Adaptive learning through temporal dynamics of state representation

Niloufar Razmi[1], Matthew R. Nassar[2,3]

[1]Psychiatry and Behavioral Sciences Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

[2]Robert J. & Nancy D. Carney Institute for Brain Science, Brown University, Providence RI 02912-1821, USA

[3]Department of Neuroscience, Brown University, Providence RI 02912-1821, USA

Abstract

Humans adjust their learning rate according to local environmental statistics, however existing models of this process have failed to provide mechanistic links to underlying brain signals. Here, we implement a neural network model that uses latent variables from Bayesian inference to shift a neural context representation that controls the "state" to which feedback is associated. Within this model, behavioral signatures of adaptive learning emerge through temporally selective transitions in active states, which also mimic the evolution of neural patterns in orbitofrontal cortex. Transitions to a previous state after encountering a one-off outlier reduce learning, as observed in humans, and provide a mechanistic interpretation for bidirectional learning signals, such as the p300, that relate to learning differentially according to the source of surprising events. 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.

Introduction

People and animals exist in complex environments and are often required to update behavior in the face of new information. While standard supervised learning or reinforcement learning models have shown great success in performing particular tasks and explaining general trends in behavior, they lack the flexibility of biological systems, which seem to adjust the influence of new information dynamically, especially in environments that evolve over time (Behrens, Woolrich, Walton, & Rushworth, 2007; Donahue & Lee, 2015; Farashahi, Donahue, Hayden, Lee, & Soltani, 2019; Li, Nassar, Kable, & Gold, 2019; Massi, Donahue, & Lee, 2018; Nassar & Gold, 2010). Recent advances in understanding these adaptive learning behaviors have relied on probabilistic modeling to better understand the computational problems that organisms face for survival in their everyday life (Soltani & Izquierdo, 2019).

Bayesian probability theory has been extensively applied to describing adaptive learning algorithms in changing environment to provide normative accounts for learning behavior. These computational models prescribe learning that is more rapid in the case of volatility and slower during periods of stability in the environment (Adams & MacKay, 2007; Behrens et al., 2007; Nassar & Gold, 2010; Wilson, Nassar, & Gold, 2010) Specifically, the behavior of human subjects in probabilistic learning and decision-making tasks has been successfully modeled by using quantitative terms representing the level of uncertainty (Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015; Muller, Mars, Behrens, & O'Reilly, 2019) and the probability of detecting a changepoint (Adams & MacKay, 2007; Nassar, Wilson, Heasly, & Gold, 2010). In this framework, human brain is viewed as implementing an optimal learning algorithm that embody the statistical properties of the world it operates in (Meyniel & Dehaene, 2017; O'Reilly, 2013). While probabilistic modeling has provided a normative account for many of the adjustments in learning rate observed in humans and animals (Behrens et al., 2007; Nassar, Bruckner, & Frank, 2019; Nassar & Gold, 2010), it can be closely approximated by many qualitatively different algorithms (Bernacchia, Seo, Lee, & Wang, 2011; Farashahi et al., 2017; Iigaya, 2016; Mathys, Daunizeau,

Friston, & Stephan, 2011; Nassar et al., 2010; Wilson, Nassar, & Gold, 2013; Yu & Dayan, 2005), making the search for its neural basis difficult. Furthermore, while a number of neural mechanisms for adaptive learning in changing environments have been proposed, many of these models are limited in the range of behaviors they can explain. For example, reservoir of time constants and metaplasticity approaches are capable of explaining slow adjustments in learning behavior, but not rapid adjustments that occur after fundamental changes in the environment (Bernacchia et al., 2011; Farashahi et al., 2017; Mathys et al., 2011). But even models that explicitly adjust learning according to the surprise elicited by such environmental changes (Iigaya, 2016; Nassar & Gold, 2010) would need to use fundamentally different rules to capture behavior in environments where surprise is evoked by one-off outlier events (oddballs) rather than fundamental changes (changepoints)(D'Acremont & Bossaerts, 2016). To make matters worse, brain signals that predict more learning in changing environments (Behrens et al., 2007; Jepma et al., 2016; McGuire, Nassar, Gold, & Kable, 2014; Nassar et al., 2010; O'Reilly et al., 2013), do not do so consistently across different statistical conditions (D'Acremont & Bossaerts, 2016), in some cases corresponding to more learning in changing environments but less learning in those containing oddballs (Nassar, Bruckner, et al., 2019). Thus, while there has been a good deal of work exploring potential neural mechanisms for adaptive learning, to date the algorithms that have been explored have lacked the generality necessary to capture human behavior and have failed provide an understanding of corresponding brain signaling.

In order to capture the broader scope of adaptive learning behaviors, we consider the framework of structure learning, which has been applied extensively to understand how learning is attributed to different features of stimuli. Within this framework, various stimuli can be thought of as emerging from a smaller number of latent states, which in turn provide a more efficient substrate for learning in the brain (A. G. E. Collins & Frank, 2013; Gershman & Niv, 2010). In principle, effective latent state representations could facilitate transfer of information from one situation to another analogous one, but also prevent information from unrelated sources to lead to interference (A. Collins & Koechlin, 2012). Previous studies have identified candidate neural implementations of this latent state representation, most notably within the orbitofrontal cortex (Schuck, Cai, Wilson, & Niv, 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014). While learning structure in time itself could provide a useful framework for thinking about adaptive learning, structure learning has typically focused on understanding relationships between stimuli, rather than over time, and with few exceptions there has been very little crosstalk between adaptive learning and structure learning fields (A. Collins & Koechlin, 2012; Franklin, Norman, Ranganath, Zacks, & Gershman, 2020; Gershman, Blei, & Niv, 2010; Lloyd & Leslie, 2013; Wilson et al., 2014).

Nonetheless, the question naturally arises: could latent states that represent temporal structure explain adaptive learning behavior? If so, what sort of neural representations would it require? Here we examine these questions using a feed-forward neural network model that maps state representations onto a continuous action space in order to performs a predictive inference task. We show that the effective learning rate of the model is increased when state representations evolve over time, and that better model performance can be achieved when state transitions are elicited by surprising events. Furthermore, we show that state transitions can speed learning after changepoints, or slow them after oddball events, under the assumption that appropriate state transitions occur between trials. In addition to providing a more general mechanistic understanding of behavior, our model provides insight into neural implementation. Efficient state

transition dynamics in the model mimic patterns of representational change that have been observed in orbitofrontal cortex (Nassar, McGuire, Ritz, & Kable, 2019), an area previously suggested represent latent states. The state transition signals required by our model also provide the first mechanistic explanation of prominent bidirectional learning signals, such as the feedback-locked p300, which signal greater learning in the presence of changepoints but less learning in the presence of oddballs. Taken together, our results support a latent-state transition model of adaptive learning behavior and suggest more generally that adjustments in learning might more accurately be described as structure learning in time.

Results

Predictive inference task to measure adaptive learning:

In order to test whether changes to latent state representations can facilitate adaptive learning behavior we modeled a predictive inference task previously performed by human participants (figure 1). In the task a helicopter, which is hidden behind clouds, drops visible bags containing valuable contents from the sky (figure 1a, right). On each trial, subjects move 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 in order to update and improve their prediction (figure 1a, left; figure 1b orange arrow).

In the "changepoint" variant of the task, bag locations are sampled from a Gaussian distribution centered on the helicopter, which occasionally relocates to a new position on the screen. Such abrupt transitions in helicopter location lead to changes in the statistical context defining the bag locations (context shifts), which can be inferred by monitoring the size of prediction errors (figure 1b, red arrow). Therefore, the helicopter position functions as a latent variable that needs to be inferred from noisy observations (i.e. dropped bags) on each trial to yield optimal task performance.

Human behavior is well described by a normative learning model that forms and updates beliefs about a latent variable (helicopter position) according to prediction errors experienced with each new observation (bag position – belief about helicopter position) (Nassar & Gold, 2010; Nassar et al., 2012). The degree to which these prediction errors update beliefs, also referred to as the learning rate, is adjusted from trial-to-trial according two variables: changepoint probability and relative uncertainty. Changepoint probability (CPP) is the Bayesian probability that the helicopter has changed to a new location since last trial, and thus increases when prediction errors are large than expected. Relative uncertainty tracks uncertainty on the current helicopter position, and is thus maximal on the trial following a changepoint and decreases as a function of time thereafter as the model collects more information about the likely helicopter position (See methods). Figure 1b shows the responses of that model compared to an example human subject. Note that both the model and subjects go through periods of stable bucket position within contexts but update position rapidly at context changes.

Figure 1

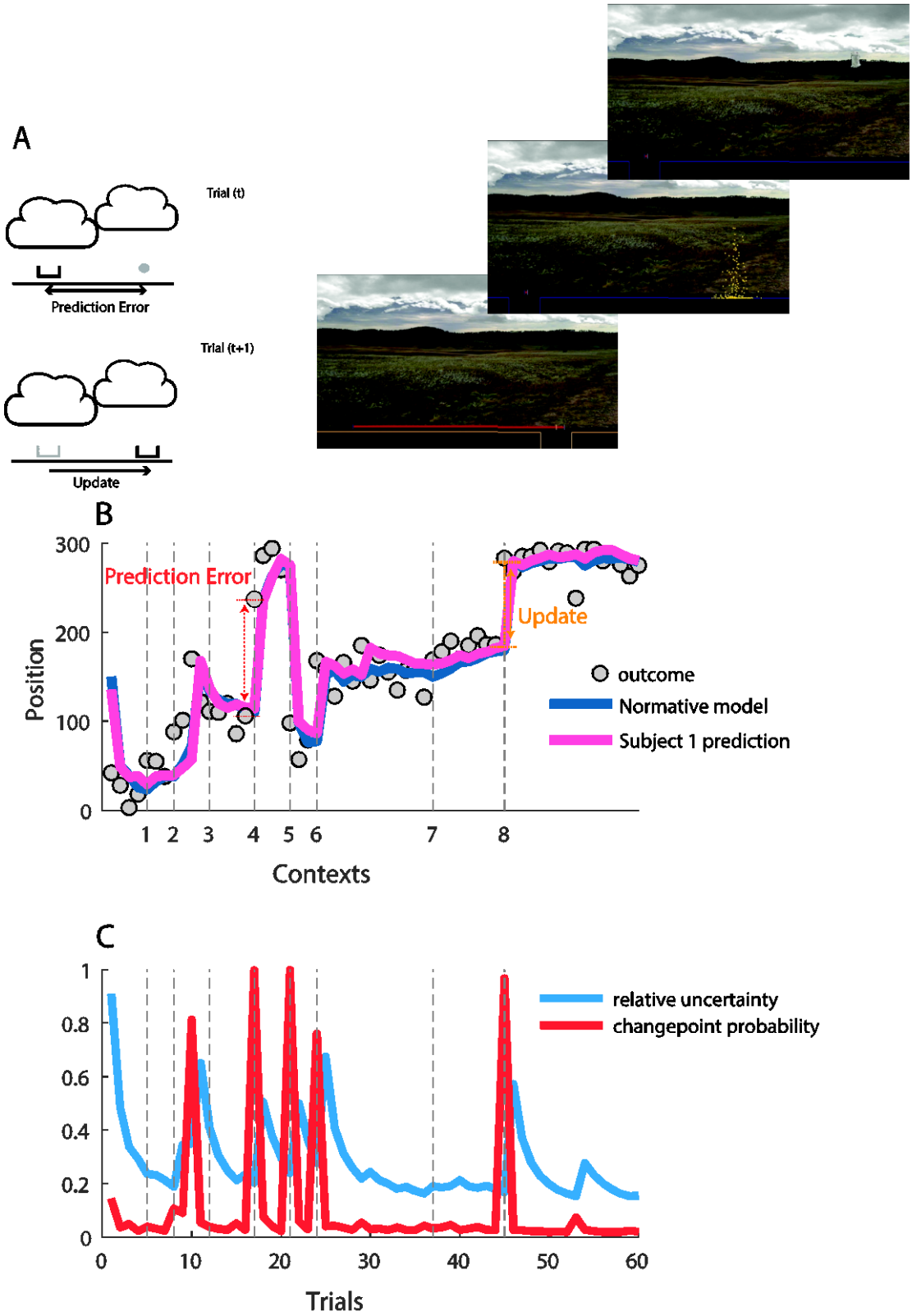
A
Trial (t)
Prediction Error
Trial (t+1)
Update
B
Prediction Error
Update
Position
300
200
100
0
1 2 3 4 5 6 7 8
Contexts
outcome
Normative model
Subject 1 prediction
C
1
0.8
0.6
0.4
0.2
0
0 10 20 30 40 50 60
Trials
relative uncertainty
changepoint probability

Figure 1: Predictive inference task to measure dynamics of adaptive learning.

A) Schematic Illustration of task design and screenshots of the task. Human subjects place a bucket at the bottom of the screen to catch a bag of coins subsequently dropped from a hidden helicopter. After observing the outcome at the end of each trial (prediction error), subject could improve their response by moving the bucket to a new location most likely to receive the bag in the subsequent trial (update). In the changepoint condition, the helicopter on occasion moves to a new location. B) The sequence of bag locations (outcome) is plotted over context shift designated by sudden changes in outcome mean (dotted vertical lines). Bucket placements made by a subject (pink) and normative model (navy) are shown with a representation of an example prediction and outcome. [Prediction = outcome (t) – estimate (t) and Update = estimate (t+1) – estimate (t)]. (C) The learning rate describing update magnitude of normative model depends on two factors, changepoint probability (CPP; red) and relative uncertainty (RU; blue), which combine to prescribe learning that is highest at changepoints (CPP) and decays slowly thereafter (RU).

Context shifts facilitate faster learning in a neural network

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 (figure 2b; see 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 on each 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 the neural network provided a flexible test bed for how different task representations might affect performance of the model. The input layer activity is modeled as a Gaussian activity bump with a mean denoting the position of the neuron with the strongest activity and a standard deviation denoting the width of the activity profile. The standard deviation is set to a relatively small constant value such that the majority of neurons in the input layer were inactive on each trial. The primary goal of this work was to understand how changes to the mean of the activity bump, across trials, affect learning performance.

The input layer could be flexibly controlled to provide different types of task representations that could, in theory, be used to guide supervised learning in the predictive inference task. In the simplest case, the input layer could represent the background stimulus present on the task, which did not differ across trials, and thus could be modeled as a mean activity bump in the input layer that was constant across all trials. The input layer could also potentially represent time; if the mean of the activity bump changes a fixed amount from one trial to the next, its activity would drift over time such that trials occurring close in time would be represented more similarly than those further separated in time. In the most complex case we considered, the input layer could represent the current "context" (figure 1b), or period of helicopter stability. In this case, the mean of the activity bump would only transition at changes of the helicopter position, and thus could be thought of as representing the latent state of the helicopter. We first modeled each of these possible input representations to better understand how they impact learning performance.

To examine how fixed stimulus and timing representations affect task performance, we examined performance of a set 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 where there was no overlap between the input layer representations on successive trials (individual 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 colored nodes in 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 zero (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 making the neuron corresponding to position of the helicopter the most heavily weighted in the decision. We observed that moderate shifts of in the input layer (context shifts) led to the best performance in our task (figure 2e), and that the effective learning rate describing the models behavior monotonically scaled with context shift (figure 2d). The performance advantage of moderate context shifts might be thought of intuitively as a tradeoff between different conditions in the task. As long as the helicopter is in the same location (i.e. same context) minimal network context shifts should lead to improved performance since they effectively average information over a greater number of bag locations within a context to more closely approximate their mean. However, after the helicopter moves to a new location, using the previously trained weights would bias bucket positions and lead to a larger prediction error, and thus larger context shifts would be preferred under such circumstances. This idea is reminiscent of the trade-off between stability and flexibility in classic delta-rule learning models where a high learning rate is desired in a volatile environment but not in a stable environment (figure 2e, red line).

Figure 2

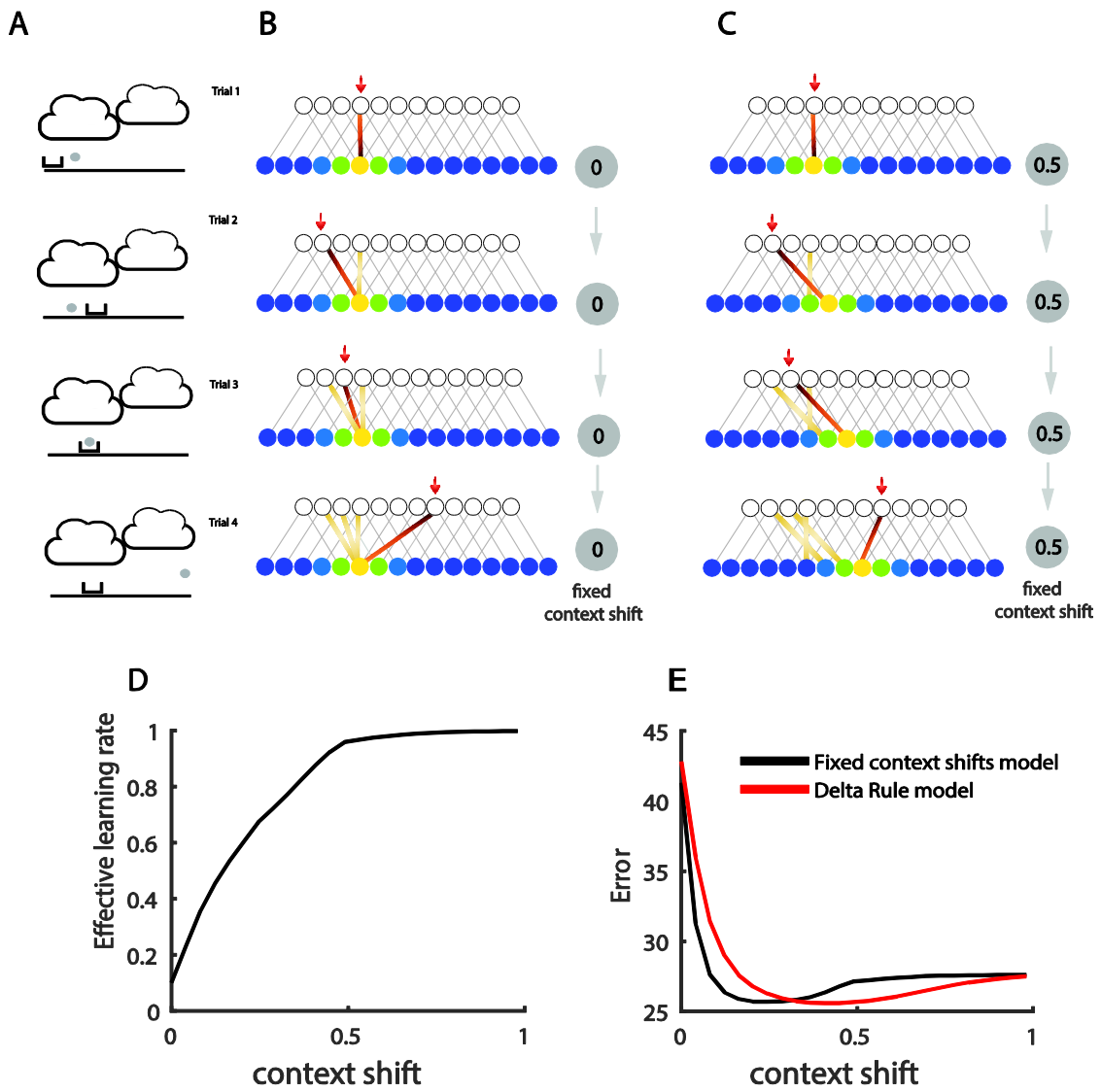

Figure 2: A neural network with fixed context shifts can approximate any constant learning rate. A-C) Network structure and weight updates for two fixed context shift models (B, C) are depicted across four example trials of a predictive inference task (A). For all networks, feedback was provided on each trial corresponding to the observed bag position (circle in panel A, red arrow in B&C) and weights of network were updated using supervised learning. Only a subset of neurons (circles) and connections between them (lines) are shown in neural network schematic. Activation in the input layer was normally distributed around a mean value that shifted rightward by a fixed amount on each trial (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 prediction error (ordinate) was minimized by neural network models (black line) that incorporated a moderate rate of fixed context shifts (abscissa). Mean error of a simple delta rule model

using various learning rates is shown in red (x-axis values indicate the context shift equivalent to the fixed delta rule learning rate derived from panel D).

Dynamic context shifts can improve task performance

Inspired by the observed relationship between context shift and accuracy, we next sought to build a modified version of the model that uses a dynamic context shift 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 (eg. allow the input layer to represent the latent state). Since such a model requires pre-existing knowledge of the outcome of a given trial, we refer to it as the ground truth context shift model (figure 3, top). Indeed, we observed that the ground truth model performance is better than any of the fixed context shift models after changepoints (figure 3c) as well as during periods of stability (figure 3d) leading to better overall performance (figure 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, utilizing information that the brain does have access to? To answer this question, we use the normative learning model, which takes the prediction errors and returns two latent variables changepoint probability and relative uncertainty. We hypothesized that since these computational variables have previously been shown to be correlated with BOLD responses in cortical regions as well as neuromodulatory systems (McGuire et al., 2014; Nassar et al., 2012), it is reasonable to believe the brain could make use this available information to further optimize its behavior. Thus, we built the dynamic context shifts model, which on each trial would shift the context proportional to its desired learning rate computed according to Bayesian estimates of changepoint probability (CPP) and relative uncertainty (RU). This model achieved good performance after changepoints by elevating context shifts (figure 3c) and during periods of stability by reducing context shifts (figure 3d). Across all trials, the dynamic context shift model achieved better performance than the best fixed context shift model ($t = 17.8,\ df = 31,\ p < 10^{-16}$), though understandably the performance of this model was not as good as the ground truth model (figure 3e). This result is consistent across different noise conditions (figure 3 – figure supplement 1).

Figure 3

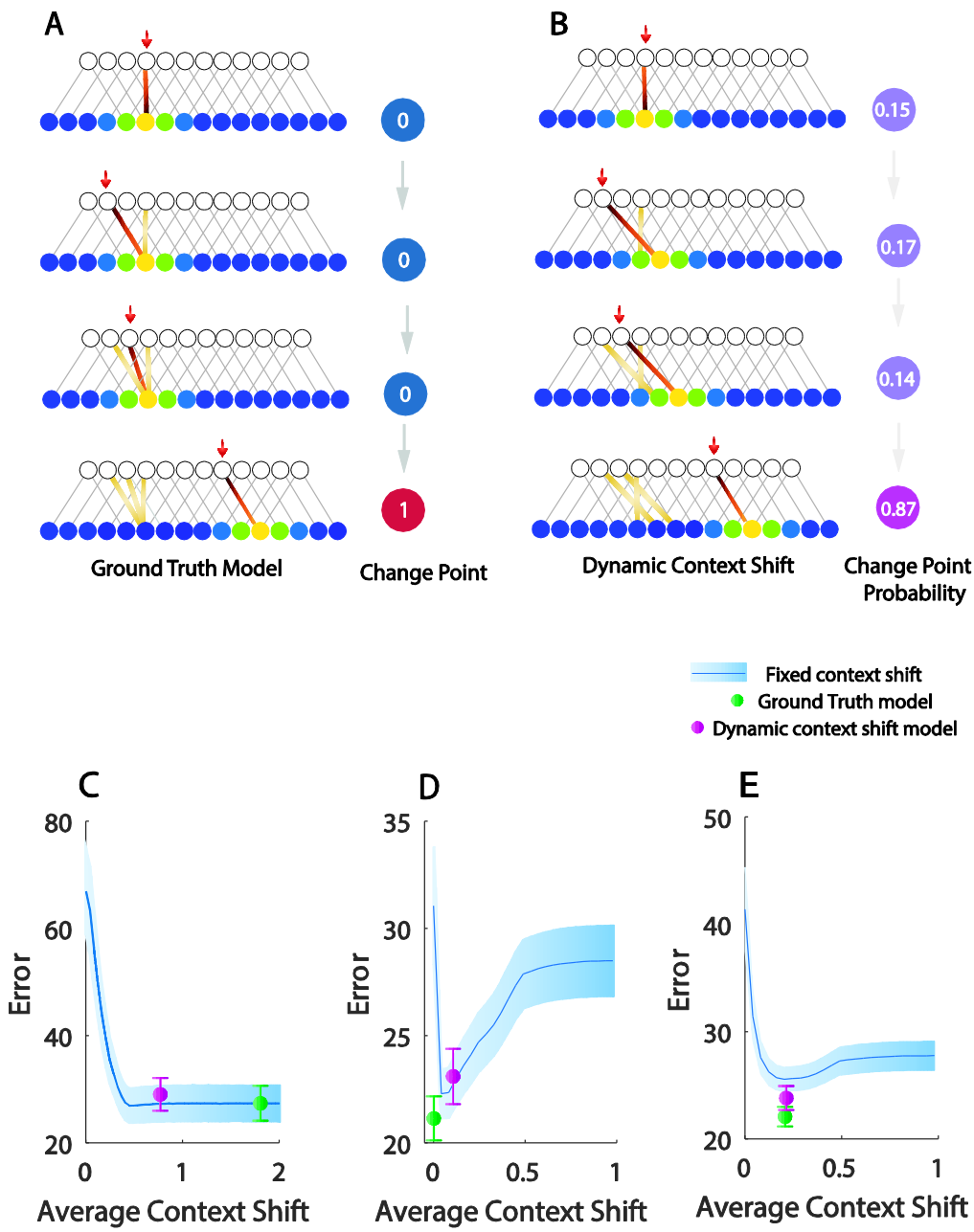

Figure -3: Dynamic context shifts facilitate better performance.

A) Schematic diagram of ground truth model network which is provided with objective information about whether a given trial is changepoint or not and uses that knowledge to shift the context only on change point trials. B) The dynamic context shift network uses subjective measures of CPP and RU from a normative learning model to adjust its context shift on each trial. Both models shift context to a greater degree on changepoint trials (bottom row) than on non-changepoint trials (top 3 rows). C) Performance on trials immediately following a changepoint was best for models that had the largest context shifts, but dynamic and ground truth context shift models were also able to achieve this level of performance. D) Performance on trials during periods of stability (> 5 trials after changepoint) was best for models with less context

shifting, but could also be achieved by the ground truth and dynamic context shift models. E) Overall, comparison of the best fixed context shift model and dynamic context shift model shows a better average performance of dynamic context shift model ($t = 17.8, df = 31, p < 10^{-16}$) while the best overall performance is achieved by the ground truth model.

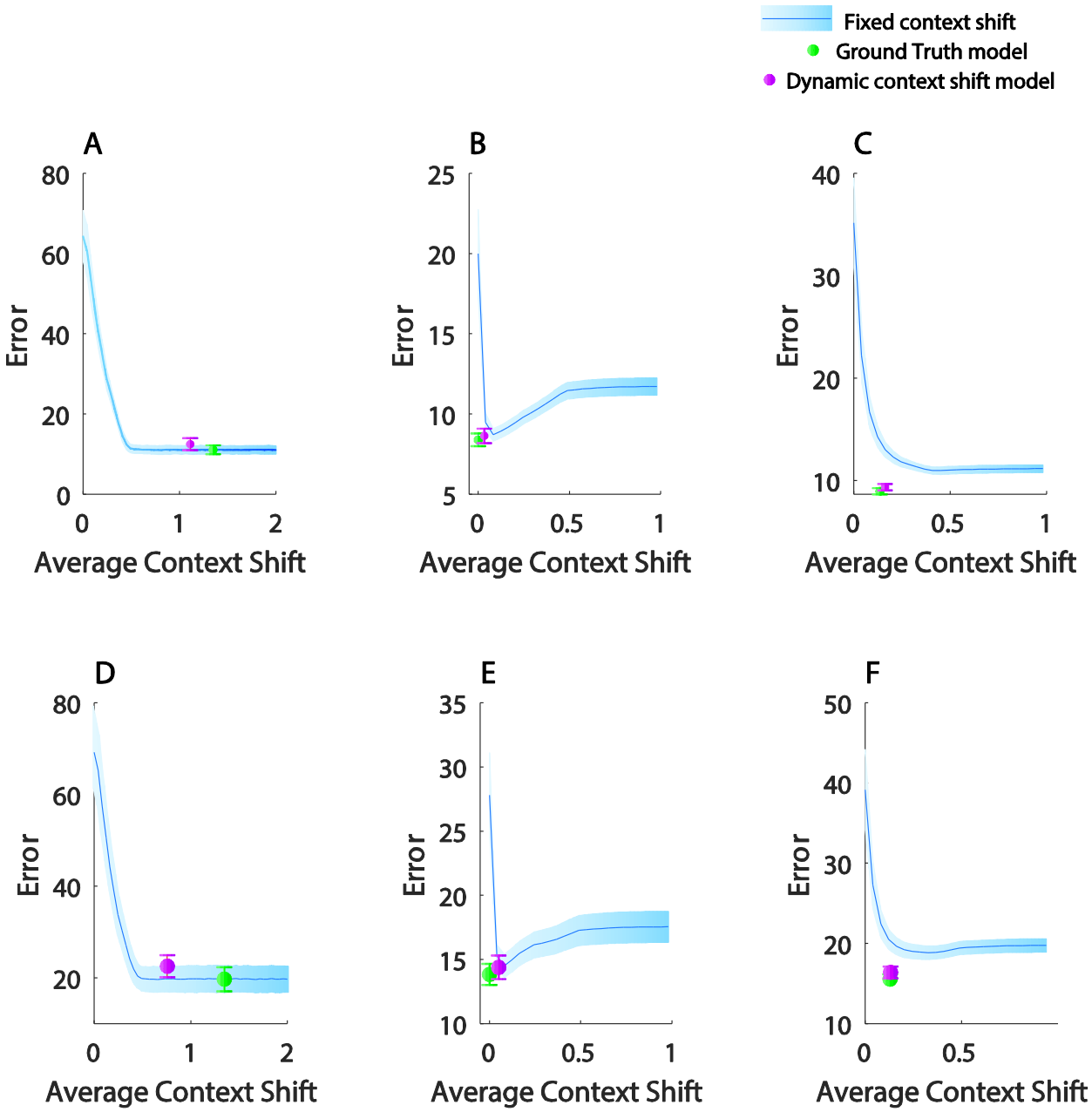

Figure 3- figure supplement 1. The same analysis as figure 3 c-e with different noise levels. Performance of the three types of models with a lower noise parameter (standard deviation of dropped bags locations around helicopter position), here noise=10 on A) trials immediately following a changepoint B) trials during periods of stability (> 5 trials after a changepoint) C) Averaged across all trials, the

performance of the dynamic context shift model is better than the best fixed context shift model ($t = 29.5, df = 31, p < 10^{-16}$ D-F). Similar analysis with a combination of the two different noise levels in different blocks (noise= 10 or 25) the same as what was used in the behavioral tasks with human participants in (McGuire et al., 2014). The same analysis for alternating noise levels condition ($t = 31.1, df = 31, p < 10^{-16}$) shows that the predictions of the dynamic model are consistent across different noise levels.

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 humans performing the task. Trial-by-trial updates made by the model were conditionally related the prediction errors it made, such that the model updated more according to prediction errors when the likelihood of a changepoint was high (figure 4a) or when the location of the helicopter was most uncertain (figure 4b). The slope of the relationship between updates and prediction errors provides us with an effective learning rate that can be used to quantitatively compare the behavior of the different models. Looking at this effective learning rate in more detail, we observe that immediately after a changepoint learning rate becomes maximal for the ground truth model and dynamic context shift models while gradually decreasing during the more stable periods. The effective Learning rate for fixed context shift model is not dependent on the number of trials after a changepoint. (figure 4c). A regression analysis that determined the contribution of changepoint probability and relative uncertainty to updates in each model (figure 4d) indicated that, like human subjects, the dynamic context shift model learned more rapidly when during periods of change or uncertainty (figure 4d). Taken together, these results suggest that our dynamic context shift model captures the primary behavioral features of adaptive learning in changing environments.

Higher trial-to-trial learning rates in humans are accompanied by rapid changes in the patterns of activation in orbitofrontal cortex (OFC; Nassar, McGuire, et al., 2019). Given that the dynamic context shift model accomplishes faster learning by changing the context representations, and that orbitofrontal cortex is thought to reflect a representations of internal context to guide reinforcement learning, we asked whether our context layer might show trial-to-trial patterns of representation consistent with what has been observed in OFC. We used the RSA approach applied in (Nassar, McGuire, et al., 2019) to create a dissimilarity matrix reflecting differences in the input layer activation across pairs of trials for our dynamic context shift model (figure 5a&b). Our model produces a pattern of dissimilarity for adjacent and nonadjacent trials that is nearly identical to the RSA matrix that was used to identify abrupt representational changes in periods of faster learning in the orbitofrontal cortex of humans (figure 5c&d). Taken together, these results are consistent with OFC encoding the dynamic context signal that is used in our model to adjust the rate of learning.

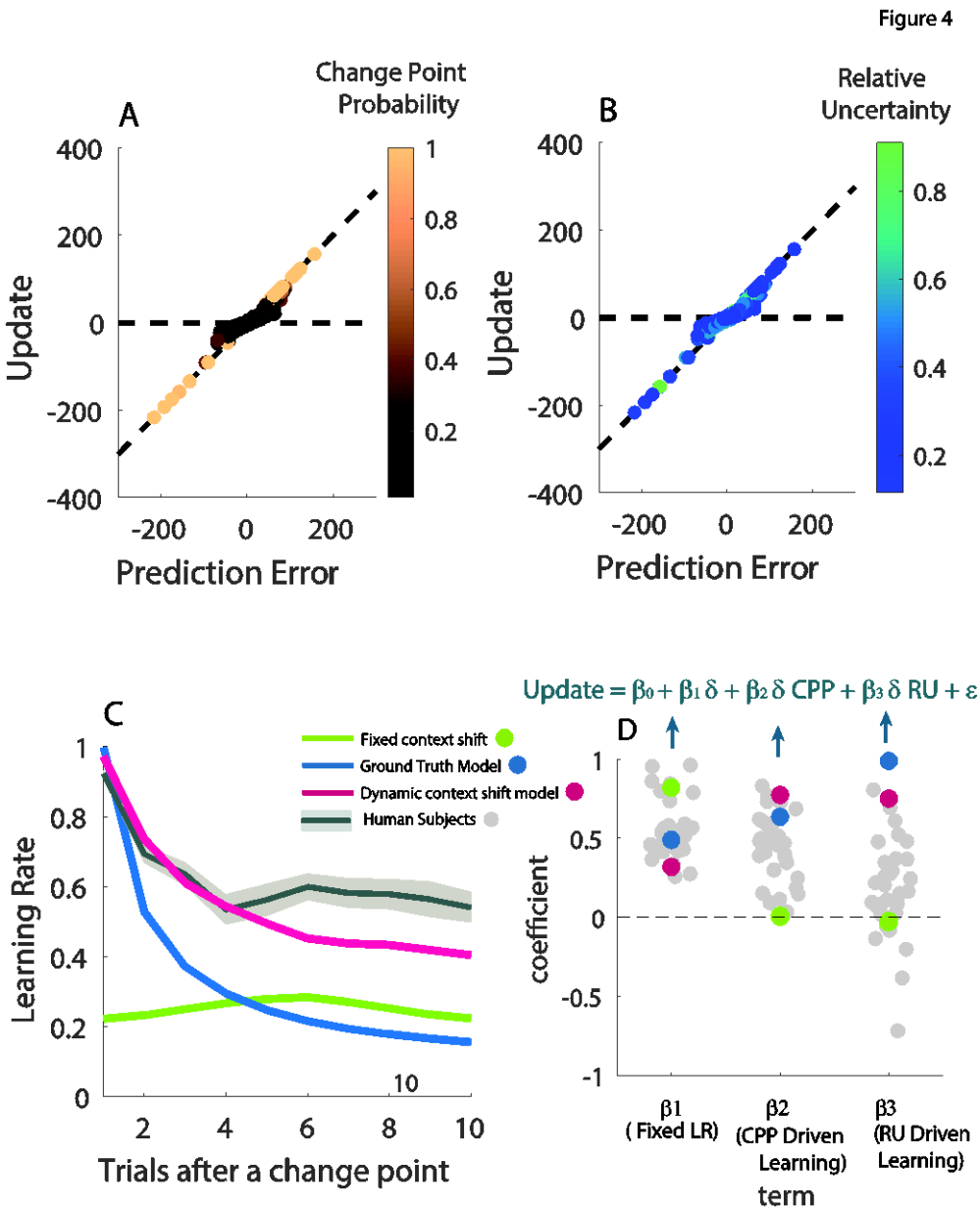

Figure 4: Dynamic context shifts facilitate adaptive learning. A&B: Dynamic context shift model single trial update (ordinate) is plotted against prediction error (abscissa) for each single trial of a simulated session with points colored according to the normative changepoint probability (A) and relative uncertainty (B). C) Effective learning rate as function of trial after change point. Both ground truth and dynamic context shifts models show adjustments in their effective learning rate relative to changepoints, with the dynamic context shift model closely approximating human subjects (gray). D) Coefficients from a regression model (top equation) fit to single trial updates in order to characterize the degree of overall learning (fixed LR), adjustments in learning at likely changepoints (CPP Driven Learning), and adjustments in learning according to normative uncertainty (RU driven learning). The grey circles represent data from human subjects doing the same task.

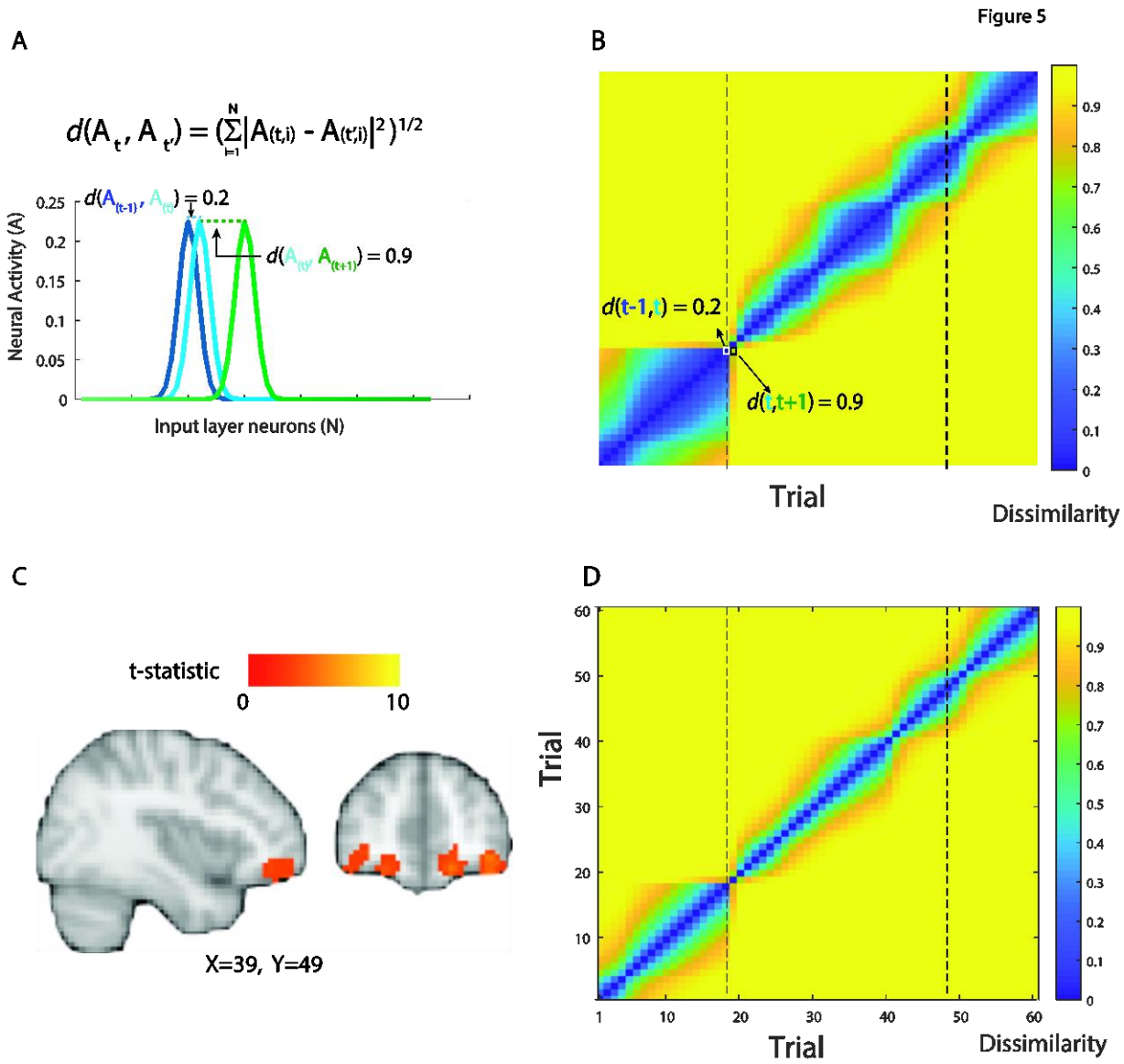

Figure 5: Dissimilarity in input layer representations across trials is increased across changepoints and resembles patterns of activity observed in orbitofrontal cortex.

(A) Dissimilarity in the input representation between pairs of trials was computed according to the Euclidean distance between those trials in the space of population activity. Trial (t) is an example changepoint. B) The dissimilarity matrix represent that input layer activity for trials within the same context (i.e. trial (t-1) and trial (t)) are more similar than for consecutive trials belonging to two different contexts (trial t and trial (t+1)). C) Regions within OFC display patterns of voxel activation that transition more rapidly during periods of rapid learning (reproduced from Nassar et al. 2019, Journal of Neuroscience). D) Dissimilarity matrix used for fMRI analysis that identified the brain regions in (C) for the same example block that were used

to characterize the dissimilarity in the input layer of the dynamic context shift model (B). Note the similarity between the two matrices (B&D), and in particular that in both cases trials separated by a changepoint (depicted by vertical dotted lines) are maximally dissimilar to one another. Across subjects the correlation between adjacent trial elements in the two matrices was consistently positive (mean[SEM] Pearson correlation = 0.884[0.011]).

Dynamic context shifts can facilitate less learning from oddball events

In order to understand how dynamic context shifts, analogous to dynamic inference of latent state, might be employed to improve learning in alternate statistical environments we considered a set of "oddball" generative statistics that have recently been employed to investigate neural signatures of learning (D'Acremont & Bossaerts, 2016; Nassar, Bruckner, et al., 2019). In the oddball condition, the mean of the output distribution does not abruptly change but instead gradually drifts at a constant rate. However, on occasion a bag might drop in an unexpected location uniformly sampled across the width of the screen, and is thus is outlier unrelated to both past and future outcomes. Thus, in the presence of oddballs, large prediction errors should lead to less, rather than more, learning. This normative behavior has been observed in human subjects (D'Acremont & Bossaerts, 2016; Nassar, Bruckner, et al., 2019).

In order to extend our model to this condition we created a network that adjusted active input units based on ground truth information regarding whether or not a given trial is an oddball (figure 6a right). Throughout the experiment, the model would shift the context with a small constant rate corresponding to the drift rate but on oddball trials, the model would undergo a large context shift to ensure that the oddball trial produced learning in a non-overlapping set of input layer neurons. The model was also endowed with knowledge of the transition structure of the task, which includes that oddballs tend to be followed by non-oddball trials, and as such, the input layer activity bump would transition to its previous non-oddball location subsequent to learning from the oddball outcome. Consequently, the results of the oddball trials would not be stored in the same context as the ordinary trials, but predictions would always be made from a non-oddball context.

Similar to our changepoint condition, we created versions of the model in which oddballs were inferred using a Bayesian inference model and their probabilities were used to guide transitions of the active input layer units. In the most direct reflection of the ground truth model, 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 non-oddball event (Dynamic *latent state context shift*). This model achieved better overall performance than the best fixed context shift model ($t = 5.8,\ df = 31,\ p = 1.7 \times 10^{-6}$), similar performance to the ground truth model, and better performance than two alternate models that used oddball probability and relative uncertainty to govern adjustments in context (Fig 6d).

Dynamic context shifts explain bidirectional learning signals observed in 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 & Ullsperger, 2013; Jepma et al., 2018, 2016), but negatively predicts behavioral adjustments in the oddball condition that we describe above (Nassar, Bruckner, et al., 2019). Here we hypothesize that the P300 serves to partition information collected in unrelated contexts by promoting dynamic context shifts. In our framework, the P300 would need to respond to surprising events that suggest the need for a new latent state representation and then promote the context shift in the input layer necessary to provide a blank slate for learning. Thus, a stronger P300 signal would produce a larger context shift, which would result in a stronger partition between current learning and previously learned associations. In changing environments, such a partition would 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 formalized this idea by measuring the effective learning rate in the dynamic context shift models for bins trials sorted according to the context shift that was experienced on them. As displayed in figure 6e, the measurement from the normative model which we used to control the size of our model context shifts reflect the same relationship with effective learning rate as the P300 signal in human behavioral results: context shifts in the changepoint condition correspond to faster learning while the same context shift signal in the oddball condition negatively corresponds to learning rate (figure 6f). Thus, our modeling suggests that the P300 may relate to learning indirectly by signaling or promoting transitions in an internal representation of latent state.

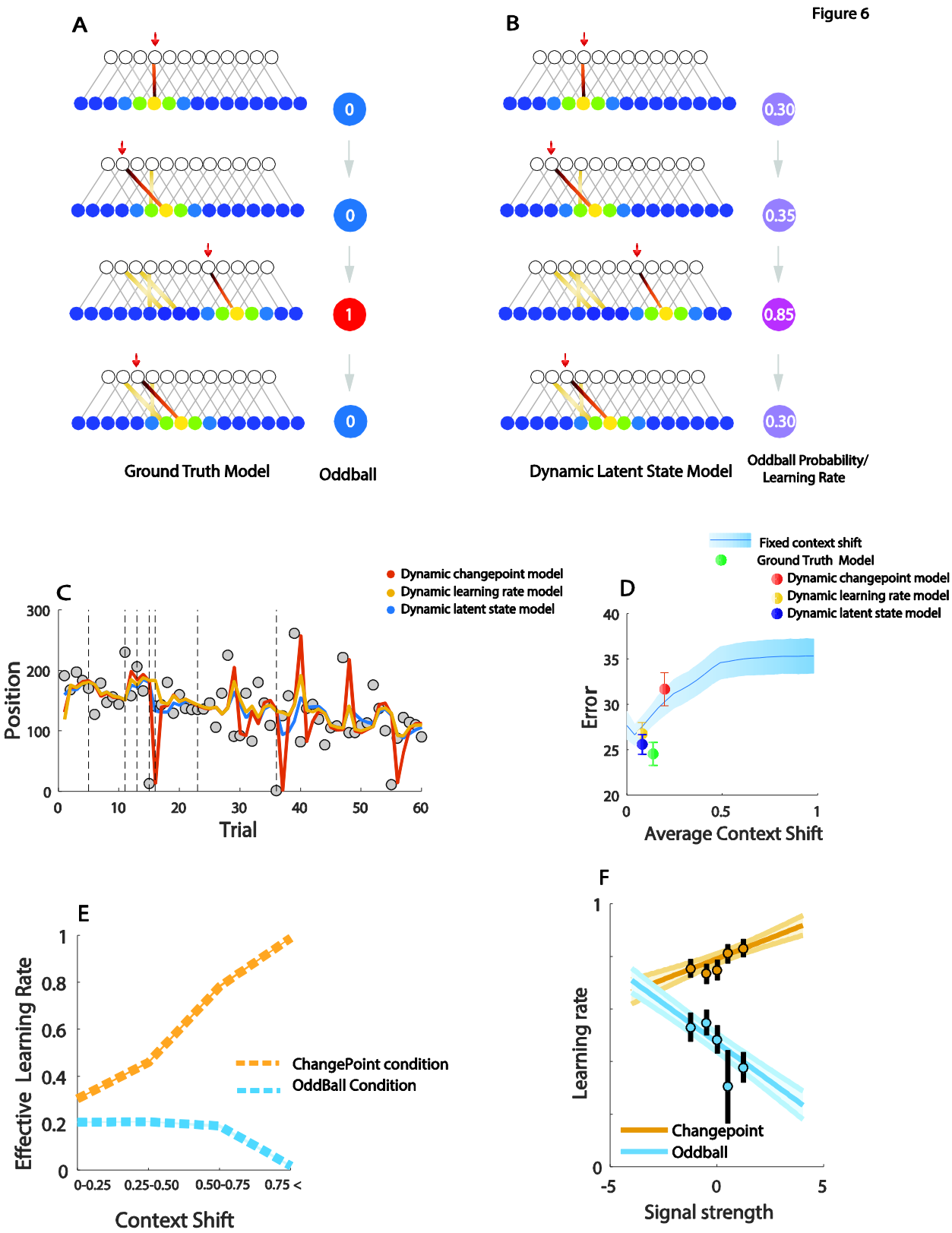

Figure 6 – Neural network structure is able to perform the oddball condition. A) Schematic representation of the ground truth model, which has a constant context shift proportionate to

the environmental drift. On oddball trials (third row) there is a large context shift, but context on next trial returns to its pre-oddball activity pattern. B) Schematic of the dynamic latent state context shift model, which on each trial shifts the context according to α_t term the same as the changepoint condition: OPP + RU – (ODD * RU) but after receiving the supervised learning signal from the outcome, returns back the context to where it was before and shift the context again according to the α_t specific to oddball condition: RU – (ODD * RU) Thus the effective context shift on each trial dictated by relative uncertainty on most trials, however a high oddball probability will cause the supervised learning signal to be stored in a completely separate context. C) Example predictions of the three candidate dynamic models (ordinate) across 60 trials (abscissa) of the oddball condition. Outcomes (eg. bag locations) are shown in grey circles. Note that the latent state formulation (blue), which parametrically changes context according to the probability of oddball events but transitions back to its previous context on the next trial, is less influenced by extreme outcomes than the other candidate models. D) The dynamic latent state model was more accurate than the dynamic learning rate context shift model ($t = 6.9,\ df = 31,\ p = 7.8 \times 10^{-8}$), the dynamic changepoint context shift model ($t = 19.5,\ df = 31,\ p < 10^{-16}$) and the best fixed context shift model ($t = 6.5,\ df = 31,\ p = 2.8 \times 10^{-7}$) but not as good as the ground truth model, in terms of performance across all simulations. E & F) dynamic latent state context shift model shows the same positive and negative relationship of effective learning rate and measures of context shifts in changepoint and oddball trials respectively as the relationship between learning rate and P300 surprise signal in human subjects during the same task.

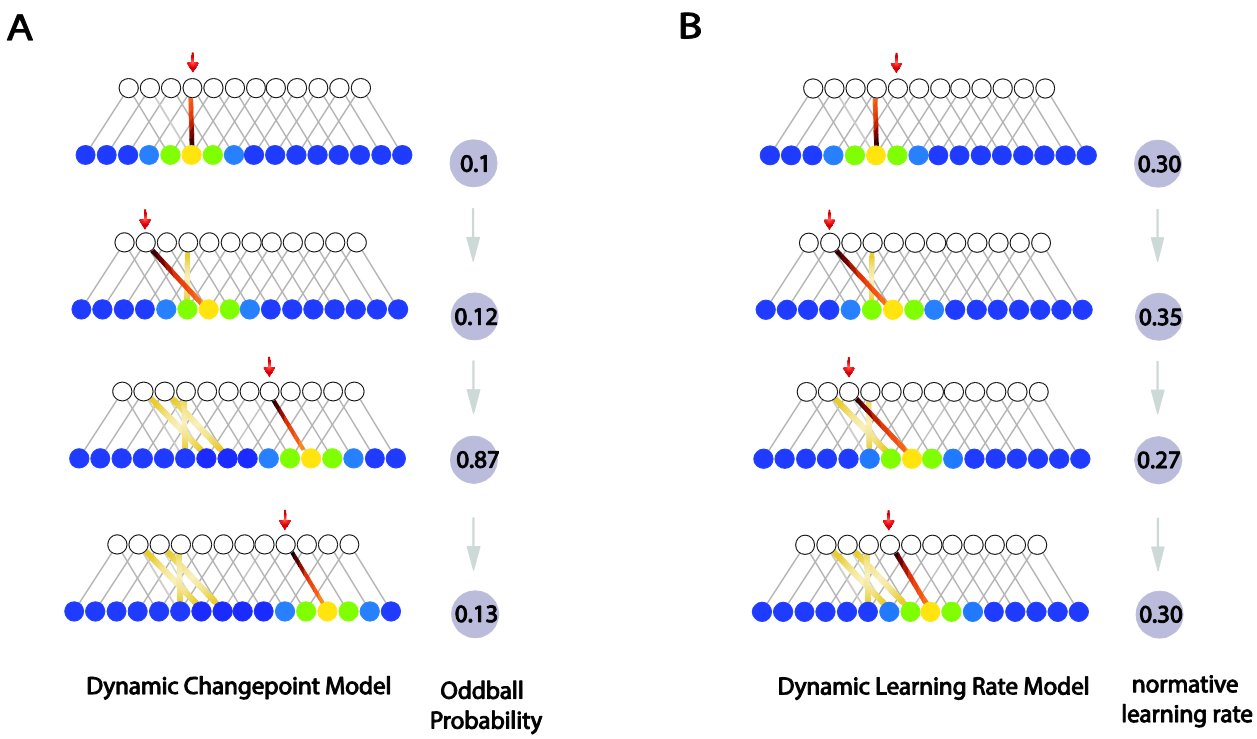

Figure 6—figure supplement 1: Schematics of alternative versions of dynamic models used in the oddball condition. A) Dynamic changepoint context shift model which shifts the context on each trial according to a normative learning rate similar to dynamic context shift model in changepoint condition (figure 3b). B) Dynamic learning rate context shift model which uses a measure of normative learning rate specific for oddball condition to shift the context, making it ignore the oddballs as they happen. Both models show poorer performance compared to the dynamic latent state context shift model (figure 6d) which is able to store the oddball observation in a different context by shifting and returning back the to previous context.

Discussion

The dynamic context shift model uses temporally dynamic state representations in a neural network to produce an adaptive learning rate. We showed that this model is able to perform two distinct behavioral tasks: 1) learn to separate information from different latent states into different contexts and 2) store the result of an outlier observation in a different context than the main context. Our model relies on variables from a Bayesian normative model, which are reflected by neural activity in multiple brain regions, to determine transitions in context.

The idea of using contexts for flexible learning builds from the notion of latent states in representation learning. However most of previous theories of state representation have typically focused on abstract and high-level cognitive models of partitioning learned associations into different contexts (A. G. E. Collins & Frank, 2013; Gershman & Niv, 2010). 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 (A. G. E. Collins & Frank, 2013; Gershman et al., 2010; 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. We further analyzed the predictions of our model to show that dynamic context shift model produce the same pattern of input layer activity as human subjects OFC in fMRI where there is a stronger similarity in neural activity between trials within the same context than trials from different contexts (Nassar, McGuire, et al., 2019).

Latent states in the current model were represented through separate neural activations in the input layer of our model. We illustrated that a ground truth context shift model, which has an all-or-none property of transitioning from one context to another, made near optimal predictions and captured adaptive learning behaviors consistent with normative models of learning. We extended this discrete notion of dynamic latent states by developing a probabilistic version of the model, the dynamic context shift model, which updates internal context representations according to statistical measures that have previously been observed to correlate with neural activity. This model facilitates performance improvements in two different statistical environments (figure 3e, figure 6d), captures key characteristics of human behavior (figure 4c&d), and provides mechanistic interpretations for two brain signals observed during adaptive learning (figure 5&6). The ability of our model to account for changes in the pattern of OFC BOLD responses provides some support for our assumption that associations are controlled through changes in the pattern of active input units over time (e.g. figure 3), rather than alternative accounts in which

associations are selectively attributed to only a subset of active units through synchronization (Verbeke & Verguts, 2019), although these two mechanisms need not be mutually exclusive.

Our model description shares some mechanistic similarities with temporal context models (TCM) of episodic and sequential memory recall. (DuBrow, Rouhani, Niv, & Norman, 2017; Franklin et al., 2020; Howard & Kahana, 2002; Kornysheva et al., 2019; Polyn, Norman, & Kahana, 2009; Shankar, Jagadisan, & Howard, 2009). In temporal context models, there is a gradual change in input layer which is attributed to passage of time and resembles the fixed context shift models we initially introduced in this study, however our dynamic model relies on discontinuous changes in context more analogous to the underlying latent state dynamics and provides a normative rationale for such abrupt transitions at surprising events, namely that such transitions promote pooling of relevant information within a context (figure 3d) and partitioning of information across contexts (figure 3c) in order to improve inference in complex and dynamic environments (figure 3e & figure 6d).

More recently, studies on temporal context models of memory have suggested the existence of event boundaries which cause discontinuity in temporal context (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). The emergence of these boundaries has been attributed to errors in predictions which, analogous to detected outliers in our model, cause the subsequent observations to be stored in a different context (DuBrow & Davachi, 2013; Rouhani, Norman, Niv, & Bornstein, 2020). Such segmented events also lead to more dissociable representations in fMRI (Antony et al., 2020; Baldassano et al., 2017; Lositsky et al., 2016). While these interpretations of discontinuity in memory are closely related to our model, we take a step further by assigning a key role to such segmentations. It is useful to segment internal context representation after a surprising event in order to improve predictions by facilitating rapid learning after a changepoint or preventing interference from oddball events.

An important question here is how to quantitatively control transitioning to new contexts, particularly when such context transitions are not overtly signaled. In previous computational models of event segmentation, surprise has been suggested as the main factor controlling such transition probabilities (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013) however the need to quantify surprise online, inevitably lead us to use Bayesian theories of learning. Our dynamic context shift model uses surprise and uncertainty estimates from a Bayesian learning model calibrated to the environmental structure in order to control the context shifts. So in a sense, inputs of our network require extrinsic manipulation according to a "model" of transition dynamics in the world. The two model-based variables of the normative learning model related to latent state inference (probability of changepoint/oddball, uncertainty) are thought to be represented in both cortical and subcortical regions of the brain (D'Acremont & Bossaerts, 2016; Kao et al., 2020; McGuire et al., 2014; Meyniel & Dehaene, 2017; Nassar, McGuire, et al., 2019; Nassar et al., 2012; O'Reilly et al., 2013; Yu & Dayan, 2005). In addition to these variables, our model requires knowledge of the transition structure between trials, in particular facilitating a return to the previous internal state representation after an oddball event. We speculate that this transition structure might be provided by a separate set of neural systems. More specifically, 1) our state representations mirror the dynamics of representations in orbitofrontal cortex (figure 5), 2) OFC receives strong inputs from the medial temporal lobe (MTL) (Wikenheiser & Schoenbaum, 2016), and 3) MTL plays an important role in model based learning (Mattar & Daw,

2018; Vikbladh et al., 2019) and planning, and thus we would speculate that the model-based transition structure required as input to our network could be provided through MTL inputs.

Previous work has shown that when faced with a surprising event, humans respond by adjusting their learning rate in a range of behavioral tasks. For example, data from past neuroimaging studies has shown that variations in fMRI activity of anterior cingulate cortex provide a measure of subjective learning rate in a volatile environment (Behrens et al., 2007; O'Reilly et al., 2013). However, this signal is not an indicator of learning rate uniformly across all conditions (D'Acremont & Bossaerts, 2016), suggesting that it may perform a function that simply correlates with learning rate under certain conditions.

In fact, a recent study on P300 signal (Nassar, Bruckner, et al., 2019), another neural correlate of surprise in humans, has shown that this signal has the same relationship with learning rate as the context shift variable in our dynamic model during a similar probabilistic inference task. In the present study, learning was defined as how much each observation influence the next prediction: after receiving a prediction error the model stores this information in a new context; if the next prediction comes from this same context, the shift in context results in larger learning rate. However, if after storing the information in a new context the model shifts to yet another new context, or back to previous context as in oddball condition, the shift in the context results in no learning (in terms of an immediate impact on behavior). The results from P300 signal experiments empirically support predictions from our dynamic context shift model where a greater signal not only corresponds with greater adjustment of the adaptive learning rate but the direction of this change is dependent on the statistical context (figure 6e).

This bidirectional adjustment of learning rate is a key prediction of our model. We propose that other neural or physiological measures of surprise that have previously been related to learning, such as the BOLD signal in anterior cingulate cortex or pupil diameter, should also provide the same results when tested in environments with different sources of surprising outcomes. Moreover, our model predicts that such a surprise signal should 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.

By introducing context shifts in our model we were able to build a mechanistic role for surprise in a learning algorithm that can explain the conditional nature of heretofore identified learning rate signals: they are actually signaling state transitions. This is also what distinguishes our model from other similar computational models (Iigaya, 2016) that have used surprise detection signals to control learning rate at the synaptic level, while not contradicting the existence of other lower level mechanisms such as metaplasticity.

Our model opens the door for a number of future investigations. We catered our model to the behavioral experiments of Nasser et al 2019 (Nassar, Bruckner, et al., 2019; Nassar, McGuire, et al., 2019), 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,

Jensen, Johnson, & Kurth-nelson, 2007), which might be thought of as part of a more general model-based inference framework as described above (Franklin et al., 2020; Whittington et al., 2019).That is to say, in order 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 which occur nearby in time pool learning in our current model.

In summary, we suggest that flexible learning can emerge from dynamic state representations and different observed relationships between learning rate and state transition depend on environmental conditions. Our results also predict that other neural or physiological signals of learning rate (Behrens et al., 2007; Jepma et al., 2016; Nassar et al., 2012) would be negatively correlated with learning in conditions when changes are attributed to outliers i.e. oddball condition . Based on current model proposal, when a changepoint or an oddball happens, there is a marked dissimilarity between units active during choice and during feedback, or in other words, what drives the decision is different from what gets reinforced. Furthermore, any perturbations of systems involved in resetting state representation should have opposite effects in different environmental contingencies.

Methods

Experimental task:

We examine human and model behavior in a predictive inference task that has been described previously (McGuire et al., 2014). The predictive inference task is a computerized task in which an animated helicopter drops bags in an open field. In the pre-training session, human subjects would learn to move a bucket with a joystick underneath of the helicopter to catch the bag and its contents. During the main phase of the experiment, the helicopter would be completely occluded by clouds and the participants were forced to infer the helicopter position on each trial based on the locations of bags that had previously been dropped.

The majority of our simulations focus on dynamic environments in which surprising events often signal a change in the underlying generative structure (changepoint condition; figures 1-5). In the chanagepoint condition, bag locations were drawn from a distribution centered on the helicopter with a fixed standard deviation of 25. The helicopter would remain stationary on most trials, but would occasionally and abruptly change 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 ($H = 0.1$).

We also consider a complementary generative environment in which surprising events tend to be unrelated to the underlying generative structure (oddball condition; figure 6). In the oddball condition, the helicopter would gradually move in the sky according to a Gaussian random walk with a constant drift rate ($DR = 10$). 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 the

oddball bag is sampled from a uniform distribution across the entire screen. The probability of an oddball bag is controlled by the hazard rate ($H = 0.1$).

Normative learning model:

The simplest delta rule model in the predictive inference task states that:

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

$$\delta = Bag\ Position(t) - B_t(t)$$

here B is belief regarding helicopter position on each trial , δ is the prediction error 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 delta rule learning under certain approximations, leading to normative prescriptions for learning rate that are adjusted dynamically (Nassar, Bruckner, et al., 2019; Nassar et al., 2010). The resulting normative learning model takes information which human subjects would normally obtain during the pre-training sessions including Hazard rate and standard deviation, but also computes two latent variables, by using the trial-by-trial prediction error: 1) changepoint probability which is computed after a prediction is made and indicate the probability that the observed outcome has been a changepoint, and 2) relative uncertainty which is computed before making the next prediction and indicates the models uncertainty about the location of the helicopter. Detailed information regarding how CPP and RU are calculated can be found in appendix-1.

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

$$\alpha_t = CPP + RU - CPP \times RU$$

Where CPP is changepoint probability and RU is relative uncertainty. Using these two latent variables, which both 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 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:

$$\alpha_t = RU - OBP \times RU$$

Where OBP is the models posterior probability estimate that an outcome was an oddball event and RU reflects the model's uncertainty about the current helicopter location.

Neural network models:

In order 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 feed forward 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 16 We implemented several versions of this model depending on how the mean 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 input layer:

$$r_j = \sum_{i=1}^{N_{in}} x_i w_{ij}$$

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:

$$estimate = \sum_{j=1}^{N_{out}} L_j \, r_j$$

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

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

Where y_j is the probability on a normal distribution centered on the observed outcome evaluated at L_j, and η is a constant synaptic learning rate controlling the weight changes of the neural network and was set to 0.1 for all models simulations.

The fixed context shifts models:

In the first models we consider, *fixed context shift* models, The mean m is computed on each trial as follows:

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

Here, Δm_f takes a fixed value for all trials throughout the simulation (figure 2b&c). We considered 50 different Δm_f values ranging from 0 to 2 to study the effect of context shift size on model performance. The word "context" refers to the subpopulation of input layer neurons that are firing above the threshold on each trial. By incrementally increasing the mean of response distribution of the input layer, we can think of this context being changed 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. In order to minimize interference from previous visits to a set of context neurons we implemented weight decay (WD) on each time step according to the following rule:

$$W_{t+1}(x_t < threshold) = W_t(x_t < threshold) \times WD$$

$$WD = 0.1$$

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 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.

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	
Weight Decay Threshold	0.01	
Weight Decay Rate (*WD*)	0.1	
Task Parameter:		
Hazard Rate (*H*)	0.1	Probability of a changepoint/oddball trial
Noise (σ_N)	25	Standard Deviation of random process generating outcomes
Standard Deviation of Drift Rate (σ_{drift})	10	Standard Deviation of the random process generating drift rate in oddball condition

Ground Truth context shifts neural network model:

To leverage the benefits of different context shifts which we observed in the fixed context shifts models we designed a model that uses optimal context shift for each trial. The ground truth context shifts model has the same design of a fixed context shift model except instead of the constant term Δm, the model uses a function for computing Δm:

$$\Delta m = \begin{cases} max(\Delta m_f), & if \;\; t \; is \; changepoint \\ 0, & otherwise \end{cases}$$

Dynamic context shifts neural network model:

In order to make the ground truth context shifts model more biologically realistic, we tried implementing variables from the normative learning model. The dynamic context shift model uses a function of the normative learning rate to compute dynamic Δm . The α_t variable, which has been referred to as learning rate earlier in the Methods, is used extensively in previous studies to model the dynamic learning rate of humans in similar tasks. Here, we see α_t as an estimate of how much context shift needs to happen on a trial. What is consistently referred to as the

"effective learning rate" in the present study, is the fraction of prediction error used to update the next response of model.

$$\alpha_t = CPP + RU - CPP\,.RU$$

$$\Delta m = f(\alpha_t)$$

On each trial the dynamic context shift model uses CPP and RU from the normative model to compute normative learning rate on that trial, and updates context according to a function of that learning rate. The function, f, provides a fixed level of context shift according to the desired learning rate by inverting the relationship between context shift and effective learning rate observed in the fixed context shift models and 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 α_t. The performance of the all models is averaged over 32 simulations.

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 includes the same helicopter which is gradually changing its position in the sky with a constant drift rate. On occasions, a bag might drop in an unexpected random location.

The fixed context shift models were the same as changepoint condition. However, since bag locations are sampled from the nearly the same distribution before and after an oddball observation (although there is a drift rate of the mean), we modified the ground truth context shift model for oddball condition so that on each trial, input activity mean m would be changed by 1) adding a constant value (0.05) proportional to the drift rate of the condition, 2) using a function for changing the mean based on whether or not a trial is an oddball and 3) After observing the result of an oddball trial, subtracting the added value from the mean so that context return to where it previously was.

Thus after a prediction is made on trial context mean changes according to:

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

$$m_{t+1} = m_t + \Delta m_1$$

After the model receives the supervised signal and stores it the new context, context transition back to:

$$m_{t+1} = m_t - \Delta m_1 + \Delta m_2$$

Where Δm_2 is a constant (here 0.05) proportional to the drift rate . This causes the information from oddball trial to be stored in a different context that won't influence the future predictions of the model.

For the dynamic context shift model we employed three different versions. First we used a model, referred to as the dynamic changepoint context shift model, in which context shifts faithfully implemented the exact same equations that were used in the changepoint condition. Specifically, context shifts (Δm) were implemented as a monotonically increasing function of the learning rate that *would* be appropriate for the changepoint condition:

$$\alpha_1 = OP + RU - OBP\,.RU$$

$$\Delta m_1 = f(\alpha_1)$$

$$m_{t+1} = m_t + \Delta m_1$$

An obvious problem with this model is that, although it involves a direct transcription of the equations from the changepoint model, oddballs should demand less learning, rather than more.

In our second model, the dynamic learning rate context shift model, we maintained the relationship between context shift and normative learning, rather than implementing the same equations as the changepoint condition exactly. Specifically, the model changes the context according to relative uncertainty, but changes context less when confronted with a likely oddball event:

$$\alpha_2 = RU - OBP \times RU$$

$$\Delta m_2 = f(\alpha_2)$$

$$m_{t+1} = m_t + \Delta m_2$$

Here because the second model is using relative uncertainty for shifting the context instead of oddball probability, the model in general makes smaller shifts to outliers.

One shortcoming of this second model is that, while context changes preserve their relationship to learning, they do not partition information as effectively in the oddball condition. In the extreme case where 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, we implemented a third version, dynamic latent state context shift model, that was inspired by the ground truth model and included both shifting the context forward and back in case of oddballs. This model changes the context first according to the oddball probability or Δm_1 in above equation, after storing the supervised learning signal in the new context, the model transition back to its previous context by subtracting the first context shift term $\Delta m_1$and move the context according to the second model mentioned above Δm_2. The Δm_1 term causes significant shifts on oddball trials, but after that the model transition back to previous context and shifts according to the Δm_2 which would not be influenced by oddball trials.

Representational similarity analysis:

We computed a trial-by-trial dissimilarity matrix where each cell in the matrix represent the number corresponding to the dissimilarity between two trials. The dissimilarity matrix (D) of the dynamic context shifts model uses Euclidean distance and is computed by:

$$D_{ij} = \sqrt{\sum_{q=1}^{N_{in}} (Act_{(i,q)} - Act_{(j,q)})^2}$$

Where $Act_{(i,q)}$ and $Act_{(j,q)}$ are the activities of neuron q of the input layer on trials i and j respectively. The dissimilarity matrix of normative learning model is constructed by using normative learning rates on each trial:

$$M_{ij} = 1 - \prod_{i=t}^{j-1} 1 - \alpha_t$$

Where the similarity between two trials i and j is inversely correlated with the product of the learning rates α_t of all trials between them. Subtracting similarity from 1 gives us the trial-by-trial dissimilarity matrix (M). Using the RSA approach we were able to directly compare the results of our neural network input layer activity to the dynamic learning rate computed from CPP and RU in a normative model.

Behavioral analysis:

Behavioral analyses are aimed at understanding the degree to which we revise our behavior in response to new observations. In order to quantify this, we define an "effective learning rate" as the slope of the relationship between trial-to-trial predictions errors and trial-to –trial updates (eg. the change in bucket position from one trial to the next). The adjective "effective" is chosen here so that this learning rate won't be mistaken by the reader with two other learning rates used in this paper: 1) the fixed learning rate of the neural network 2) the normative learning rate from the reduced Bayesian model. To measure this relationship, we use a standard regression model which contained update (UP) and prediction error (PE) on each trial:

$$UP = \beta_0 + \beta_1 \times PE$$

we also performed a more extensive regression analysis that included terms for 1) prediction error 2) prediction error times changepoint probability 4) prediction error times relative uncertainty. (figure4-D)

Comparison to P300 analysis:

For analyzing the effect of context shift size on effective learning rate we fit trial update with prediction error for 4 bins of context shift size separately, calculated individually for changepoint and oddball condition. The corresponding figure (figure 6e) of P300 signal and learning rate are from ref (Nassar, Bruckner, et al., 2019).

Data Availability

All analysis and modeling code (including code for generating the figures) has been made available on GitHub: https://github.com/NiloufarRazmi/dynamicContextShift .

Acknowledgments

We thank Linda Yu, Robert Wilson & Cristian Buc Calderon for helpful discussion and comments. This work was supported by R00AG054732 to M.R.N.

Competing interests

The authors declare that no competing interests exist.

References:

Adams, R. P., & MacKay, D. J. C. (2007). *Bayesian Online Changepoint Detection*. Retrieved from http://arxiv.org/abs/0710.3742

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2020). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *BioRxiv*, 2020.03.26.008714. https://doi.org/10.1101/2020.03.26.008714

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). *Learning the value of information in an uncertain world*. *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bernacchia, A., Seo, H., Lee, D., & Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*(3), 366–372. https://doi.org/10.1038/nn.2752

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

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

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

D'Acremont, M., & Bossaerts, P. (2016). Neural Mechanisms behind Identification of Leptokurtic Noise and Adaptive Behavioral Response. *Cerebral Cortex*, *26*(4), 1818–1830. https://doi.org/10.1093/cercor/bhw013

Donahue, C. H., & Lee, D. (2015). Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nature Neuroscience*, *18*(2), 295–301. https://doi.org/10.1038/nn.3918

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277–1286. https://doi.org/10.1037/a0034024

Farashahi, S., Donahue, C. H., Hayden, B. Y., Lee, D., & Soltani, A. (2019). Flexible combination of reward information across primates. *Nature Human Behaviour*, *3*(11), 1215–1224. https://doi.org/10.1038/s41562-019-0714-3

Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., & Soltani, A. (2017). Metaplasticity as a Neural Substrate for Adaptive Learning and Choice under Uncertainty. *Neuron*, *94*(2), 401-414.e6. https://doi.org/10.1016/j.neuron.2017.03.044

Fischer, A. G., & Ullsperger, M. (2013). Real and Fictive Outcomes Are Processed Differently but Converge on a Common Adaptive Mechanism. *Neuron*, *79*(6), 1243–1255. https://doi.org/10.1016/j.neuron.2013.07.006

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327–361. https://doi.org/10.1037/rev0000177

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, Learning, and Extinction. *Psychological Review*, Vol. 117, pp. 197–209. Retrieved from https://nivlab.princeton.edu/sites/default/files/nivlab/files/gershmanetal2009.pdf

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256. https://doi.org/10.1016/j.conb.2010.02.008

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *ELife*, *5*, e18073. https://doi.org/10.7554/eLife.18073

Jepma, M., Brown, S. B. R. E., Murphy, P. R., Koelewijn, S. C., de Vries, B., van den Maagdenberg, A. M., & Nieuwenhuis, S. (2018). Noradrenergic and Cholinergic Modulation of Belief Updating. *Journal of Cognitive Neuroscience*, *30*(12), 1803–1820. https://doi.org/10.1162/jocn_a_01317

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLOS Computational Biology*, *12*(10), e1005171. Retrieved from https://doi.org/10.1371/journal.pcbi.1005171

Kao, C.-H., Khambhati, A. N., Bassett, D. S., Nassar, M. R., McGuire, J. T., Gold, J. I., & Kable, J. W. (2020). Functional brain network reconfiguration during learning in a dynamic environment. *Nature Communications*, *11*(1), 1682. https://doi.org/10.1038/s41467-020-15442-2

Li, Y. S., Nassar, M. R., Kable, J. W., & Gold, J. I. (2019). Individual Neurons in the Cingulate Cortex Encode Action Monitoring, Not Selection, during Adaptive Decision-Making. *The Journal of Neuroscience*, *39*(34), 6668 LP – 6683. https://doi.org/10.1523/JNEUROSCI.0159-19.2019

Lloyd, K., & Leslie, D. S. (2013). Context-dependent decision-making: a simple Bayesian model. *Journal of The Royal Society Interface*, *10*(82), 20130069. https://doi.org/10.1098/rsif.2013.0069

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., … Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *ELife*, *5*, e16070. https://doi.org/10.7554/eLife.16070

Massi, B., Donahue, C. H., & Lee, D. (2018). Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron*, *99*(3), 598-608.e4. https://doi.org/https://doi.org/10.1016/j.neuron.2018.06.033

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(19), E3859–E3868. https://doi.org/10.1073/pnas.1615773114

Muller, T. H., Mars, R. B., Behrens, T. E., & O'Reilly, J. X. (2019). Control of entropy in neural models of environmental state. *ELife*, *8*, 1–30. https://doi.org/10.7554/eLife.39404

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *ELife*, *8*, 1–26. https://doi.org/10.7554/eLife.46975

Nassar, M. R., & Gold, J. I. (2010). *Supplementary Material for : Bayesian On-line Learning of the Hazard Rate in Change-Point Problems*. *22*(9), 2452–2476.

Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable forms of uncertainty-driven representational change across the human brain. *Journal of Neuroscience*, *39*(9), 1688–1698. https://doi.org/10.1523/JNEUROSCI.1713-18.2018

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*(7 JUN), 1–10. https://doi.org/10.3389/fnins.2013.00105

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38). https://doi.org/10.1073/pnas.1305373110

Redish, A. D., Jensen, S., Johnson, A., & Kurth-nelson, Z. (2007). *Reconciling Reinforcement Learning Models With Behavioral Extinction and Renewal : Implications for Addiction , Relapse , and Problem Gambling*. *114*(3), 784–805. https://doi.org/10.1037/0033-295X.114.3.784

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269. https://doi.org/https://doi.org/10.1016/j.cognition.2020.104269

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), 635–644. https://doi.org/10.1038/s41583-019-0180-y

Verbeke, P., & Verguts, T. (2019). Learning to synchronize: How biological agents can couple neural task modules for dealing with the stability-plasticity dilemma. *PLOS Computational Biology*, *15*(8), e1006604. Retrieved from https://doi.org/10.1371/journal.pcbi.1006604

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2019). The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *BioRxiv*, 770495. https://doi.org/10.1101/770495

Wikenheiser, A., & Schoenbaum, G. (2016). Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, *17*. https://doi.org/10.1038/nrn.2016.56

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural Computation*, *22*(9), 2452–2476. https://doi.org/10.1162/NECO_a_00007

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2013). A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Computational Biology*, *9*(7),

e1003150–e1003150. https://doi.org/10.1371/journal.pcbi.1003150

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

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

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, Vol. 133, pp. 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Appendix – 1

Changepoint condition:

The probability that a changepoint has occurred in the current trial is computed as the likelihood of current trial observation coming from a new helicopter position, which is randomly chosen from a uniform distribution between 0 and 300, over the sum of the likelihoods of the current observation coming from a non-changepoint distribution (Gaussian distribution centered around the most recent belief about helicopter positions) and changepoint distribution:

$$\mathrm{CPP} = \Omega_t = \frac{U(X_t|0{,}300)\,H}{U(X_t|0{,}300)H + N(\delta_t; 0, \sigma_\mu^2 + \sigma_N^2)(1-H)}$$

Where Ω_t is changepoint probability, H is the hazard rate and is the same hazard rate used in the generative process for simulating outcomes, δ_t is the difference between the predicted and actual outcome, σ_μ^2 is the variance on predictive distribution, and σ_N^2 is the variance of the distribution of bags around the helicopter.

Relative uncertainty is computed as estimation uncertainty about bag locations as a fraction of total uncertainty that is sum the of estimated uncertainty and noise.

$$RU = \tau_{t+1} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_N^2}$$

Where is σ_μ^2 is estimation uncertainty, σ_N^2 is noise (variance on the bag distribution) and τ_{t+1} is the relative uncertainty. It should be noted that, here, estimation uncertainty (the numerator term) originates from insufficient knowledge about the helicopter position and will decrease as the model accumulates more information from observed bag locations. This is different from the noise (second term in denominator) which captures the uncertainty inherent in the probabilistic nature of bag drops and can't be reduced by more training.

Estimation uncertainty is computed recursively after observing an outcome and using the changepoint probability of the current trial:

$$\sigma_\mu^2 = \Omega_t \sigma_N^2 + (1 - \Omega_t)\, \tau_t \sigma_N^2 + \Omega_t (1 - \Omega_t)(\tau_t + B_t(1 - \tau_t) - X_t)$$

Where B_t is the current belief about helicopter position and X_t is the current trial outcome.

Oddball condition:

The probability of an observation being an oddball is computed in a similar manner using the prediction error of the model on each trial:

$$\mathrm{OBP} = \Omega_t = \frac{U(X_t|0{,}300)\, H}{U(X_t|0{,}300)H + N(\delta_t; 0, \sigma_\mu^2 + \sigma_N^2)(1 - H)}$$

Where H is the hazard rate of the oddball condition. Here, the numerator is the likelihood of current outcome being an oddball and the second term in the denominator is the likelihood of current outcome coming from a normal distribution with the mean of predicted outcome and variance of relative uncertainty.

Again similar to changepoint condition, relative uncertainty is computed as estimation uncertainty over the total uncertainty:

$$RU = \tau_{t+1} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_N^2}$$

Although for the oddball condition there is a few changes in the way estimation uncertainty is computed:

$$\sigma_\mu^2 = \Omega_t \frac{\sigma_N^2 \tau_t}{1 - \tau_t} + (1 - \Omega_t)\, \tau_t \sigma_N^2 + \Omega_t (1 - \Omega_t)(\delta_t \tau_t)^2 + \sigma_{drift}^2$$

Where the first term represents the contribution of an oddball observation to the uncertainty, the second term is the contribution of a nonoddball observation to the uncertainty, the third term is uncertainty proportional to the prediction error received from either an oddball or a nonodball trial and the last term reflects the uncertainty expected from the drift rate of the helicopter in between two trials.