A COMPUTATIONAL ROLE FOR AROUSAL IN OPTIMAL INFERENCE

Matthew Robert Nassar

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

_____
Joshua Gold
Professor, Neuroscience

Graduate Group Chairperson

_____
Joshua Gold
Professor, Neuroscience

Dissertation Committee:

Yale Cohen, Associate Professor, Otorhinolaryngology
Javier Medina, Assistant Professor, Psychology
Vijay Balasubramanian, Professor, Physics
Jonathan Cohen, Professor, Psychology

A COMPUTATIONAL ROLE FOR AROUSAL IN OPTIMAL INFERENCE

ACKNOWLEDGEMENTS

The work described within evolved over the course of my PhD thesis with Dr. Joshua Gold at the University of Pennsylvania. Josh provided ideas, direction, and advice throughout my thesis years and without his guidance this work would not have been possible. In addition, the atmosphere of intellectual exchange fostered by Josh in his lab and by the University of Pennsylvania Neuroscience Graduate Group in general was instrumental in expanding my very focused original ideas into a general account of the neural computations underlying learning. In particular, this work would not have been possible without insights provided by other members of the Gold lab, especially Chi-Tat Law, Ben Heasly, Ching-Ling Teng, Long Ding and Robert Wilson who were highly influential on my early graduate work and Yin Li and Takahiro Doi who helped to shape my thinking later on. In addition to intellectual contributions, the Gold lab provided considerable technical support that allowed me to complete this project. The work included here depended critically on the efforts of Katherine Rumsey, Ben Heasly, Robert Wilson and Kinjan Parikh.

In addition to consistent feedback and guidance from Josh and other lab members, a very active and intellectually engaging thesis committee also helped to shape my ideas. Thus the insights provided by Yale Cohen, Vijay Balasubramanian, Javier Medina and Jon Cohen should not go without mention.

Outside of the academic arena, much credit goes to those in my life who have supported my graduate work less specifically. Many thanks are due to the friends who have torn me free during moments of research-induced myopia and to my family, especially my parents, who have supported and encouraged me in research and in life. Most thanks of all are due to my wife Joy.

ABSTRACT

A COMPUTATIONAL ROLE FOR AROUSAL IN OPTIMAL INFERENCE

Matthew Nassar

Joshua Gold

Making accurate predictions is one of the most critical functions of the brain. Whether made by a monkey deciding where to forage, a deer deciding which way to run, or a wall-street broker deciding how to invest, decisions are informed by expectations about possible future outcomes. These expectations are learned over time through experience and are rapidly adjusted when they fail to match observations. Here I propose and support the thesis that learning systems in the brain optimize the accuracy of predictions in a changing world, even though this necessitates becoming insensitive to incoming sensory information under some conditions. Furthermore I propose a biologically inspired model for achieving accurate predictions and suggest a novel role for the arousal system in optimally adjusting the influence of incoming sensory information. I support these theses with a series of experiments that utilize computational modeling, as well as behavioral and pupillometric measurements in humans.

Table of contents

List of Figures

CHAPTER 1

Understanding the brain: levels of analysis.

The human brain contains approximately 100 billion neurons interconnected by 100 trillion synapses (Williams and Herrup, 1988). This tremendous complexity enables feats of information processing that humble even the greatest achievements of artificial intelligence and computer vision. However, this complexity also poses a formidable challenge to anyone wishing to understand the how the system functions. Not only is measuring each cog in the machine technically impossible, it is also not clear what one would do with perfect descriptions of each of the components. David Marr best formalized the issue in terms of perception as follows:

"[T]rying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense" (Marr, 1982) (p. 27)

Marr suggests three complementary levels of analysis necessary for completely understanding a system. The top level, which he refers to as the computational

level, requires a normative approach. That is, one must determine the critical problem being solved by the system and ask how the problem could be optimally solved. The normative approach does not necessarily provide any information about "how" the brain might solve a certain problem, but it will likely provide a set of rules, which must be obeyed for any possible solution to the problem, much like aerodynamics provides for flight. The second level of analysis proposed by Marr is the algorithmic or representational level: how does the brain represent the variables necessary to solve the problem. What are the actual algorithms employed to achieve the desired function? Marr's third level of analysis is one of implementation: how does the system realize the algorithm in physical hardware (Marr, 1982).

My dissertation explores how the brain learns using each of these levels of analysis. The following sections aim to provide a coherent introduction to the concepts relevant to my theses at each level. The first section examines a possible normative framework for understanding learning in terms of prediction. The second section examines the constraints on animal and human learning algorithms revealed through behavioral studies. Finally, the third section discusses the biological architecture available for mediating those algorithms.

Learning as predictive inference.

Learning, or experience dependent change in behavior, is one of the most robust behavioral phenomena observed across species. It has previously been suggested that some forms of learning serve to provide predictions for the future, which in turn can be used for appropriate behavioral modifications (Preuschoff and Bossaerts, 2007;Courville et al., 2006). However, a point that has been underappreciated is the extent to which optimal predictive behavior depends on the exact nature of the environment for which it is designed. Here I examine some features of a series of optimal prediction algorithms designed for increasingly complex environments. The comparison of these different predictive models will reveal hallmarks of optimal inference in dynamic environments that are very different from what one sees in optimal inference models tailored to static or continuously changing environments.

The outcome of future actions can often be predicted due to regularities in the process by which outcomes are generated. For example, sticking ones finger in an electrical outlet leads to a fairly unambiguous and consistent result. This makes the problem of predicting future electrical socket-related outcomes fairly simple; it takes only one such experience to recognize that all such future actions are likely to lead to negative outcomes. Future decisions can then be biased away from actions that involve self-electrocution.

Some processes lead to far less reliable results. Take for example a monkey attempting to predict the caloric yields he might attain by choosing one of several foraging locations. The yield attained on one day may differ from that on the next, as there is a fair amount of variability that cannot be controlled by the monkey. Within this document I will refer to a stable source of irreducible variability as noise. To specifically define noise, here I will assume that the foraging values at a given location are drawn independently on each observation from a normal distribution with mean μ and standard deviation σ_n:

$$X_t \sim \mathcal{N}(\mu,\ \sigma_n)$$

Noise does not prohibit predictions, but it does provide an upper bound on prediction accuracy. The prediction minimizing squared errors simply becomes the mean of the distribution, μ. Furthermore, noise changes the optimal strategy for updating those predictions over time. Unlike the electrical outlet example, each daily yield provides a fairly unreliable estimate of the true underlying distribution of possible daily yields. The best possible strategy for forming a prediction involves pooling all of the pertinent data, which can be done by simply taking the average of all previous yields, which provides the best possible approximation of μ. This strategy can also be implemented in a Markov form, such that the observer need not store all previous outcomes in memory. One such strategy for maintaining and updating predictions efficiently is the delta rule, which was simultaneously

developed in the fields of animal behavior and machine learning (Rescorla and Wagner, 1972;Sutton and Barto, 1998):

$$B_{t+1} = B_t + \alpha_t \times \delta_t$$

Where B_{t+1} is the updated belief, which serves as a prediction for time t+1. The prediction error, δ_t, is the difference between the observed outcome (X_t) and the the predicted outcome (B_t) at time t:

$$\delta_t = X_t - B_t$$

The learning rate, α_t, determines the extent to which each new observation influences the updated prediction. When α_t is equal to 1 predictions are set equal to the most recent observation, whereas when α_t is equal to 0 the updated prediction is simply equal to the prediction on the previous time-step, irrespective of the new observation. For the case of the average over all previous outcomes, which is the optimal strategy for updating predictions in the presence of noise, α_t depends on the total number of observations (including the current one):

$$\alpha_t = \frac{1}{N_t}$$

As is demonstrated in figure 1.1, this strategy for updating predictions rapidly converges on the true mean of the underlying distribution. In addition, the learning rate term, which determines the extent to which predictions are influenced by new data, decays to zero.

Although the average of all data is the best prediction of a noisy but stable variable, it does not perform well under situations where the variable of interest changes in time. For example, it might be the case that certain foraging locations are slowly becoming more fruitful, whereas other foraging locations are becoming more barren. Statistically, we can model a continuous change by assuming that a random variable, D_t, is added to the mean of the outcome distribution at each time-step (μ_t) to produce the mean for the next trial:

$$\mu_{t+1} = \mu_t + D_t$$

Here we will assume that D_t is drawn from a normal distribution with mean μ_d and variance σ_d:

$$D_t \sim \mathcal{N}(\mu_d,\ \sigma_d)$$

When σ_d is large, past observations rapidly become meaningless, as the μ_t is an uncertain predictor of μ_{t+1} and an even more uncertain predictor of μ_{t+n}. Such circumstances require relying more on recent observations, as these observations

are more accurate predictions of the mean on the current time-step. The optimal strategy for updating predictions under such conditions is referred to as the Kalman filter and is depicted in figure 1.2. The Kalman filter does a relatively good job of estimating the mean of the distribution (ie best possible prediction) even though the mean is changing at each time-step. Unlike the optimal updating algorithm in the absence of change, the Kalman filter uses a learning rate, referred to as the Kalman gain, that decays to a non-zero asymptote that depends on the drift and noise variances (σ_d and σ_n).

Although the Kalman filter can provide optimal and efficient predictions in a continuously changing environment, it does not account for abrupt and discontinuous changes (ie. the fruit tree at a certain foraging location dies and stops bearing fruit). Change-points can render past information irrelevant to the problem of predicting future outcomes and thus pose a major problem to standard learning algorithms. Optimal predictions in a discontinuously changing environment have been derived according to Bayes rule and rely on the intuition that optimal inference after a change-point simply requires taking the average of all observations since the most recent change-point (Wilson et al., 2010;Adams and MacKay, 2007;Fearnhead and Liu, 2007). Since change-point locations are unknown, they must be inferred from the data themselves (ie lack of fruit at a previously high yield location). In order to do this, an optimal predictive model must consider all possible run lengths and the distribution of outcomes predicted by these separate possible

models of the world. Each run length has a separate predictive distribution over possible outcomes and thus the probabilities of each possible run length can be computed recursively according to Bayes rule by taking into account the likelihood with which each possible run length would produce the new outcome (Wilson et al., 2010). Predictions made by such a model accurately reflect the mean during stable periods but rapidly adjust to the new mean after a change-point (see figure 1.3).

One interesting feature of optimal inference amid change-points is that not all data are equally influential. Where the inference model rapidly adjusts predictions in response to some observations (such as those subsequent to change-points) it is relatively unaffected by other observations (such as those occurring after a long run of stable data). This effect is visible in the learning rates in figure 1.3 b. Thus, environments with change-points demand an optimal agent to perform frequent online adjustments of the influence of new observations on predictions, while environments with only noise and continuous drift prescribe observation influence to decay to some asymptotic value and then remain constant.

The thesis that I will support in the ensuing chapters is that predictive learning systems in the brain are attempting to optimize predictions in discontinuously changing environments. Following directly from this thesis is the prediction that the influence of information on predictions should depend heavily on the structure of outcomes including the presence and recency of change-points. To test this idea of

developed a predictive inference task in which subjects directly report predictions allowing direct measurement of the influence of each outcome on the predictions of the observer. I demonstrate that influence depends critically on features characteristic of a discontinuously changing environment, in particular probability and recency of change-points.

Algorithms underlying learning.

Over the last 50 years, behavioral psychology has gained substantial insight into the exact rules that guide how humans and animals update expectations in response to experience. Some of the earliest learning studies were performed in classical conditioning paradigms where predictions were measured in terms of an implicit response to an innocuous (conditioned) stimulus that was previously paired with an aversive or rewarding (unconditioned) stimulus. Behavior in such paradigms suggests that the transfer of implicit responding is greatest when the absolute difference between the expected and actual outcome valence is greatest. These findings gave rise to the Rescorla-Wagner model for classical conditioning, which bears a notable resemblance to the delta rule described above (Rescorla and Wagner):

$$\Delta V_x^{n+1} = \alpha_x \beta(\lambda - V_{tot})$$

$$V_x^{n+1} = V_x^n + \Delta V_x^{n+1}$$

where V_x is the strength of the association between the conditioned stimulus (x) and the unconditioned stimulus, V_{tot} is the total associative strength of all conditioned stimuli, α and β are rate parameters specific to the conditioned and unconditioned stimulus, and λ is the maximum conditioning possible. A slight rearrangement of the equations reveals that they are identical to the delta rule format described above, albeit with two separate rate terms.

In contrast to classical conditioning, operant conditioning probes the extent to which the research subject alters choice behavior based on outcome history. The delta rule family of models, including an actor critic implementation based on biological architecture, has been used to describe behavior in a broad range of operant tasks across a broad range of species (Daw et al., 2011). However, until very recently such models have been assumed to contain a learning rate that is constant for all trials of a given task performed by a given subject. As described above, this constraint does not allow optimal learning under a large subset of circumstances. In particular, such models are not capable of performing well when learning to predict outcomes that can change discontinuously. Equally important is the functional implication of this idea: if learning rates are constant then each observation should affect stored beliefs equivalently. That is to say, there is no

necessity for a mechanism to amplify the impact of some pieces of sensory information.

I contend that human learning might be better described by a delta rule where the learning rate is not constant, but rather adjusted according to the statistics of recent observations. To support this hypothesis I propose such a model to describe the behavior of human subjects in a predictive inference task designed to probe the influence of new observations on the predictions of the observer. The model, which contains a learning rate that is adjusted according to Bayesian estimates of change-point probability and uncertainty, provides an improved description of subject behavior over a fixed-learning model, as well as achieving better predictive performance in a dynamic environment.

Implementation of delta-rule updating in the brain.

The delta rule is a strong candidate for a neural belief updating algorithm because of its computational simplicity, effectiveness for a wide range of problems, and relationship to known brain mechanisms. For example, neurons with activity reflecting decision-related beliefs have been reported in several prefrontal areas, including anterior cingulate cortex (ACC) (Kennerley and Wallis, 2009), orbitofrontal cortex (OFC) (Padoa-Schioppa and Assad, 2006), and lateral pre-

frontal cortex (LPFC) (Kennerley et al., 2009). Prediction error-like signals have been reported most notably in the ascending dopaminergic system (Schultz et al., 1997), but also in the lateral habenula (Matsumoto and Hikosaka, 2007) and the ACC (Kennerley et al., 2011;Matsumoto et al., 2007). Although neural correlates of learning rate have remained relatively unexplored relative to prediction errors, two recent fMRI studies identified an area in dorsal ACC with BOLD activity related to learning rate. Specifically, activity in dorsal ACC correlates with volatility, a statistical estimate of the rate at which the reward contingencies are changing (Behrens et al., 2007). In a Bayesian belief-updating model, this volatility estimate determined the influence of new outcomes on the adjusted belief. More recently, a BOLD response in the same region was shown to correlate with trial-by-trial learning rates used by a model fit to subject behavior (Krugel et al., 2009).

Although human fMRI studies suggest a cortical representation of learning rate, rodent behavioral studies have suggested that learning rate might also depend on the LC, a brainstem nucleus that provides the noradrenergic (NA) modulation of cortical and thalamic circuitry. LC is reciprocally connected to prefrontal cortex (ACC and OFC), and noradrenaline is thought to modulate processing related to attention and action monitoring in these regions (Aston-Jones and Cohen, 2005). LC activity and prefrontal NA are greatest after the action-outcome contingency is altered in a manner similar to the environmental change-points discussed previously (Bouret and Sara, 2004;Dalley et al., 2001). These increased prefrontal

NA levels are thought to facilitate behavioral adaptation. This idea is supported by behavioral experiments involving set-shifting, in which an animal is forced to switch from a behavioral strategy that depends on one sensory cue to a new behavioral strategy that depends on a different sensory cue. The ability of rodents to adapt in such experiments is enhanced by pharmacological activation of LC (Devauges and Sara, 1990). This facilitation of behavioral adaptation can be blocked by direct application of α1 antagonists to medial prefrontal cortex (mPFC) (Lapiz and Morilak, 2006), the evolutionary precursor to the cortical region thought to encode learning rate in humans (ACC). NA deafferentation in the mPFC also leads to impairment of adaptive set-shifting behavior (McGaughy et al., 2008;Tait et al., 2007). Such behavioral effects are also seen when NA levels are modulated through manipulation of the NA transporter, NET. Inhibition of NA re-uptake leads to increased prefrontal NA and enhanced performance of rodents and monkeys performing tasks that require reversal of a previously learned action-outcome contingency. This performance gain was specifically attributed to a decrease in errors of perseveration, suggesting that NA plays a role in controlling the rate of behavioral adaptation (Seu et al., 2009). Many of these results can be accounted for by a computational model in which the LC responds to environmental change-points, thereby modulating prefrontal cortical processing via NA release such that unexpected outcomes lead to greater behavioral adjustment (Yu and Dayan, 2005;Yu and Dayan, 2003).

Although measuring LC activity directly is technically difficult, there is a recent move to establish pupil diameter as a proxy for LC activity. Although direct confirmation is still needed, this idea is supported by several lines of evidence, including 1) a compelling example of simultaneous measurements of locus coeruleus activity and pupil diameter in a monkey that are closely correlated (Aston-Jones and Cohen, 2005), 2) similar modulations of pupil diameter and locus coeruleus activity under certain task conditions such as changes in utility that affect behavioral engagement (Jepma and Nieuwenhuis, 2010;Gilzenrat et al., 2010) and 3) a proposed anatomical substrate involving common activation from the nucleus paragigantocellularis, which contributes to both locus coeruleus and sympathetic nervous system function (Nieuwenhuis et al., 2010;Aston-Jones et al., 1986).

I examined whether LC activation might dictate an adaptive learning rate to allow optimal predictions in dynamic environments by using pupillometry to measure arousal levels, and by proxy LC activity, while subjects made predictive inferences in a dynamic environment. Subject pupils were larger during periods of uncertainty after change-points and increased in diameter during change-point trials. Trial-by-trial learning rates used by subjects could be predicted based on pupil response both within and across subjects. In addition, a task irrelevant manipulation that caused a robust increase in pupil diameter also systematically altered learning rates, suggesting that the pupil-linked arousal system plays a causal role in setting the adaptive learning used to optimize inference in dynamic environments.

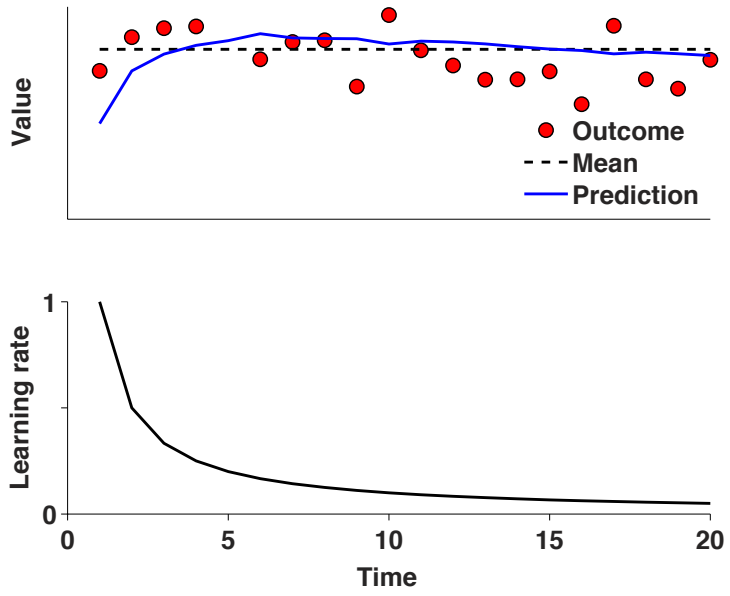


Figure 1.1 Optimal inference in a stable but noisy environment. A) Optimal inference in a noisy but stable environment. B) Influence of each successive observation on updated prediction measured in units of the learning rate from a delta-rule model.

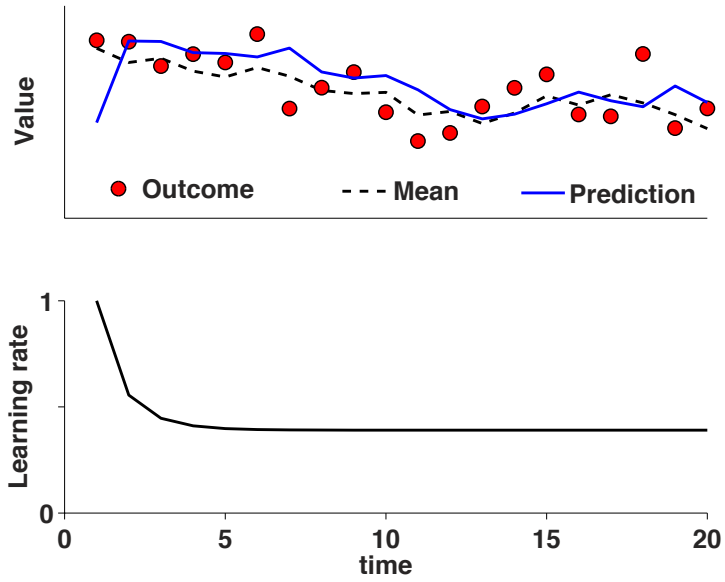


Figure 1.2 Optimal inference in a noisy and continuously drifting environment. A) Optimal inference in a noisy and continuously drifting environment. B) Influence of each successive observation on updated prediction measured in units of the learning rate from a delta-rule model.

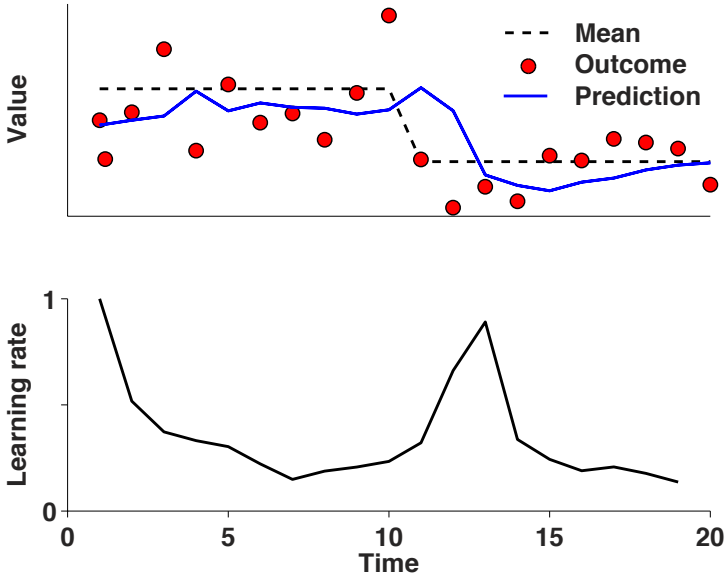


Figure 1.3 Optimal inference in a discontinuously changing environment with unknown change-point locations. A) Predicted (blue) and actual (red) outcomes over time (ordinate). B) Estimation of the influence of each observation on the updated prediction. Although the optimal algorithm cannot be represented as a simple delta rule, here we compute the learning rate for each trial that would allow a delta rule to reproduce the behavior of the optimal model exactly.

‘

CHAPTER 2

An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment

Matthew R. Nassar, Robert C. Wilson, Benjamin Heasly, and Joshua I. Gold.
Journal of Neuroscience, 2010, 30:12366-78

Abstract

Maintaining appropriate beliefs about variables needed for effective decision-making can be difficult in a dynamic environment. One key issue is the amount of influence that unexpected outcomes should have on existing beliefs. In general, outcomes that are unexpected because of a fundamental change in the environment should carry more influence than outcomes that are unexpected because of persistent environmental stochasticity. Here we use a novel task to characterize how well human subjects follow these principles under a range of conditions. We show that the influence of an outcome depends on both the error made in predicting that outcome and the number of similar outcomes experienced previously. We also show that the exact nature of these tendencies varies considerably across subjects. Finally, we show that these patterns of behavior are consistent with a computationally simple reduction of an ideal-observer model. The model adjusts the influence of newly experienced outcomes according to ongoing estimates of uncertainty and the probability of a fundamental change in the process by which outcomes are generated. A prior that quantifies the expected frequency of such environmental changes accounts for individual variability, including a positive

relationship between subjective certainty and the degree to which new information influences existing beliefs. The results suggest that the brain adaptively regulates the influence of decision outcomes on existing beliefs using straightforward updating rules that take into account both recent outcomes and prior expectations about higher-order environmental structure.

Introduction

Behavior often depends on the ability to predict future outcomes from past experiences. In an unchanging environment, beliefs that underlie effective predictions are typically stable. However, in a dynamic environment the past does not always predict the future, and beliefs must therefore sometimes adapt rapidly, particularly after unexpected outcomes (Rushworth and Behrens, 2008). One common and effective algorithm for describing such adaptation is the delta rule (Sutton and Barto, 1998;Williams, 1992):

$$B_{t+1} = B_t + \alpha_t \times \delta_t \qquad [1]$$

where a new belief at time t+1 (B_{t+1}) depends on the previous belief (B_t) and the error made in predicting the most recent outcome (δ_t). The influence of the new outcome is controlled by the learning rate (α_t). When α_t=0, the updated belief reflects the previous belief but not the most recent outcome. When α_t=1, the updated belief reflects the most recent outcome but not the previous belief.

Assigning influence to new outcomes in a dynamic environment is difficult because the source of prediction errors is generally unknown (Behrens et al., 2007;Yu and Dayan, 2005). One source of error is stochastic fluctuations in an otherwise stable action-outcome relationship ("noise"). Noise can make each outcome a bad

predictor of the next, implying that new outcomes should affect beliefs only minimally. Another source of error is a fundamental change-point in the action-outcome relationship ("volatility"). Change-points can render historical outcomes irrelevant, implying that new outcomes should influence beliefs strongly.

Previous work has shown that, on average, human subjects elevate learning rates during periods of volatility on probabilistic decision tasks. Such behavior can be fit by both a Bayesian model for optimal belief updating and a computationally frugal extension of delta-rule updating (Behrens et al., 2007;Krugel et al., 2009). Our goal was to build on these studies and, instead of relying on model fitting to average behavior on simple choice tasks, directly measure the learning rates used by subjects in noisy and volatile environments. We also sought to reconcile these data with both the Bayesian and delta-rule models to better understand the underlying neural computations.

We developed a novel task that required subjects to predict the next numerical value to be presented in a sequence (Fig. 1A). The values were chosen randomly from a Gaussian distribution with a mean that changed occasionally, giving rise to both noisy and volatile prediction errors. The subject updated each prediction as a fraction of the current prediction error, equivalent to setting the learning rate (α_t). Thus, the task provided a trial-by-trial measurement of outcome influence.

We present several new findings. First, subjects recognized change-points from unexpectedly large prediction errors, which temporarily increased prediction uncertainty and the influence of subsequent outcomes. Second, there were strong individual differences, including some subjects who were highly influenced by new outcomes and others who generally ignored them. Third, these behaviors were consistent with a modified delta-rule model, derived from a systematic reduction of the Bayesian ideal observer (Wilson et al., 2010;Adams and MacKay, 2007;Fearnhead and Liu, 2007), in which individual differences were attributed to different expectations about the rate of occurrence of change-points. The results provide a novel, quantitative framework describing the dynamics of belief updating in a changing environment.

Materials and Methods

Behavioral tasks

Human subject protocols were approved by the University of Pennsylvania internal review board. Thirty subjects (13 female, 17 male; mean age = 25.2 years, range = 19 – 31 years) participated in the study after providing informed consent. Twenty-seven subjects completed both the estimation and confidence tasks (see below), in that order. One subject completed only the estimation task, and two subjects completed only the confidence task.

Estimation task. This task required subjects to predict each subsequent number to be presented in a series of numbers. For each trial *t,* a single number (X_t) was presented that was a rounded pick sampled independently and identically from a Gaussian distribution whose mean (μ_t) changed at unsignaled change-points and whose standard deviation (σ_t) was fixed for each of the four experimental blocks of 200 trials (5, 15, 25, or 35, presented blockwise in ascending order for 14 subjects and descending order for 14 subjects); that is, $X_t \sim \mathcal{N}(\mu_t, \sigma_t)$. Change-points in the mean of the generative distribution occurred after at least 5 trials plus a random pick from an exponential distribution with a mean of 20 trials. Thus, the true rate of change-points, or hazard rate (H, in units of change-points/trial) was 0 for the first 5 trials after a change-point and 0.05 for all trials thereafter. The average hazard rate of a change-point across all trials was 0.04.

The display showed a line representing the range of possible numbers (0 to 300), a bar representing the current estimate, a bar representing the most recent number presented, and a line between these bars representing the current prediction error (Fig. 1A). The subject updated his or her prediction on each trial to an integer value between the previous prediction and the newly generated number (ensuring that learning rates would fall between zero and one) using a video gamepad. Each subject first performed two training blocks (standard deviations of 3 and 20). Each session consisted of four test blocks.

Subjects were told that the numbers were generated from a noisy process that would change over the course of the task. They were instructed to minimize their prediction errors, on average, across all blocks of the task; i.e., minimize $\langle|\delta_t|\rangle$. Payout depended on how well they achieved this goal. Because prediction errors depended substantially on the specific sequence of numbers generated for the given session, we computed two benchmark error magnitudes to help determine payout. The lower benchmark (LB) was computed as the mean absolute difference between sequential generated numbers, $\langle|X_t - X_{t-1}|\rangle$. The higher benchmark (HB) was the mean difference between mean of the generative distribution on the previous trial and the generated number, $\langle|X_t - \mu_{t-1}|\rangle$. Payout was computed as follows:

$\langle\|\delta_t\|\rangle$ > LB	= \$8
LB > $\langle\|\delta_t\|\rangle$ > 2/3 LB + 1/3 HB	= \$10
2/3 LB + 1/3 HB > $\langle\|\delta_t\|\rangle$ > 1/2 (LB + HB)	= \$12
$\langle\|\delta_t\|\rangle$ < 1/2 (LB + HB)	= \$15

The reduced Bayesian model, when given the true hazard rate (0.04), was capable of achieving the maximum payout for all task sessions.

Confidence task. This task was similar to the estimation task, except subjects also indicated their confidence in each prediction. A series of numbers was generated as above (3 blocks of 200 trials with standard deviations 10, 20, and 30). Subjects were

instructed not only to make a prediction on each trial, as described above, but also to indicate a symmetric window around the prediction that they believed, with 85% confidence, would contain the next number. Subjects earned "points" on each trial in which the generated number fell within the specified window. Feedback included a sound to indicate when the generated number fell within the specified window and a running tally of points earned by the subject.

Point values were chosen to incentivize confidence windows that were 85% likely to contain the next number in the sequence, as follows. The expected value of points earned across all possible window sizes was defined by a Gaussian distribution with a mean equal to the minimum range capable of including 85% of the probability density under the generative distribution. The number of points at stake for a given window size was computed by dividing the expected value of that window size by the probability that the new outcome would fall within this window (assuming the window is centered on the actual mean of the generative distribution). Thus, total points earned at the end of the session depended both on the ability to correctly estimate the mean, but also the use of windows that approximated 85% confidence intervals. Points earned by subjects (SP) were compared to the number points that would be earned by the two benchmark strategies described above, if those strategies used confidence-window sizes that maximized expected point value (LBP & HBP). Payout was computed as follows:

SP < LBP	= \$8
LBP < SP < 2/3 LBP + 1/3 HBP	= \$10
2/3 LBP + 1/3 HBP < SP < 1/2 (LBP + HBP)	= \$12
SP > 1/2 (LBP + HBP)	= \$15

Data analysis. Prediction errors were computed by subtracting the subject's prediction (B_t in Eq. 1) from the actual outcome (X_t) on each trial. Learning rates were calculated for each trial according to Eq. 1: the current update, $B_{t+1}–B_t$, was divided by the current prediction error, δ_t. Trial-by-trial error z-scores were computed by dividing the absolute error magnitude by the standard deviation of the generative distribution. Error-independent learning rates were computed by first fitting a sigmoid-shaped, cumulative Weibull function (with four parameters, governing shape, offset, lower bound, and upper bound) to learning rate as a function of error z-score. The residuals to this fit represented learning rates that were relatively independent of error magnitude. Relative uncertainty was computed by taking the z-score of confidence window size for a given generative standard deviation.

Models

Optimal task performance requires knowledge about the probability distribution $p(X_{t+1} \mid X_{1:t})$, which is the predictive distribution over possible outcomes on trial $t+1$, $p(X_{t+1})$, given all previous samples, $(X_{1:t})$. Optimal performance on the estimation

task requires specifying the mean of this predictive distribution, whereas optimal performance on the confidence task requires knowledge about the width of this predictive distribution, as well. Computing the mean of the current predictive distribution is difficult because of unsignaled change-points in the generative process. If the most recent change-point was known to have occurred r_t trials ago, the predictive mean could be computed simply by taking the mean of the last r_t outcomes:

$$\hat{\mu}_{t+1}(r_t) = \frac{1}{r_t} \sum_{t-r_t+1}^{t} x_i \qquad [2]$$

However, because change-points are unsignaled, the optimal solution must be reformulated in terms of all possible run-lengths, which describe the number of data points that could have been generated from the current distribution:

$$p(X_{t+1}|X_{1:t}) = \sum_{r_t} p(X_{t+1}|r_t)p(r_t|X_{1:t}) \qquad [3a]$$

where $p(X_{t+1}| r_t)$ is the predictive distribution in X, conditional on run length, which is computed from the previous r_t samples treated as if they were generated by the current distribution*:*

$$p(X_{t+1}|r_t) = p(X_{t+1}|X_{t-r_t+1:t}) = \int d\mu d\sigma \quad p(x_{t+1}|\mu,\sigma)p(\mu,\sigma|x_{t-r_t+1:t}) \qquad [3b]$$

and $p(r_t \mid X_{1:t})$ is the distribution of possible run lengths, given all previous data. Thus, the mean of the predictive distribution can be described in terms of r_t:

$$\mu_{t+1} = \sum_{r_t} \mu_{t+1}(r_t)p(r_t|X_{1:t}) \qquad [4]$$

We applied two different classes of model to our task: a Bayesian ideal-observer model that computes the full run-length distribution, and a reduced Bayesian model that approximates the run-length distribution using only its first moment.

Full Bayesian model. The full Bayesian model computes the entire run-length distribution recursively to generate the predictive distribution (Fearnhead and Liu, 2007;Adams and MacKay, 2007). An alternative but mathematically equivalent approach, which does not use run length explicitly but instead maintains representations of probability distributions over all possible values of the parameters of the generative process (Behrens et al., 2007), is also possible, but we do not use it here. Both approaches depend strongly on the hazard rate, which specifies the prior probability of a change-point. When the hazard rate is known, the full, recursive solution of the run-length-based model uses the message-passing algorithm depicted in Fig. 6A. After t trials, the model updates predictive distributions (in X_{t+1}) for each of the t+1 possible run lengths, as well as the

probability distribution over those run lengths. When the hazard rate is unknown, like for our subjects, the optimal solution is more complicated. It requires maintaining a distribution over not only possible run lengths, but also possible hazard rates, thus at least $(t+1)^3$ separate predictive distributions are required for inference at time t (Wilson et al., 2010). To make this algorithm more tractable computationally, we implemented a pruning algorithm previously shown to reduce computations with a minimal loss of performance (Wilson et al., 2010).

Reduced Bayesian model. We also developed an even more computationally tractable and neurally feasible inference algorithm that is based on a systematic reduction of the full Bayesian model. In this model, the predictive distribution is not computed across all possible run lengths but instead with respect to a single, expected run length ($\hat{r}_t$). On each trial, the model considers two possibilities: that a change-point did or did not occur. Accordingly, the probability of a change-point (*cp*) on a given trial, Ω, can be computed using Bayes' rule:

$$
\begin{aligned}
p(cp|X_t) = \Omega_t &= \frac{p(X_t|cp)p(cp)}{p(X_t)} \\
&= \frac{p(X_t|cp)p(cp)}{p(X_t|cp)p(cp) \;+\; p(X_t|\text{-}cp)p(\text{-}cp)} \\
&= \frac{U_{\text{-}}(X_t|0,\ 300)H}{U_{\text{-}}(X_t|0,\ 300)H + \mathcal{N}_{\text{-}}(X_t|\hat{\mu}_t, \hat{\sigma}_t^2)(1-H)}
\end{aligned}
\qquad [5]
$$

where $U(X_t|0, 300)$ is the uniform distribution from which X_t is generated (independent of the previous generative distribution) if a change-point occurred, $\mathcal{N}(X_t|\hat{\mu}_t, \hat{\sigma}_t^2)$ is the predictive distribution if a change-point did not occur (and thus depends on both $\hat{r}_t$ and recent outcomes), and H is the hazard rate (set to 0.04, the average value for the task).

The variance of the predictive distribution depends on both the run length and the expected amount of noise from the generative distribution:

$$\hat{\sigma}_t^2 = N^2 + \frac{N^2}{\hat{r}_t} \qquad [6]$$

where N is the standard deviation of the generative distribution; see below for an alternative model in which this quantity is inferred from the data. In Eq. 6, the first term on the right-hand side reflects uncertainty about the outcome for the given μ, and the second term reflects uncertainty about the actual location of μ. As run length increases, uncertainty about the location of μ decreases, but uncertainty implicit in the stochasticity of the generative process (noise) remains.

The expected (mean) value of the predictive distribution is based on two possibilities, one that a change-point occurred and thus only the most recent data point is relevant:

$$\mu_t^{cp} = X_t \quad [7a]$$

and a second possibility that a change-point did not occur and thus the mean is updated to take into account the new data point:

$$\mu_t^{\neg cp} = \frac{X_t + \hat{r}_t \times \hat{\mu}_{t-1}}{\hat{r}_t + 1} \quad [7b]$$

The mean of the posterior distribution is an average of these two possibilities, weighted by the probability that a change-point occurred:

$$\hat{\mu}_t = \frac{(X_t + \hat{r}_t \times \hat{\mu}_{t-1})(1 - \Omega_t)}{\hat{r}_t + 1} + \Omega_t X_t \quad [7c]$$

An advantage of this approach is that this update equation can be rearranged as a delta rule:

$$\hat{\mu}_t = \hat{\mu}_{t-1} + \alpha_t \times \delta_t \quad [7d]$$

where δ_t is the prediction error $(X_t - \hat{\mu}_t)$ and α_t is the learning rate:

$$\alpha_t = \frac{1 + \Omega_t r_t}{r_t + 1} \quad [7e]$$

Similarly, the expected run-length is updated on each trial according to the two possible generative scenarios and their respective probabilities:

$$\hat{r}_{t+1} = (\hat{r}_t + 1)(1 - \Omega_t) + \Omega_t \quad [8]$$

Computing best-fitting hazard rates. To test whether prior expectations about hazard rate could account for across-subject variability, we fit the reduced model to data from each subject with the hazard rate as a free parameter. The model was applied separately to each block, with *N* (Eq. 6) fixed to the true generative standard deviation for that block. The best-fitting hazard rates were determined using a constrained search algorithm (fmincon in MATLAB, min/max hazard=0/1) that found the value of *H* that minimized the total squared difference between model and subject predictions.

We considered two possible implementations of the reduced Bayesian model. The first made predictions as the mean of the current predictive distribution ($\hat{\mu}_t$). The second made predictions as the mean of the distribution at time *t*+1. This quantity depends on not only the current predictive distribution, but also the uniform prior distribution, because there is a possibility that a change-point might occur and thus the next number would come from a new distribution. All analyses were done with the first implementation, which provided better fits to the behavioral data (the ratio of Bayesian information criteria of fits using the first versus the second model had a

median [interquartile range] value across task blocks of 0.93 [0.86-0.97], paired Wilcoxon test for H_0: median=0, $p < 0.001$).

Inferring noise using the reduced model. Because subjects were not told explicitly the amount of noise (the standard deviation of the distributions used to generate the numbers), we also developed a version of the reduced model that included an algorithm to infer the amount of noise from the data. This model computes a quantity whose expectation is equal to the generative noise:

$$\hat{N}^2_{t+1} = \hat{N}^2_t + \alpha_{t(N)} \times \left(\frac{\hat{r}_t \delta_t^2}{\hat{r}_t + 1} - \hat{N}^2_t \right) \quad [9]$$

where $\hat{N}^2$ is the inferred variance, which is updated according to a delta rule that depends on both the run length and prediction error. The expected value of the prediction-error term (in parentheses) is zero for non-change-point trials.

The learning rate, $\alpha_{t(N)}$, affects the extent to which new prediction errors influence the noise estimate and was assumed to be proportional to the probability that the trial contained information about variance (i.e., was not a change-point trial) and inversely proportional to the the amount of such information previously collected:

$$\alpha_{t(N)} = \frac{1 - \Omega_t}{\sum_{1}^{t}(1 - \Omega_t)} \qquad [10]$$

Thus, $\alpha_{t(N)}$ goes to zero if a change-point is likely to have occurred or as the number of previous non-change-point trials goes to infinity.

Although this algorithm is capable of inferring noise, it uses learning rates that tend toward zero after only a few trials and thus seem unlikely to be used by subjects. We therefore modeled the possibility that learning rates used to infer noise were related to those used to infer μ. Specifically, we instituted a minimum $\alpha_{(N)}$ that depends on the hazard rate (H), the model parameter that dictates the average learning rate (see Fig. 8B):

$$\alpha_{(N)}^{MIN} = \kappa H \times (1 - \Omega_t) \qquad [11]$$

where κ is a scaling constant. For Fig. 9C,F,I, κ was set to 0.5 (results were similar using values ranging from 0.2 to 1).

Reduced Bayesian model with under-weighted likelihood information. To more closely match our measured behavioral data, we revised the reduced model to reduce the weight of likelihood information in change-point detection. Thus, in lieu of Eq. 5, this version computed Ω_t as:

$$\Omega_t = \frac{U(X_t|0,\ 300)^{\lambda} H}{U(X_t|0,\ 300)^{\lambda} H + \mathcal{N}(X_t|\hat{\mu}_t, \hat{\sigma}_i^2)^{\lambda}(1-H)} \quad [12]$$

where the likelihood weight, λ, is a fractional term (0…1) that limits the use of likelihood information in change-point detection. When λ=0, the model becomes a fixed learning rate delta-rule model in which the learning rate is determined by *H*. When λ=1, the model is equivalent to the reduced Bayesian model discussed above. This model was fit to subject data with λ and *H* as free parameters, using a constrained search algorithm to minimize the squared difference between subject and model predictions.

Reduced Bayesian model with drifting mean. A final alternative model used a generative framework that assumed that the mean of the generative distribution drifted from trial to trial. Although such drift did not actually occur, we wanted to test whether subjects behaved as if it did. This kind of drift is often accounted for using a Kalman filter, which provides an efficient means for updating beliefs based on noisy samples from a drifting process. However, this approach performs poorly in environments with discontinuous changes, such as in our task. Conversely, the pure change-point model provides an efficient algorithm for updating beliefs when the world changes only at discreet change-points. We therefore combined these approaches, as follows. The drift was assumed to be $N \sim (0, D^2)$, where *D* is the drift

rate. This generative framework prescribes more uncertainty about the location of the true mean, which leads to a wider predictive distribution (to replace Eq. 6):

$$\hat{\sigma}_t^2 = \hat{\sigma}^{*2}_t + \frac{\hat{\sigma}^{*2}_t}{\hat{r}_t} + D^2 \qquad [13]$$

To consolidate uncertainty about the mean into a single variable and allow correct computation of the learning rate (Eq. 7e), we re-computed the run length to reflect the total uncertainty about the mean of the distribution:

$$\hat{r}_{t*} = \frac{N^2}{\sigma_t^2} \qquad [14]$$

This adjusted run length was used for the learning rate (Eq. 7e) and update (Eq. 8) equations. This model was fit to subject data with *N*, *D*, and *H* as free parameters.

Results

We used a novel estimation task to quantify how human subjects update beliefs in the face of both noise and volatility. Below, we first describe the task and show that subjects tended to use different learning rates to update beliefs under different conditions. Second, we show that the choice of learning rate depended on the degree to which estimation errors were larger than expected, the recency of such an

unexpectedly large error, and the relative uncertainty of the subject. Third, we introduce a novel model, which is a form of Bayesian ideal observer reduced to implement delta-rule updating, that captures many key aspects of the data. Fourth, we use the model to show that individual differences in performance suggest differences in whether errors tend to be interpreted as either noise or volatility. Fifth, we introduce several model variants that even more closely match human behavior.

Learning rate varied from trial to trial. Thirty subjects performed the estimation and confidence tasks in 57 total sessions. The tasks required the subject to sequentially update a belief about the next number in a series. The numbers were picked from a Gaussian distribution with a mean that changed at random intervals (change-points) and a standard deviation (noise) that was stable over each block of 200 trials (Fig. 1A). Subjects were instructed to estimate the next number that would be generated by the computer and to minimize the error on these estimates. Visual feedback consisted of a bar that reflected the difference between the subject's estimate and the most recently generated number shown on each trial and the mean absolute error shown at the end of each 200-trial block. Payment scaled inversely with the mean absolute error for the session.

In principle, payout maximization required basing estimates on the median (in this case also the mean) of the generative distribution. However, information about the

generative distribution was not given to subjects explicitly. Therefore, they were required to infer properties of this distribution based on the previously observed numbers. The behavioral data were consistent with a sequential-updating strategy that approximated the central tendency of the generative distribution (data from an example session are shown in Fig. 1B). Estimates tended to approximate the mean during periods of stability and then change relatively rapidly at change-points in the generative distribution to re-settle at the new mean.

In theory, a delta-rule algorithm might generate qualitatively similar, adaptive behavior even when the learning rate is fixed to a constant value, because update magnitude would be proportional to error magnitude. However, such a fixed learning-rate model was not a valid description of behavior for this task (Fig. 1D). The subjects used learning rates that differed from trial to trial and spanned the allowed range from 0 to 1. Moreover, although the learning rates used by different subjects varied considerably (the mean learning rate per subject ranged from 0.07 to 0.71), the particular sequence of learning rates chosen by each subject provided better predictions than randomly ordered sequences of the same values (the median [95% confidence intervals] value, computed across subjects, of the difference in mean absolute error between 1000 randomized sequences versus the actual sequence per subject = 2.59 [2.46 2.72], Wilcoxon test for H_0:median=0, $p < 0.001$). Thus, subjects made effective predictions by assigning some outcomes more

influence than others. The remaining analyses aimed to understand the rules that governed how this assignment of influence was made.

Learning rate depended on surprising outcomes. One important factor that governs the magnitude of the chosen learning rate is the occurrence of change-points in the mean of the generative distribution. In general, when a change-point occurs, information obtained prior to the change-point is no longer useful in making predictions, and thus the learning rate should increase to emphasize newly arriving information. Consistent with this idea, subjects typically used higher learning rates on change-point trials (the first trial of a new mean of the generative distribution) than on other trials (Fig. 2A).

Change-point locations were unknown to the subjects and thus must have been inferred from statistical features of the sequential trial outcomes. One such feature is the magnitude of error (δ) relative to expected errors. Change-points are likely to correspond to a surprisingly large error, where surprise is defined with respect to the expectation of $|\delta|$. Consistent with this idea, the overall positive relationship between α and $|\delta|$ depended heavily on the standard deviation of the generative distribution (Fig. 2B,C). A given absolute error magnitude tended to lead to a higher learning rate for less noisy distributions, when such an error was less expected. To further quantify this effect, we normalized absolute prediction errors by the standard deviation of the generative distribution. This "z-scored error" was

predictive of learning rate, relatively independent of the noise magnitude (Fig. 2C; Spearman's ρ across all subjects was 0.15, permutation test for H_0: ρ=0, $p < 0.001$). We also note that this basic trend was consistent but varied considerably in magnitude across subjects (Fig. 2D), a finding that we analyze in more detail below.

The effect of a change-point on the choice of learning rate persisted for many trials beyond the occurrence of the change-point. In the trials following a change-point, prediction errors tended to decrease sharply, as subjects adjusted their estimates to match the new distribution (Fig. 3A, gray). In contrast, learning rates tended to decrease more gradually following a change-point (Fig. 3A, black). This gradual decay in learning rate did not depend on the magnitude of the relative (z-scored) prediction error: after adjusting for the relationship between learning rate and z-scored error (see Fig. 2D), there were still changes in learning rate that persisted for many trials after a change-point. The peak value in this adjusted learning rate tended to occur on the first trial following a change-point and then decay gradually (Fig. 3B).

Learning rate magnitude was related to confidence. Ideal-observer theory suggests that any information acquired after a change-point should be highly influential because the observer is uncertain about the current belief (Wilson et al., 2010;Yu and Dayan, 2003). Conversely, subsequent acquisition of information from a stable environment should lead the observer to become more confident and less

influenced by each new outcome. To examine this relationship between confidence and learning rate and test how well it could explain the slowly decaying learning rates shown in Fig. 3, we trained subjects on a task that required specification of an 85% confidence window. This task probed not only the central tendency of the subject's belief about the generative distribution, but also uncertainty that subjects had in their own estimates. The example session in Fig. 4A shows estimates (solid blue) and the 85% confidence windows (dashdot blue) specified by a subject over the course of a full session.

There was a systematic relationship between the size of the confidence window and the standard deviation of the generative distribution, with greater uncertainty corresponding to higher noise (Fig. 4B). Moreover, subjects tended to make trial-by-trial adjustments to the confidence window to reflect changes in uncertainty, particularly after a change-point. On average, confidence windows were largest after a change-point and gradually became smaller as subjects collected more data from the new distribution (Fig. 4C). This effect was largest when there was less noise and change-points were most easily detectable. The time course of this decay is similar to the error-independent decay in learning rate (compare 4C and 3B).

In addition to these general trends across subjects, there was considerable individual variability in the choice of confidence-window size (e.g., whiskers in Fig. 4B) that was related to learning rate. This relationship is typified by the behavior of

two example subjects, shown in Fig. 5A & B. Subject SG (Fig. 5A) used small learning rates and tended to specify large confidence windows, indicating high uncertainty (Fig. 5A). In contrast, subject LY tended to use large learning rates and small confidence windows (Fig. 5B). In addition to these differences in mean learning rate and uncertainty between these two subjects, there was also a difference in the relationship between the two variables. Subject SG, who tended to use small learning rates overall, also tended to use relatively larger learning rates on trials in which she was most uncertain about her previous estimate. In contrast, subject LY, who tended to use large learning rates overall, also tended to use smaller learning rates on trials in which she was most uncertain about her previous estimate.

Across subjects, mean confidence-window size was negatively correlated with mean learning rate (Fig. 5C). This relationship implies that subjects who tended to use large learning rates and thus be highly influenced by new information (like subject LY) also tended to be more confident in their estimates. Moreover, the mean learning rate used by a given subject across all conditions was predictive of how that subject's learning rate related to the confidence-window size from the previous trial (Fig. 5D). Subjects who tended to use small learning rates (like subject SG) chose larger learning rates following trials in which they specified a large confidence window, suggesting that these subjects were most influenced by outcomes when they were most uncertain. In contrast, subjects who tended to use large learning rates (like subject LY) chose larger learning rates following trials in

which they specified a small confidence window, suggesting that these subjects were most influenced by outcomes when they were most certain.

The overall negative relationship between confidence window size and learning rate might seem at first to contradict ideal-observer theory. As noted above, an ideal observer should make extensive use of new information and therefore use high learning rates when uncertainty is high. However, as we show in the next section there are at least two sources of uncertainty, which for this task have potentially different effects on an ideal observer. Taking into account these multiple sources of uncertainty can help to clarify the relationship between actual and optimal behavior.

A reduced Bayesian delta-rule model. Optimal prediction in a discontinuously changing environment is a computationally demanding problem (Yu and Dayan, 2005;Wilson et al., 2010). A solution to this problem requires maintaining a set of nodes, each of which maintains the predictive distribution for a possible duration of stability, or run length (*r*; Adams and MacKay, 2007;Fearnhead and Liu, 2007). Optimal predictions are made on each trial by taking a weighted average of these nodes. However, in this approach the number of nodes scales linearly with the number of observations if the rate at which change-points occur, or hazard rate, is known (Fig. 6A) or with the number of observations cubed if the hazard rate is unknown (Wilson et al., 2010). Thus, the optimal solution to our task must maintain

and update likelihood estimates for thousands of predictions based on different possible generative scenarios.

Our goal was to test models that could at least approximate optimal performance while using more plausible mechanisms. We therefore considered a particular reduction of the full Bayesian ideal-observer model (Fig. 6B). Instead of maintaining information about each possible value of *r*, this model maintains only a single "expected run length" ($\hat{r}$) node. On each trial, the model considers two possible generative scenarios: that the newly generated number came from the same distribution as the previous one, or that the new number came from a new distribution. Probabilities of these possible scenarios are computed according to Bayes' rule, and $\hat{r}$ is updated accordingly. A compelling feature of this complexity reduction is that the new model implements a form of delta rule (Eq. 7). The learning rate depends on both $\hat{r}$ and the probability that a change-point occurred (Eq. 7e). In the limit as the probability of a change-point goes to zero, the model prescribes a learning rate equal to $1/(\hat{r}+1)$ (Fig. 6C). However, as the probability of a change-point goes to one, the learning rate increases linearly toward one, consistent with a discarding of historical information that is unlikely to pertain to the new environment. The reduced Bayesian model achieves similar performance to that of the full model, and both models performed better than a delta rule that used a fixed learning rate that minimized absolute errors over a session (Fig. 6D).

The reduced Bayesian model exhibited many of the same characteristics as human subjects on the estimation task (Fig. 7). Like for the psychophysical data, the model's choice of learning rate tended to increase as a function of error magnitude, with larger increases when the standard deviation of the prior, stable distribution was small (Fig. 7A–C). Moreover, the model tended to have higher learning rates on the trial after a change-point, which then decayed gradually over many trials (Fig. 7D & E). In the model, this gradual decay is caused by the decay in uncertainty occurring over the same period (Fig. 7F). Despite these overall trends that matched the subjects' behavior, the model tended to perform much better and in fact closely matched the performance of the full Bayesian model (Fig. 6D).

A straightforward manipulation of the model could also reproduce much of the across-subject variability. A key parameter of the model is the hazard rate (H), which describes the expected rate of change-points. This parameter has been shown to differ across subjects in change-point detection tasks (Steyvers and Brown, 2006). We fit the model to data from each subject separately for each different standard deviation of the generative process with the hazard rate as a single free parameter. This procedure allowed us to test whether the reduced model could explain not only the trends in subject learning rates, but also whether differences across subjects could be explained by varying expectations about the instability of the generative environment.

Subjects that tended to use higher learning rates were best fit by higher hazard rates (Fig. 8A). This effect is due largely to the fact that higher hazard-rate models tend to use higher learning rates (Fig. 8B) because they infer change-points more frequently. The fit hazard rates tended to be much larger than the actual hazard rate of change-points in our task, which, averaged across all conditions, was equal to 0.04 (vertical dashed line in Fig. 8A). Thus, the model suggests that subjects tended to overestimate the frequency with which changes occur, to a degree that varied considerably across subjects. Moreover, the different fit values of the hazard rate affected model performance in a manner that at least qualitatively matched across-subject differences, including the dependence of learning rate on z-scored error (compare Figs. 2D and 7C).

Models with inferred noise better matched behavior. We extended the reduced Bayesian model to account for our finding that subjects who tended to be most confident in their estimates were also the quickest to update those estimates given new information (Fig. 5C). This finding seems counterintuitive to the notion that learning rate should be largest when confidence is lowest (and thus new information should be highly informative). However, two main types of uncertainty exist within the task that have opposite effects on the learning rate (Eq. 6). One type of uncertainty is related to run length: when the run length is small, few samples contribute to the estimate of the mean of the generative distribution, making that estimate uncertain and therefore imposing higher learning rates (Fig. 6C). The

second type of uncertainty is related to the expected standard deviation of the generative distribution, or noise: when the estimate of noise is high, the model tends to underestimate the probability of a change-point, leading to a decrease in learning rate. We propose that this second form of uncertainty has a strong effect on the choice of learning rates.

To examine this idea, we extended the model to include different forms of noise estimation (Fig. 9) and compared performance of each form of the model to the behavioral data presented in Fig. 5. The simplest form used estimates of noise that were fixed within a block (Fig. 9A). In this case, overall uncertainty, like learning rate, declined with run length (Eq. 6). Higher hazard-rate models inferred lower run lengths, on average, leading to a strong, positive relationship between mean uncertainty and learning rate across simulated sessions (Fig. 9D). There was also a strong, positive relationship between uncertainty and learning rate across simulated trials that tended to decline as a function of the mean learning rate, but never to below zero (Fig. 9G). Thus, this model did not match the behavioral data.

The second model used a sequentially updated estimate of noise (Eq. 9). When applied to the same task conditions that the subjects experienced, this model generated estimates of noise that were highly unstable early in each session but then stabilized as more information was collected (Fig. 9B). However, even these stabilized estimates tended not to match the value of the true generative noise (the

ratio of estimated to actual noise ranged from 0.5 to 1.2 after 200 simulated trials, where hazard rate was set to the value that best fit performance of each individual subject). The model's dependence on hazard rate (in particular via biased values of $\hat{\tau}$ in the prediction-error term in Eq. 9) gave rise to a negative relationship between hazard rate and noise estimates, because with high hazard rates, errors tended to be interpreted as change-points rather than noise. Because high hazard rates correspond to larger learning rates, on average, these effects resulted in a negative relationship between overall uncertainty and learning rate, like in the behavioral data (Fig. 9E). There was also a strong, positive relationship between uncertainty and learning rate across simulated trials that tended to decline as a function of the mean learning rate, but never to below zero (Fig. 9H). Thus, this model also did not match the behavioral data.

The third model used a more realistic, sub-optimal strategy for inferring noise (Eq. 11). This model assumed that beliefs about the noise of the generative distribution, like beliefs about its mean, were updated using learning rates that varied substantially across subjects. In particular, this model assumed that beliefs about noise were updated using learning rates proportional to those used to update beliefs about the mean of the distribution. This procedure led to more variable estimates of noise than the other two models (Fig. 9C) and, like the second model, a strong, negative relationship between overall uncertainy and learning rate across simulated sessions (Fig. 9F). Moreover, unlike the second model and like the behavioral data,

this model showed both positive and negative correlations between trial-by-trial uncertainty and learning rate that depended on hazard rate (Fig. 9I). Specifically, high hazard rates corresponded to a negative correlation between learning rate and total uncertainty, whereas low hazard rates corresponded to a positive correlation between learning rate and uncertainty. These results imply that subjects use an imperfect noise-inference algorithm that updates beliefs about noise rapidly and in proportion to the rate at which they update beliefs about the mean, μ. This algorithm leads subjects who expect more changes to see less noise and can account for inter-subject variability in the relationship between uncertainty and learning rate.

Thus, the hazard rate is central to an account of the across-subject variability in learning rates, uncertainty, and the relationship between the two. This account suggests a strategic tradeoff that was navigated in different ways by different subjects (Fig. 10). Subjects who were fit by high hazard rates tended to perform relatively well in the first few trials after a change-point but relatively poorly during periods of stability. Conversely, low-hazard subjects tended to perform relatively poorly after change-points but well during periods of stability. Thus, the choice of hazard rate reflected a tradeoff between successful prediction amid noise and successful adaptation after change-points.

*Models that under-weigh errors better matched behavior***.** Above we used a model with only a single free parameter, the hazard rate, to describe the main trends in updating behavior for individual subjects and the population. However, this model was quantitatively inconsistent with subject performance. In particular, subjects did not react to change-points as effectively as the model. Subjects tended to use higher learning rates after change-points than on other trials, but to a lesser extent than the model (Fig. 11A). This sub-optimal behavior of human subjects reflected a relationship between learning rate and z-scored error that was too flat (Fig. 11B).

One explanation for this difference might be that subjects underuse likelihood information when assessing whether a change-point occurred on a given trial. Adding a parameter (λ in Eq. 12) to the reduced model that allows for such sub-optimal computation lets the model range from a fixed learning rate delta-rule model (λ=0) to the reduced-Bayesian model (λ=1). Fits of this parameter indicate that all subjects fall between the two extremes, and that most of the subjects seemed to adjust learning rates only modestly when compared to the reduced-Bayesian model (Fig. 11C).

A second possible explanation for the shallowness of the relationship between learning rate and z-scored error is that subjects maintain inaccurate beliefs about environmental statistics other than hazard rate. For example, subjects might expect the mean of the generative distribution to drift from trial to trial. This possibility can

be modeled by adding drift variance (*D* in Eq. 13) to the variance on the predictive distribution after each timestep. This model can be applied to subject data with drift (*D*), hazard rate (*H*), and expected noise (*N*) all fit as free parameters (Eqs. 13 and 14), producing predictions that have a more shallow relationship between learning rate and z-scored error (Fig. 11B). This model described subject behavior better than either the reduced-Bayesian model with only the hazard rate as a free parameter (for 30 out of 30 subjects) or a delta-rule model with a fixed learning rate (for 28 of 30 subjects). The reduced-likelihood model was similarly effective at describing subject behavior relative to the reduced-Bayesian model with only the hazard rate as a free parameter (for 30 out of 30 subjects) or a delta-rule model with a fixed learning rate (for 29 of 30 subjects; Fig. 11D).

Discussion

The goal of this work was to examine quantitatively the influence of sequential outcomes on the beliefs of human subjects in a dynamic environment with both noise and abrupt, unsignaled change-points. Unlike previous studies (e.g., (Behrens et al., 2007;Krugel et al., 2009;Corrado et al., 2005)), we used a task that allowed for a trial-by-trial measurement of the learning rate (Fig. 1), which reflects the degree to which a new outcome influences an existing belief. This approach allowed us to identify two primary relationships between learning rates and the outcomes that gave rise to them. The first was that the learning rate tended to increase as a

function of the absolute magnitude of the most recent prediction error, scaled by the expectation of noise. The second was that the learning rate, along with uncertainty, also tended to rise immediately then decay slowly following a change-point.

To account for these results, we developed a simplified version of a Bayesian ideal-observer model. The model's learning rates are analytically tractable and depend on only two variables: change-point probability and run length. For a given run length, change-point probability is monotonically related to the magnitude of the absolute error, scaled by the noise of the generative distribution. By relating learning rate to change-point probability, the model simulates the positive relationship between learning rate and absolute error in our behavioral data (compare Figs. 2C and 7B). Thus, the model, like the subjects, resets beliefs when they are no longer applicable to the current environment.

In contrast to change-point probability, run length is inversely related to both learning rate (Fig. 6C) and uncertainty (Eq. 6). When the model recognizes a change-point, run length is reset to one, leading to increased uncertainty and driving any subsequent outcome to carry more influence (Fig. 7E). Run length increases as a function of trials after a change-point, leading to a narrower predictive distribution and smaller learning rates, consistent with our behavioral data (compare Fig. 3B with 7E, and 4C with 7F). Thus, the model, like the subjects, relies more heavily on historical outcomes when more pertinent outcomes have been observed.

Our reduced model shares commonalities with a number of relatively simple models developed previously to describe animal and human learning behavior. Several models of classical conditioning, including Rescorla-Wagner, a straightforward form of delta rule, and Pearce-Hall, which describes changes in associability between stimuli, learn from surprising outcomes (Pearce and Bouton, 2001). However, unlike our approach, these models do not distinguish between noisy and volatile errors. Such a mechanism has been incorporated into a recently proposed extension to the delta rule, in which recent errors are compared to older ones (Krugel et al., 2009). This comparison allows the model to react to change-points with increased learning rates, but not in a manner that scales with noise and without a notion of uncertainty.

Bayesian approaches to belief updating, although often computationally demanding, can provide such a notion of uncertainty by assessing the probabilities of many possible generative scenarios. Such models can effectively describe human behavior on armed-bandit tasks in which the reward structure either drifts (Daw et al., 2006) or changes discontinuously (Behrens et al., 2007). We showed that a reduced version of the optimal belief-updating algorithm, formulated as a delta rule, can effectively model behavior when it includes elements of both the true generative environment (discontinuous change) and a non-existent element (drift). This result suggests that subjects adjust learning according to perceived generative processes

that do not necessarily match the actual generative processes, an idea that likely extends to armed-bandit tasks in which subjects are uncertain about the exact reward structure.

Such differences between actual and perceived generative models might also explain the substantial variability across subjects in the extent to which individuals updated existing beliefs based on new information. Some subjects tended to maintain existing beliefs under nearly all conditions (i.e., had learning rates near zero). In contrast, other subjects tended to adjust their beliefs dramatically in response to each new outcome (i.e., had learning rates near one). This variability was related to subjective certainty, in that subjects who used higher learning rates were also more confident in their predictions and tended to show more negative relationships between uncertainty and learning rate.

The reduced Bayesian model can account for this individual variability by adjusting the prior probability of change-points, or hazard rate. Increasing the hazard rate leads to higher estimates of change-point probability and thus higher learning rates, on average. Under these conditions, a larger proportion of errors are attributed to change-points, rather than noise. This attribution leads to a chronic underestimation of noise and accounts for the otherwise counterintuitive, negative relationship between average uncertainty and learning rate. Thus, the model suggests that individual variability reflects a form of perceptual bias about how errors are interpreted.

Such a perceptual bias might be useful if it reflects the true probability of change-points in the current environment, particularly if new information is scarce. However, we found that most subjects behaved as if they substantially overestimated the true hazard rate (Fig. 8). Thus, individuals appear to have preconceived strategies for coping with probabilistic environments. Given the computational complexity of existing models for online inference of hazard rate (Wilson et al., 2010), it seems plausible for such higher-order policies to develop over a longer time, either through experience on the developmental timescale or perhaps even evolutionary selection. However, this still leaves open the question of why such diverse policies exist across our subject pool.

The answer to this question might involve a fundamental trade-off inherent in selecting a hazard rate. Using a high hazard rate implies high sensitivity to change-points, but over-sensitivity to noisy outcomes during periods of stability. In contrast, lower hazard rates provide less sensitivity to noisy outcomes but also less sensitivity to change-points. Sensitivity to either change-points or noise might have different consequences under different conditions or for different individuals, giving rise to the diversity of predispositions about hazard rate that we observed. One potential genetic substrate of this predisposition is a polymorphism in monamine catabolism enzyme COMT that leads to lower learning rates in reversal tasks but improved performance in working-memory tasks (Krugel et al., 2009;Bruder et al., 2005). Our task is, to our knowledge, the first to demonstrate both the advantages and disadvantages of hazard-rate policy and thus may serve as a valuable tool for determining whether COMT or other polymorphisms play a role in navigating this trade-off.

A strong motivation for the form of reduced Bayesian model that we used was its relationship to delta-rule models of learning, whose biological underpinnings have been studied extensively (Niv, 2009). Among the strongest biological evidence is the discovery of signals in the brainstem dopaminergic system that encode a form of reward-prediction error (δ in Eq. 1; (Schultz, 1998)). More recent work has begun to link these prediction-error signals to activity in anterior cingulate cortex (ACC), a brain area thought to encode information related to subjective beliefs used for decision-making. ACC neurons encode subjective beliefs about outcome probability and value and action cost (Kennerley et al., 2009). Single neurons in monkey ACC also encode prediction errors, a finding that is corroborated by human fMRI and EEG data (Matsumoto et al., 2007;Hayden et al., 2009;Debener et al., 2005). Ablation of ACC in macaques leads to impaired use of outcome history in the guidance of action selection, further suggesting a role in belief updating (Kennerley et al., 2006).

Despite these advances in understanding neural substrates for delta-rule learning in terms of prediction errors (δ in Eq. 1), less is known about the learning rate (α in Eq. 1). The learning rate regulates the relative contributions of stored information about previous outcome history and the new sensory information about the current outcome. One possible implementation involves interactions between top-down cognitive control and bottom-up sensory processing and thus might be related to similar mechanisms of attention (Dayan et al., 2000;Posner, 2008). However,

nothing is known about how those mechanisms relate to the learning rate we examined in this study.

Our model provides several insights that might help identify some of the underlying mechansims. The first is that learning rate depends critically on the estimated change-point probability. Change-point probability is related to absolute prediction-error magnitude, scaled by expected uncertainty. Absolute prediction-error signals are encoded by neurons in monkey ACC, the same area thought to encode decision-relevant beliefs and prediction errors related to those beliefs (Matsumoto et al., 2007). Thus, the ACC might also contain at least one of the necessary variables to compute learning rate. Consistent with this idea, fMRI measurements of the ACC in human subjects engaged in a dynamic probabilistic task correlated with a model parameter (volatility) that reflected an optimal assessment of the rate at which reward contingencies were likely to be changing and learning rates fit to subjects (Behrens et al., 2007;Krugel et al., 2009). This signal might also include subjective hazard-rate biases, because subjects who were best fit by high learning-rate models tended to show larger ACC BOLD responses to new outcomes than subjects fit by low learning-rate models.

Another prediction of the model is that learning rates are computed according to run length. It is unknown whether the ACC encodes run length, however it would provide a parsimonious solution to the compartmentalization of belief updating

machinery within the brain. Theoretical work has also suggested that an uncertainty signal inversely related to run length might be encoded by a more global neuromodulatory system, such as the locus coeruleus-norepinephrine system (Yu and Dayan, 2005). Our task and model provide a framework for testing this possibility.

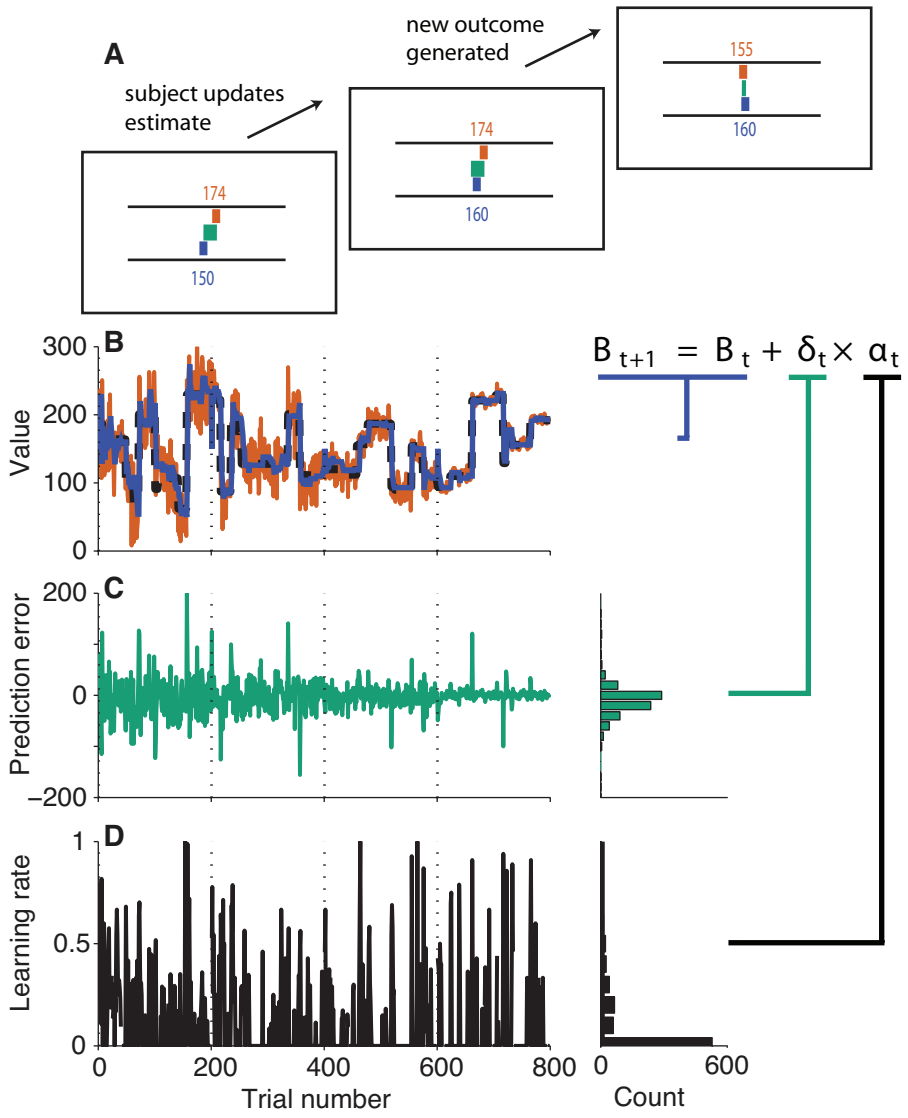


Figure 1. Estimation task and its relationship to prediction errors and learning rate. **A**, Schematized trial of the estimation task. The subject makes a prediction (blue) and is then shown the outcome (red) and the error made in predicting the outcome (teal). After the subject updates his prediction as a fraction of the error, a new outcome is generated. **B**, An example session. Numbers (red line) are generated from a normal distribution with a variance that is constant within blocks of 200 trials (vertical, dotted lines) and a mean (dashed black line) that changes at random times. The subject's trial-by-trial predictions are shown in blue. **C**, Trial-by-trial prediction errors from the session in **B** (actual in red minus prediction in blue). Histogram (right) shows the distribution of prediction errors made over the course of the entire session. **D**, Trial-by-trial learning rates from the session in **B**, computed as the fraction of the prediction error used to update the next prediction using a

delta rule, as shown. Histogram (right) shows the distribution of learning rates across the entire session.

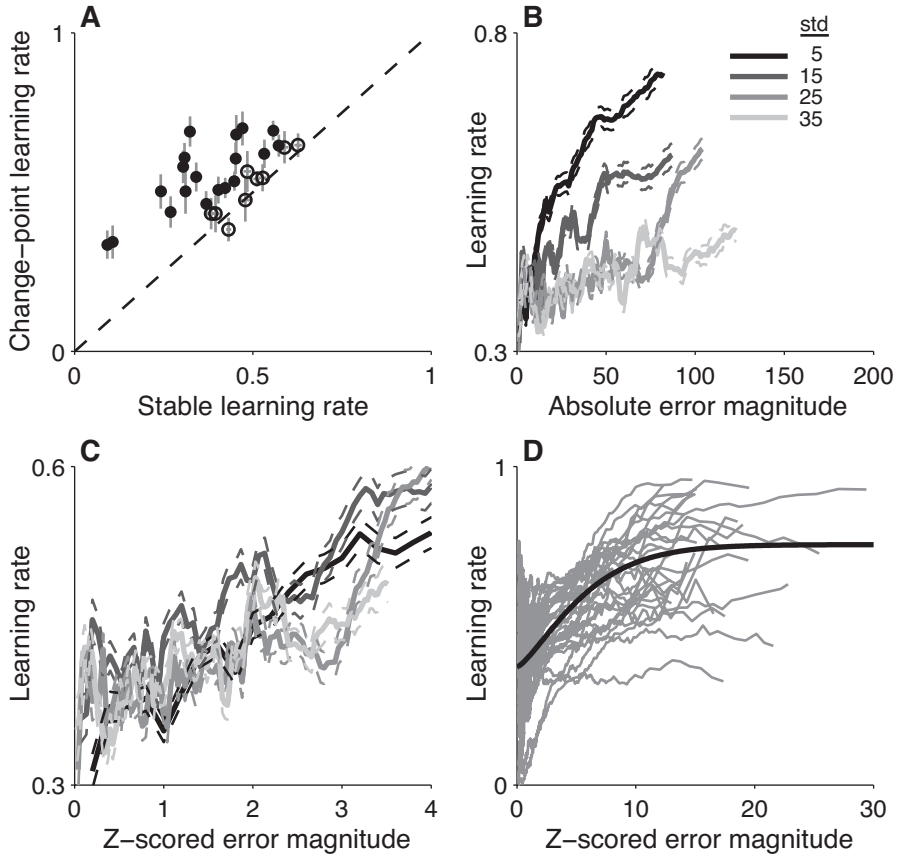


Figure 2. Learning rates increased after unexpected errors. **A**, Mean±SEM learning rates on trials in which the mean of the generative distribution changed (ordinate) versus on other trials (abscissa; error bars are obscured by the points). Points are data from individual subjects. Filled symbols indicate Wilcoxon test for H_0: equal

median learning rates on change-point and non-change-point trials, $p<0.05$. **B**, Learning rate plotted as a function of median absolute error magnitude, averaged using running bins of 150 trials, for four different standard deviations of the generative distribution, as indicated. Data averaged across all subjects. Solid and dashed lines indicate mean and SEM, respectively. **C**, Learning rate plotted as a function of median relative error magnitude, plotted as in **B**. Relative error magnitude was computed by dividing the absolute error magnitude by the standard deviation of the generative distribution. **D**, Individual subject learning rates plotted as a function of relative absolute error magnitude (gray lines). Black line indicates a cumulative Weibull function fit to data from all subjects.

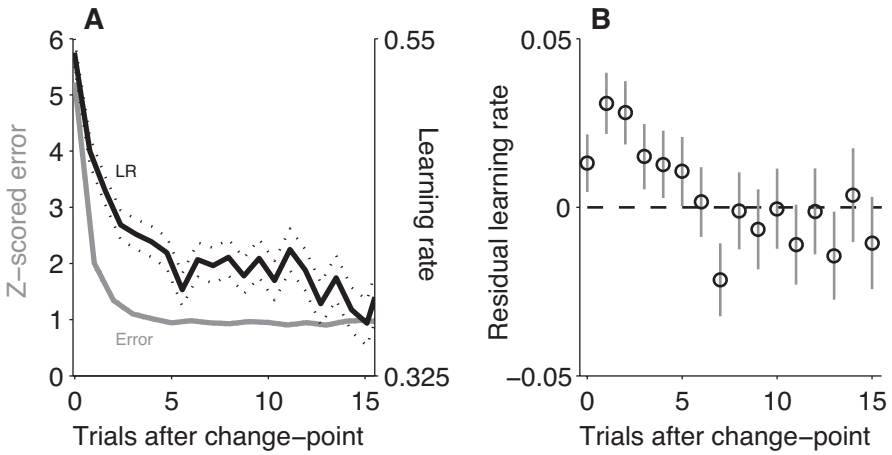


Figure 3. Learning rates decayed slowly after change-points. **A**, Prediction errors (gray, left ordinate) and learning rates (black, right ordinate) plotted as a function of trials after a change-point. Solid lines indicate mean across all subjects and all conditions, dotted lines indicate SEM. **B**, Learning rate residuals plotted as a function of trials after a change-point. Residuals were computed by subtracting the learning rates predicted by the cumulative Weibull fit shown in Fig. 2D from the actual learning rates, and thus reflect the portion of learning rate that was not explained by relative error magnitude. Points and errorbars are mean±SEM across all subjects.

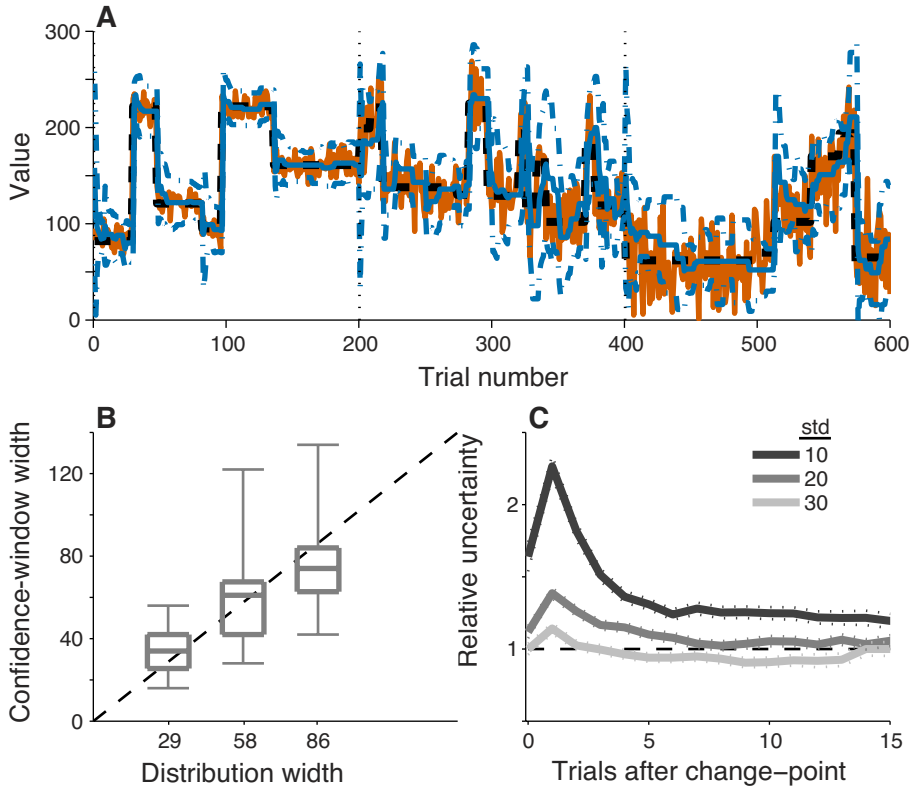


Figure 4. Subjective confidence measurements. **A**, An example session of the confidence task. Subjects specified a symmetric window (dashed blue lines) around their estimate (solid blue line) that they were 85% certain would contain the next number (red) generated using the current mean (dashed black line) and standard deviation (stable in blocks, indicated by the vertical, dotted lines). **B**, Box-and-whisker plot (central line is median, box is interquartile range, and whiskers are the data range) of the distribution of the mean width of the 85% confidence window computed per subject for each standard-deviation condition. **C**, Relative uncertainty as a function of trials after a change-point. Relative uncertainty was computed by dividing the specified confidence window size by the size of the smallest window capable of including 85% of the probability density in the actual generative distribution (x-axis markers in **B**). Solid and dotted lines indicate mean and SEM, respectively.

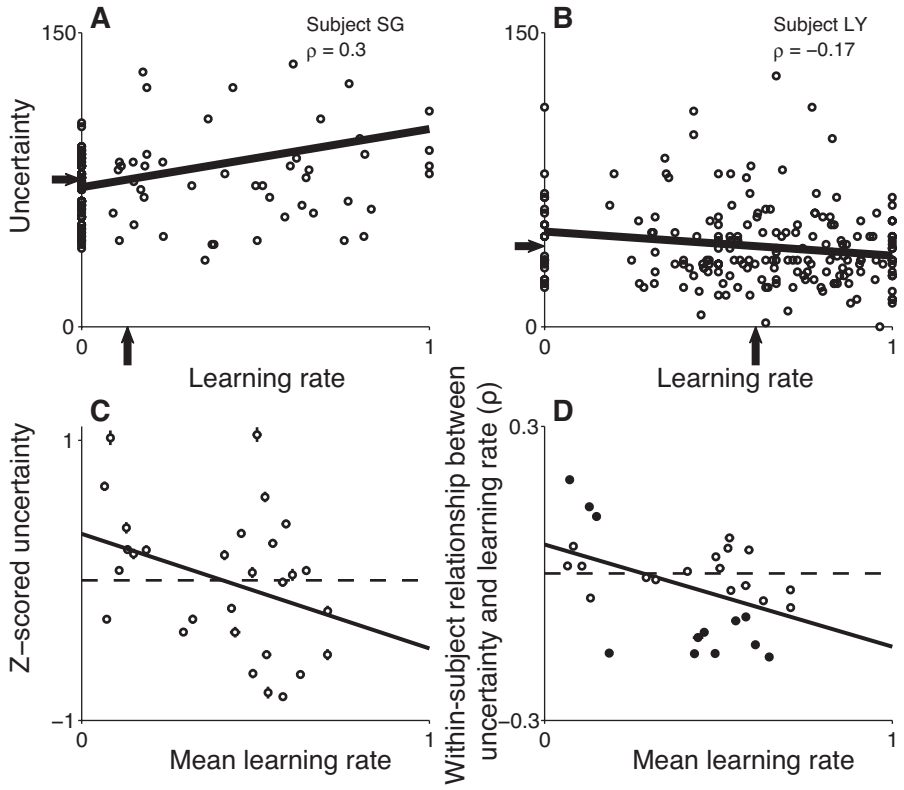


Figure 5. Relationship between confidence and learning rate. **A**, **B**, Trial-by-trial learning rates plotted as a function of uncertainty (confidence-window width) for an example task block (std=20) for two different subjects. Solid lines are linear fits. Arrows indicate the mean values of the confidence-window width and learning rate. **C**, Mean relative uncertainty (computed as the z-scored confidence-window width across all conditions per subject) plotted as a function of mean learning rate. Symbols and error bars are mean±SEM per subject. Solid line is a linear fit (r=-0.38, H_0: r=0, p=0.04). The negative correlation implies that subjects who used higher learning rates tended to be more certain about their predictions. **D**, Trial-by-trial relationship between relative uncertainty and learning rate per subject (ordinate, computed as Spearman's ρ as in **A** and **B**, with filled symbols indicating H_0: ρ=0, p<0.05; a positive/negative value indicates that the subject tended to use

higher/lower learning rates on trials in which they were more uncertain about their previous prediction) plotted as a function of the average learning rate used by that subject. Symbols and errorbars are mean±SEM per subject. Solid line is a linear fit (r=-0.44, H_0: r=0, p=0.02). The negative correlation implies that subjects who used lower learning rates tended, on average, to have more positive trial-by-trial relationships between uncertainty and learning rate.

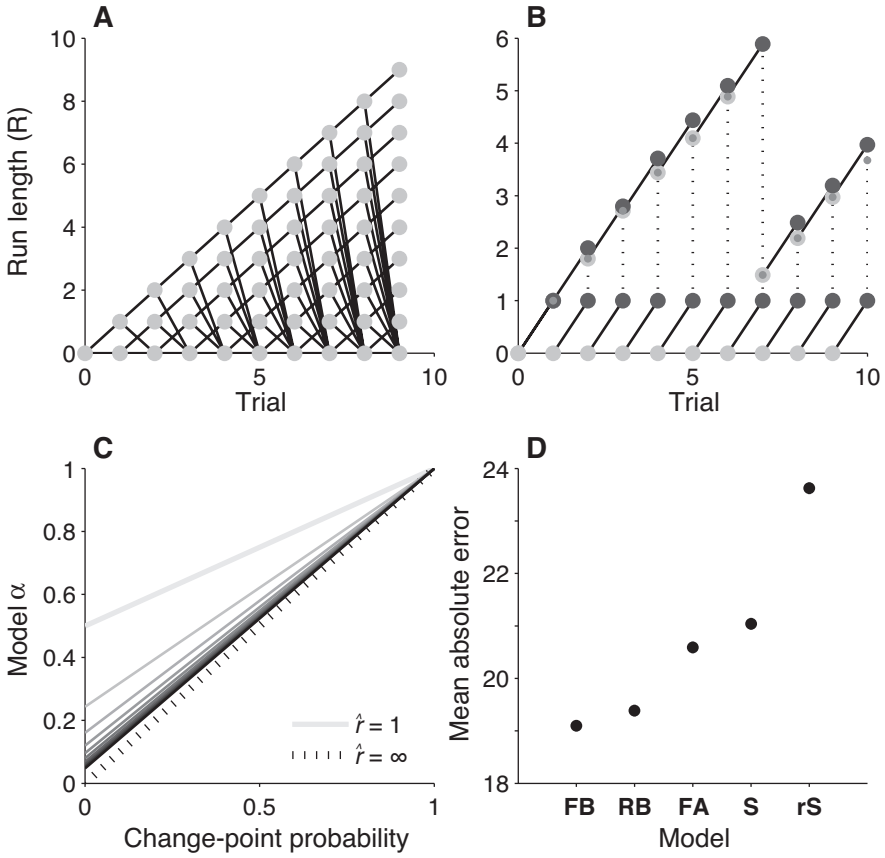


Figure 6. Bayesian model. **A**, Message-passing algorithm for the full model. Run length (*r*) refers to the number of data points obtained previously from the current generative distribution. On each trial, the distribution either changes, and *r* is set to zero, or the generative distribution does not change, and *r* is increased by one. After *t* trials, the algorithm must maintain and update *t*+1 predictive distributions (one for each possible *r*) and the probability distribution across these possible values of *r*. **B**, Message-passing algorithm for the reduced model. Instead of considering all possible values of *r*, the model considers only the possibility that a change-point did occur (represented by solid lines from *r*=0 to *r*=1) or did not occur (represented by all other solid lines). Posterior probabilities of these alternatives are computed

according to Bayes' rule, then combined by taking the expected value of the run-length distribution, $\hat{r}$ (small gray filled circles). Only a single, approximate predictive distribution is maintained and updated on a trial-by trial basis. This approach massively reduces complexity and leads the algorithm to take the form of a delta rule (see Methods). **C**, Learning rates used by the reduced Bayesian model can be described analytically in terms of $\hat{r}$ and change-point probability. Lines indicate relationships between learning rate and change-point probability for a given $\hat{r}$ (increasing for darker lines). The dotted black line reflects the theoretical limit of the function as $\hat{r}$ goes to infinity. **D**, Performance of subjects and models. Mean absolute errors made by the full Bayesian model (FB), the reduced Bayesian model (RB), a delta-rule model using the best fixed learning rate possible for each session (FA), subjects (S), and a delta-rule model using subject learning rates in random order (rS).

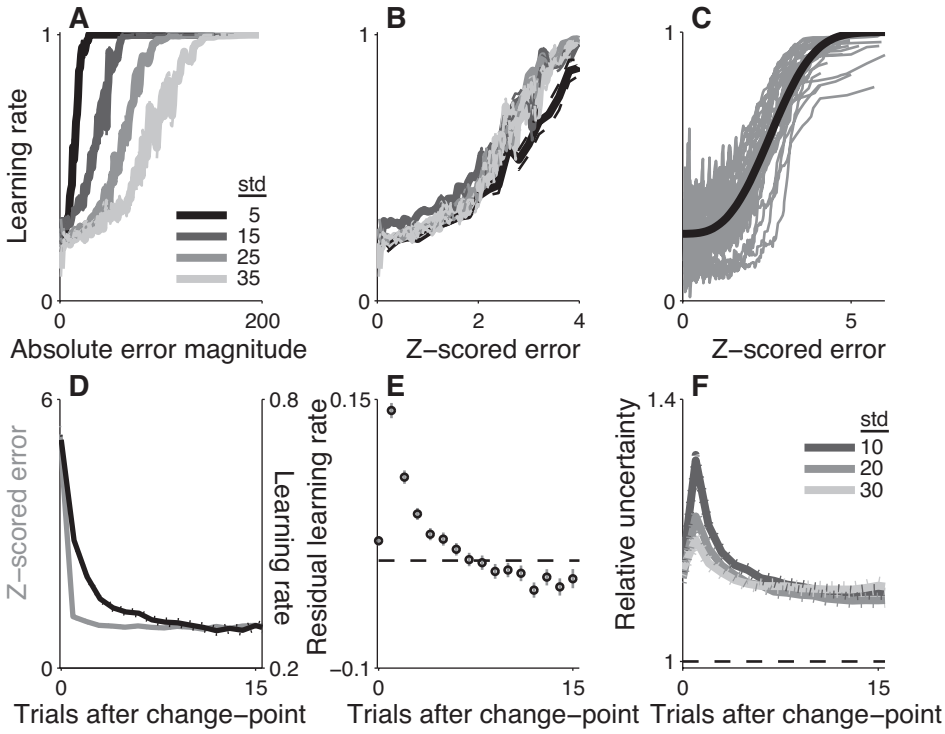


Figure 7. The reduced Bayesian model qualitatively reproduces belief-updating behavior. All plots in this figure depict simulated data using the reduced Bayesian model. One model parameter, the hazard rate, was fit for each block to minimize the difference between model and subject predictions. **A**, Learning rate as a function of absolute error magnitude for different standard deviations of the generative distributions, as shown. Compare to Fig. 2B. **B**, Learning rate as a function of z-scored error, plotted as in **A**. Compare to Fig. 2C. **C**, Across-subject variability in the relationship between learning rate and z-scored error, simulated by fitting data from different subjects with different hazard rates (gray lines). Black line is cumulative Weibull fit. Compare to Fig. 2D. **D**, Z-scored error (gray, left ordinate) and learning rate (black, right ordinate) plotted as a function of trials after a change-point. Solid and dashed lines are mean±SEM. Compare to Fig. 3A. **E**, Learning rate residuals plotted as a function of trials after a change-point. Residuals were computed by subtracting the learning rates predicted by the cumulative Weibull fit shown in **C** from the actual learning rates, and thus reflect the portion of learning

rate that was not explained by relative error magnitude. Points and errorbars are mean±SEM across all simulated data. Compare to Fig. 3B. **F**, Relative model uncertainty (computed as the minimal window containing at least 85% of the probability density in the predictive distribution specified by the model divided by the 85% width of the true generative distribution) plotted as a function of trials after a change-point. Grayscale reflects the standard deviation of the given task block, as indicated. Compare to Fig. 4C.

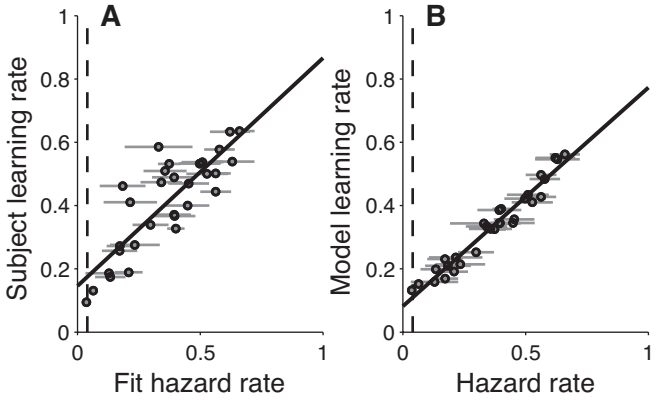


Figure 8. Relationship betwen learning rate and hazard rate. **A**, Variability in subject learning rates can be described by the hazard rate in the model. Subjects that are fit best by high hazard rate versions of the reduced Bayesian model use higher learning rates, on average. The dashed line indicates the actual average hazard rate for the task. Points and errorbars represent the mean and standard error or the mean, respectively. The solid line is a linear fit (r=0.84, p<0.001). **B**, Higher hazard rate models tend to use higher learning rates. Points and errorbars represent the mean and standard error of the mean for all fits to a given subject (across all task blocks). The solid line is a linear fit (r=0.98, p<0.001).

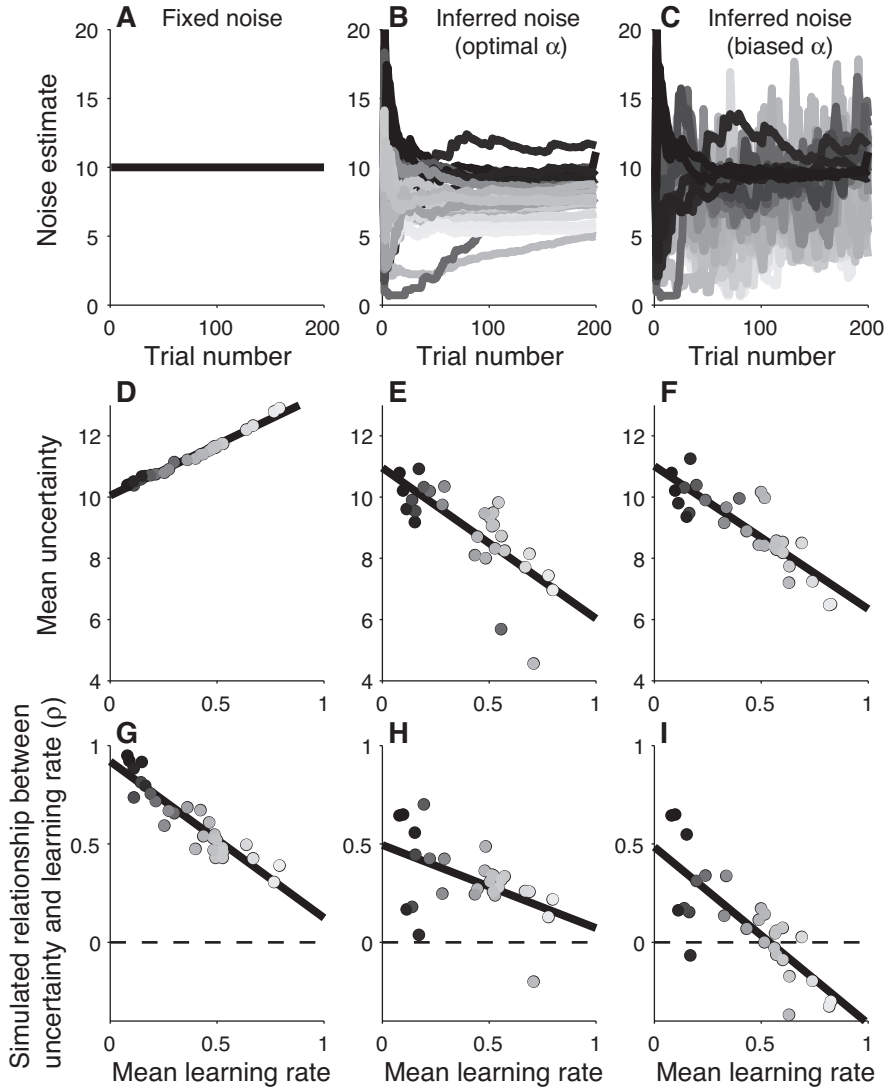


Figure 9. On-line noise inference. Individual variability was simulated by using models that employed the hazard rates fit to individual subject data (see text; in all panels, grayscale represents the different hazard rates, with lighter shades for higher rates). Three models that differed only in their method for computing noise were used to

simulate performance. The first, simplest model (left column) used the actual standard deviation of the generative distribution. The second model (middle) inferred noise using an on-line algorithm with learning rates that assumed noise was constant over each block of 200 trials (Eqs. 9, 10). The third model (right) inferred noise using the same algorithm as the second model, but with a minimum learning rate that depended on hazard rate (Eqs. 9, 11). **A, B, C**, Noise estimates from each model over the course of each 200-trial block in which the standard deviation of the generative distribution was equal to 10. **D,E,F**, The mean uncertainty estimate for each simulated block of trials plotted as a function of the mean learning rates used in that simulation. Lines are linear fits. Negative relationships in **E** and **F** reflect the fact that individuals modeled with higher hazard rates tended to use higher learning rates and thus infer less noise. **G,H,I**, Correlations between uncertainty and learning rate within single simulated task blocks plotted as a function of the mean learning rate simulated for that subject. Lines are linear fits. All models show a negative relationship, but only the third model matches the behavioral data, with low mean learning rates typically corresponding to positive relationships between learning rate and uncertainty and high mean learning rates typically corresponding to negative relationships between learning rate and uncertainty.

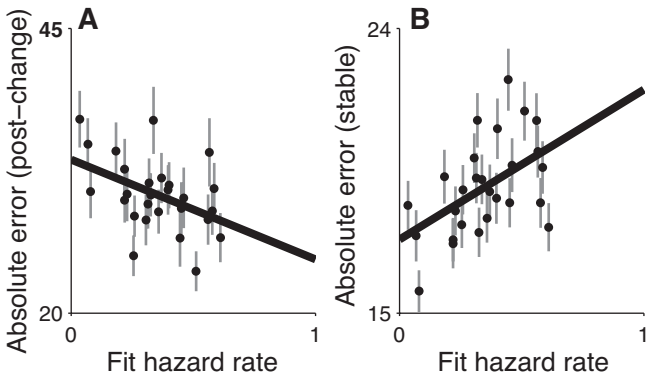


Figure 10. Hazard rate trade-off. **A**, Average absolute errors made by subjects 1–5 trials after a change-point plotted as a function of the fit hazard rate from the reduced Bayesian model for each subject (points). Line is a linear regression (r=-0.43, p = 0.02). The negative relationship implies that subjects who used higher hazard rates made better predictions after change-points. **B**, Average absolute errors made by subjects 6+ trials after a change-point plotted as a function of the fit hazard rate for each subject (points). Line is a linear regression (r=0.51, p < 0.01). The positive relationship implies that subjects who used lower hazard rates made better predictions during periods of stability.

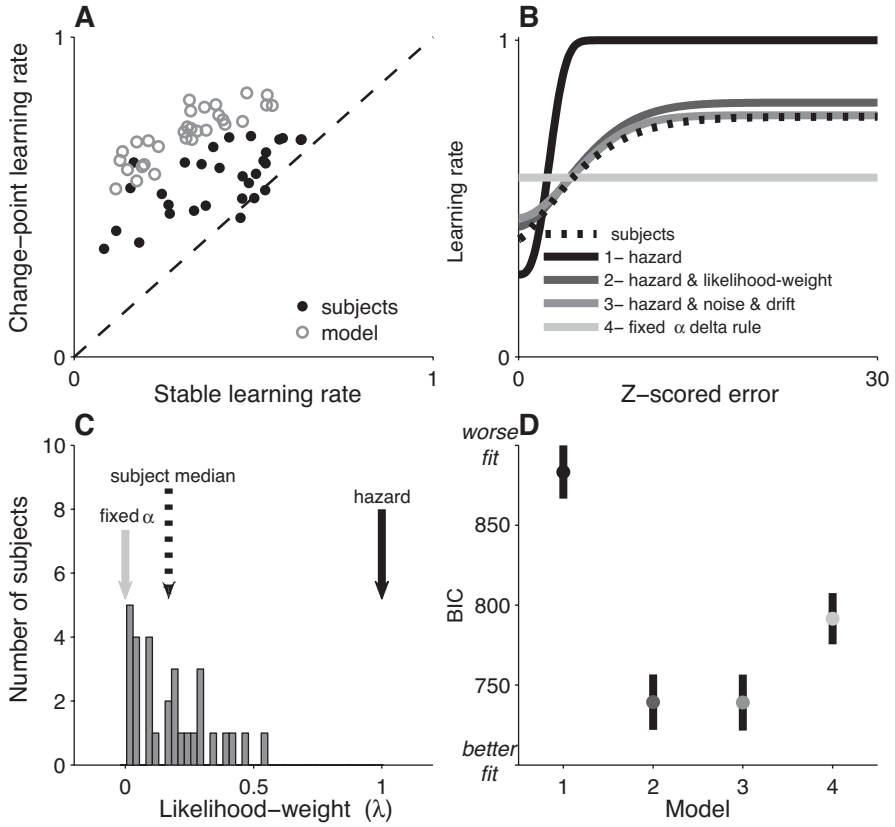


Figure 11. Better descriptive models to capture sub-optimal performance. **A**, Although subjects (filled symbols; data plotted as in Fig. 2A) and the reduced Bayesian model (open symbols) both used higher learning rates after change-points than during a stable period, the model tends to show a larger effect. **B**, Relationship between learning rate and relative error magnitude for subjects (dotted line, the fit from Fig. 2**D**) and several models fit to subject behavior, as indicated. **C**, Histogram of the average likelihood weight fit to each subject (λ in Eq. 12). When λ =0, the model updates beliefs according to a fixed learning rate delta rule. When λ=1, the model is the reduced Bayesian model. All subjects fell between these two extremes. **D**, Bayesian information criterion (BIC) for all models in **B** fit to subject data. Lower values imply better fits, including penalties for additional parameters. Points and errorbars are mean±SEM across subjects. Grayscale and model numbers are as in panel **B**.

CHAPTER 3

Rational regulation of learning dynamics by pupil–linked arousal systems

Matthew R. Nassar, Katherine M. Rumsey, Robert C. Wilson, Kinjan Parikh, Benjamin Heasly and Joshua I. Gold. *Nature Neuroscience*, 2012, 15:1040-6

Abstract

The ability to make inferences about the current state of a dynamic process requires ongoing assessments of the stability and reliability of data generated by that process. We found that these assessments, as defined by a normative model, were reflected in non–luminance–mediated changes in pupil diameter of human subjects performing a predictive–inference task. Brief changes in pupil diameter reflected assessed instabilities in a process that generated noisy data. Baseline pupil diameter reflected the reliability with which recent data indicated the current state of the data–generating process and individual differences in expectations about the rate of instabilities. Together these pupil metrics predicted the influence of new data on subsequent inferences. Moreover, a task– and luminance–independent manipulation of pupil diameter predictably altered the influence of new data. Thus, pupil–linked arousal systems can help regulate the influence of incoming data on existing beliefs in a dynamic environment.

Introduction

Many decisions, from foraging to financial, depend on the ability to infer a state of the world from both historical and newly arriving information. Such inferences are particularly challenging when they must account for multiple sources of uncertainty. When the uncertainty results from noise, reflecting random fluctuations in the information generated by an otherwise stable state, the average over all historical information is most predictive of future observations. In contrast, when the uncertainty results from a change in the state itself, only the most recent information pertains to the new state. Thus, historical information should be discounted and beliefs should be updated rapidly to maximize their predictive power. Under certain conditions, human subjects appear to encode and respond appropriately to these different forms of uncertainty when making inferences in a dynamic environment (Behrens et al., 2007;Nassar et al., 2010;Yu and Dayan, 2005). Here we examined whether this ability is governed, at least in part, by arousal systems that affect pupil diameter, which are thought to include the noradrenergic brainstem nucleus locus coeruleus (Nieuwenhuis et al., 2010;Aston-Jones and Cohen, 2005;Jepma and Nieuwenhuis, 2010;Gilzenrat et al., 2010).

Non–luminance–mediated changes in pupil diameter have long been used as

indicators of clinical, cognitive, and arousal states (Krugman, 1964;Granholm and Steinhauer, 2004;Schmidt and Fortin, 1982;Kahneman and Beatty, 1966). One interpretation of these pupil changes is that they reflect the amount of cognitive effort exerted at a given time, which can be related to task uncertainty (Kahneman and Beatty, 1966). Accordingly, changes in pupil diameter can be elicited via manipulations of the uncertainty associated with possible actions in certain choice tasks (Jepma and Nieuwenhuis, 2010;Richer and Beatty, 1987) . Changes in pupil diameter can also reflect perceived changes in the world, including perceptual switches during perceptual rivalry, detection of targets in oddball or near–threshold tasks, responses to low–probability go signals in a go/no–go task, and perceived changes in task utility that can affect task engagement (Gilzenrat et al., 2010;Richer and Beatty, 1987;Hakerem et al., 1964;Einhäuser et al., 2008;van Olst et al., 1979) .

These kinds of uncertainty– and change–related signals are thought to contribute to rational inference in a dynamic environment, including helping to regulate the relative influence of historical and newly arriving information on existing beliefs (Nassar et al., 2010;Yu and Dayan, 2005). Such regulation is a key feature of cognitive flexibility and can be equivalent to adjusting the learning rate in a reinforcement–learning framework (Behrens et al., 2007;Sutton and Barto, 1998). Our goal was to determine how such learning–rate adjustments relate to pupil–linked arousal systems. We show that the arousal system and possibly the

locus coeruleus can play important and computationally complex roles in rationally regulating the influence of incoming information on beliefs about a dynamic world.

Results

We measured pupil diameter in thirty human subjects while they performed an isoluminant version of a predictive–inference task [(Nassar et al., 2010)]. Below we describe task performance, summarize a nearly optimal model that captures key features of performance, demonstrate that certain aspects of pupil diameter encode key variables in the model that can be used to predict performance, and finally show that a task–independent manipulation of arousal and pupil diameter can lead to predictable changes in task performance.

Behavior

The predictive–inference task required subjects to minimize errors in predicting the next number (outcome) in a series. The outcomes were picked from a Gaussian distribution with a mean that changed at random intervals (change points) and a standard deviation (set to either 5 or 10) that was stable over each block of 200 trials (Fig 1). After each prediction was recorded, the new outcome was shown using an iso–luminant display for 2 s, during which time the subject maintained fixation and pupil diameter was measured (Fig. 1). After this interval,

the outcome disappeared and the previous prediction reappeared, to be updated for the subsequent trial. Payment scaled inversely with the subject's mean absolute error during the session (Nassar et al., 2010).

We quantified the extent to which each new outcome influenced the subsequent prediction as the learning rate in a simple delta–rule model (Eq. 3) (Nassar et al., 2010). The learning rate was equal to the magnitude of change in the prediction expressed as a fraction of the error made on the previous prediction. Thus, a learning rate of one indicated abandonment of the previous prediction in favor of the most recent outcome. A learning rate of zero indicated maintenance of the previous prediction despite a non–zero prediction error.

Subjects tended to use variable learning rates that spanned the entire allowed range, from zero to one. Within this range, learning rates tended to be higher for larger errors, scaled by the noise of the generative distribution (Fig. 2A). Learning rates also tended to be highest on the trial after a change point and then decay for several trials thereafter (Fig. 2B). These basic trends were similar across subjects, although individual subjects used dramatically different distributions of learning rates (Fig. 2C).

Reduced Bayesian model

The learning rates used by subjects were consistent with both a full and a

simplified version of the optimal (Bayesian) model (Nassar et al., 2010;Adams and MacKay, 2007;Fearnhead and Liu, 2007;Wilson et al., 2010). One advantage of the reduced Bayesian model is that it updates beliefs according to a delta rule in which the learning rate is computed according to only two parameters computed per trial: change–point probability and relative uncertainty (Fig. 3A).

Change–point probability approximates the posterior probability that the mean of the generative distribution changed since the previous trial, given all previous data. If the mean did change, then previous outcomes should be unrelated to future ones and not contribute to an updated prediction. Accordingly, the model uses learning rates that scale linearly towards one (thus discarding historical information) as change–point probability approaches one (Fig. 3A). Change–point probability is computed by comparing the probability of each new outcome given either the current predictive distribution or the occurrence of a change point (Eq. 5). Its value increases monotonically as a function of the absolute difference between predicted and actual outcome, scaled according to the standard deviation of the generative distribution (Eq. 6, Fig. 3B).

Relative uncertainty is a function of total uncertainty, which in our task arises from two sources. The first source, noise, reflects the unreliability with which a single sample can be predicted from a distribution with a known mean. The

second source reflects the unreliability of the current estimate of the mean, which decreases as more data are observed from a distribution. Relative uncertainty is the magnitude of this second form of uncertainty as a fraction of total uncertainty, analogous to the gain in a Kalman filter. Relative uncertainty determines the learning rate when change–point probability is zero and sets the *y*–intercept of the relationship between change–point probability and learning rate otherwise (Fig. 3A). The effects of relative uncertainty on model learning rates are greatest on the trials following a change point, when its value peaks at 0.5 and then decays over several trials (Eq. 7; Fig 3C).

Like the human subjects, the model tended to compute learning rates that were highest just following a change point in the mean of the generative distribution and then decayed for several trials independently of noise. When applied to the exact same outcome sequences as the subjects, the model also tended to produce similar learning rates (Fig. 3D).

We related change–point probability and relative uncertainty computed in the model to the mean pupil diameter ("pupil average") and change in pupil diameter ("pupil change") measured during the 2–s outcome–viewing period (Fig. 1 inset), using two linear regression models. The first, simpler model had four parameters: change–point probability and relative uncertainty computed from the reduced Bayesian model, the standard deviation of generative distribution, and a binary

variable describing whether or not the prediction error was exactly zero. The second model included all of these parameters, as well as several potential confounding factors such as eye position and velocity (see Methods). The models are complementary: the first avoids potential interactions between large numbers of parameters and thus has coefficients that are more readily interpretable, whereas the second avoids missing out on the many factors that in principle could affect our pupil measurements. Both models captured a significant amount of variability in the pupil data (For pupil average/pupil change data, an *F*–test rejected the null model relative to the small model for 27/15 of the 30 subjects, and a nested *F*–test rejected the small model relative to the large model for 29/19 of the 30 subjects, $p<0.05$).

Below we first report the most prominent effects from these regression analyses, which were similar for the two models and include roughly monotonic relationships between pupil change and change–point probability and between pupil average and relative uncertainty. We later show that these relationships were in fact slightly more complicated and included a dependence on baseline pupil diameter that helps us to interpret the results in terms of known properties of the arousal system.

Pupil change reflected change–point probability

The change in pupil diameter during the outcome–viewing period, like change–

point probability in our model, tended to increase as a function of error magnitude, scaled as a function of noise (Fig. 4A; compare to Figs. 3B). Accordingly, when computed by the model using the same sequence of outcomes experienced by each subject, change–point probability tended to be positively predictive of z–scored pupil change (Fig. 4B ordinate). The complement was also true: change–point probability varied systematically as a function of pupil change for data pooled across the population (Fig. 4C). In contrast, there was no consistent relationship between change–point probability and pupil average (Fig. 4B abscissa).

One notable exception to the positive relationship between pupil change and error magnitude occurred for trials in which the error was exactly zero, which corresponded to relatively large pupil changes (left–most data in Fig. 4A). Accordingly, a binary variable added to the linear model that described whether or not the subject correctly predicted the outcome was related to pupil change (the mean value of the regression coefficient was 0.180 z_{PC} for the four–parameter regression model and 0.156 z_{PC} for the larger model; $p<0.05$ for H_0: mean=0 for each model) but not pupil average (mean regression coefficient=–0.076 and –0.092 z_{PA} for the smaller and larger regression models, respectively, $p>0.05$). Thus, pupil change reflected not only change–point probability, but also whether or not the subject correctly predicted the observed outcome.

Average pupil diameter reflected belief uncertainty

The average pupil diameter during the outcome–viewing period, like relative uncertainty in our model, tended to peak on the trial after a change point and then diminish in magnitude as more relevant information reinforced the existing belief (Fig 5A; compare to Figs. 2B and 3C). Accordingly, when computed by the model using the same sequence of outcomes experienced by each subject, relative uncertainty tended to be positively predictive of pupil average (Fig. 5B abscissa). This result did not simply reflect differences in motor output following change points (e.g., longer button presses to choose a learning rate near one), because similar results were obtained in a control experiment in which subject predictions were reset using a learning rate of 0.5 on each trial, thus requiring the same motor act to choose a learning rate of either zero or one (mean regression coefficient=0.30 and 0.35 $z_{PA/RU}$ for the smaller and larger regression models, respectively, $p<0.05$). The complement was also true: relative uncertainty varied systematically as a function of pupil average for data pooled across the population (Fig. 5C). In contrast, there was no consistent relationship between relative uncertainty and pupil change (Fig. 5B ordinate).

Overall uncertainty in our task depends on not only relative uncertainty but also noise, which we manipulated by varying the standard deviation of the generative distribution in blocks (STD=5 or 10). Consistent with our model, in which noise is only used to compute change–point probability (Eqs. 5 and 6), these

manipulations of noise were reflected in pupil change but only insofar as pupil change represented change–point probability (Fig. 4A). These manipulations of noise did not have any other systematic effects on either pupil change or pupil average (p>0.1 for H_0: a mean value of zero for the regression coefficient describing the influence of noise on the given pupil measurement for both regression models). Thus, for this task pupil average did not appear to reflect overall uncertainty about a future outcome but rather a specific form of uncertainty that arises after change points and signals the need for rapid learning.

Pupil metrics reflected individual learning differences

As noted above (Fig. 2C), there was a great deal of variability in the average learning rates used by individual subjects. These individual differences are thought to reflect biases that govern the extent to which subjects tend to interpret the cause of prediction errors in terms of either noise or change points (Nassar et al., 2010). One advantage of our reduced model is that it can simulate these individual differences in terms of the subjective hazard rate, which is the expected rate at which change points will occur. Accordingly, fitting the model to behavioral data from individual subjects with subjective hazard rate as a single free parameter yielded fit values that varied systematically with average learning rates (r=0.93, H_0: r=0, p<0.001; Fig 6A).

These individual differences in the inferred (fit) subjective hazard rates corresponded to individual differences in both the temporal dynamics and magnitude of outcome–locked pupil responses. We quantified the temporal dynamics using an index that related the pupil response on a given trial to a mean–subtracted version of the template shown in Fig. 6B. This template describes the strength of the across–subject, linear relationship between pupil diameter and hazard rate in a sliding time window. This relationship was strongest soon after outcome onset, thus likely reflecting prior expectations about the newly arriving outcome. There was a positive relationship between the mean value of this index and fit hazard rate for individual subjects (r=0.51, p<0.01). In addition, there was a positive relationship between pupil average and fit hazard rate for individual subjects (r=0.40, p<0.05).

Based on these relationships, we constructed a linear regression model using the temporal–dynamics index and pupil average to explain individual differences in task performance. The model yielded strong, pupil–based predictions of per–subject values of both fit hazard rate (r=0.59, p<0.001) and average learning rate (r=0.59, p<0.001; Fig 6C). Thus, individual differences in average learning rate, which can be described computationally as differing expectations about the rate of change–points, could be predicted from the temporal dynamics and average magnitude of pupil diameter measured during outcome viewing.

Pupil metrics predicted trial–by–trial learning rates

The relationships between pupil metrics and parameters of the reduced Bayesian model suggest that measurements of pupil diameter during the outcome–viewing period can be used to predict the subsequent learning rate. For example, we found positive relationships between pupil change and change–point probability (Fig. 4) and between pupil average and relative uncertainty (Fig. 5). Thus, observing relatively high values of either pupil metric on a given trial should indicate that the subject will use a larger–than–average learning rate when adjusting beliefs according to the outcome observed on that trial. We tested this idea directly, as follows.

First, we examined the relationship between pupil change, pupil average, and learning rate for individual subjects. We used a regression model to describe learning rate (z–scored per subject) in terms of pupil change and pupil average. On average, this linear regression computed per subject yielded a positive coefficient for pupil change (mean=0.108 z_{LR}/z_{PC}, $p<0.05$ for H_0: mean=0) and a smaller, not statistically significant, positive coefficient for pupil average (mean=0.085 z_{LR}/z_{PA}, $p=0.13$; Fig. 7A).

Second, we used a simple, weighted sum of pupil change and pupil average to assess their combined predictive power across subjects. Using weights equal to the mean value of the per–subject regression coefficients from the previous

analysis (Fig. 7A), the weighted sum was moderately predictive of learning rate across all subjects (r=0.067, p<0.001). However, this analysis did not take into account a systematic, negative dependence of the sum of these per–subject coefficients (which is related to the overall ability of the weighted sum to account for learning rate) on subjective hazard rate predicted by pupil dynamics (Fig. 7B). Subjects with low pupil–predicted hazard rates had pupil responses that were good predictors of learning rate. Subjects with increasingly high pupil–predicted hazard rates had pupil responses that were increasingly less predictive, and in some cases negatively predictive, of learning rate.

Third, we used a more complicated linear model that also included across–subject differences in pupil dynamics that related to subjective hazard rates, which markedly improved our overall ability to use pupil metrics to predict learning rates. This model had three terms: 1) the sum of pupil change and pupil average computed per trial, weighted according to average regression coefficients in Fig. 7A; 2) the pupil–predicted hazard rate, computed per subject (see Fig 6C); and 3) the multiplicative interaction between these two variables. Using this model, pupil measurements could effectively predict learning rates for all data from all subjects (r=0.38, p<0.001). These predictions accounted for variations in learning rates both across (Fig. 6B) and within (Fig. 7C) subjects.

Task–independent pupil manipulation altered behavior

To examine whether the correlations between pupil measures and learning behavior might reflect an underlying causal process, we used an arousal manipulation that affected pupil diameter and measured its effects on learning behavior. In particular, we occasionally and without warning switched the auditory cue that preceded fixation. Subjects were told that these auditory–cue switches were unrelated to the task and they therefore should ignore the specific sounds. Nevertheless, this manipulation led to increases in both pupil average and pupil change on trials in which the fixation cue was switched (Fig 8A; *t*–test for H_0: mean effect size=0, $p<0.001$ for both pupil average and pupil change). Thus, we caused consistent changes in the pupil measures that were correlated with the computational variables needed to solve the task.

This manipulation caused systematic changes in task performance that depended on baseline pupil diameter (Fig. 8B). For trials with relatively small baseline diameter (i.e., less than its per–subject median value), individual subjects tended to use larger learning rates on auditory–switch trials than otherwise (Fig 8B abscissa; mean across subjects=0.113, *t*–test for H_0: mean=0, $p<0.01$). For trials with relatively large baseline diameter, subjects used slightly smaller learning rates on auditory–switch trials than otherwise, although this trend was not statistically significant (Fig 8B ordinate; mean=–0.037, $p=0.35$). The average difference in the size of these effects from small– versus large–diameter trials was >0, implying that the effects of this manipulation depended on

baseline pupil diameter (Fig 8B diagonal; paired *t*–test, $p<0.001$). These effects did not result from systematic differences in task conditions for switch versus non–switch trials, because the same three analyses yielded no effects when applied to learning rates computed by our reduced Bayesian model ($p>0.5$).

This dependence on baseline pupil diameter is suggestive of the Yerkes–Dodson "inverted U" relationship between arousal and learning. According to that idea, learning is highest for moderate levels of arousal and lowest for either overly high or overly low levels of arousal (Yerkes and Dodson, 2004). Our subjects appeared to be consistently engaged during task performance, implying that we were probably not sampling overly low or high arousal states. Nevertheless, in a narrower range and assuming a correspondence between arousal state and baseline pupil diameter, we found that the relationships between learning behavior and our arousal manipulation were qualitatively consistent with an "inverted U." In particular, auditory–switch trials tended to correspond to the largest increases in learning rate when baseline pupil diameter was relatively low (steepest ascent in the "inverted U") and the largest decreases in learning rate when baseline pupil diameter was relatively high (steepest descent in the "inverted U"; Fig. 8C, open circles).

This "inverted U" relationship was also apparent in our previous pupil measurements, in two ways. First, across subjects, those with larger average

pupil diameters during outcome viewing tended to use learning rates that were less, or even negatively, predicted by fluctuations in pupil metrics relative to other subjects (Fig 7B). Second, subjects that had lower pupil–predicted hazard rates used learning rates that were positively correlated with pupil metrics when their baseline pupil diameter was low but negatively correlated when their baseline pupil diameter was high (Fig 8C, filled circles). Thus, results from both our pupil-manipulation and pupil-measurement experiments were consistent with an important role for the arousal system in the rational regulation of learning.

Discussion

We examined the relationship between pupil diameter, which is related to arousal and autonomic state, and learning rate, which describes the extent to which new information is used to adjust existing cognitive beliefs. Consistent with previous work (Nassar et al., 2010;Behrens et al., 2007;Krugel et al., 2009), we found that human subjects performing a predictive–inference task were most heavily influenced by outcomes that occurred shortly after a change point in the outcome–generating process. One possible mechanism for this effect is a dynamic regulation of the relative influence of incoming information on cortical processing (Yu and Dayan, 2005). Insights into the computations required for such a regulator are provided by a reduced model that approximates the ideal observer for the task, describes subject behavior, and bases learning rates on

two parameters that we found to be represented in pupil measurements: change–point probability and relative uncertainty.

In our model, change–point probability depends on the absolute value of the most recent prediction error and drives increased learning after surprisingly large errors. We found that change–point probability was positively correlated with changes in pupil diameter. This relationship is consistent with early pupillometry studies that showed an inverse relationship between stimulus–evoked pupil responses and stimulus probability, as well as more recent work interpreting outcome–locked pupil responses in terms of the surprise associated with errors in judging uncertainty, called the risk prediction error (Raisig et al., 2010;Friedman et al., 1973;Preuschoff et al., 2011). We also found that pupil change was not always directly related to change–point probability, with particularly large pupil changes on trials with exactly zero error that might have been surprisingly rewarding and/or reflected an association with an atypical consequence (i.e., no possibility of updating the next prediction).

Relative uncertainty, the second parameter in our model, represents uncertainty about the true underlying mean and drives learning from outcomes that occur after a change point. We found that relative uncertainty was correlated with average pupil diameter. We also found that changes in another form of uncertainty that should not drive learning (i.e., changes in the standard deviation

of the generative process in our task) did not lead to similar effects on pupil diameter. These results are complementary to a recent finding that pupil diameter tends to increase during exploratory decisions that occur during periods of uncertainty about the best available option (Jepma and Nieuwenhuis, 2010). These findings suggest that pupil–linked arousal systems encode an uncertainty signal that facilitates both learning and information–seeking behaviors.

We also found strong individual differences in task behavior that could be captured by fitting a prior expectation about the rate of change points (hazard rate) to behavioral data. We found that subjects who were fit by higher hazard–rate models tended to have larger pupil dilations during the outcome–viewing period. This physiological difference arose early in the viewing period, consistent with the idea that these individual differences reflected a prior expectation about the source of the upcoming error.

We used these relationships between pupil metrics and change–point probability, relative uncertainty, and the hazard–rate prior to predict the extent to which subjects were influenced by each new outcome. We also manipulated pupil diameter using a task–irrelevant auditory manipulation that resulted in changes in task performance that were consistent with our measured relationships between pupil metrics and key task variables. These results provide new insights into the specific computations that are reflected in pupil diameter and establish their

causal role in belief updating.

These computations likely involve, at least in part, neural activity in the locus coeruleus. One intriguing possibility is that the two key variables from our model are encoded by two distinct modes of locus coeruleus activation (Aston-Jones and Cohen, 2005): change–point probability, reflected in pupil change, is encoded by phasic activation of the locus coeruleus, whereas relative uncertainty, reflected in pupil average, is encoded by tonic activation of the locus coeruleus. Although direct confirmation is still needed, this idea is supported by several lines of evidence, including: 1) a compelling example of simultaneous measurements of locus coeruleus activity and pupil diameter in a monkey that are closely correlated (Aston-Jones and Cohen, 2005); 2) similar modulations of pupil diameter and locus coeruleus activity under certain task conditions, such as changes in utility in that affect behavioral engagement (Jepma and Nieuwenhuis, 2010;Gilzenrat et al., 2010); and 3) a proposed anatomical substrate involving common activation from the nucleus paragigantocellularis, which contributes to both locus coeruleus and sympathetic nervous system function (Nieuwenhuis et al., 2010;Aston-Jones et al., 1986). The consequence of locus coeruleus involvement would be the task–related release of norepinephrine throughout the nervous system. Consistent with our results, norepinephrine release is thought to permit or facilitate changes in behavior that follow unexpected changes in the environment and learning in general, possibly by modulating experience–

dependent neural plasticity (Yu and Dayan, 2005;Sara et al., 1994;Aston-Jones et al., 1997;Tully and Bolshakov, 2010;Harley, 1987;Corbetta et al., 2008;Bouret and Sara, 2005).

More generally, our results are consistent with the idea that brain areas that regulate the influence of newly arriving information on existing beliefs are also strongly linked to arousal and autonomic function (Behrens et al., 2007;Yu and Dayan, 2005;Jepma et al., 2010;Gilzenrat et al., 2010;Preuschoff et al., 2011;Critchley et al., 2001;Critchley, 2005). These areas likely include not just the locus coeruleus but also the anterior cingulate cortex (ACC), which has strong reciprocal connections with the locus coeruleus and whose activity encodes several signals closely related to change–point probability, including unsigned prediction errors and learning rates (Behrens et al., 2007;Aston-Jones and Cohen, 2005;Krugel et al., 2009;Matsumoto et al., 2007). This arousal system appears to govern not simply overall alertness or other non–specific factors that might affect overall task performance, but rather a computationally sophisticated process that rationally regulates the influence of new sensory information in a dynamic environment. These computations take into account both ongoing processing of task–relevant variables like change–point probability and relative uncertainty and state variables including prior expectations about the rate of change. These factors are combined in a manner that is consistent with the Yerkes–Dodson “inverted U” relationship between arousal level and learning

rate (Fig. 8C) (Yerkes and Dodson, 2004).

In summary, our work suggests a relationship between arousal state and learning rate that is likely a result of a coordinated learning–arousal network including the locus coeruleus and ACC. The representation of normative learning variables in this network suggests that subtle changes in arousal might reflect rational regulation of the influence of new information on ongoing inferences about a dynamic world.

Methods

Predictive–inference task. Human subject protocols were approved by the University of Pennsylvania Internal Review Board. Thirty subjects (19 female, 11 male; age range = 19–29 years) participated in the primary study and an independent sample of 29 subjects (17 female, 12 male; age range = 19–25 years) participated in the arousal manipulation study after providing informed consent. Both studies used a predictive–inference task that required subjects to predict each subsequent number to be presented in a series (Nassar et al., 2010). For each trial t, a single integer (X_t) was presented that was a rounded pick sampled independently and identically from a Gaussian distribution whose mean (μ_t) changed at unsignaled change points and whose standard deviation (σ_t) was fixed to either 5 or 10 within each block of 200 trials. Change points

occurred with a probability of zero for the first three trials following a change point and 0.1 for all trials thereafter.

To facilitate measurements of non–luminance–mediated effects on pupil diameter, we used a different visual display and task timing than in our previous study[2]. Subjects were shown a numeric representation of their current prediction at a central location on a CRT monitor. Background screen pixels were a checkerboard of light and dark pixels (mean±STD luminance in a circle with radius 6.5 cm= 0.457±0.010 cd/m^2). Numbers were drawn in an intermediate gray color (0.445±0.005 cd/m^2). When viewed passively by a control group of four subjects outside of the context of the predictive–inference task, no individual stimulus (number) had a significant effect on average pupil diameter or evoked changes in pupil diameter (*t*–test for H_0: equal means between each stimulus and all others, $p>0.3$ for all stimuli after correcting for multiple comparisons), nor did the number of digits contained within the stimulus affect either pupil variable ($p>0.4$).

For each trial, the subject indicated his or her updated prediction using a video gamepad. Each prediction was constrained to be between the previous prediction and the most recent outcome, thus limiting learning rates to between zero and one. After the new prediction was chosen, the numeric representation of this prediction disappeared, an auditory cue was played, and a numeric

representation of the new outcome was shown. Subjects were instructed to fixate centrally for 2 s at this point; failure to do so (within a square window, 9° per side) resulted in a tone indicating a fixation error. After 2 s the new outcome disappeared, the prediction re–appeared, and an auditory cue was played to indicate that the prediction should be updated. Fourteen subjects also participated in a control version of the task in which the prediction was reset after viewing the new outcome to reflect an update equivalent to a learning rate of 0.5. For this task, the same motor output (in terms of number or duration of button presses) was required to use a learning rate of either zero or one on each trial.

Subjects were told that the numbers were generated from a noisy process and that several discreet change points would occur over the course of the task. They were instructed to make a prediction on each trial (B_t) such that the average error made on all predictions, $\langle |B_t - X_t| \rangle$, would be minimized. Payout depended on how well they achieved this goal, as described previously (Nassar et al., 2010).

The pupil–manipulation task was identical to primary version of the task, except that the auditory cue played at the beginning of fixation was occasionally switched to another sound from a library of 31 sound effects downloaded from an online library. Sounds were 0.09–1.4 s in duration (mean±STD = 0.72±0.42 s) and played at 56–70 dB (*A*–weighted; mean±STD = 62.5±3.9 dB). Switch trials occurred at random, with a probability of 0.1 on the 9 trials following a switch, 0.8

thereafter. On switch trials, the given sound was played, on average, 7 dB louder than otherwise. Seven of 29 subjects completing the pupil–manipulation task were excluded from further analyses because of an excessive number of fixation errors (blinks or lost fixation on >40 percent of trials).

Pupil–diameter measurements. Pupil diameter was sampled at 120 Hz and recorded throughout the task using an infrared video eye–tracker (ASL, Inc.). Blinks were identified using a custom blink filter based on pupil diameter and vertical and horizontal eye position, then removed by linear interpolation of values measured just before and after each identified blink. Blink–filtered diameter was low–pass filtered using a Butterworth filter with a cutoff frequency of 3.75 Hz. These filtered measurements were then *z*–scored within each session.

All analyses excluded trials in which blinks or fixation errors during outcome viewing were detected online (these events were followed by a beep to remind the subject to minimize their occurrence). The first 20 trials from each block were also excluded to avoid possible changes in average luminance at block boundaries. Pupil average was computed for each trial by taking the mean of all 240 z–scored pupil measurements from the 2 s–long outcome–viewing period of the trial. Pupil change was computed for each trial by subtracting the average pupil measurement from early in the outcome–viewing period (0–1 s after

outcome presentation) from the average pupil measurement from late in the outcome–viewing period (1–2 s after outcome presentation). Trials that included blinks that were detected offline (but not online) were used to compute pupil average by interpolating values from just before and just after the blink. These trials were not used to compute pupil change, which was much more sensitive to the timing of blinks.

Reduced Bayesian model. Optimal performance on the predictive–inference requires inferring the probability distribution over possible outcomes on the next timestep, given all previous data and the process by which those data were generated: $p(X_{t+1}|X_{1:t})$. Because the relationship between the data on the next timestep is independent of all previous data conditioned on the mean of the current distribution (μ), the solution can be formulated in terms of μ:

$$p(X_{t+1}|X_{1:t}) = \sum_{\mu_t} p(X_{t+1}|\mu_t)p(\mu_t|X_{1:t}) \quad [1]$$

and the probability distribution over possible means given previous data can be inverted according to Bayes' rule:

$$p(\mu_t|X_{1:t}) = \frac{p(X_{1:t}|\mu_t)p(\mu_t)}{p(X_{1:t})} \quad [2]$$

Although computationally tractable solutions to this problem exist, these solutions specify learning rates that are complicated functions of either the probability distribution over all possible means[1] or over all possible "runs" of non–change–point trials [19]. To simplify the algorithm, the reduced model computes the posterior probability distribution over possible means as described above but maintains only the first two moments of this distribution. This assumption massively reduces the number of required computations but has minimal effects on performance [2]. An added advantage of this model is that it can be formulated as a delta rule:

$$B_{t+1} = B_t + \alpha_t \times \delta_t \qquad [3]$$

$$\delta_t = x_t - B_t$$

where *B* is the belief about the mean of the underlying distribution; α is the learning rate; and δ is the prediction error, which is the difference between the actual and predicted outcome. The learning rate depends on two variables that are updated on each trial:

$$\alpha_t = \tau_t + (1 - \tau_t)\Omega_t \quad [4]$$

where change–point probability (Ω) reflects the probability that μ_t is not equal to μ_{t-1}, and relative uncertainty (τ) reflects the variance on the predictive distribution in μ (i.e., uncertainty about the location of the mean) divided by the variance on the predictive distribution in X (i.e., total uncertainty about the location of the next outcome).

Performance of the reduced Bayesian model also depends on an expectation about the prior probability on change points, or the hazard rate. Specifically, hazard rate directly influences the computation of change–point probability on each trial:

$$\Omega_t = \frac{U(X_t|0,300)H}{U(X_t|0,300)H + \mathcal{N}(X_t|B_t,\sigma_t^2)(1-H)} \quad [5]$$

Where U and N represent uniform and normal distributions, respectively; H is the hazard rate; B_t is the model's prediction on trial t; and σ^2 is the total variance on the predictive distribution, which is discussed below. We incorporated hazard rate into the model in two ways: 1) using the true generative hazard rate for trials in which a change point did not recently occur (0.1) or 2) by fitting the model to behavior by minimizing the total squared difference between subject and model

predictions using a constrained search algorithm (fmincon in MATLAB) with hazard rate as a free parameter.

The total variance on the predictive distribution in the model comes from two sources:

$$\sigma_t^2 = N^2 + \frac{\tau_t N^2}{1 - \tau_t} \quad [6]$$

The first source is the standard deviation on the outcome–generating distribution (*N*). The second source is uncertainty about the mean of that distribution and depends on both *N* and relative uncertainty (τ). Here we set *N* to be the actual experimental standard deviation, but we update τ after each outcome according to the variance on the predictive distribution over possible means:

$$\tau_{t+1} = \frac{N^2\Omega_t + (1-\Omega_t)(\tau_t N^2) + \Omega_t(1-\Omega_t)(\tau_t + B_t(1-\tau_t) - X_t)}{N^2\Omega_t + (1-\Omega_t)(\tau_t N^2) + \Omega_t(1-\Omega_t)(\tau_t + B_t(1-\tau_t) - X_t) + N^2} \quad [7]$$

such that if a change point occurs, relative uncertainty is reset to 0.5 (first term in numerator); if a change point does not occur, relative uncertainty is reduced (second term in numerator); and if the model is uncertain about whether a change point occurred, relative uncertainty is increased to reflect this uncertainty

(third term in numerator).

Statistical analyses. Trial–by–trial values of pupil average and pupil change were each z–scored for the full session (z_{PA} and z_{PC}, respectively) and then fit with a linear regression model using four parameters: 1) change–point probability, computed by the reduced Bayesian model for each trial; 2) relative uncertainty, computed by the reduced Bayesian model for each trial; 3) noise, the standard deviation of the outcome–generating distribution; and 4) a binary vector specifying whether or not the subject correctly predicted the outcome on that trial. We also used a larger model that, in addition to the above four parameters, included: the average horizontal and vertical eye position and the change in horizontal and vertical eye position measured during the outcome–viewing period; the subject's prediction and the computer–generated outcome from the current trial; the pupil change measured on the previous trial; and the trial number within the block and within the session.

Pupil–predicted hazard rates were derived from pupil measurements and the reduced Bayesian model as follows. First, we inferred the subjective hazard rate used by each subject by fitting his or her behavioral data to the reduced Bayesian model with hazard rate (H) as the only free parameter. Next, we fit a linear regression model explaining H in terms of pupil measurements. That model had two terms, computed per subject: 1) the mean value of pupil average, and 2) an

index of pupil dynamics. The index was computed as the mean value of the dot product of trial–by–trial pupil measurements and the mean–subtracted curve shown in Fig. 6B. Finally, we used the coefficients from a linear fit that excluded the data from an individual subject to combine the mean pupil average and pupil–dynamics index (from the excluded subject) into a pupil–predicted hazard rate for that subject.

Pupil–predicted learning rates were computed according to the relationships between pupil metrics and model parameters. Linear fits to the relationship between pupil average and relative uncertainty were computed for each subject, and these fits were used estimate relative uncertainty for each trial–by–trial measurement of pupil diameter. Linear fits to the relationship between pupil change and change–point probability were computed for each subject, and these fits were used to estimate change–point probability for each trial–by–trial measurement of pupil change. To compute predicted learning rates, the two predicted model quantities were combined according to Eq. 4. We also used a more complex linear model that took into account pupil–predicted hazard rates; see text for details.

Arousal-induced learning effects for the inverted–U analyses were computed separately for sound–manipulation and non–manipulation sessions. For sound–manipulation sessions, learning rates were fit to a cumulative Weibull as a

function of error magnitude for each subject and noise condition, to account for the relationship shown in Fig. 4A. Residuals from this fit, which reflected error-independent variability in learning rate, were z–scored per subject. Initial pupil diameter, as measured by the average diameter during the first 100 ms of the outcome phase, was also z–scored per subject. Data were binned across subjects according to the initial diameter z–score. The effect of the sound manipulation was computed as a signed *d'* describing the difference in the z–scored residual learning rates used on auditory shift versus non–auditory shift trials. For non–manipulation sessions, the relationship between pupil metrics and learning rate was characterized only for subjects with low pupil–predicted hazard rates (<0.6). Subjects with high pupil–predicted hazard rates tended to have small or negative relationships between pupil metrics and learning rate and thus were omitted from this analysis. Arousal effect size was computed as the correlation coefficient between the weighted sum of pupil metrics and learning rate, each z–scored per subject (positive/negative values indicate that learning rates tended to increase/decrease as pupil effects increased) for equally sized bins of baseline pupil diameter (z–scored per subject).

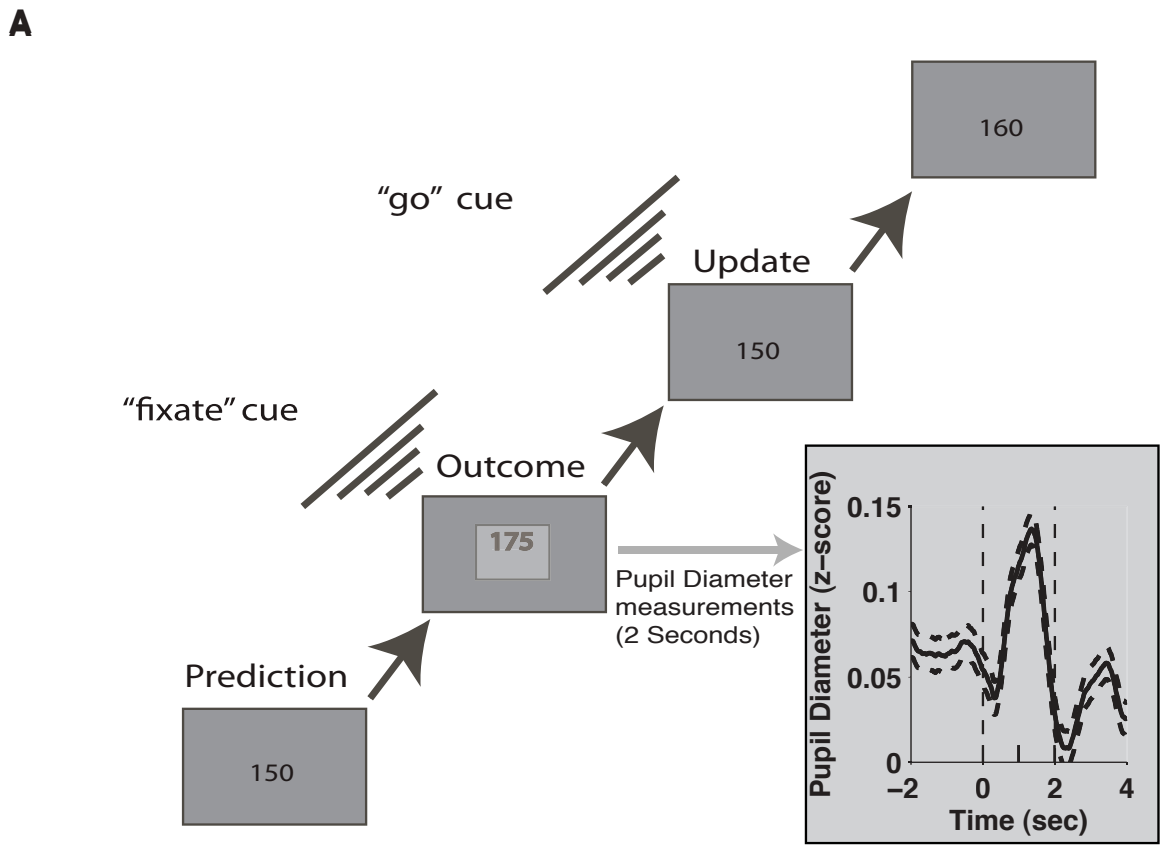


Figure 1. Predictive–inference task sequence and pupillometry. Learning rate was computed by dividing the difference in the prediction from one trial to the next by the difference between the current outcome and the current prediction. Inset: mean±SEM pupil diameter, averaged across z–scores computed per subject, aligned to outcome presentation (time=0). Pupil average was computed for each trial as the mean pupil diameter, z–scored by subject, across the entire 2–s fixation window (vertical dashed lines). Pupil change was computed for each trial as the difference in mean diameter, z–scored by subject, measured late (time=1–2s) versus early (time=0–1s) during fixation.

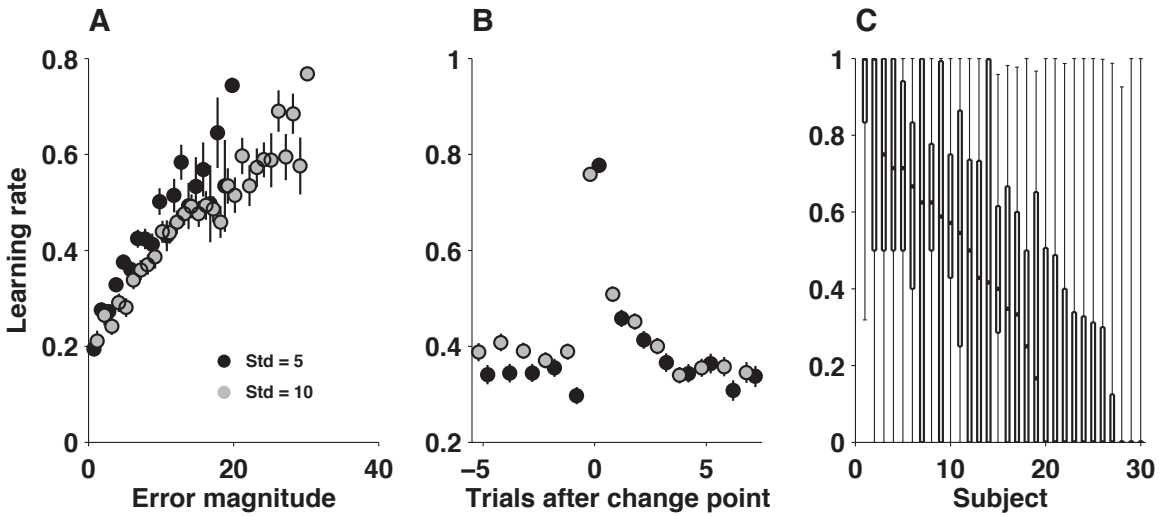


Figure 2. Task performance. A, Learning rates were highest after subjects made larger errors, scaled by noise (as indicated). Points and errorbars are mean±SEM from all subjects. B, Learning rates were highest on change–point trials and decayed thereafter, similarly for both noise conditions. Points and errorbars are mean±SEM from all subjects. C, Learning–rate distributions across all trials from each of the 30 subjects (abscissa), sorted by median learning rate. Horizontal line, box, and whiskers indicate median, 25^{th}/75^{th} percentiles, and 5^{th}/95^{th} percentiles, respectively.

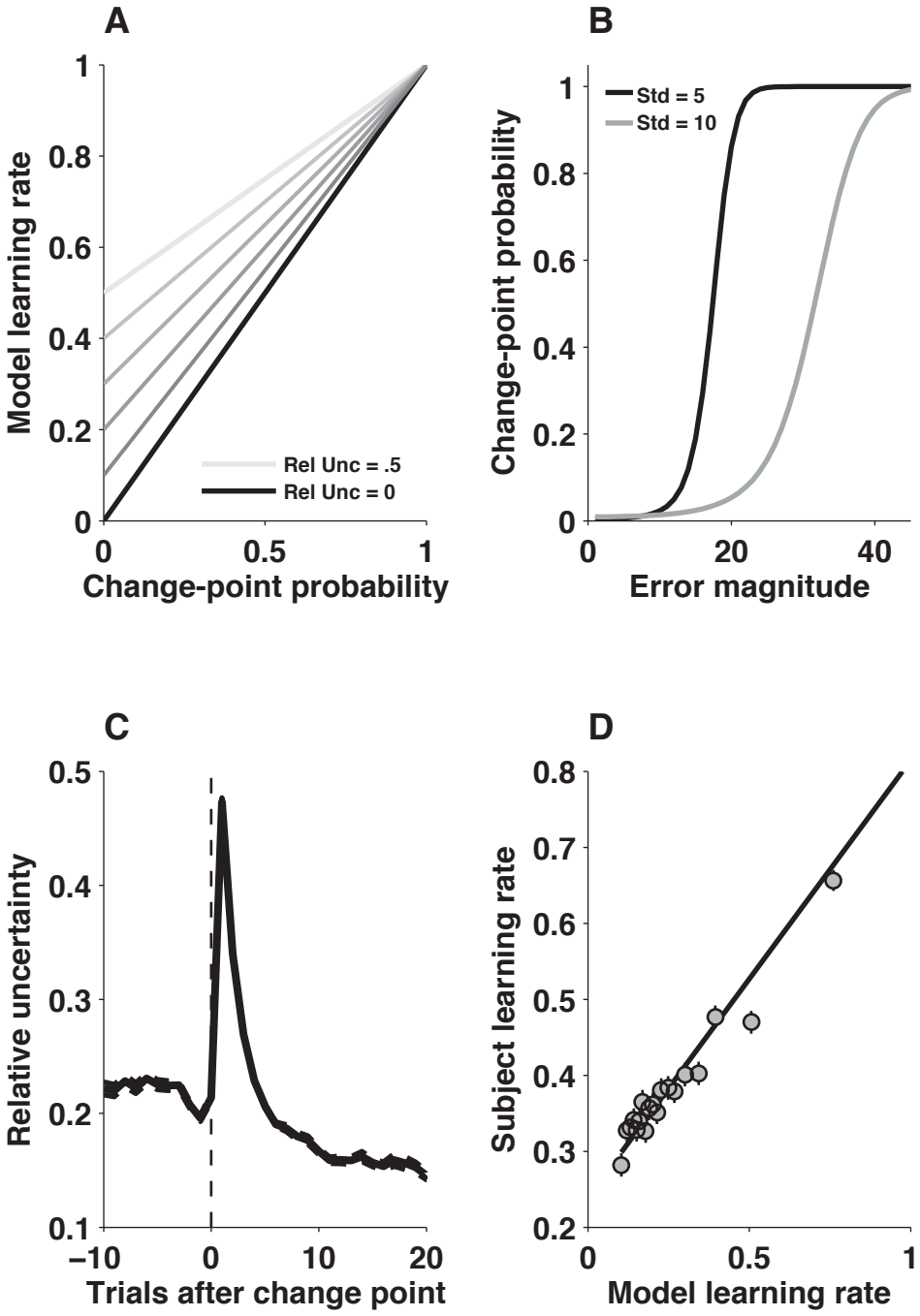


Figure 3. Reduced Bayesian model. A, Learning rate as a function of change–point probability (abscissa) and relative uncertainty (line shading), as computed by the model. B, Change–point probability computed by the model as a function

of error magnitude (abscissa) for the two different noise conditions, as indicated, computed for a given relative uncertainty (equal to 0.02 for this figure). C, Mean±SEM relative uncertainty computed by the model aligned to change points from all sequences experienced by the subjects for the two different noise conditions. D, Trial–by–trial comparison of subject and model learning rates. Model learning rates were computed using the same sequence of outcomes experienced by each subject. Points and error bars are mean±SEM data from all subjects grouped into 20 five–percentile bins according to the corresponding model learning rate. The solid line is a linear fit to the unbinned data (r=0.33, p<0.001).

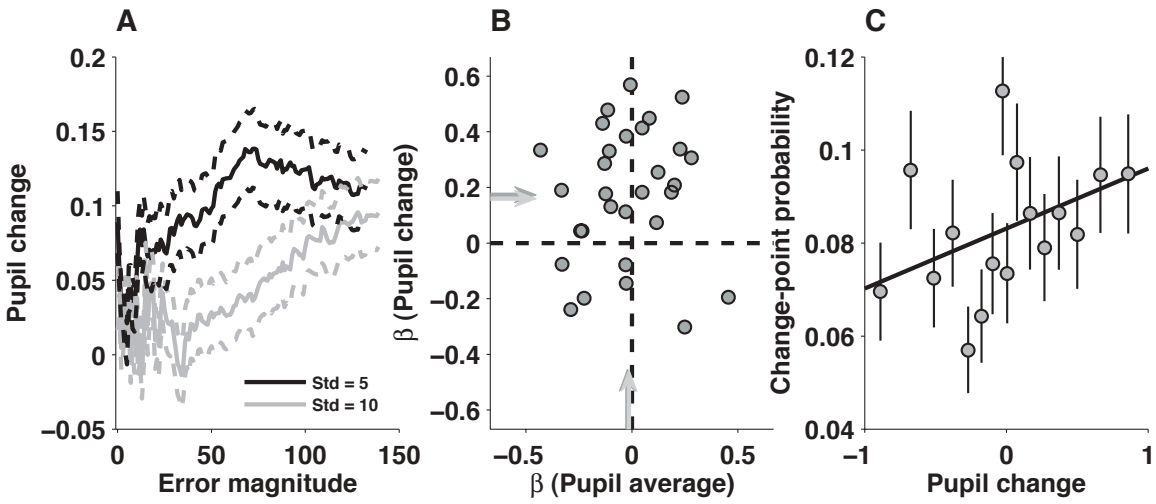


Figure 4. Relationship between pupil change and change–point probability. A, Mean±SEM pupil change from all trials and all subjects for running bins of 150 trials, binned according to the absolute prediction error and sorted by noise, as indicated. B, Regression coefficients describing the linear relationship between change point–probability (p_{CH}) and z–scored pupil change (z_{pc}, ordinate) versus the regression coefficients describing the linear relationship between p_{CH} and z–scored pupil average (z_{PA}, abscissa). Points are regression coefficients computed for each subject individually, using the four–parameter regression model. Arrows indicate mean values from this model (dark, equal to 0.174 z_{PC}/p_{CH}, t-test for H_0: mean=0, p<0.001 for the ordinate, –0.022 z_{PA}/p_{CH}, p=0.58 for the abscissa) or from the full model (light, equal to 0.148 z_{PC}/p_{CH}, p<0.001 for the ordinate, –0.014 z_{PA}/p_{CH}, p=0.70 for the abscissa). Dark arrows are partially occluded by light ones. C, Change–point probability from the reduced Bayesian model versus pupil change. Points and error bars are mean±SEM data from all subjects grouped into 20 five–percentile bins. The solid line is a linear fit to the unbinned data (slope = 0.012 p_{CH}/z_{PC}, p<0.001 for H_0: slope=0).

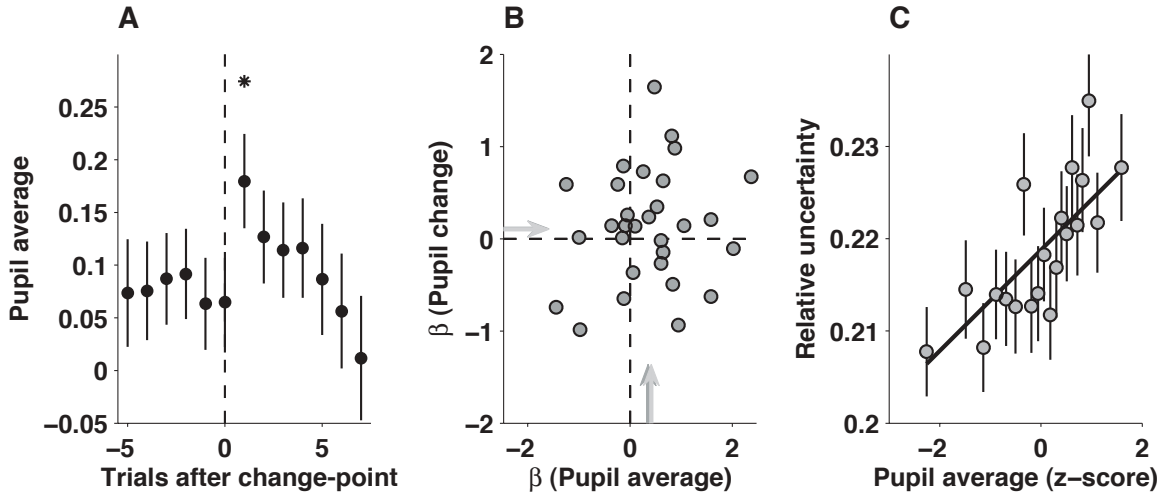


Figure 5. Relationship between pupil diameter and relative uncertainty. A, Mean±SEM pupil average from all subjects as a function of trials relative to task change points. Asterisk indicates trials differing significantly from all other trials (permutation test for H_0: equal means after correction for multiple comparisons, $p<0.05$). B, Regression coefficients describing the relationship between relative uncertainty (*RU*) and *z*–scored pupil change (z_{pc}, ordinate) versus the regression coefficients describing the relationship between *RU* and *z*–scored pupil average (z_{PA}, abscissa). Points are regression coefficients computed for each subject individually, using the four–parameter regression model. Arrows indicate mean values from this model (dark, equal to 0.135 z_{PC}/RU, *t*-test for H_0: mean=0, $p=0.28$ for the ordinate, 0.35 z_{PA}/RU, $p<0.05$ for the abscissa) or from the full model (light, equal to 0.127 z_{PC}/RU, $p=0.24$ for the ordinate, 0.40 z_{PA}/RU, $p<0.01$ for the abscissa). Dark arrows are partially occluded by light ones. C, Relative uncertainty from the reduced Bayesian model versus pupil average. Points and error bars are mean±SEM data from all subjects grouped into 20 five–percentile bins. The solid line is a linear regression to unbinned data (slope = 0.0055 RU/z_{PA}, $p<0.001$ for H_0: slope=0).

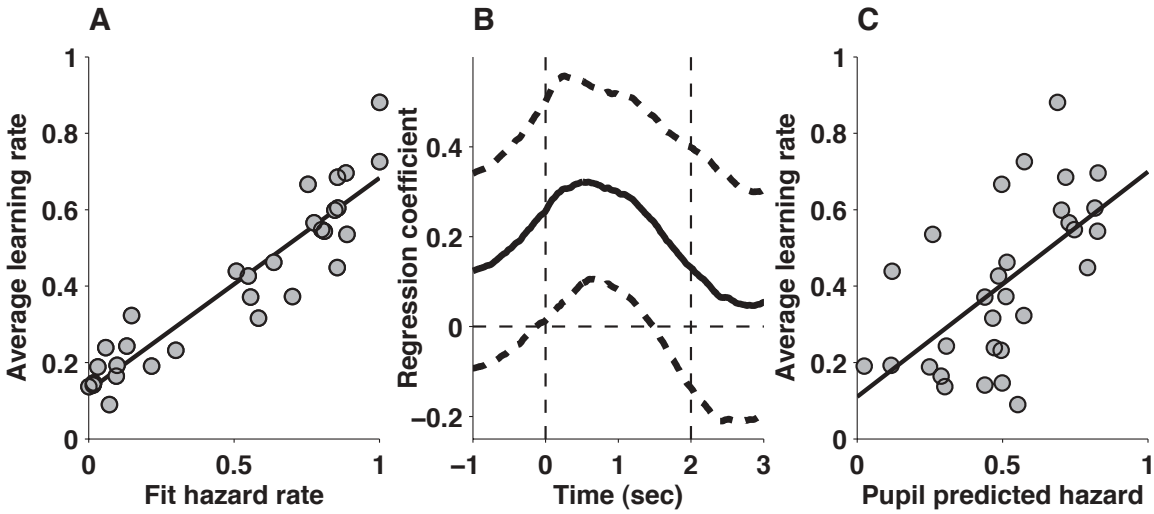


Figure 6. Individual differences in learning rate, hazard rate, and pupil diameter. A, Mean learning rate per subject versus the hazard rate of the reduced Bayesian model that best fit that subject's performance (points). The solid line is a linear fit (r=0.93, p<0.001). B, Regression coefficients describing the relationship between fit hazard rates and bin–by–bin pupil measurements across subjects, computed in sliding 8.3–ms bins and aligned to outcome presentation (time=0). Dotted lines indicate 95% confidence intervals. C, Relationship between pupil–predicted hazard rate and average learning rate for each subject (points). Pupil–predicted hazard rates were computed using a linear regression model that included both shape and magnitude of the average pupil response for each subject (see Methods). The solid line is a linear fit (r=0.59, p<0.001).

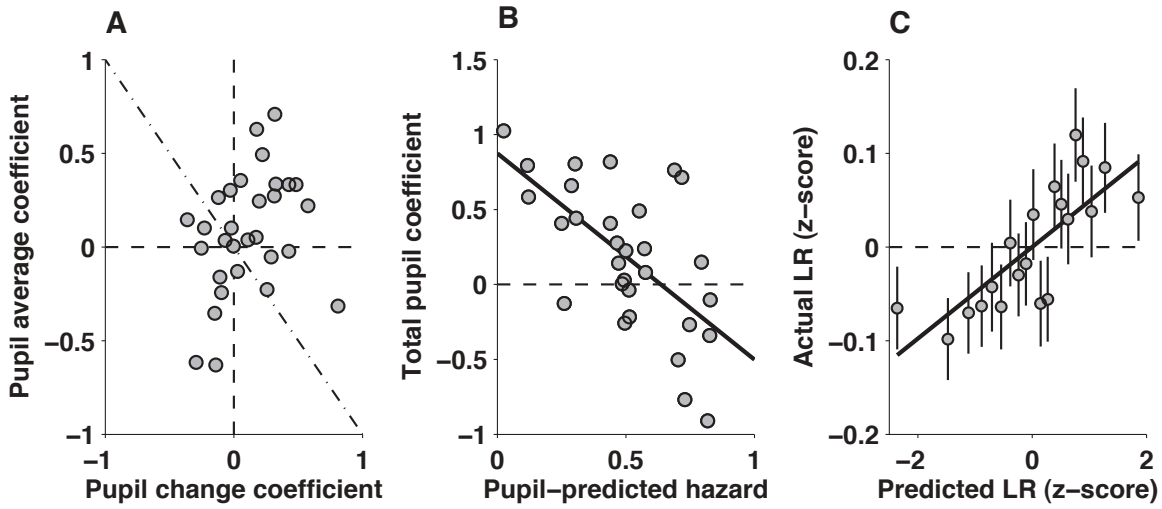


Figure 7. Pupil metrics predict learning rate. A, Regression coefficients describing the linear, trial–by–trial relationships between pupil change and the subsequent learning rate (ordinate) and between pupil average and the subsequent learning rate (abscissa). Points are regression coefficients computed for each subject individually, using a four–parameter regression model that also included trial number and block number as covariates. B, The relationship between learning rate and pupil parameters depended on the subject's baseline pupil response. For each subject, the sum of the regression coefficients from panel A are plotted as a function of the pupil–predicted hazard rate from Figure 6C. The line is a linear fit (r= –0.059, p<0.001). C, Predicted versus actual learning rate. Both values are z–scored per subject. Data from all subjects are grouped into 20 equally sized bins of predicted learning rate. The line is a linear fit to the unbinned data (Slope = 0.052 zActual/zPredicted, p<0.001 for H_0: slope=0).

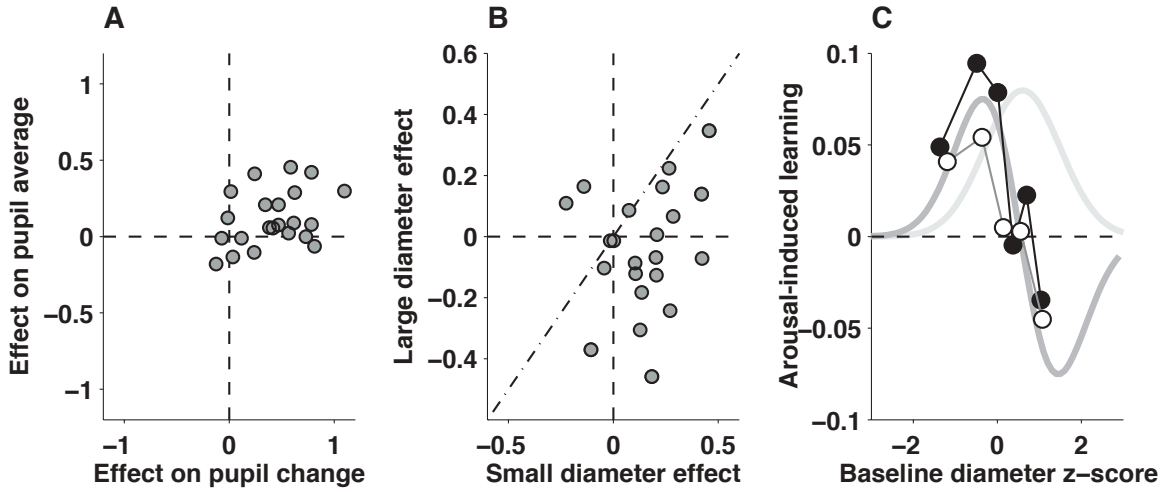


Figure 8. Effects of the pupil manipulation. A, Evoked changes in pupil diameter. For each subject, pupil average (ordinate) and pupil change (abscissa) were z–scored across all trials. Each point represents the difference in the mean z–scores for auditory switch versus non–switch trials for an individual subject. Positive values indicate larger values on switch trials. B, Evoked changes in learning behavior. For each subject, learning rate was z–scored across all trials and fit to a cumulative Weibull as a function of error magnitude for each noise condition, to account for the relationship shown in Fig. 4A. Each point represents the difference in the mean value of the residuals from these fits for auditory switch versus non–switch trials for an individual subject, separated by trials in which the initial pupil diameter was smaller (ordinate) or larger (abscissa) than its median value. Positive values indicate larger learning rates on auditory switch trials. C, A possible relationship between learning and arousal based on an "inverted U" (light gray, modeled as Gaussian). A given change in learning for a given a change in pupil metrics (ordinate), plotted as a function of baseline pupil diameter (abscissa), is shown for: 1) the hypothesized Gaussian (its derivative is shown in dark gray), 2) the measured effects of the auditory manipulation (open points), and 3) the measured relationship between pupil metrics and learning rate during non-manipulation sessions. See Methods for details.

CHAPTER 4

A healthy fear of the unknown: perspectives on the interpretation of parameter fits from computational models in neuroscience

Matthew R. Nassar and Joshua I. Gold

Abstract

Computational models are commonly used to infer the latent factors responsible for generating behavior. However, the complexity of many behaviors can handicap the interpretation of such models. Here we provide perspectives on problems that can arise when interpreting parameter fits from models that provide incomplete descriptions of behavior. We illustrate these problems using commonly used and neurophysiologically motivated reinforcement-learning models fit to simulated behavioral data sets from learning tasks. These models can pass a host of standard goodness-of-fit tests and other model-selection diagnostics even when they do not include a complete description of behavior. We show that such incomplete models can be misleading by yielding biased estimates of the parameters explicitly included in the model. This problem is particularly pernicious when the neglected factors are unknown and therefore not easily identified by model comparisons and similar methods. An obvious conclusion is that a

parsimonious description of behavioral data does not necessarily imply an accurate description of the underlying computational mechanisms. Moreover, general goodness-of-fit measures are not a strong basis to support claims that a particular model can provide a generalized understanding of the computational factors that govern behavior. To help overcome these challenges, we advocate the design of tasks that provide direct reports of the computational variables of interest. Such direct reports complement computational modeling approaches by providing a more complete, albeit possibly more task-specific, representation of the factors that drive behavior. Computational models then provide a means to connect such task-specific results to a more general algorithmic understanding of the brain.

The use of models to infer the neural computations that underlie behavior is becoming increasingly common in neuroscience research, especially for cognitive and perceptual tasks involving decision-making and learning. As their sophistication and usefulness expand, these models become increasingly central to the design, analysis, and interpretation of experiments. We consider this to be generally a positive development but provide here some perspectives on the challenges inherent to this approach, particularly when behavior might be driven by unexpected factors that can complicate the interpretation of model fits. Our goal is

to raise awareness of these issues and present complementary approaches that can help ensure that that our understanding of the brain does not become overly conditioned on the quality of existing models fit to particular data sets.

We illustrate these challenges using a set of models that describe the ongoing process of learning values to guide actions and are used extensively in the field of cognitive neuroscience (Beeler et al., 2010;Doll et al., 2011;Frank et al., 2009;Jepma and Nieuwenhuis, 2010;Walton et al., 2010;Sul et al., 2011;Seo and Lee, 2008;Strauss et al., 2011;Nassar et al., 2010;Luksys et al., 2009;Daw et al., 2006;Behrens et al., 2007;Krugel et al., 2009). These models adjust expectations about future outcomes according to the difference between actual and predicted outcomes, known as the prediction error. Originally developed in parallel in both animal- and machine-learning fields (Rescorla and Wagner, 1972;Bertsekas and Tsitsiklis, 1996;Sutton and Barto, 1998), this relatively simple form of reinforcement-learning algorithm, often referred to as a "delta rule" because the prediction error is typically represented as a the Greek symbol delta (∂) in the equations, has: 1) provided efficient solutions to a broad array of biologically relevant problems (Sutton and Barto, 1998); 2) accounted for many, but not all, learning phenomena exhibited by both human and non-human subjects (Sutton and Barto, 1998); 3) provided a generative architecture that has been used to predict behavior across tasks, compare brain activity to learning variables within a single task, and explore the range of possible behaviors that one might expect to find in a

variable population (Miller et al., 1995;Dayan and Daw, 2008); and 4) guided an understanding of the neural computations expressed by the brainstem dopaminergic system (Schultz et al., 1997). These successes have led to the proposal that the interpretation of delta-rule model parameters fit to behavioral data from human subjects performing simple learning tasks might serve as a more precise diagnostic tool for certain mental disorders than existing methods (Huys et al., 2011;Huys et al., 2009;Maia and Frank, 2011). Thus reinforcement-learning models are becoming highly influential in guiding and filtering our understanding of normal and pathological brain function.

Here we focus on the interpretation of a term in most delta-rule models called the learning rate. The learning rate, α, determines the amount of influence that the prediction error, δ, associated with a given outcome has on the new expectation of future outcomes, E:

$$E_{t+1} = E_t + \alpha \times \delta_{t+1}$$

EQ 1

As its name implies, the learning rate determines how slowly or quickly the model adapts to errors. A fixed value near zero implies that expectations are updated slowly, essentially averaging over a long history of past outcomes. In contrast, a fixed value near one implies that expectations are updated quickly to match the

most recent outcomes. Thus, the learning rate can be interpreted as the amount of influence each unpredicted outcome exerts on the subsequent expectation.

Recent work has highlighted the advantages of using learning rates that, instead of remaining fixed, are adjusted adaptively according to environmental dynamics (Nassar et al., 2010;Behrens et al., 2007;Krugel et al., 2009;Yu and Dayan, 2005;Preuschoff and Bossaerts, 2007;Mathys et al., 2011). For example, adaptive learning rates can help ensure that expectations remain stable during periods of stability but change rapidly in response to abrupt environmental changes. Consistent with this idea, human behavior on tasks containing abrupt changes conforms to models in which the influence of each outcome depends on the statistics of other recent outcomes (Nassar et al., 2010;Behrens et al., 2007;Krugel et al., 2009). Such rational adjustments of learning rate are most prominent after changes in action-outcome contingencies that lead to surprisingly large prediction errors (Nassar et al., 2010;Krugel et al., 2009).

Here we consider in detail two of these change-point tasks. The first, an estimation task, requires subjects to predict the next in a series of outcomes (randomly generated numbers) (Nassar et al., 2010). Each outcome is drawn from a normal distribution with a fixed mean and variance. However, the mean of this distribution is occasionally re-sampled, producing abrupt change-points in the series of outcomes. Learning rates can be measured directly on a trial-by-trial basis, using

predictions and outcomes plugged into Eq. 1. Previous work showed that subjects performing this task tended to choose learning rates that were consistent with predictions from a reduced form of a Bayesian ideal-observer algorithm, including a positive relationship between error magnitude and learning rate. However, the details of this relationship varied considerably across individual subjects. Some subjects tended to use highly adaptive learning rates, including values near zero following small errors and values near one following surprisingly large prediction errors. In contrast, other subjects used a much narrower range of learning rates, choosing similar values over most conditions. This across-subject variability was described by a flexible model that could generate behaviors ranging from that of a fixed learning-rate delta rule to that of the reduced Bayesian algorithm, depending on the value of a learning rate "adaptiveness" parameter.

The second task is a four-alternative forced-choice task that includes occasional, unsignaled change-points in the probabilistic associations of monetary rewards for each choice target (Krugel et al., 2009). Learning rates are not measured directly, as they can be for the estimation task, but rather inferred from model fits. Like for the estimation task, previous studies suggested that learning rates adapted to recent outcomes, particularly following large, unexpected errors. These learning-rate dynamics also varied across individual subjects in a manner that, interestingly, was related to allelic variants of the COMT enzyme, which is involved in synaptic clearance of dopamine in the prefrontal cortex.

The existence of this kind of across-subject variability can have dramatic effects on the interpretation of behavioral data fit by models with simpler, fixed learning-rate delta rules. To demonstrate these effects, we simulated performance for both the estimation task and the four-choice task using a broad range of learning-rate adaptiveness levels and fit the simulated data to fixed learning-rate models. In all cases, the simpler, fixed learning-rate model was preferred over a null model constituting random choice behavior even after penalizing for additional complexity (e.g., using BIC or AIC; see Supplemental Materials for details). Despite passing these model-selection criteria, we highlight two misleading conclusions that might be drawn from these fits: biased estimates of learning rates and of exploratory behavior.

The problem of mis-estimating learning rates is depicted in Fig. 1 A&B. Panel A shows simulations based on the estimation task, for which we measured the learning rate directly from the simulated behavioral response on each trial (black circles and error bars reflect median and interquartile range, respectively, across 800 simulated trials). Panel B shows simulations based on the four-choice task, for which we determined the learning rate on each trial based on its value in the internal, generative process used in the simulations. In both cases, increasing the adaptive nature of the learning rate led to learning rates that were increasingly variable, as expected. However, these learning rates also tended to become smaller

in magnitude. This reduction in average magnitude reflected the design of the simulated tasks, which included relatively few change-points that, in the adaptive model, are associated with higher learning rates.

However, the best-fitting values of the learning-rate parameter in a fixed learning-rate model tell a different story (Fig. 1 A &B, gray points). When behavior was simulated using a fixed learning rate (learning-rate adaptiveness = 0), the best-fitting models naturally captured the appropriate value. However, when behavior was simulated using increasingly adaptive learning rates, the fixed learning-rate models returned systematically larger estimates of learning rate than were actually used by the simulated subjects. Thus, learning rate fits from a fixed-learning rate model were not a good measure of the true influence of outcomes on subsequent predictions for a subject that used adaptive learning rates.

The problem of mis-estimating exploratory behavior is depicted in Fig. 1 C&D. In machine learning, the inverse-temperature parameter of a soft-max function is often used to optimize the tradeoff between exploiting actions known to be valuable in the present (emphasized at higher inverse temperatures) and exploring actions that might be valuable in the future (emphasized at lower inverse temperatures) (Sutton and Barto, 1998;Ishii et al., 2002). Similarly, reinforcement-learning models often include an inverse-temperature parameter that is used to characterize the extent to which subjects explore alternative actions, rather than exploiting the one

thought to be most valuable (Daw et al., 2006;Luksys et al., 2009). Accordingly, when we simulated behavior on either the estimation task or the four-choice task using a fixed learning rate and an action-selection process governed by an inverse-temperature parameter, fits from a model with a fixed learning rate and an inverse-temperature process returned appropriate estimates of the inverse temperature used in the generative process (left-most circles in Fig. 1C&D, corresponding to learning-rate adaptiveness=0).

However, when the simulated subjects used increasingly adaptive learning rates, inverse-temperature fits from a fixed learning-rate model substantially overestimated the true variability in action selection (circles in Fig 1 C&D: inferred inverse temperature decreases as learning-rate adaptiveness increases). These biased parameter estimates were not simply a problem with the fixed learning-rate model. Fitting an alternative model that used optimal (maximally adaptive) learning rates (Nassar et al., 2010;Wilson et al., 2010) to the behavior of the same simulated subjects yielded the opposite pattern of results: the model accurately inferred the level of exploratory action selection for simulated subjects that choose learning rates adaptively but overestimated this quantity for subjects that used simpler strategies of less-adaptive, or even fixed, learning rates (squares in Fig 1C: inferred inverse temperature decreases as learning-rate adaptiveness decreases). For both models, these problems were not apparent from standard analyses of best-fitting parameter values, which had similar confidence intervals and covariance estimates

for biased and unbiased fit conditions (see Supplemental Materials for details). These problems also did not simply reflect difficulties in estimating model parameters when the inverse temperature was low and behavior was more random, because the problem was also apparent when the inverse temperature was high. Thus, subtle differences in learning that were not accounted for by the inference model caused underestimation of the inverse-temperature parameter, which might be misinterpreted as increases in exploratory action selection.

Diagnosis of these kinds of problems is difficult, especially when the subtle aspect of behavior that is missing from the model is unknown. Model-selection practices that compare likelihoods of various models (after either cross validation or penalization of parameter numbers) are useful for identifying the better of two or more models. However, these practices require *a priori* knowledge of the models to be tested, and they cannot provide any insight into whether the best of the tested models provides a complete description of behavior. One might be tempted to interpret likelihoods directly and set a criterion for what might be considered to be a "good" model. However, these metrics cannot say whether or not a model is correct. For example, consider a test of the suitability of a fixed learning-rate model for simulated subjects that can vary in terms of learning-rate adaptiveness and exploratory behavior. Similar values of AIC, BIC, and other likelihood-based quantities are obtained for fixed delta-rule models fit to two very different subjects: one who uses a fixed learning rate, which is consistent with the model, and relatively high exploration;

and another who uses a highly adaptive learning rate, which is inconsistent with the model, and relatively low exploration. Interpretation of parameter fits from the latter case would be misleading, whereas parameter fits from the former would be unbiased and informative.

To overcome these limitations, it is sometimes effective to look for indications that a model is failing under specific sets of conditions for which behavior is heavily influenced by the assumptions of the model. For the case of adaptive learning, fixed learning-rate models fail to address adaptive responses to inferred change-points in the action-outcome contingency. Thus, it can be instructive to examine the likelihoods of these models computed for choice data collected shortly after change-points. For the case of the estimation task, a fixed learning-rate model shows an obvious inability to account for data from trials just after a change-point for all but the least adaptive simulated subjects (Fig 2A; dip in log-likelihood at trial 1). However, this approach is not effective for the four-choice task (Fig 2B).

Another potentially useful approach for diagnosing misinterpreted learning-rate adaptiveness is to compute parameter fits using subsets of data according to their timing relative to change-points. For the estimation task, eliminating data from trials immediately following change-points has dramatic effects on fits for both learning rate (Fig 2C) and inverse temperature (Fig 2E). However, this diagnostic approach is far less effective for the four-choice task (Fig. 2 D&F). Thus, for tasks

like the estimation task that provide explicit information about the subject's underlying expectations, the insufficiency of the fixed learning-rate model can be fairly simple to diagnose. However, for tasks like the four-choice task in which information about the subject's expectations is limited to inferences based on less informative choice behavior, parameter biases are still large (Fig. 1B, D) but model insufficiency is far less apparent.

A sobering conclusion that can be drawn from these examples is that even when the parameter fits from a computational model are reasonably likely to produce a dataset, and even when this likelihood is robust to perturbations in the specific trials that are fit or the settings of other parameters in the model, the model might still be missing specific features of the data. Missing even a fairly nuanced feature of the data (such as adaptive learning) can lead the parameters in the model to account for the feature in surprising ways. These unexpected influences can lead to parameter fits that, if interpreted naïvely, might suggest computational relationships that are unrelated to, or even opposite to, the true underlying relationships. Here we use an example from reinforcement learning, but the lessons apply to any model fitting procedure that requires the interpretation of best-fitting parameter values. Certain parameters, like the inverse-temperature parameter in many reinforcement-learning models, seem particularly susceptible to this problem, because they are sensitive to many forms of behavioral variability that might or might not have alternative explanations.

These challenges highlight the narrow wire on which the computational neuroscientist walks. On one hand, we seek to generalize a wide array of physiological and behavioral data from different tasks onto a tractable set of computational principles. On the other hand, the results that we obtain from each experiment are conditioned on assumptions from the particular model through which they are obtained. We believe that the goals of computational neuroscience are possible even in the face of this contradiction. Obtaining generalizable results depends on not only good modeling practices (Daw, 2009) but also the extensive use of model-free analysis to dissect and interpret data from both experiments and simulated model data. For example, the estimation task described above was designed to allow learning rates from individual trials to be computed directly and not inferred via model fits to resulting choice behaviors. This approach revealed clear task-dependent effects on adaptive learning (Nassar et al., 2010). In principle, congruence between such model-free analyses and fit model parameters can help support interpretations of those parameters and has the advantage of testing modeling assumptions and predictions directly rather than via comparisons of different model sets (Ding and Gold, 2012;Walton et al., 2010;Frank et al., 2004). In contrast, inconsistencies between model-free analyses and fit model parameters can help guide how the model can be modified or expanded – keeping in mind, of course, that adding to a model's complexity can improve its overall fit to the data but often

by overfitting to specious features of the data and making it more difficult to interpret the contributions of individual parameters (Ding and Gold, 2012).

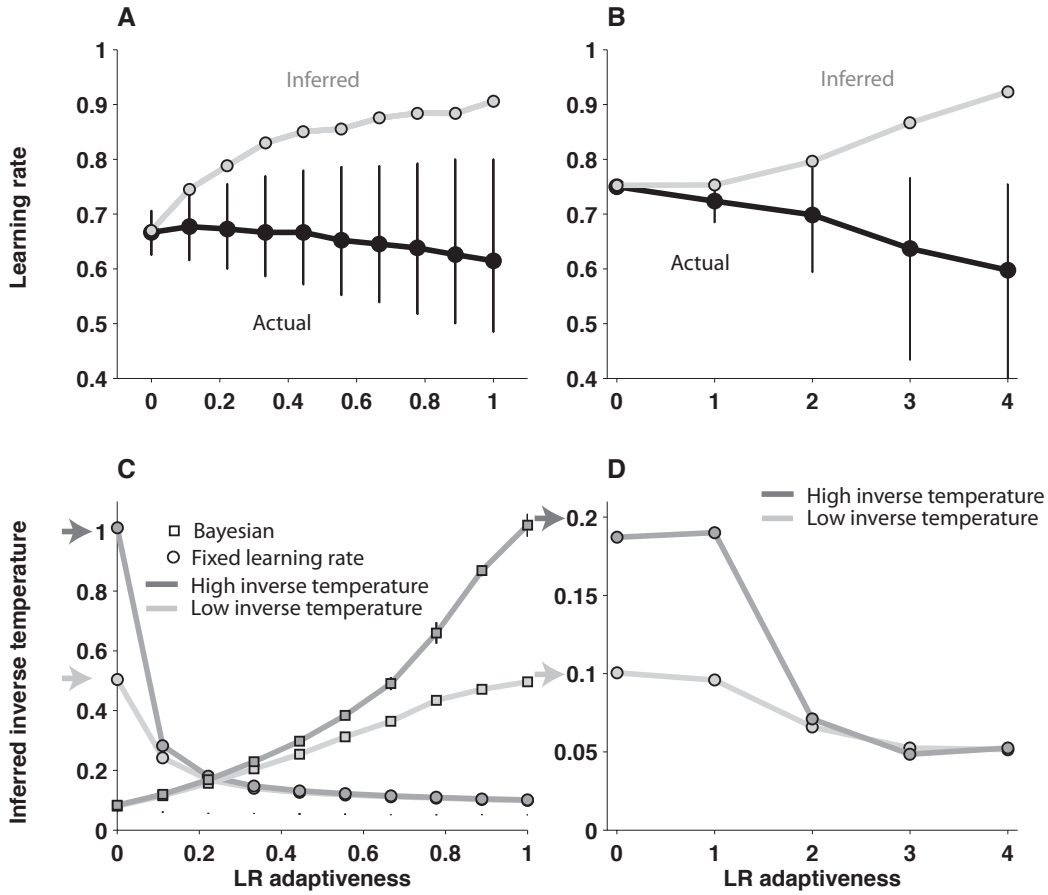


Figure 1. Learning-rate adaptiveness can be misinterpreted as elevated learning rates and decreased inverse temperatures for the estimation (A,C) or four-alternative (B,D) tasks (see text). In all panels, the abscissa represents learning-rate adaptiveness (0 is equivalent to using a fixed learning rate; higher numbers indicate higher adaptiveness to unexpected errors). A & B. Actual (black)

and model-inferred (gray) learning rates used by agents with different levels of learning-rate adaptiveness. Points and error bars represent the median and interquartile range, respectively, of data from six simulated sessions. C & D. Best-fitting values of the inverse-temperature parameter, intended to describe exploratory behavior, inferred using a fixed delta-rule (circles) or approximately Bayesian (squares) model. Shades of gray indicate the level of exploratory behavior of the simulated agent, as indicated. Arrows indicate the actual value of the inverse temperature parameter used in the generative process. Points and error bars (obscured) represent the mean and standard error of the mean, respectively, of data from six simulated sessions.

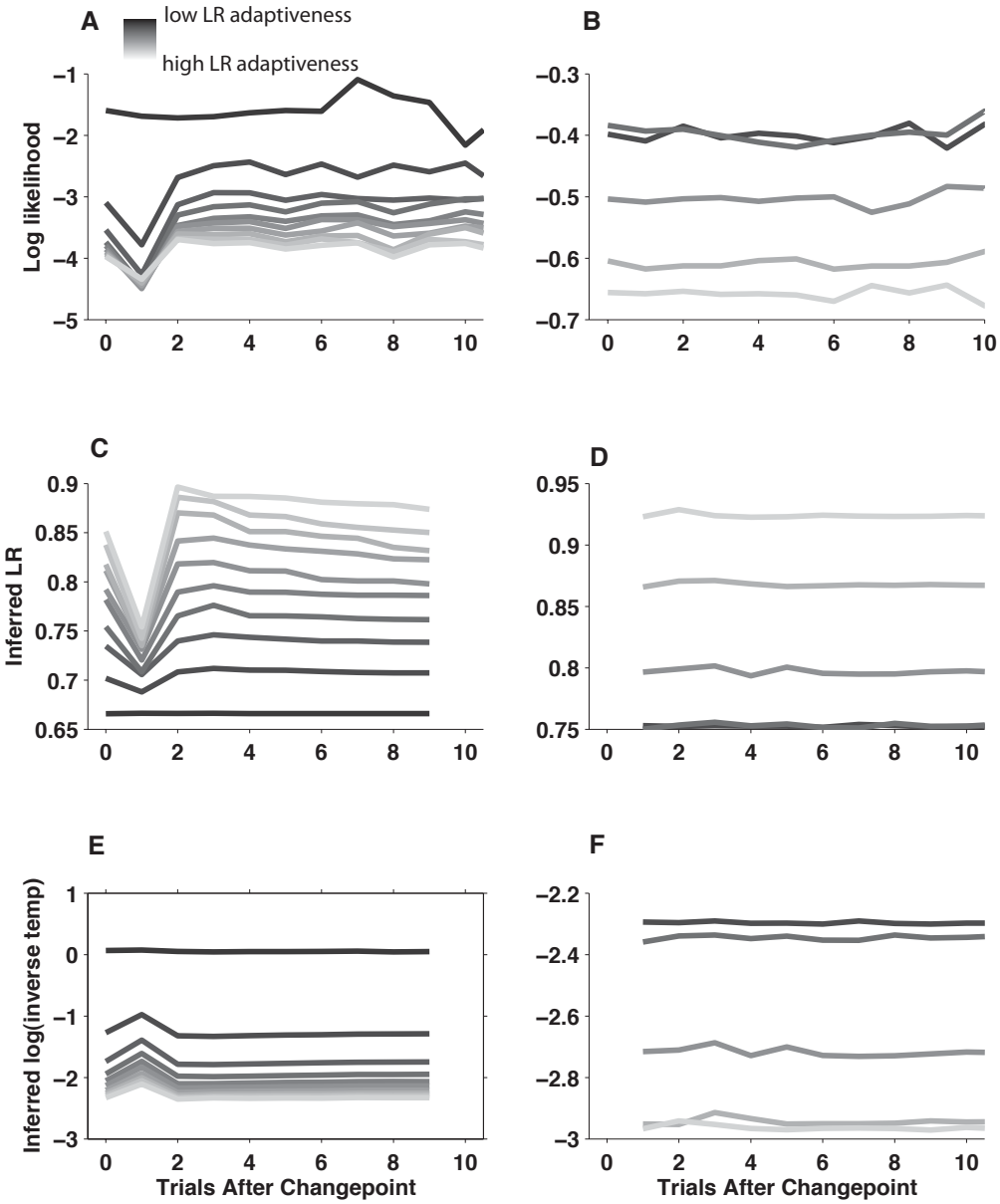


Figure 2. Poor fits from models that ignore learning-rate adaptiveness are easily identified in the estimation, but not the four-choice, task. A & B. Mean log likelihood associated with a fixed learning-rate model, per simulated trial from the estimation (A) or four-choice (B) task, aligned to change-points in the generative process. Lighter shades of gray represent data from simulated agents with higher

levels of learning rate adaptiveness. C–F. Learning rates (C and D) or inverse temperatures (E and F) inferred from model fits that exclude log-likelihood information from trials occurring 0–10 trials after change-points (abscissa) for estimation (C and E) and four choice (D and F) tasks. The transient changes in A, C, and E evident for all but the least adaptive simulated agents reflect the fixed learning-rate model's inability to account for behavior just following change-points on the estimation task; no comparable effects are evident for the four-choice task.

Chapter 4: Supplemental material

Predictive-inference task simulations.

Task design. The subject's task was to predict the value of each subsequent outcome presented in a sequence. Outcomes were generated by rounding picks from a normal distribution with a standard deviation equal to 35 (values between 5 and 40 gave similar results) and a mean that was initiated as a random value picked from a uniform distribution ranging from zero to 300. For each trial, a weighted coin flip determined whether the mean of the distribution would remain the same as on the previous trial (p=0.7, non-change-point trials) or whether the mean would be re-picked from a uniform distribution ranging from zero to 300 (p=0.3, change-point trial). Each sequence consisted of 800 outcomes.

Simulated behavior. Task performance was simulated using the computational model that was best able to describe the range of behaviors of human subjects described previously (Nassar et al., 2010). The model updates beliefs about the mean of the generative distribution after observing each new outcome according to the error made in predicting that outcome:

S. Eq. 1: $$E_{t+1} = E_t + \alpha \times \delta_{t+1}$$

where E is the expected value of the distribution and δ is the difference between the actual outcome and the predicted one (E). The learning rate, α, is adjusted from

trial-to-trial in accordance with estimates of uncertainty and change-point probability with a set of equations derived from the Bayesian ideal observer for the task. These inference equations for this model (see Nassar et al., 2010, for details) include two meta-parameters: hazard rate and LR adaptiveness (previously termed likelihood weight). Hazard rate controls the subjective expectation on the prior probability of a change-point, which in human subjects tends to overshoot the actual value and in our simulations was, accordingly, set to 0.5. LR adaptiveness determines the extent to which unlikely outcomes are used to recognize change-points and in turn adjust learning rates. A LR adaptiveness value of zero is equivalent to a fixed learning rate, whereas a LR adaptiveness value of one is consistent with optimal belief updating. To model the heterogeneity of human subjects in this regard to this parameter, we simulated ten subjects evenly spaced across the allowable range from zero to one.

We simulated behavior using the inference model, described above, in tandem with a probabilistic action-selection process using an inverse-temperature parameter. This process was implemented by computing the probability of choosing each option, p(x), according to a softmax function:

S. Eq. 2: $$p(x) = \frac{e^{\beta \cdot V_x}}{\sum_{n=0}^{300} e^{\beta \cdot V_n}}$$

where Vx is inversely proportional to the distance between the potential prediction (x) and the estimate derived from the inference model described above (Et), and β is

the inverse temperature, which determines the variability in action selection and has previously been used as an indicator of exploratory behavior. Here we used inverse temperatures ranging from 0.2 to 1. For each set of parameters, the simulated subjects completed five task sessions.

Model fitting. We fit simulated behavior from the predictive-inference task using a fixed learning-rate model. This model updated beliefs according to S. Eq. 1, albeit with a fixed learning rate (α) for all trials from a given session. This model also used the same action-selection mechanism described above (S. Eq. 2). This model was fit to simulated behavior with learning rate and inverse temperature as free parameters, using a constrained search algorithm (fmincon implemented in Matlab) to minimize the negative log likelihood of the model relative to the simulated behavioral data.

Four-alternative forced-choice simulations.

Task design. The four-alternative forced-choice task simulated here was previously developed and used by Krugel and colleagues (Krugel et al., 2009). Subjects were asked to choose between four possible alternatives according to perceived value. After choosing an alternative, the subject was shown the value of the outcome associated with that choice. There were two possible outcome values: a high value (250 pts) and a low value (50 pts). For each trial, one (best) alternative is the most likely to yield a large reward. To maximize our ability to achieve reliable model fits, we simulated sessions with 10,000 trials in which outcomes were assigned to each

possible choice by the following process:

1) A weighted coin flip determined whether the best target would remain in the same location as on the previous trial (the probability of a change = 0.1).

If Change: a new best target is sampled at random from all targets.

Otherwise: the best target remains in the same position as previously.

2) Outcome values were chosen at random (from the two possible values) for each alternative with p(high value) = 0.8 for the best alternative and 0.2 for all other alternatives.

Simulated subject behavior. Behavior was simulated according to the adaptive learning-rate model used by Krugel and colleagues that was best capable of describing the behavior of human subjects (2). In brief, choices were selected according to a softmax action-selection rule that depended on a value function (q) and an inverse temperature term (3):

S. Eq. 3

$$p(x) = \frac{e^{\beta \cdot E_x t}}{\sum_{i=1}^{4} e^{\beta \cdot E_i t}}$$

After each trial, the value of the chosen option for the current timestep (EI,t) was updated according to the reward prediction error on that trial:

S. Eq. 4

$$E_{i,t+1} = E_{i,t} + \alpha_t \delta_t$$

where δ reflects the difference between the actual outcome value, and α is the learning rate. The learning rate was adjusted on each trial according to the slope of the recent, unsigned prediction errors (m):

S. Eq. 5

$$if \quad m > 0 \qquad \alpha_t = \alpha_{t-1} + f(m_t) \cdot (\alpha_{t-1})$$

$$if \quad m < 0 \qquad \alpha_t = \alpha_{t-1} + f(m_t) \cdot (1 - \alpha_{t-1})$$

Thus, learning rate increased if the absolute value of the most recent prediction errors was large but decreased if the absolute value of recent prediction errors was small. The form of f(m) was a double-sigmoid transfer function:

S. Eq. 6

$$sign(m) \cdot (1 - e^{-(m\lambda)^2})$$

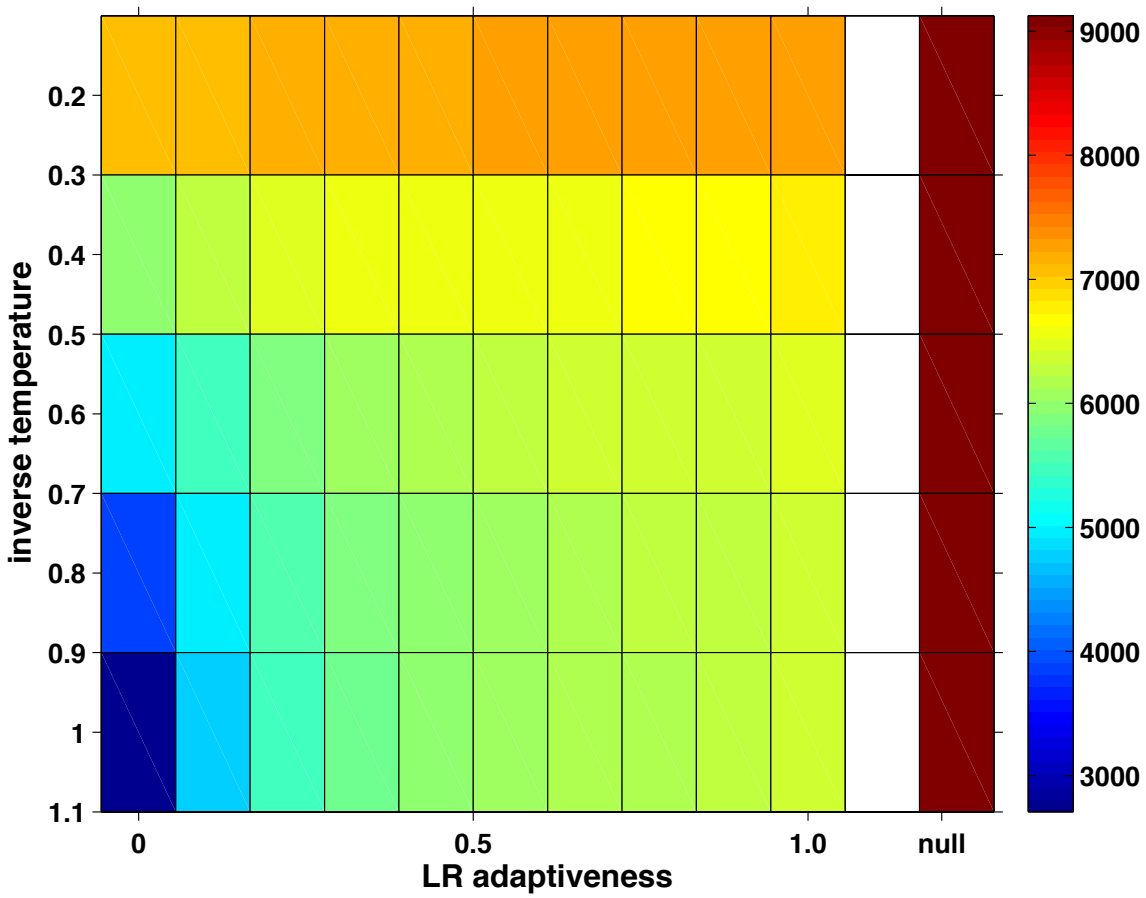
where the learning rate adaptiveness parameter, λ, determines the extent to which learning rates are altered according to recent absolute errors. When λ is equal to zero, learning rates become stable and thus maintain their initial value (which in our simulations was set to zero). When λ takes larger values, the learning rate becomes increasingly dependent on the slope of recent absolute prediction errors.

Model fitting. We fit simulated behavior with a model that included the same action-selection (S. Eq. 3) and learning mechanisms (S. Eq. 4) described above but used a fixed learning rate for all trials (instead of S. Eqs. 5 and 6). Thus, the model had two free parameters (learning rate and inverse temperature), which were fit

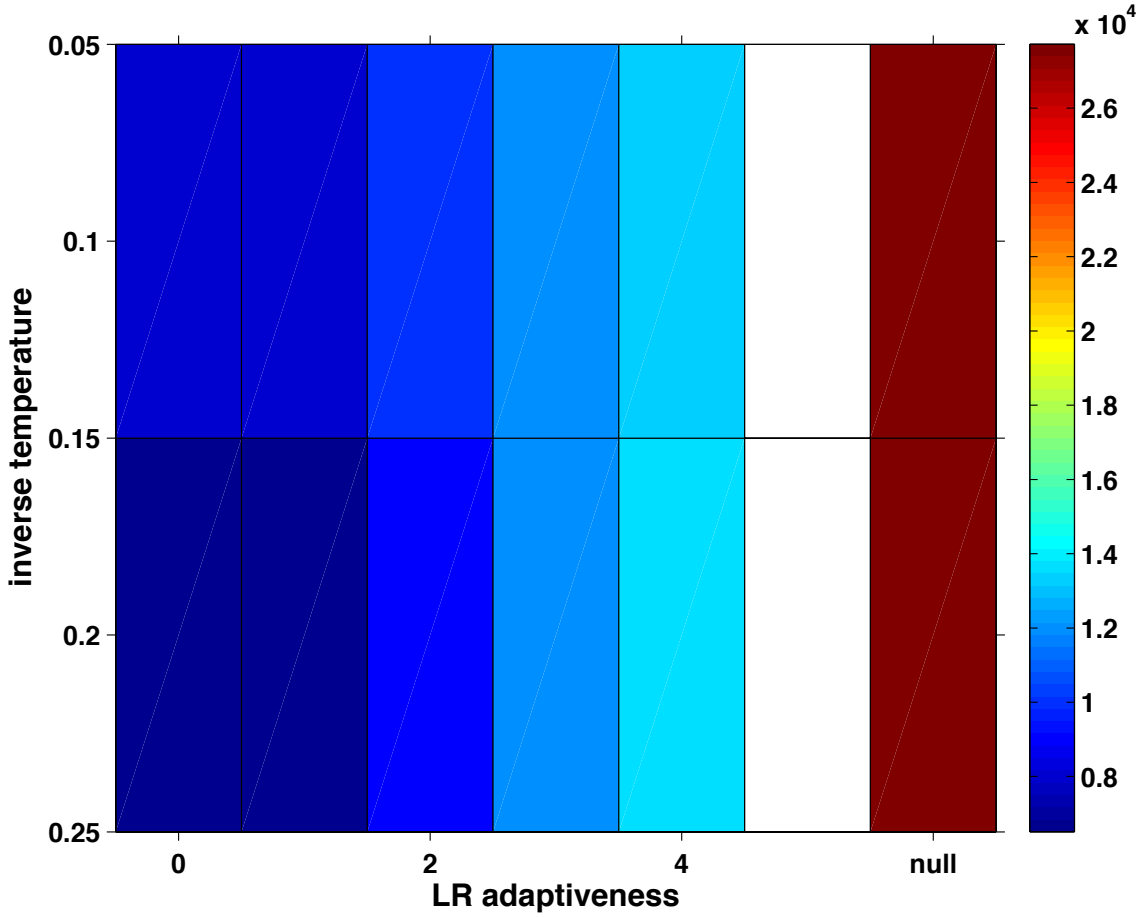
simultaneously to simulated behavioral data by minimizing negative log likelihood using the Matlab function fmincon.

Standard model selection tests.

All model fits described in the main text were better descriptors of behavior than null models that reflected random choice behavior for the respective tasks, as measured by BIC or AIC (BIC values are shown in S. Figs. 1 and 2). That is, even the most ill-suited models (e.g., fixed learning-rate models fit to adaptive learning behavior) would not be rejected on the basis of appropriate likelihood-based tests.



S. Fig 1: BIC values for all models fit to simulated predictive inference data. BIC values are represented in color (see legend to the right: hotter colors reflect higher BIC values; that is, worse fits) for each simulated value of inverse temperature and LR adaptiveness when fit by a fixed learning-rate model. For comparison, the BIC of a null model that reflects random choice activity is included (column on far right).



S. Fig 2: BIC values for all models fit to simulated four-choice task data. BIC values are represented in color (see legend to the right: hotter colors reflect higher BIC values; that is, worse fits) for each simulated value of inverse temperature and LR adaptiveness when fit by a fixed learning-rate rate model. For comparison, the BIC of a null model that reflects random choice activity is included (column on far right).

CHAPTER 5

Ongoing work and future directions

In chapters 2-4 I show that human subjects conform to the basic tenants of a normative model for making predictions in a dynamic environment, that this behavior can be simulated using a slight extension to a biologically inspired delta-rule model, and that the key variables of this model seem to be represented in the pupil-linked arousal system and driving belief-updating behavior, and that the subtle extensions of this model have substantial impact on standard model fitting procedures. In this final chapter I will discuss the significance of specific results from these chapters, identify some important open questions, and describe some additional experiments that I have embarked on to answer these questions.

Mechanism of adaptive learning rate. The pupillometric studies described in chapter 3 revealed that pupil-linked arousal systems reflected the two variables necessary for computing learning rate according to a modified delta rule capable of near-optimal inference in dynamic environments. The sound manipulation experiment highlighted the behavioral importance of this signal by demonstrating that an unexpected auditory stimulus capable of causing a change in pupil diameter also led to systematic changes in learning rate. This relationship was explained in terms of known and theorized effects of noradrenaline released from the locus

coeruleus (LC), the levels of which are thought to covary with baseline arousal measures including pupil diameter (Aston-Jones and Cohen, 2005).

One intriguing question stemming from this work is how, exactly, a global neuromodulatory signal such as the one that LC might affect the extent to which new observations are incorporated into an updated prediction about the world. One interesting possibility is that the change in learning rate reflects a boost in the extent to which sensory information propagates through cortico-thalamic circuitry toward association cortex where abstract beliefs are represented. This possibility is supported by several neurophysiological studies that show enhanced throughput of sensory information relative to noise during noradrenergic modulation (Waterhouse et al., 1998;Hurley et al., 2004;Devilbiss and Waterhouse, 2000;Devilbiss and Waterhouse, 2004). One possible mechanism through which signal amplification might be achieved is a change in the gain of the non-linear activation function of sensory neurons (Servan-Schreiber et al., 1990).

This type of radical change to the input-output function controlling activity of neurons in sensory cortex should lead to a drastic shift in the flow of information through the brain. One strong prediction made by this model is that fMRI BOLD signals in sensory cortex should covary more with those in prefrontal regions when NE levels are high. As a collaborative project related to my thesis work, I have worked with Dr. Joe McGuire and Dr. Joe Kable to conduct an experiment in which

16 subjects completed the a variant of the predictive inference task described in chapters 2-4 in an fMRI scanner. Our initial analyses have focused on identifying specific regions that have enhanced BOLD responses to various conditions that tend to drive learning and have identified an area of interest in the dorsal cingulate cortex. However, our future plans include examining whether correlations between BOLD responses in the occipital and prefrontal cortices depend on subject learning rate (and by extension LC activity).

Another potential mechanism for the arousal-induced changes in learning is amplification of feedback signals mediating behavioral updating. This behavioral updating signal is thought to take the form of a reward prediction error. Several areas of the brain including anterior cingulate cortex (ACC), the habenula, and most famously the ascending dopaminergic system have been shown to contain cells the fire in proportion to reward prediction errors (Matsumoto and Hikosaka, 2007;Matsumoto et al., 2007;Schultz et al., 1997). Since fMRI work from our collaboration as well as others indicated a relationship between fMRI BOLD activity in ACC and learning rate (Behrens et al., 2007;Krugel et al., 2009), I designed an experiment to look directly at feedback signals in ACC of rhesus macaques in a task where optimal behavior requires adaptive learning.

The task prompts the monkey to choose one of ten possible targets. The correct target is then revealed to the subject, and after a delay before either receiving a juice

reward (if he chose the rewarded target) or beginning the next trial (otherwise). The process by which rewarded targets is determined contains both noise (in the form of a spatial probability distribution across all targets) and change-points (as the best target is re-picked on a small proportion of trials).

Adaptive learning can be measured by analyzing switch behavior as a function of change-point probability, which can be inferred through the spatial distance between chosen and rewarded targets, and uncertainty, which is related to the number of trials since the last change-point. Like human subjects, both of the monkeys trained on this task display adaptive learning that is greatest after surprising outcomes or shortly after a change in outcome contingency. Our preliminary recordings do not demonstrate an increase in overall firing of ACC neurons during high learning trials, which could be one simple interpretation of the BOLD response. Rather, there seems to be a trend toward enhanced signaling of outcome (ie error or correct) in single units in ACC on trials where learning rate was high. In principle such a signal enhancement could give rise to enhanced updating on these trials, however this dataset is still preliminary and confirmation of this idea will require more neural recordings, which will be completed by Yin Li, a neuroscience graduate student in the Gold Lab over the coming year.

Origins of individual differences. A striking feature of the behavioral and pupillometric data reported in chapters 2 and 3 is the incredible physiologic and

behavioral diversity across subjects. The finding that individual differences in learning were related to individual differences in pupil response suggests the possibility that these differences might depend on baseline neuromodulatory state. One strategy for testing this possibility is to identify groups differing in underlying neuromodulatory state and determine whether these groups differ in behavior on the predictive inference task.

I have taken this approach in two collaborative projects that relate directly to my thesis work. The first such project relies on the differences in dopamine signaling in old and young adults (Li et al., 2001). Although I postulate that noradrenergic signaling is mediating the enhanced learning after change-points, dopamine and noradrenaline share many antecedent conditions and can serve redundant roles in some forms of learning (Ouyang et al., 2012). To examine whether age-related differences in learning behavior I have embarked on a collaborative project with Dr. Ben Eppinger and Dr. Shu-Chen Li at the Max Planck Institute for human development in Berlin. The study will include 60 young and 60 old subjects that will be genotyped with respect to the DAT1 and DRP32 polymorphisms, which affect functional dopamine signaling. Although genotyping is not yet complete, the behavioral data from an initial cohort of 30 young and old subjects show a modest group difference with older subjects using significantly reduced learning rates (see figure 5.1).

A second collaborative project is underway examining whether schizophrenic patients, who have increased D2 dopamine signaling but diminished D1 dopamine signaling, differ from I.Q. matched controls in behavior on a variant of the predictive inference task. Data is being collected Dr. David Leitman and Dr. Bruce Turetzky in the department of Psychiatry at the University of Pennsylvania. Although initial data from schizophrenic subjects also suggest a decrease in learning rate in this group, in principle differences in more subtle aspects of predictive inference behavior (ie. relative uncertainty and hazard rate best describing subject behavior) between the schizophrenic and aged groups might provide insight into distinct roles that different receptor subtypes might have in setting baseline learning behavior.

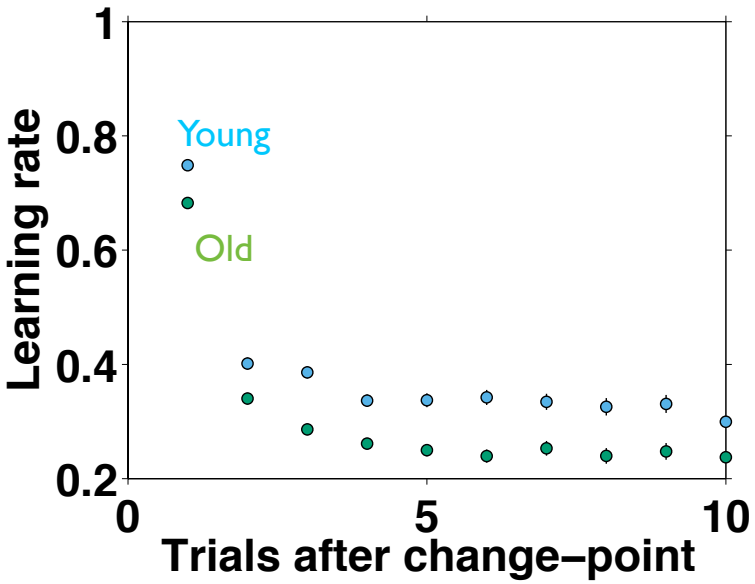


Figure 5.1 Age differences in learning rate. Average learning rate for younger (age 20-30 years, blue) and older (age 60-70 years, green) adults as a function of trials after a change-point in a variant of the predictive inference task described in chapter 2.

Change-points in perception. Predictions about future outcomes are not only useful for guiding behavior, but also for interpreting ambiguous sensory stimuli. Previous work suggests that humans and animals combine sensory information with expectations about the probability of potential stimuli in a roughly Bayesian fashion (Knill and Pouget, 2004). This combination process allows smaller perceptual errors on average in a stable regime, but also biases all perceptual estimates toward the expected stimuli. One potential issue with biasing perceptions according to expectations is that under many circumstances contexts can change leading to expectations that are violated. The use of such violated expectations would decrease perceptual accuracy and thus a system designed to minimize perceptual errors in a dynamic environment should immediately reject prior expectations that are inconsistent with incoming sensory information.

In order to test this idea directly I designed an auditory localization task that built on the main features of the predictive inference task described in chapter 2. The task was instantiated by Shilpa Sarode and Kamesh Krishnamurthy and used to collect preliminary data that was recently presented at the Society for Neuroscience conference. Human subjects were instructed to predict (stimulus expectation) and then indicate (stimulus perception) on each trial the virtual source location of a binaurally presented noise burst, filtered using a standard head-related transfer function associated with different frontal, azimuthal source locations. The locations were drawn independently for each trial from a normal distribution (the "source

context"), but the mean of this distribution was re-picked on a random subset of trials according to a change-point process. After each trial, the subject was shown a visual representation of the true stimulus location.

We characterized the influence of prior expectations on perception in terms of the relationship between prediction errors (stimulus expectation) and perceptual errors (stimulus perception) measured on each trial. Consistent with optimal inference in a dynamic environment, the influence of prior expectations on perception was smallest just after a change-point, even on the first stimulus from the new distribution. The influence of prior expectations increased gradually as subjects encountered more stimuli from the new distribution and the expectations became more reliable. The results suggest that the brain can rapidly calibrate the relative influence of prior expectations and incoming sensory information according to ongoing assessments of their reliability to guide perception. Thus it appears that the dynamic re-weighting of prior information occurs on a perceptual timescale much faster then that necessary for the behavior described in chapters 2 and 3.

CHAPTER 6

Conclusions: generality of findings

Previous chapters developed the notion of influence in learning, demonstrated the computational factors determining influence, and mapped these factors onto measurements of a pupil-linked arousal network. Although the notion of influence is quite general, our examinations of computational and physiological mechanisms of influence were made specifically in a predictive inference task developed explicitly for that purpose. In this section I will bridge the findings from previous chapters to a broader understanding of mechanisms of learning in the brain.

Although there are many distinct forms of learning that can be measured with specific tasks there is some evidence that these disparate types of learning might engage a few separate but interacting learning systems. One such system is a network including the ascending dopaminergic system (ventral tagmental area and substantia nigra pars compacta) as well as dorsal and ventral striatum. This network is thought to implement an actor/critic form of reinforcement learning that uses state representations supplied by prefrontal cortices to inform expectations of action values in striatum. Action values in dorsal striatum are used to select actions, whereas value representations in ventral striatum serve to supply expectations that are combined with sensory feedback in the ascending dopaminergic system in order to provide a reward prediction error signal used to train the state-action mappings

stored in connection weights in striatum. When positive prediction errors are signaled (through enhanced dopamine release in the striatum) synaptic weights of neurons mapping recently encountered states onto recently chosen actions are enhanced (Takahashi et al., 2008). This allows the network to learn to choose actions when they are valuable without having an explicit model of the how an action in one state might map onto the next, and thus is often called "model-free" (Daw et al., 2011).

This model free learning network shares features with the delta rule model employed in chapters 2-4 for updating inference based on outcomes. Both models employ the use of prediction errors to instruct learning. Both models contain a learning rate term that essentially controls the influence of new prediction errors on expectations maintained either as an abstract belief (in the reduced-Bayesian model) or as a striatal connectivity matrix determining state-action mapping. Amplification of the learning rate in the physiological model-free learning network could be accomplished by amplifying phasic dopamine signals that encode reward prediction errors. Although it is currently unknown whether these signals are modulated by the computational factors that govern learning rate, projections of locus coeruleus to dopaminergic nuclei and the existence of a sub-population of dopaminergic neurons that encodes salience rather than reward prediction error may provide a means for incorporating arousal encoded learning computations into

the reinforcement signal (Mejías-Aponte et al., 2009;Matsumoto and Hikosaka, 2009) .

An obvious difference between the two systems is that the reduced Bayesian model represents beliefs and prediction errors on an abstract space, whereas the striatal/dopaminergic system seems to explicitly represent values and reward prediction errors in a valence space where positive values and reward prediction errors are represented by higher firing rates of striatal and dopaminergic neurons respectively. This allows the output of the network, in terms of the firing rates of striatal neurons, to provide a signal proportional to the probability with which a particular action should be chosen (Takahashi et al., 2008). Although such a network can efficiently incorporate reward information to reinforce chosen actions, it is not clear how such a network would represent the type of outcome information provided in our predictive inference task. The outcomes in the predictive inference task specify the action that would have provided the most reward, so it is possible that the same network could incorporate this information by simulating the action that would have provided the most reward and then reinforcing the connectivity matrix through a fictive reward prediction error signal.

While it is unknown whether the striatum has access to such fictive learning signals, such signals have been shown to exist in anterior cingulate cortex (ACC) an area of prefrontal cortex that is heavily innervated by dopaminergic nuclei (Briand et al.,

2007;Kennerley et al., 2011) . It could be that these signals measured in ACC are reflecting a more global neuromodulatory signal broadcast by DA neurons, in which case the striatum might have access to the same signal, allowing it to effectively learn about both chosen and un-chosen options. However, it is also possible that anterior cingulate cortex, which is known to support a number of the necessary computations for predictive inference in our task (prediction errors, learning rates) and be highly involved in many forms of behavioral updating might perform model-free inference directly (Kennerley et al., 2011;Behrens et al., 2007).

Regardless of where these algorithms are implemented, one general concern raised by the sound manipulation experiment (chapter 3) is to the specificity with which learning rates can be selectively modified. For example, the brain might be simultaneously maintaining and updating beliefs about several variables, say the quality of a certain restaurant and the safety of a certain neighborhood. When the brain obtains surprising data in one of these domains (say a terrible meal at the restaurant) it seems at first glance that the brain should reset its beliefs in that domain (amplify learning about the restaurant) while maintaining beliefs in the other (stable beliefs about neighborhood). However, finding that a surprising sound could alter the influence of numerical outcomes on updated beliefs suggests that the brain does not completely compartmentalize adjustments in learning to a particular stimulus-relevant domain. One possible explanation for this effect relies on the underlying structure of change-points encountered in the world. If change-points

tend to be correlated over dimensions, then observing a surprising stimulus in one dimension should, in fact, prescribe rapid learning in the other dimensions. For example, sudden economic hardship might lead a neighborhood to become unsafe and a restaurant owner to tend toward lower quality ingredients. Thus observation of a bad hamburger indicates the possibility of economic decline, which in turn leads to uncertainty about he safety of the neighborhood and in turn rapid learning about that variable. It is unknown to what extent real-world change-points might have this sort of correlation structure or to what extent correlations in learning rate across dimensions match real-world statistics, however work addressing these questions will be critical to understanding the true optimality of arousal induced modulation of learning rate.

Hard-wired assumptions about latent structure of change-points incorporated in the arousal driven learning system might account for some implicit expectations about the latent structure of the world, however it is clear that the brain also learns such latent structures explicitly through experience and incorporates this knowledge into inferences about the world (Daw et al., 2011). This type of learning is referred to as model-based, as it requires building an explicit probabilistic model of how various states map onto one another. In contrast to model-free learning, which is thought to take place in the striatum, model-based learning is thought to occur largely in prefrontal regions including dorsal lateral prefrontal cortex (Gläscher et al., 2010). Although some aspects of the reduced Bayesian model rely on model free

(prediction error) signals, other aspects require an understanding of how states evolve over time, or the probabilistic structure of the generative environment. In particular, change-point probability calculations are based on the probabilistic mapping of a latent variable (mean of distribution) to an observable one (actual outcome). This mapping is likely learned over time; subjects used more adaptive learning rates, as well as hazard rates better matching the experimental conditions, under conditions where they had more training (compare performance in chapter 3 to that in chapter 2). One mechanistic explanation for this might be that model-based learning is used to develop finely tuned probabilistic expectations, which are in turn used to calculate learning rates that are broadcast through the noradrenergic system and then used to amplify learning signals in a model-free learning network. Although interactions between model-free and model-based learning systems have been observed in the striatum, it is unclear to what extent these interactions depend on arousal systems or reflect the optimization process described above (Daw et al., 2011).

Concluding remarks

The brain is exquisitely evolved to collect sensory information and use it to inform future actions. However, in a dynamic and stochastic world, a single sensory snapshot does not provide perfect information regarding the best possible future action, and a strategy for combining snapshots over time is required. The best

strategy for incorporating new sensory information in a changing world requires dynamically adjusting the influence of new snapshots according to the predictive quality of older ones. Here I have shown that human subjects conform to this strategy when assigning influence to abstract information in a predictive inference task and that such behavior could be achieved with a simple model-free reinforcement-learning rule, albeit with some model-based assessments of stimulus probabilities. Interestingly, under stable conditions this near-optimal model prescribes becoming relatively insensitive to new sensory information. At first glance this prediction seems surprising; why would the brain spend so much energy maximizing the informational content in each sensory snapshot only to ignore them?

The answer is that even perfect sensors are only as informative as the external environment. The informational content of an observation can be defined as the negative log probability of that observation, such that improbable events are highly informative and completely predictable ones are uninformative. Since outcomes become predictable during a stable contingency they also become less informative. The extent to which the information content drops off during a stable period depends critically on exactly what type of information is measured: while information about the present is always provided by observations, information about the future approaches zero after several observations in a stable regime (see figure 6.1). Through this lens optimal inference can be seen as appropriately gating

sensory experience according to its relevant informational content, where relevance explicitly requires pertaining to future events.

Arousal has long been thought to play a role in controlling the flow of sensory information and arousal systems including locus coeruleus are most responsive to improbable (ie. highly informative) stimuli (Pfaff, 2006;Aston-Jones et al., 1994). Through diffuse projections locus coeruleus has the capability of influencing stimulus representations across modalities and at different levels of abstraction. This dissertation demonstrates that arousal systems play a role in controlling the influence of abstract sensory observations on higher order beliefs according to the relevant information provided by those observations. This work not only bolsters a burgeoning view of generalized brain arousal as physical implementation of information based sensory gating (Pfaff, 2006), but also addresses the larger question as to why low arousal states exist in the first place. Where previous work has discounted decrements in arousal as lapses in a fallible attention system, the findings developed here suggest a normative explanation: stable variability in our environment can allow unexpected stimuli to contain no information relevant to future decisions. By reducing sensory flow under such conditions, we minimize the extent to which we are misled by distracting and uninformative stimuli. Thus decrements in arousal provide a means for the brain to resist learning from unpredicted but uninformative stimuli.

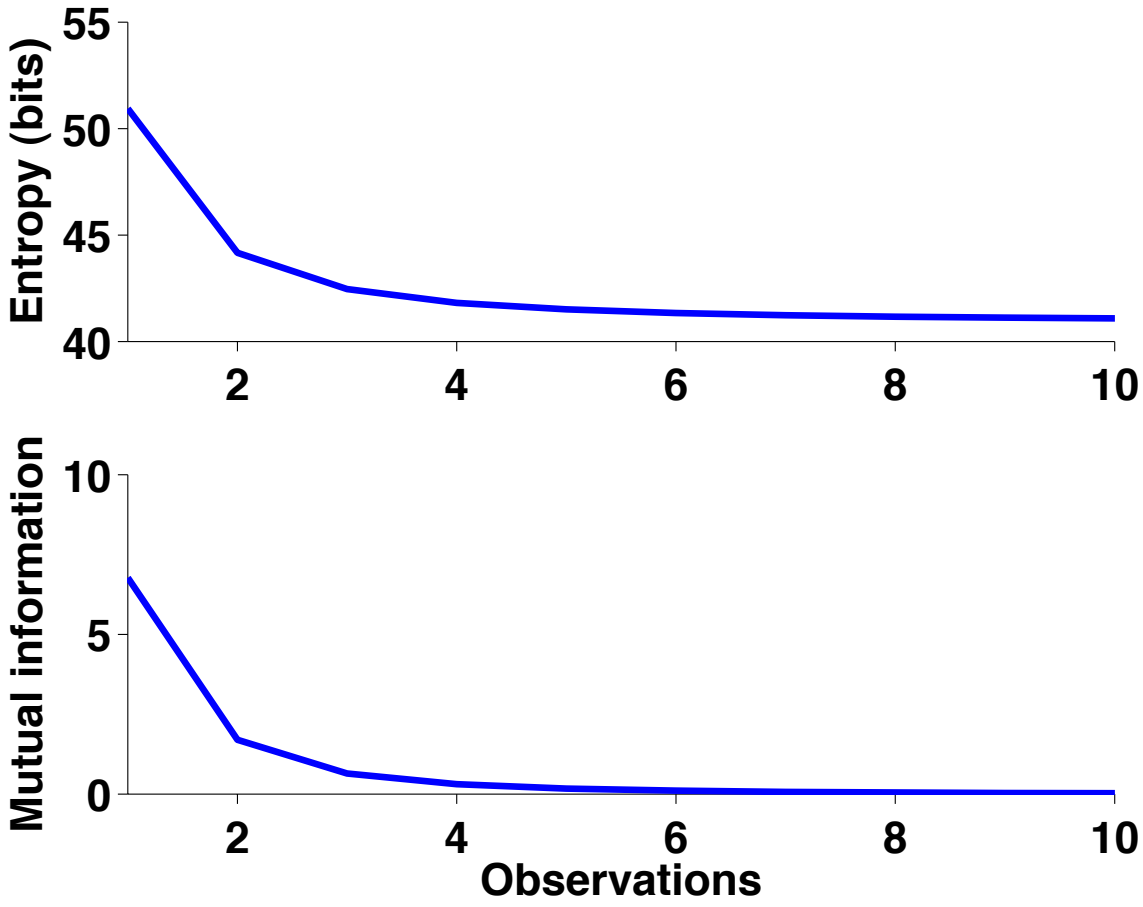


Figure 6.1 Decay in information from data generated by a noisy process. A) Entropy (expected information, in bits) computed as the expectation of negative log probability for each outcome where probabilities are computed by discretizing the predictive distribution from the optimal inference algorithm at each time-step. B) Mutual information (in bits) contained in subsequent observations from the same noisy process. Mutual information is computed as the entropy over a given observation (as above) minus the entropy over the next observation. Mutual information between two subsequent observations can be thought of the amount of information in an observation that pertains to the next (future) observation.

Reference

Adams RP, MacKay DJC (2007) Bayesian Online Changepoint Detection. University of Cambridge Technical Report.

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28:403-450.

Aston-Jones G, Ennis M, Pieribone VA, Nickell WT, Shipley MT (1986) The brain nucleus locus coeruleus: restricted afferent control of a broad efferent network. Science 234:734-737.

Aston-Jones G, Rajkowski J, Kubiak P (1997) Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. Neuroscience 80:697-715.

Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T (1994) Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. J Neurosci 14:4467-4480.

Beeler JA, Daw N, Frazier CR, Zhuang X (2010) Tonic dopamine modulates exploitation of reward learning. Front Behav Neurosci 4:170.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214-1221.

Bertsekas DP, Tsitsiklis JN (1996) Neuro-dynamic programming. Belmot, MA: Athena Scientific.

Bouret S, Sara SJ (2004) Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. Eur J Neurosci 20:791-802.

Bouret S, Sara SJ (2005) Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. Trends Neurosci 28:574-582.

Briand LA, Gritton H, Howe WM, Young DA, Sarter M (2007) Modulators in concert for cognition: modulator interactions in the prefrontal cortex. Prog Neurobiol 83:69-91.

Bruder GE, Keilp JG, Xu H, Shikhman M, Schori E, Gorman JM, Gilliam TC (2005) Catechol-O-methyltransferase (COMT) genotypes and working memory:

associations with differing cognitive operations. Biol Psychiatry 58:901-907.

Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. Neuron 58:306-324.

Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-Nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84:581-617.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends Cogn Sci 10:294-300.

Critchley HD (2005) Neural mechanisms of autonomic, affective, and cognitive integration. J Comp Neurol 493:154-166.

Critchley HD, Mathias CJ, Dolan RJ (2001) Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron 29:537-545.

Dalley JW, McGaughy J, O'Connell MT, Cardinal RN, Levita L, Robbins TW (2001) Distinct changes in cortical acetylcholine and noradrenaline efflux during contingent and noncontingent performance of a visual attentional task. J Neurosci 21:4908-4914.

Daw, N. (2011) Trial-by-trial data analysis using computational models. In *Attention and performance* XXIII. (Eds. Phelps, E., Robbins, T., Delgrado, M.) Oxford University Press, Oxford.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204-1215.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cognitive, Affective, & Behavioral Neuroscience 8:429-453.

Dayan P, Kakade S, Montague PR (2000) Learning and selective attention. Nat Neurosci 3 Suppl:1218-1223.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25:11730-11737.

Devauges V, Sara SJ (1990) Activation of the noradrenergic system facilitates an

attentional shift in the rat. Behav Brain Res 39:19-28.

Devilbiss DM, Waterhouse BD (2000) Norepinephrine exhibits two distinct profiles of action on sensory cortical neuron responses to excitatory synaptic stimuli. Synapse 37:273-282.

Devilbiss DM, Waterhouse BD (2004) The effects of tonic locus ceruleus output on sensory-evoked responses of ventral posterior medial thalamic and barrel field cortical neurons in the awake rat. J Neurosci 24:10773-10785.

Ding L, Gold JI (2012) Separate, Causal Roles of the Caudate in Saccadic Choice and Execution in a Perceptual Decision Task. Neuron 75:865-874.

Doll BB, Hutchison KE, Frank MJ (2011) Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J Neurosci 31:6188-6198.

Einhäuser W, Stout J, Koch C, Carter O (2008) Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proc Natl Acad Sci U S A 105:1704-1709.

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

Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat Neurosci 12:1062-1068.

Frank MJ, Seeberger LC, O'Reilly R C (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940-1943.

Friedman D, Hakerem G, Sutton S, Fleiss JL (1973) Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. Electroencephalogr Clin Neurophysiol 34:475-484.

Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD (2010) Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci 10:252-269.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585-595.

Granholm E, Steinhauer SR (2004) Pupillometric measures of cognitive and emotional processes. Int J Psychophysiol 52:1-6.

Hakerem G, Sutton S, Zubin J (1964) Pupillary reactions to light in schizophrenic patients and normals. Ann NY Acad Sci 105:820-82.

Harley CW (1987) A role for norepinephrine in arousal, emotion and learning?: limbic modulation by norepinephrine and the Kety hypothesis. Prog Neuropsychopharmacol Biol Psychiatry 11:419-458.

Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. Science 324:948-950.

Hurley LM, Devilbiss DM, Waterhouse BD (2004) A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. Curr Opin Neurobiol 14:488-495.

Huys QJ, Moutoussis M, Williams J (2011) Are computational models of any use to psychiatry? Neural Netw 24:544-551.

Huys QJM, Vogelstein J, Dayan P (2009) Psychiatry: insights into depression through normative decision-making models. Advances in neural information processing systems 21:729-736.

Ishii S, Yoshida W, Yoshimoto J (2002) Control of exploitation-exploration meta-parameter in reinforcement learning. Neural Netw 15:665-687.

Jepma M, Nieuwenhuis S (2010) Pupil Diameter Predicts Changes in the Exploration-Exploitation Tradeoff: Evidence for the Adaptive Gain Theory. J Cogn Neurosci.

Jepma M, Te Beek ET, Wagenmakers EJ, van Gerven JM, Nieuwenhuis S (2010) The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. Front Hum Neurosci 4:170.

Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science 154:1583-1585.

Kennerley SW, Behrens TE, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581-1589.

Kennerley SW, Dahmubed AF, Lara AH, Wallis JD (2009) Neurons in the frontal lobe

encode the value of multiple decision variables. J Cogn Neurosci 21:1162-1178.

Kennerley SW, Wallis JD (2009) Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci 29:2061-2073.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940-947.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27:712-719.

Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci U S A 106:17951-17956.

Krugman HE (1964) Some applications of pupil measurement. J Marketing Res 1:15-19.

Lapiz MD, Morilak DA (2006) Noradrenergic modulation of cognitive function in rat medial prefrontal cortex as measured by attentional set shifting capability. Neuroscience 137:1039-1049.

Li SC, Lindenberger U, Sikström S (2001) Aging cognition: from neuromodulation to representation. Trends Cogn Sci 5:479-486.

Luksys G, Gerstner W, Sandi C (2009) Stress, genotype and norepinephrine in the prediction of mouse behavior using reinforcement learning. Nat Neurosci 12:1180-1186.

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154-162.

Marr D (1982) Vision: A computational approach. San Francisco: Freeman & Co.

Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A bayesian foundation for individual learning under uncertainty. Front Hum Neurosci 5:39.

Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447:1111-1115.

Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:837-841.

Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647-656.

McGaughy J, Ross RS, Eichenbaum H (2008) Noradrenergic, but not cholinergic, deafferentation of prefrontal cortex impairs attentional set-shifting. Neuroscience 153:63-71.

Mejías-Aponte CA, Drouin C, Aston-Jones G (2009) Adrenergic and noradrenergic innervation of the midbrain ventral tegmental area and retrorubral field: prominent inputs from medullary homeostatic centers. J Neurosci 29:3613-3626.

Miller RR, Barnet RC, Grahame NJ (1995) Assessment of the Rescorla-Wagner model. Psychol Bull 117:363-386.

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.

Nieuwenhuis S, De Geus EJ, Aston-Jones G (2010) The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophysiology.

Niv Y (2009) Reinforcement learning in the brain. J Math Psychol 53:139-154.

Ouyang M, Young MB, Lestini MM, Schutsky K, Thomas SA (2012) Redundant catecholamine signaling consolidates fear memory via phospholipase C. J Neurosci 32:1932-1941.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.

Pearce JM, Bouton ME (2001) Theories of associative learning in animals. Annual review of psychology 52:111-139.

Pfaff DW (2006) Brain arousal and information theory: neural and genetic mechanisms. Harvard University Press.

Posner MI (2008) Measuring alertness. Ann N Y Acad Sci 1129:193-199.

Preuschoff K, Bossaerts P (2007) Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci 1104:135-146.

Preuschoff K, 't Hart BM, Einhäuser W (2011) Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Front Neurosci 5:115.

Raisig S, Welke T, Hagendorf H, van der Meer E (2010) I spy with my little eye: detection of temporal violations in event sequences and the pupillary response. Int J Psychophysiol 76:1-8.

Rescorla R, Wagner A (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning (Black AH, Prokasy WF, eds), pp64-99. New York: Appleton-Century-Crofts.

Rescorla RA, Wagner ARA theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning II (Black AH, Prokasy WF, eds), pp64-99. Appleton-Century-Crofts.

Richer F, Beatty J (1987) Contrasting effects of response uncertainty on the task-evoked pupillary response and reaction time. Psychophysiology 24:258-262.

Rushworth MFS, Behrens TEJ (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11:389-397.

Sara SJ, Vankov A, Hervé A (1994) Locus coeruleus-evoked responses in behaving rats: a clue to the role of noradrenaline in memory. Brain Res Bull 35:457-465.

Schmidt HS, Fortin LD (1982) Electronic pupillography in disorders of arousal. In: Sleeping and waking disorders: indication and technique (Guilleminault C, eds). Menlo park, CA: Addison-Wesley.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Seo H, Lee D (2008) Cortical mechanisms for reinforcement learning in competitive games. Philos Trans R Soc Lond B Biol Sci 363:3845-3857.

Servan-Schreiber D, Printz H, Cohen JD (1990) A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. Science 249:892-895.

Seu E, Lang A, Rivera RJ, Jentsch JD (2009) Inhibition of the norepinephrine transporter improves behavioral flexibility in rats and monkeys. Psychopharmacology (Berl) 202:505-519.

Steyvers M, Brown S (2006) Prediction and change detection. Advances in neural information processing systems 18:1281.

Strauss GP, Frank MJ, Waltz JA, Kasanova Z, Herbener ES, Gold JM (2011) Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. Biol Psychiatry 69:424-431.

Sul JH, Jo S, Lee D, Jung MW (2011) Role of rodent secondary motor cortex in value-based action selection. Nat Neurosci 14:1202-1208.
Sutton RS, Barto AG (1998) Reinforcement learning: An introduction. Cambridge, MA: MIT Press.

Tait DS, Brown VJ, Farovik A, Theobald DE, Dalley JW, Robbins TW (2007) Lesions of the dorsal noradrenergic bundle impair attentional set-shifting in the rat. Eur J Neurosci 25:3719-3724.

Takahashi Y, Schoenbaum G, Niv Y (2008) Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. Front Neurosci 2:86-99.

Tully K, Bolshakov VY (2010) Emotional enhancement of memory: how norepinephrine enables synaptic plasticity. Mol Brain 3:15.

van Olst EH, Heemstra ML, Ten Kortenaar T (1979) Stimulus significance and the orienting reaction. In: The orienting reflex in humans (Kimmel HD, van Olst EH, Orlebeke JF, eds), pp521-547. Hillsdale, NJ: Erlbaum.

Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron 65:927-939.

Waterhouse BD, Moises HC, Woodward DJ (1998) Phasic activation of the locus coeruleus enhances responses of primary sensory cortical neurons to peripheral receptive field stimulation. Brain Res 790:33-44.

Williams RJ (1992) Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning 8:229-256.

Williams RW, Herrup K (1988) The control of neuron number. Annu Rev Neurosci 11:423-453.

Wilson RC, Nassar MR, Gold JI (2010) Bayesian online learning of the hazard rate in change-point problems. Neural Comput 22:2452-2476.

Yerkes RM, Dodson JD (2004) The relation of strength of stimulus to rapidity of habit-formation. Journal of comparative neurology and psychology 18:459-482.

Yu A, Dayan P (2003) Expected and unexpected uncertainty: ACh and NE in the neocortex. Advances in neural information processing systems:173-180.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.