust learning dramatically at changepoints but also participants who learn relatively consistently. A summary of the fitted parameters for the winning model (the model with three free parameters) is shown in Table 2. The participants also used a higher hazard rate than the true hazard rate of the task but close to the hazard rate used by our network-based dynamic context shift model. It is noteworthy that the fixed-context shift model provided the worst quantitative fit to participant data, suggesting that dynamics played a critical role in matching participant behavior.

Simulated data from our fit models reproduced the key behavioral hallmarks of human adaptive learning. In particular, the winning model used higher effective learning rates immediately after changepoints, recapitulating human learning dynamics, whereas the fixed-context shift model used the same effective learning rate across all trials (Fig. 4C). Previous work has captured human learning dynamics using a regression model that includes separate parameters to model fixed learning (β_1), increased learning after likely changepoints (β_2), and increased learning during periods of uncertainty (β_3 ; McGuire et al., 2014). Here we performed posterior predictive checks on our winning model using this same approach and found that our best fitting model, which mixed dynamic and fixed-context shifts, could reproduce key hallmarks of learning (Fig. 4D, red points). In addition, the model could reproduce the impressive heterogeneity in learning behavior across participants, in part through different mixtures of fixed and dynamic context shifts (Fig. 4D, gray points). Together, these results suggest that a network that includes a mixture of dynamic and fixed-context shifts can reproduce hallmarks of human adaptive learning behavior.

These context adjustments provide a potential explanation for rapid changes in activity patterns, or “network resets,” that have been observed during periods of rapid learning in rodent medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC; Karlsson et al., 2012; Nassar et al., 2019a). Rodent studies previously identified neural population activity changes that occurred during periods of uncertainty when animals were rapidly shifting behavioral policies (Karlsson et al., 2012). Human neuroimaging work took a similar approach to identify patterns of activity that changed more rapidly during periods of rapid learning following changepoints, after

← simply reflected the probabilistic nature of our task, and that as the number of trials is increased, the ground truth model always achieved better performance than all other models (see additional Fig. 3 at: <https://github.com/learning-memory-and-decision-lab/dynamicStatesLearning>; Bayesian Context Shift : $t = 21.9, df = 31, p < 10^{-16}$; Network — Based Context Shift : $t = 20.48, df = 31, p < 10^{-16}$).

controlling for other factors (Nassar et al., 2019a). An important open question raised by these studies is why such representations exist at all; in both cases, the representations were not reflecting the behavioral policy, and their dynamics would not be necessary for implementing existing models of adaptive learning (Nassar et al., 2010, 2012). Given that our dynamic context shift model accomplishes adaptive learning by dynamically changing the context representations, we asked whether our input layer might give rise to population dynamics that are similar to the phenomena observed in these previous studies.

To do so, we used a representational similarity analysis (RSA) approach to create a dissimilarity matrix reflecting differences in the input layer activation across pairs of trials for our dynamic context shift model (Fig. 5A). By using the activity profile of the input layer of the dynamic context shift model, we were able to obtain a pattern of dissimilarity across all pairs of trials for each simulated task session (Fig. 5B). Examining this dissimilarity matrix reveals abrupt representational shifts at changepoints (Fig. 5B, dotted lines). To quantify the observed changes in activity pattern, we computed the dissimilarity across adjacent pairs of trials and examined how this adjacent trial similarity was affected by changepoints in the task. Consistent with empirical data, we found that representations in our context layer shifted more rapidly immediately after a changepoint (Fig. 5C; mean dissimilarity for changepoint/non-changepoint trials = 0.73/0.22, $t = -54.54$, $df = 31$, $p < 10^{-16}$). In some sense, this is not surprising, given that we built our model to achieve faster learning after changepoints by shifting the activity pattern in the input layer. Nonetheless, our model provides a potential normative explanation for why network reset phenomena are observed during periods of rapid learning: in our model, such changes in activity optimize behavior by providing a clean slate for learning after environmental change.

Dynamic context shifts can reduce learning from oddballs

To understand how dynamic context shifts might be used to improve learning in an alternate statistical environment, we considered a set of “oddball” generative statistics that have recently been used to investigate neural signatures of learning (D’Acromont and Bossaerts, 2016; Nassar et al., 2019b). In the oddball condition, the mean of the output distribution does not abruptly change but instead gradually drifts according to a random walk. However, on occasion a bag is dropped at a random location uniformly sampled across the width of the screen with no relationship to the helicopter, constituting an outlier unrelated to both past and future outcomes. In the presence of such oddballs, large prediction errors should lead to less, rather than more, learning. This normative behavior has been observed in adult human subjects (D’Acromont and Bossaerts, 2016; Nassar et al., 2019b; Nassar and Troiani, 2021).

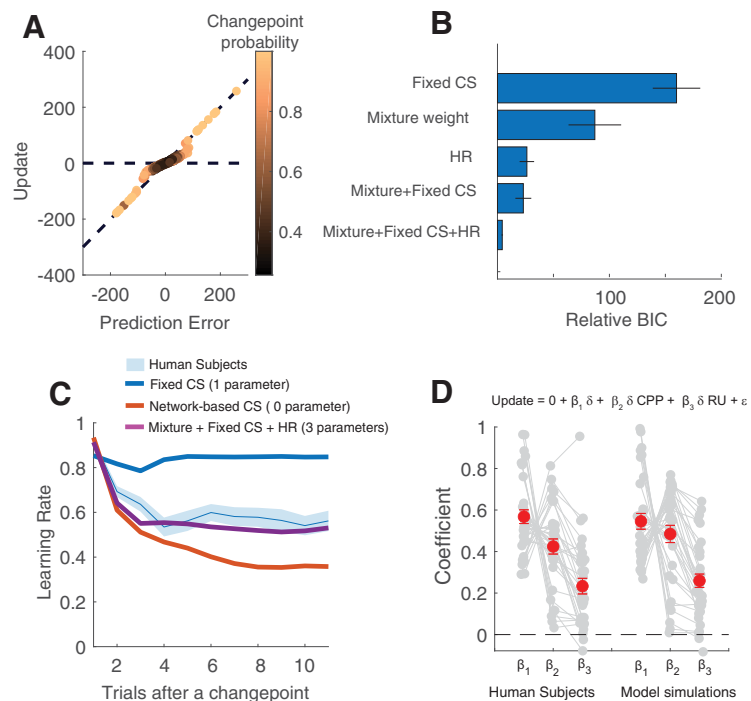


Figure 4. Dynamic context shifts describe quantitative and qualitative patterns in human learning behavior. **A**, Single-trial updates from dynamic context shift model (ordinate) are plotted against prediction errors (abscissa) for each single trial of a simulated session with points colored according to the normative changepoint probability. Note that large absolute prediction errors, corresponding to high changepoint probabilities, tend to lead to updates on the unit line, corresponding to an effective learning rate of one. **B**, Relative BIC (abscissa) is plotted for five models that were fit to human behavioral data (ordinate). Lines reflect SEM. Fixed CS, Model with the fixed-context shift as a free parameter; mixture weight, model that mixes fixed-context shift with network-based dynamic context shift in which mixture weight is the only free parameter; HR, dynamic network-based context shift model with hazard rate as a free parameter; Mixture + Fixed CS, model with mixture weight and fixed-context shift as two free parameters; Mixture + Fixed CS + HR, model with mixture weight, fixed-context shift and hazard rate as three free parameters. **C**, Effective learning rate (ordinate) is plotted for trials that differ in their alignment to the most recent changepoint (abscissa). The best fitting model (Mixture + Fixed CS + HR, purple) resembles qualitative behavior of human subjects (light blue). **D**, Coefficients from a regression model (top equation) fit to single-trial updates to characterize the degree of overall learning [β_1 : Fixed LR (learning rate)], adjustments in learning at likely changepoints (β_2 : CPP Driven Learning), and adjustments in learning according to normative uncertainty (β_3 : RU Driven Learning). Gray circles represent fits to individual human subjects (left) and simulated data from the best model (right), whereas red circles/bars reflect mean/SEM coefficients across individuals.

Table 2. A Summary of the fitted parameters for the best model

Fitted parameter	Fixed-context shift	Mixture weight	Hazard rate
Mean	0.49	0.65	0.48
SD	0.28	0.24	0.28

Mean and SD of the fitted parameters for the model with three free parameters (fixed-context shift, mixture weight and hazard rate) across all human participants.

To examine whether dynamic context transitions could afford adaptive learning in the oddball condition, we created a network analogous to the ground truth model described above, but active input units were adjusted according to the oddball condition transition structure (Fig. 6A). Specifically, on each trial, the model would shift the context with a small constant rate, corresponding to the drift rate in the generative process (i.e., the helicopter position slowly drifting from trial to trial). On oddball trials, the model would undergo a large context shift, ensuring that the oddball outcome would be associated with a nonoverlapping set of input layer neurons, in much the same way as for changepoint observations in our previous model. However, the model was also endowed with knowledge of the transition structure

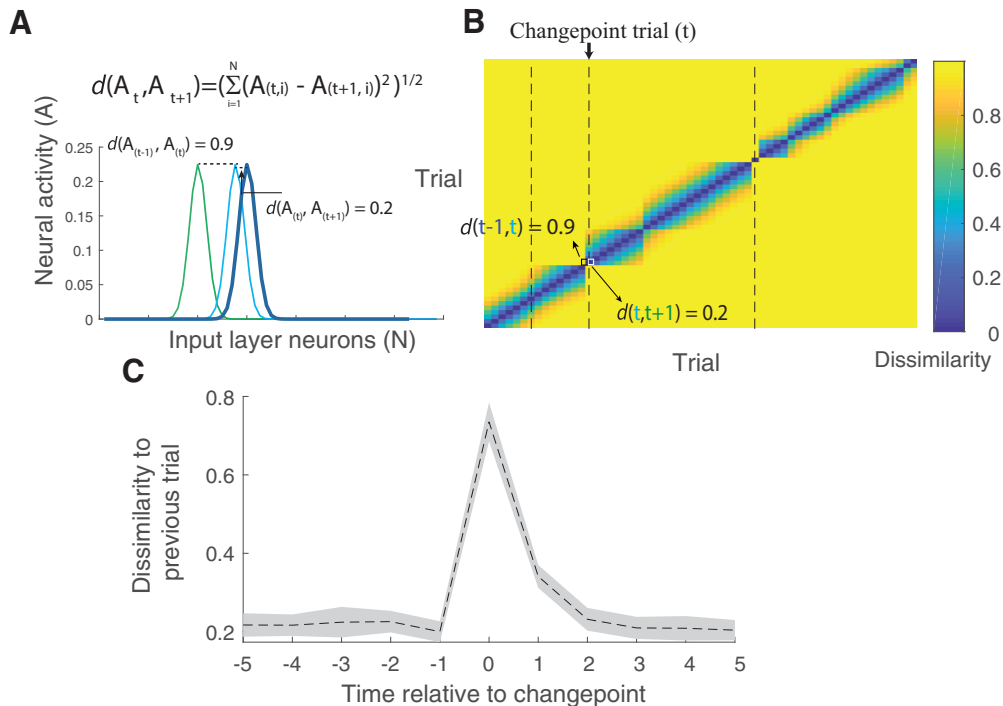


Figure 5. Input layer representations change rapidly at changepoints. **A**, Dissimilarity (d) in the input representation between pairs of trials was computed according to the Euclidean distance between those trials in the space of population activity (A ; here exemplified in terms of three trials, where trial t is an example changepoint). Note that the cyan activity bump corresponding to trial t is shifted relative to the green bump corresponding to trial $t - 1$ (green). **B**, A dissimilarity matrix representing the dissimilarity in input layer activity for each pair of trials in a simulated task session. Dotted lines reflect changepoints, and thus trials between the dotted lines occurred in the same task context (helicopter position). Note that trials within the same context (i.e., trial t and trial $t + 1$) are more similar than for consecutive trials belonging to two different contexts (trial $t - 1$ and trial t). **C**, Mean/SEM (dotted line/shading) dissimilarity between adjacent trials (ordinate) is plotted across trials relative to changepoint events (abscissa) for 32 simulated sessions. Note the rapid change in input layer activity profiles (i.e., high adjacent trial dissimilarity) at the changepoint event, reminiscent of previously observed network reset phenomena that have been linked to periods of rapid learning in both rodents and humans.

of the task, which includes that oddballs are typically followed by nonoddball trials, and, as such, the input layer activity bump would transition to its previous nonoddball location subsequent to learning from the oddball outcome (Yu et al., 2021). Consequently, the learned associations from oddball trials would not be stored in the same context as the ordinary trials, and predictions were always made from the previous “nonoddball” context, thereby minimizing the degree to which oddballs contribute to behavior.

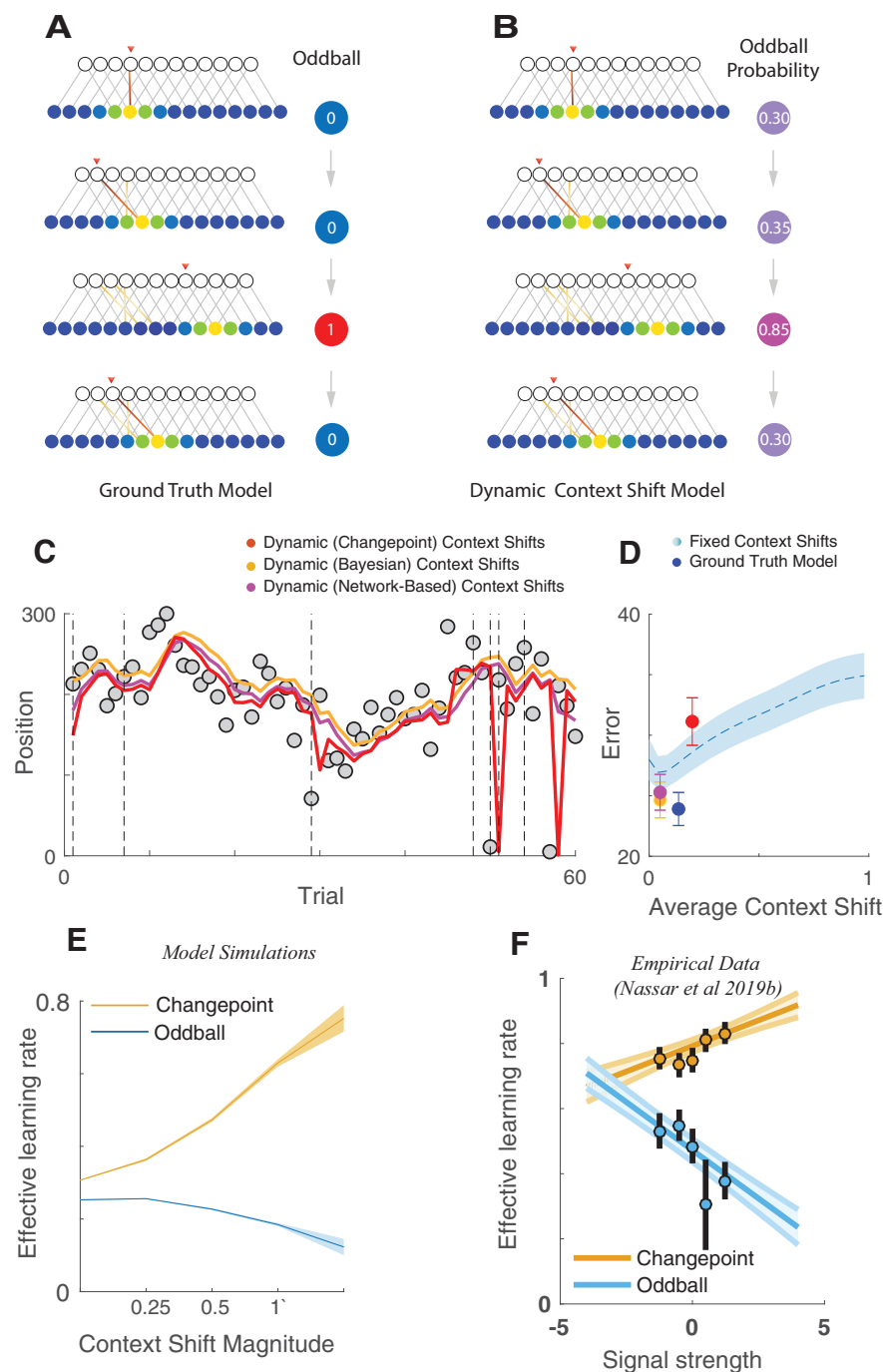
Like in the changepoint condition, we also created versions of the model in which oddballs were inferred probabilistically using either a Bayesian inference model or the activity profile of the output units. Oddball probabilities (computed either from the normative model or the output activity of the network itself) were then used to guide transitions of the active input layer units (Fig. 6B). In these models, the probability of an oddball event drove immediate transitions of the active input layer units to facilitate storage of information related to oddballs in a separate location, but subsequent predictions were always made from the input units corresponding to the most recent nonoddball event (plus a constant expected drift). These models achieved significantly better overall performance than the best fixed-context shift model and performance similar to the ground truth context shift model (Fig. 6C,D). The advantage conferred through dynamic context shifts was specific to the oddball structural assumptions, as a model that used dynamic context shifts based on the changepoint generative structure yielded worse performance than fixed-context shift models (Fig. 6C,D, red). It is noteworthy that, given the appropriate structural representation, the dynamic context shift model produced normative behavior in changepoint condition, where it increased learning by sustaining

the newly activated context, but produced normative learning in the oddball context (decreasing learning on oddball trials) by immediately abandoning the new context in favor of the more “typical” one.

Dynamic context shifts explain bidirectional learning signals observed in the brain

A primary objective in this study was to identify the missing link between the algorithms that afford adaptive learning in dynamic environments and their biological implementations. One key challenge to forging such a link has been the contextual sensitivity of apparent learning rate signals observed in the brain. For example, in EEG studies the P300 associated with feedback onset positively predicts behavioral adjustments in static or changing environments (Fischer and Ullsperger, 2013; Jepma et al., 2016, 2018), but negatively predicts behavioral adjustments in the oddball condition that we describe above (Nassar et al., 2019b). These bidirectional relationships are strongest in people who adjust their learning strategies most across conditions, and persist even after controlling for a host of other factors related to behavior, suggesting that they are actually playing a role in learning, albeit a complex one (Nassar et al., 2019b).

Here we propose an alternative mechanistic role for the P300: it reflects the need for a context shift. Our model provides an intuition for why such a signal might yield the previously observed bidirectional relationship to learning. A stronger P300 signal, corresponding to a larger context shift, would result in a stronger partition between current learning and previously learned associations. In changing environments, this could effectively increase learning, as it would decrease the degree to which prior experience is reflected in



the weights associated with the currently active input units. In the oddball environment, where context changes prevent oddball events from affecting weights of the relevant input layer units, we would make the opposite prediction. We tested this idea directly in our model by measuring the effective learning rate in the dynamic context shift model for bins of trials sorted according to the magnitude of the context shift that was used for them. The results of this analysis revealed a positive relationship between the context shift used by the model and its effective learning rate in the changepoint condition, but a negative relationship between context shift and learning rate in the oddball condition (Fig. 6E). This result is qualitatively similar to empirically observed bidirectional relationships between learning and the P300 (Fig. 6F). Thus, our results are consistent with the possibility that the P300 relates to learning indirectly, by signaling or promoting transitions in a mental context representation that effectively partition learning across context boundaries, including changepoints and oddballs.

Relationship between context shifts and pupil diameter response

One major theory of learning has suggested that adaptive learning is facilitated by fluctuations in arousal mediated by the locus coeruleus (LC)/norepinephrine (NE) system (Yu and Dayan, 2005). This idea has been supported by evidence from transient pupil dilations, which in animals are linked to LC/NE signaling (Reimer et al., 2016; Joshi and Gold, 2020) and are positively related to learning in changing environments (Nassar et al., 2012). Nonetheless, these results are difficult to interpret in light of another study that used both changepoints and oddballs, and observed the opposite relationship between pupil dilation and learning (O'Reilly et al., 2013). The contextual link between pupil diameter and learning may have a common biological origin with that of the P300 signal explored above, as the signals share a host of

simulated trials selected according to the magnitude of context shifts on those trials (abscissa) separately for changepoint (yellow) and oddball (blue) tasks. Note that larger context shifts in the changepoint condition correspond with greater learning, but in the oddball correspond with less learning. **F**, Effective learning rate of human participants (ordinate; Nassar et al., 2019b) in binned trials according to the magnitude of the feedback-locked P300 EEG signal on that trial (abscissa) computed separately for changepoint (blue) and oddball (yellow) task conditions. Note qualitative similarity between the empirical observations related to the P300 signal and the predictions of our model regarding context shift magnitude.

common antecedents and have both been proposed to reflect transient LC/NE signaling (Nieuwenhuis et al., 2011; Vazey and Aston-Jones, 2014; Joshi and Gold, 2020). In contrast to learning theories, another prominent theory has suggested that the LC/NE system plays a role in resetting ongoing context representations (network reset hypothesis; Bouret and Sara, 2005), which maps well onto the context shift signals that our model requires to adjust effective learning rates.

Here we formalize the network reset hypothesis in terms of context transitions in our model and explore the predictions of this formalization for the relationship between pupil diameter and learning. Specifically, we consider the possibility that the LC/NE system is related to the instantaneous context shifts in our model, and that pupil dilations occur as a delayed and temporally smeared version of this LC/NE signal (see Materials and Methods). In this framework, we might consider two distinct influences on the pupil diameter. First, the context shifts elicited by observations that deviate substantially from expectations, which might reflect either changepoints or oddballs depending on the statistical context (Fig. 7A, purple observation highlighted in green box). Second, the context shift required to “return” to the previous context after a likely oddball event, which must occur after processing feedback from a given trial, but before the start of the next trial (Fig. 7A, red box). Jointly considering transitions at these two discrete timepoints yields the prediction that both changepoints and oddball events should lead to pupil dilations, but that these dilations should be prolonged in the oddball condition (Fig. 7B). We regressed these pupil signals onto an explanatory matrix that included model-derived measures of learning (trial-wise empirically derived learning rate) and surprise (estimated changepoint/oddball probability) to better understand their relationship to behavior. The results from this simulation yielded a positive relationship between our modeled pupil signal and surprise, but a late negative relationship between pupil diameter and learning (Fig. 7C). These results are generally in agreement with the results of the study by O’Reilly (2013) and support the possibility that pupil diameter reflects a temporally extended indicator of the context transitions predicted by our model. We suggest that this idea can be empirically tested in future studies involving pupillometry in predictive inference tasks with two distinct generative structures (oddball and changepoint) where our predictions state that pupil dilations would be observed during both conditions but with different temporal profiles: the changepoint condition with a pupil response corresponding to only one initial context shift, but the oddball condition including two responses for each respective context shift.

Discussion

In this work, we demonstrate that abrupt transitions in context, triggered by unexpected outcomes, can facilitate improved performance in two different statistical environments that differ in the sort of adaptive learning that they require and do so in a manner that mimics human behavior. Specifically, we showed that human participants’ behavior in a predictive inference task can be explained by a mixture of fixed and dynamic context shifts. The neural network model that we proposed underwent context transitions that depended on the likelihood of the observed outcome and the hazard rate. The hazard rate that optimized performance and matched participant behavior in our simulations was considerably higher than the true rate of changepoints in the task (0.1), calling in to question whether it should really be interpreted as a hazard rate. One alternative

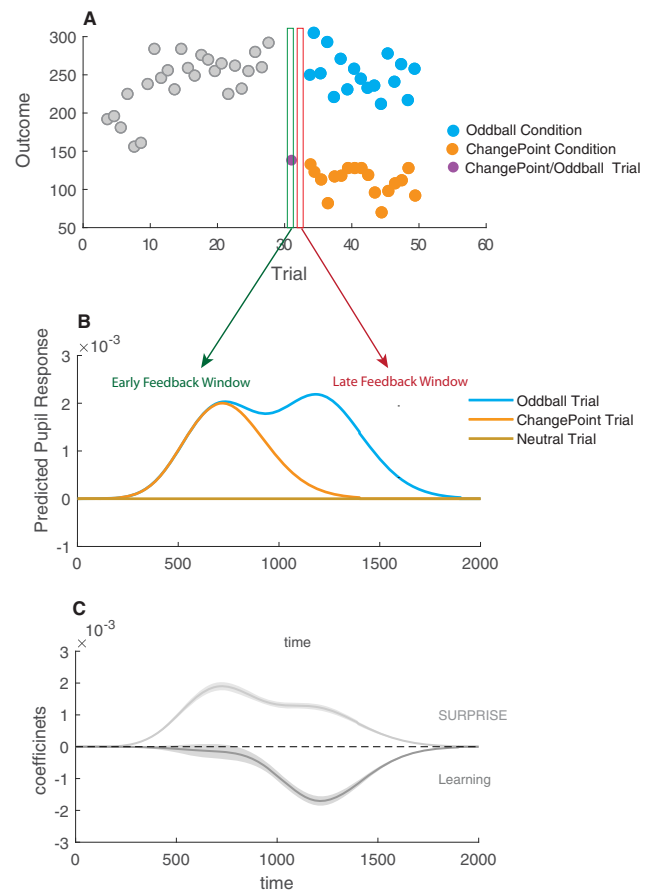


Figure 7. Pupil responses simulated to reflect context shifts at multiple timepoints within a trial positively reflect surprise and negatively reflect learning across changepoint and oddball conditions. **A**, An example set of outcomes (ordinate) over trials (abscissa) is depicted to demonstrate the key difference between the changepoint (orange) and oddball (cyan) generative structures. Based on our dynamic and network-based context shift model predictions, a surprising outcome is accompanied by a context shift in both the changepoint and oddball conditions (green box). A second context shift is predicted to happen only in the oddball condition in the intertrial interval after experiencing an oddball event (red box), corresponding to the expected return to the more typical context (cyan points). **B**, Predicted pupil responses (ordinate) are plotted over time (abscissa) for three trial types (colors). Pupil responses were simulated as the convolution of a gamma function with the expected context shift on each trial at two discrete time points. The first occurred at 400 ms after observing the outcome, and context shifts at this time point were proportional to changepoint probability/oddball probability in our model; the second time point was at 900 ms after the outcome when subjects would be expected to begin preparing a prediction for the next trial outcome, the context shifts at this time point were proportional to the intertrial interval context shifts necessary to return to the “typical” context after an oddball trial. Based on the predictions of our model, a context shift should occur at the first time point in both changepoint and oddball trials, while a context shift at the second time point should only happen at the oddball condition. **C**, Simulated pupil responses positively reflect surprise early after feedback (light gray) but negatively reflect learning during a later time window (dark gray). Coefficients for learning and surprise were obtained by regressing simulated pupil responses onto an explanatory matrix that contained regressors capturing surprise (changepoint/oddball probability) and learning (dynamic trial-by-trial learning rate), as estimated by a reduced Bayesian model.

interpretation would be to think of this parameter as controlling the threshold of a nonlinear activation function that monitors conflict in the output layer of the network, which is the mechanistic role that it plays in our network.

Context representations from this dynamic model provide a mechanistic interpretation of cortical activity patterns that abruptly change during periods of rapid learning (Karlsson et al., 2012; Nassar et al., 2019a). The context shift signal, which allows

the model to adjust context representations dynamically to afford adaptive learning behaviors, provides a mechanistic interpretation for feedback-locked P300 signals that conditionally predict learning, and may also resolve a contradiction in different studies examining the relationship between pupil dilation and learning (Nassar et al., 2019b). The true test of this model will come from future experiments to validate the new predictions of the current model.

The input layer that our model uses for flexible learning builds on the notion of latent states for representation learning. Through this lens, our work can be thought of as an extension to a larger body of research on structure learning, much of which has focused on identifying commonalities across stimulus categories (Gershman and Niv, 2010; Collins and Frank, 2013). In cases where temporal dynamics have been explored, the focus has been on the degree to which latent states allow efficient pooling of information across similar contexts that are separated in time (Gershman et al., 2010; Collins and Frank, 2013; Wilson et al., 2014). Here we highlight another advantage of using temporal dynamics to control active state representations: efficient partitioning of information in time to prevent interference. In addition to highlighting this advantage, our results highlight a shared anatomic basis for state representations across different types of tasks. Patterns of input layer activity in our model transition rapidly after changepoints to facilitate adaptive learning, much like network reset phenomena that have been observed in mPFC in rodents and OFC in humans (Karlsson et al., 2012; Nassar et al., 2019b). Rapid transitions in OFC are particularly interesting given that this area has been suggested to represent latent states for sharing knowledge across common structures (Wilson et al., 2014; Schuck et al., 2016).

The context shifts in our model, and network reset phenomena in the brain, relate to temporal context models previously explored in the episodic memory literature. In such models, the passage of time and observation of stimuli cause gradual changes in temporal context (Franklin et al., 2020; Howard and Kahana, 2002; Kornysheva et al., 2019; Polyn et al., 2009; Shankar et al., 2009) and more recently these models have been extended to include discontinuous changes at event boundaries, often inferred through surprise (Zacks et al., 2007; DuBrow and Davachi, 2013; DuBrow et al., 2017; Rouhani et al., 2020). Our dynamic context shift model takes a similar approach, using surprise, as indexed by the probability of an unexpected event (changepoint/oddball), to control context shifts. Such probabilities can be inferred using a Bayesian learning model calibrated to the environmental structure; however, we show that they could also be estimated from the output layer of our network itself. Previous work has suggested that changepoint and oddball probability are reflected by BOLD activations in both cortical and subcortical regions (Yu and Dayan, 2005; Nassar et al., 2012; O'Reilly et al., 2013; McGuire et al., 2014; D'Acremont and Bossaerts, 2016; Meyniel and Dehaene, 2017; Nassar et al., 2019a; Kao et al., 2020). While such signals have previously been interpreted as early-stage computations performed in the service of computing a learning rate, our work suggests that they serve another purpose, namely in signaling the need to change the active context representation.

Of particular interest in this regard is the feedback-locked P300 signal, an EEG-based correlate of surprise in humans (Mars et al., 2008; Kolossa, 2016; Kopp et al., 2016). A recent study showed that this signal positively related to learning in a changing environment and negatively related to learning in one containing oddballs (Nassar et al., 2019b). Here we show that the

context shift variable in our dynamic model has the exact same bidirectional relationship to learning. In our model, this reflects a causal relationship, whereby context transitions that persist in the changepoint condition lead new observations to have greater behavioral impact (i.e., more learning; Fig. 6E), and transient context transitions in the oddball condition limit the behavioral impact of oddball events by associating them with a different context from the one in which predictions are generated (i.e., less learning; Fig. 6E).

This bidirectional adjustment of learning rate is a key prediction of our model. We also predict that other physiological measures of surprise that have previously been related to learning, such as pupil diameter, should also provide similar results in environments with different sources of surprising outcomes. However, a key difference of pupil dilation predictions is that given the slow time course of the pupil signal, we predict that it will aggregate multiple state transitions that can occur on an oddball trial (i.e., the transition away from the original state to a new one, and the transition back to the original state). This aspect of the signaling predicts heightened pupil dilations on oddball relative to changepoint trials, which agrees qualitatively with previous observations (O'Reilly et al., 2013) and may help to resolve confusion in the existing literature regarding the relationship between pupil dilations and behavioral adjustment (Nassar et al., 2012; O'Reilly et al., 2013). Our model predicts that such a signal should also drive changes in state representations in OFC. This prediction, at least in part, is consistent with another recent experiment on neuromodulatory control of uncertainty (Muller et al., 2019), in which the strength of pupil dilation predicts the level of uncertainty regarding the current state of the environment, represented in medial orbitofrontal cortex. Our model predicts that these relationships should also depend on the task structure, with state transitions driving OFC representations toward an alternative state in reversal tasks (Muller et al., 2019) toward a completely new persisting state in changepoint tasks (Nassar et al., 2019a) and toward a transient state after oddball events. These relationships between state transition signals and neural representations have yet to be measured across the range of contexts that would be necessary to fully test the predictions of our model, and thus are interesting avenues for future empirical work.

Our model opens the door for a number of future investigations. We catered our analysis to the behavioral experiments of Nassar et al. (2019a,b), and therefore only considered changepoint and oddball conditions but did not study the case where context could either shift to a new context or return to a previous context it has learned before. Recognizing that a new observation actually comes from a previously learned context would involve additional pattern recognition and memory retrieval mechanisms (Redish et al., 2007), which might be thought of as part of a more general model-based inference framework as described above (Whittington et al., 2019; Franklin et al., 2020). That is to say, to solve all types of real-world problems, our model would be required to know not only that an observation is different from the recent past, but also which previously encountered state would provide the best generalization to this new situation. Doing so effectively would require organization of states based on similarity, such that similar states shared learning to some degree, in the same way that states that occur nearby in time pool learning in our current model.

Model limitations

The design of our network has several limitations that would need to be overcome to fully realize the potential of our

overarching framework. The first is that our network was endowed with knowledge of the task transition structure, raising an important question for future work as to how this structure could be learned directly from observations. In our tasks, the transition structure differed between changepoints and oddballs, with changepoints promoting persisting state representations and oddballs promoting an immediate transition back to the previous state; however, real-world learning occurs in a much more diverse set of environments, where simultaneously learning transition structure and applying it to guide behavioral adjustment would be challenging to say the least.

A second set of limitations stems from our simplified ring organization of the input (context) layer of our network. This simplification causes potential issues for the oddball condition we model, in that future contexts could rely on the same input units that were previously associated with oddball events. In our simplified network, we solved this problem through slow weight decays that slowly turn unused input units into blank slates for future learning. However, we suspect that the brain uses a different solution, namely a more complex organization of context representations—for example, if the input layer were two dimensional, with one dimension corresponding to slow drifts and the other corresponding to oddball events, an oddball context could never be encountered with any amount of drift.

Summary

In summary, we suggest that flexible learning emerges from dynamic internal context representations that are updated in response to surprising observations in accordance with task structure. Our model requires representations consistent with those that have previously been observed in orbitofrontal cortex as well as the state transition signals necessary to update them. We suggest that biological signals previously thought to reflect “dynamic learning rates” actually signal the need for internal state transitions, and our model provides the first mechanistic explanation for the context dependence with which these signals relate to learning. Together, our results support the notion that adaptive learning behaviors may arise through dynamic control of representations of task structure.

References

- Adams RP, MacKay DJC (2007) Bayesian online changepoint detection. arXiv:0710.3742.
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
- Bernacchia A, Seo H, Lee D, Wang X-J (2011) A reservoir of time constants for memory traces in cortical neurons. *Nat Neurosci* 14:366–372.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624–652.
- Bouret S, Sara SJ (2005) Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci* 28:574–582.
- Browning M, Behrens TE, Jochem G, O'Reilly JX, Bishop SJ (2015) Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 18:590–596.
- Cockburn J, Frank M (2013) Reinforcement learning, conflict monitoring, and cognitive control: an integrative model of cingulate-striatal interactions and the ERN. In: *Neural basis of motivational and cognitive control* (Mars RB, Sallet J, Rushworth MFS, Yeung N, eds), pp 310–331. Cambridge, MA: MIT.
- Collins AGE, Frank MJ (2013) Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol Rev* 120:190–229.
- d'Acremont M, Bossaerts P (2016) Neural mechanisms behind identification of leptokurtic noise and adaptive behavioral response. *Cereb Cortex* 26:1818–1830.
- Donahue CH, Lee D (2015) Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nat Neurosci* 18:295–301.
- DuBrow S, Davachi L (2013) The influence of context boundaries on memory for the sequential order of events. *J Exp Psychol Gen* 142:1277–1286.
- DuBrow S, Rouhani N, Niv Y, Norman KA (2017) Does mental context drift or shift? *Curr Opin Behav Sci* 17:141–146.
- Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, Soltani A (2017) Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron* 94:401–414.e6.
- Farashahi S, Donahue CH, Hayden BY, Lee D, Soltani A (2019) Flexible combination of reward information across primates. *Nat Hum Behav* 3:1215–1224.
- Fischer AG, Ullsperger M (2013) Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron* 79:1243–1255.
- Franklin NT, Norman KA, Ranganath C, Zacks JM, Gershman SJ (2020) Structured event memory: a neuro-symbolic model of event cognition. *Psychol Rev* 127:327–361.
- Gershman SJ, Niv Y (2010) Learning latent structure: carving nature at its joints. *Curr Opin Neurobiol* 20:251–256.
- Gershman SJ, Blei DM, Niv Y (2010) Context, learning, and extinction. *Psychol Rev* 117:197–209.
- Howard MW, Kahana MJ (2002) A Distributed Representation of Temporal Context. *J Math Psychol* 46:269–299.
- Iigaya K (2016) Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *Elife* 5:e18073.
- Jepma M, Murphy PR, Nassar MR, Rangel-Gomez M, Meeter M, Nieuwenhuis S (2016) Catecholaminergic regulation of learning rate in a dynamic environment. *PLoS Comput Biol* 12:e1005171.
- Jepma M, Brown SBRE, Murphy PR, Koelewijn SC, de Vries B, van den Maagdenberg AM, Nieuwenhuis S (2018) Noradrenergic and cholinergic modulation of belief updating. *J Cogn Neurosci* 30:1803–1820.
- Joshi S, Gold JI (2020) Pupil size as a window on neural substrates of cognition. *Trends Cogn Sci* 24:466–480.
- Kao C-H, Khambhati AN, Basset DS, Nassar MR, McGuire JT, Gold JI, Kable JW (2020) Functional brain network reconfiguration during learning in a dynamic environment. *Nat Commun* 11:1682.
- Karlsson MP, Tervo DGR, Karpova AY (2012) Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338:135–139.
- Kolossa A (2016) A new theory of trial-by-trial P300 amplitude fluctuations. In: *Computational modeling of neural activities for statistical inference*, pp 41–69. Cham, Switzerland: Springer.
- Kopp B, Seer C, Lange F, Kluytmans A, Kolossa A, Fingscheidt T, Hooijink H (2016) P300 amplitude variations, prior probabilities, and likelihoods: a Bayesian ERP study. *Cogn Affect Behav Neurosci* 16:911–928.
- Kornysheva K, Bush D, Meyer SS, Sadnicka A, Barnes G, Burgess N (2019) Neural competitive queuing of ordinal structure underlies skilled sequential action. *Neuron* 101:1166–1180.e3.
- Li YS, Nassar MR, Kable JW, Gold JI (2019) Individual neurons in the cingulate cortex encode action monitoring, not selection, during adaptive decision-making. *J Neurosci* 39:6668–6683.
- Mars RB, Debener S, Gladwin TE, Harrison LM, Haggard P, Rothwell JC, Bestmann S (2008) Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *J Neurosci* 28:12539–12545.
- Massi B, Donahue CH, Lee D (2018) Volatility facilitates value updating in the prefrontal cortex. *Neuron* 99:598–608.e4.
- Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5:39.
- McGuire JT, Nassar MR, Gold JI, Kable JW (2014) Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* 84:870–881.
- Meyniel F, Dehaene S (2017) Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc Natl Acad Sci U S A* 114:E3859–E3868.
- Muller TH, Mars RB, Behrens TE, O'Reilly JX (2019) Control of entropy in neural models of environmental state. *Elife* 8:e39404.
- Nassar MR, Troiani V (2021) The stability flexibility tradeoff and the dark side of detail. *Cogn Affect Behav Neurosci* 21:607–623.

- Nassar MR, Wilson RC, Heasly B, Gold JI (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30:12366–12378.
- Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci* 15:1040–1046.
- Nassar MR, McGuire JT, Ritz H, Kable JW (2019a) Dissociable forms of uncertainty-driven representational change across the human brain. *J Neurosci* 39:1688–1698.
- Nassar MR, Bruckner R, Frank MJ (2019b) Statistical context dictates the relationship between feedback-related EEG signals and learning. *Elife* 8:e46975.
- Nassar MR, Waltz JA, Albrecht MA, Gold JM, Frank MJ (2021) All or nothing belief updating in patients with schizophrenia reduces precision and flexibility of beliefs. *Brain* 144:1013–1029.
- Nieuwenhuis S, De Geus EJ, Aston-Jones G (2011) The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology* 48:162–175.
- O'Reilly JX (2013) Making predictions in a changing world-inference, uncertainty, and learning. *Front Neurosci* 7:1–10.
- O'Reilly JX, Schuffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc Natl Acad Sci U S A* 110:E3660–E3669.
- Polyn SM, Norman KA, Kahana MJ (2009) A context maintenance and retrieval model of organizational processes in free recall. *Psychol Rev* 116:129–156.
- Redish AD, Jensen S, Johnson A, Kurth-nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychol Rev* 114:784–805.
- Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS (2016) Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat Commun* 7:13289.
- Rouhani N, Norman KA, Niv Y, Bornstein AM (2020) Reward prediction errors create event boundaries in memory. *Cognition* 203:104269.
- Schuck NW, Cai MB, Wilson RC, Niv Y (2016) Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* 91:1402–1412.
- Shankar K, Jagadisan U, Howard M (2009) Sequential learning using temporal context. *J Math Psychol* 53:474–485.
- Vazey EM, Aston-Jones G (2014) Designer receptor manipulations reveal a role of the locus coeruleus noradrenergic system in isoflurane general anesthesia. *Proc Natl Acad Sci U S A* 111:3859–3864.
- Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TEJ (2019) The Tolman-Eichenbaum Machine: unifying space and relational memory through generalisation in the hippocampal formation. *BioRxiv* 770495. doi: [10.1101/770495](https://doi.org/10.1101/770495).
- Wilson RC, Nassar MR, Gold JI (2010) Bayesian online learning of the hazard rate in change-point problems. *Neural Comput* 22:2452–2476.
- Wilson RC, Nassar MR, Gold JI (2013) A mixture of delta-rules approximation to Bayesian inference in change-point problems. *PLoS Comput Biol* 9:e1003150.
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81:267–279.
- Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46:681–692.
- Yu LQ, Wilson RC, Nassar MR (2021) Adaptive learning is structure learning in time. *Neurosci Biobehav Rev* 128:270–281.
- Zacks JM, Speer NK, Swallow KM, Braver TS, Reynolds JR (2007) Event perception: a mind-brain perspective. *Psychol Bull* 133:273–293.