Behavioral/Systems/Cognitive

An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment

Matthew R. Nassar,[1] **Robert C. Wilson,**[2] **Benjamin Heasly,**[1] **and Joshua I. Gold**[1]

[1]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and [2]Department of Psychology, Princeton University, Princeton, New Jersey 08540

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 (Williams, 1992; Sutton and Barto, 1998):

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

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 $\alpha_t = 0$, the updated belief reflects the previous belief but not the most recent outcome. When $\alpha_t = 1$, 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 (see Fig. 1*A*). 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.

Received Feb. 15, 2010; revised May 28, 2010; accepted July 13, 2010.

This work was supported by National Institutes of Health (NIH) Grant EY015260, the McKnight Endowment Fund for Neuroscience, the Burroughs-Wellcome Fund, the Sloan Foundation (J.G.), NIH Training Grant in Computational Neuroscience T90 DA 22763-01, which is part of the NIH Blueprint for Neuroscience Research, and the Russell Ackoff Doctoral Student Fellowship from the Wharton Risk Management and Decision Processes Center (M.N.). We thank Long Ding, Chi-Tat Law, Nicole Senecal, and Ching-Ling Teng for helpful comments.

Correspondence should be addressed to Joshua Gold, Department of Neuroscience, University of Pennsylvania, 116 Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104-6074. E-mail: jigold@mail.med.upenn.edu.

DOI:10.1523/JNEUROSCI.0822-10.2010

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 (Adams and MacKay, 2007; Fearnhead and Liu, 2007; Wilson et al., 2010), 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 SD (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 five 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 five 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 (see Fig. 1*A*). 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 (SDs 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:

$$\langle|\delta_t|\rangle > \text{LB} = \$8 \tag{2}$$

$$\text{LB} > \langle|\delta_t|\rangle > 2/3\ \text{LB} + 1/3\ \text{HB} = \$10 \tag{3}$$

$$2/3\ \text{LB} + 1/3\ \text{HB} > \langle|\delta_t|\rangle > 1/2\ (\text{LB} + \text{HB}) = \$12 \tag{4}$$

$$\langle|\delta_t|\rangle < 1/2(\text{LB} + \text{HB}) = \$15 \tag{5}$$

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 (three blocks of 200 trials with SDs 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 and HBP). Payout was computed as follows:

$$\text{SP} < \text{LBP} = \$8 \tag{6}$$

$$\text{LBP} < \text{SP} < 2/3\ \text{LBP} + 1/3\ \text{HBP} = \$10 \tag{7}$$

$$2/3\ \text{LBP} + 1/3\ \text{HBP} < \text{SP} < 1/2\ (\text{LBP} + \text{HBP}) = \$12 \tag{8}$$

$$\text{SP} > 1/2\ (\text{LBP} + \text{HBP}) = \$15 \tag{9}$$

Data analysis. Prediction errors were computed by subtracting the subject's prediction (Eq. 1, B_t) 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 SD 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 SD.

Models

Optimal performance requires inferring the mean (in the estimation task) and width (in the confidence task) of the probability distribution over future outcomes, given all previous outcomes, $p(X_{t+1} \mid X_{1:t})$. We first develop a full, Bayesian solution to this problem. We then systematically reduce this Bayesian solution to a more computationally tractable algorithm that is a form of delta rule.

The Bayesian solution depends on knowledge of the underlying generative process. There are many different ways of describing this process that are mathematically equivalent and therefore could, in principle, be used to formulate the Bayesian ideal observer. For example, outcomes in our task could be generated by a process that depends on a binary variable describing whether or not a change point occurred on a given trial, or, alternatively, by a scalar variable describing the number of outcomes ("run length," or r) that occurred since the most recent change point. In either case, Bayesian inference can be accomplished by inverting the generative process (Adams and MacKay, 2007; Behrens et al., 2007; Fearnhead and Liu, 2007). Here we consider a generative process that depends on run length primarily because this process can be related to a modified delta rule in a straightforward manner.

Specifically, the generative process for our task is based on a weighted coin flip that determines whether the current run of r outcomes will continue ($r_t = r_{t-1} + 1$). If the run does continue, then the generative mean μ_t is set to the previous mean ($\mu_t = \mu_{t-1}$). If the run does not continue (which happens with probability H, which in our task is 0 if $r <$ 5 and 0.05 otherwise, but for simplicity can be thought of as a constant), then r is reset to zero, and a new mean is picked from a uniform distribution [$\mu_t \sim U(0, 300)$]. In either case, an outcome (X_t) is generated on each trial from a normal distribution [$X_t \sim \mathcal{N}(\mu_t, \sigma)$]. Within this generative framework, the predictive distribution can be expressed in terms of the possible means of the generative distribution, given all previous data:

$$p(X_{t+1} \mid X_{1:t}) = \sum_{\mu_t} p(X_{t+1}, \mu_t \mid X_{1:t}) = \sum_{\mu_t} p(X_{t+1} \mid \mu_t) p(\mu_t \mid X_{1:t}) \tag{10}$$

where all of the relevant information from previous outcomes is stored in the final term, $p(\mu_t | X_{1:t})$, which can be expressed as the marginal (with respect to run length) over the joint distribution across the current values of both the mean and run length:

$$p(\mu_t \mid X_{1:t}) = \sum_{r_t} p(\mu_t, r_t \mid X_{1:t}) \tag{11}$$

One strategy for solving this problem is to first compute the joint distribution and then marginalize across all possible run lengths. The joint distribution can be computed according to Bayes' rule:

$$p(\mu_t, r_t \mid X_{1:t}) = \frac{p(X_{1:t} \mid \mu_t, r_t) p(\mu_t, r_t)}{p(X_{1:t})} \tag{12}$$

where the likelihood term, $p(X_{1:t} | \mu_t, r_t)$, can rewritten as a joint distribution including the mean and run length from the previous trial, marginalized over these variables, and rearranged as follows:

$$p(\mu_t, r_t \mid X_{1:t}) = \frac{p(X_t \mid \mu_t, r_t) \sum_{\mu_{t-1}} \sum_{r_{t-1}} p(\mu_t \mid r_t, \mu_{t-1}) p(r_t \mid r_{t-1}) p(\mu_{t-1}, r_{t-1} \mid X_{1:t-1})}{p(X_t \mid X_{1:t-1})} \tag{13}$$

In this case, the likelihood (first) term specifies the probability of only the newest outcome given all possible values for the current mean and run length, and the prior probability of the possible means and run lengths are determined according the transition functions specified by the generative process and the probability distribution over mean and run length computed on the previous trial.

Because the prior in Equation 13 depends only on knowledge of the generative process and the posterior distribution computed on the previous trial, the process is Markovian and can be updated in a straightforward manner after each new outcome. When the hazard rate is known, this updating procedure can be implemented using the message-passing algorithm depicted in Figure 6*A*. 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, which requires either $(t + 1)^2$ or $(t + 1)^3$ separate predictive distributions depending on whether the hazard rate is constant or variable over time, respectively (Wilson et al., 2010). To make this algorithm more computationally tractable, we implemented a pruning algorithm shown previously 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:

$$\begin{aligned} p(cp \mid X_t) = \Omega_t &= \frac{p(X_t \mid cp) p(cp)}{p(X_t)} \\ &= \frac{p(X_t \mid cp) p(cp)}{p(X_t \mid cp) p(cp) + p(X_t \mid {\sim}cp) p({\sim}cp)} \\ &= \frac{U(X_t \mid 0, 300) H}{U(X_t \mid 0, 300) H + \mathcal{N}(X_t \mid \hat{\mu}_t, \hat{\sigma}_t)(1 - H)} \end{aligned} \tag{14}$$

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, $\mathcal{N}(X_t|\hat{\mu}_t, \hat{\sigma}_t)$ 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:

$$\hat{\sigma}_t{}^2 = N^2 + \frac{N^2}{\hat{r}_t} \tag{15}$$

where N is the SD of the generative distribution; see below for an alternative model in which this quantity is inferred from the data. In Equation 15, 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,

$$\hat{\mu}_t^{cp} = X_t \tag{16}$$

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,

$$\hat{\mu}^{\sim cp}_t = \frac{(X_t + \hat{r}_t \times \hat{\mu}_{t-1})}{\hat{r}_t + 1} \tag{17}$$

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

$$\hat{\mu}_t = \frac{(X_t + \hat{r}_t \times \hat{\mu}_{t-1})(1 - \Omega_t)}{\hat{r}_t + 1} + \Omega_t X_t \tag{18}$$

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

$$\hat{\mu}_t = \hat{\mu}_{t-1} + \alpha_t \times \delta_t \tag{19}$$

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

$$\alpha_t = \frac{1 + \Omega_t \hat{r}_t}{\hat{r}_t + 1} \tag{20}$$

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

$$\hat{r}_{t+1} = (\hat{r}_t + 1)(1 - \Omega_t) + \Omega_t \tag{21}$$

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. 15) fixed to the true generative SD 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 vs 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 SD 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:

$$\hat{N}_{t+1}^2 = \hat{N}_t^2 + \alpha_{t(N)} \times \left(\frac{\hat{r}_t \delta_t^2}{\hat{r}_t + 1} - \hat{N}_t^2 \right) \tag{22}$$

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 amount of such information collected previously:

$$\alpha_{t(N)} = \frac{1 - \Omega_t}{\sum_1^t (1 - \Omega_t)} \tag{23}$$

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. 8*B*):

$$\alpha_{(N)}^{\text{MIN}} = \kappa H \times (1 - \Omega_t) \tag{24}$$

where κ is a scaling constant. For Figure 9, *C*, *F*, and *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 Equation 14, this version computed Ω_t as follows:

$$\Omega_t = \frac{U(X_t | 0, 300)^{\lambda} H}{U(0, 300)^{\lambda} H + \mathcal{N}(X_t | \hat{\mu}_t, \hat{\sigma}_t^2)^{\lambda}(1 - H)} \tag{25}$$

where the likelihood weight, λ, is a fractional term (0 . . . 1) that limits the use of likelihood information in change-point detection. When $\lambda = 0$, the model becomes a fixed learning rate delta-rule model in which the learning rate is determined by H. When $\lambda = 1$, 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 $\sim\mathcal{N}(0,D)$, 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. 15):

$$\hat{\sigma}_t^2 = N^2 + \frac{N^2}{\hat{r}_t} + D^2 \tag{26}$$

To consolidate uncertainty about the mean into a single variable and allow correct computation of the learning rate (Eq. 20), we recomputed the run length to reflect the total uncertainty about the mean of the distribution:

$$\hat{r}_{t^*} = \frac{N^2}{\sigma_t{}^2} \tag{27}$$

This adjusted run length was used for the learning rate (Eq. 20) and update (Eq. 21) 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 an SD (noise) that was stable over each block of 200 trials (Fig. 1*A*). 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 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. 1*B*). Estimates tended to approximate the mean during periods of stability and then change relatively rapidly at change points in the generative distribution to resettle 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. 1*D*). 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 vs 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.

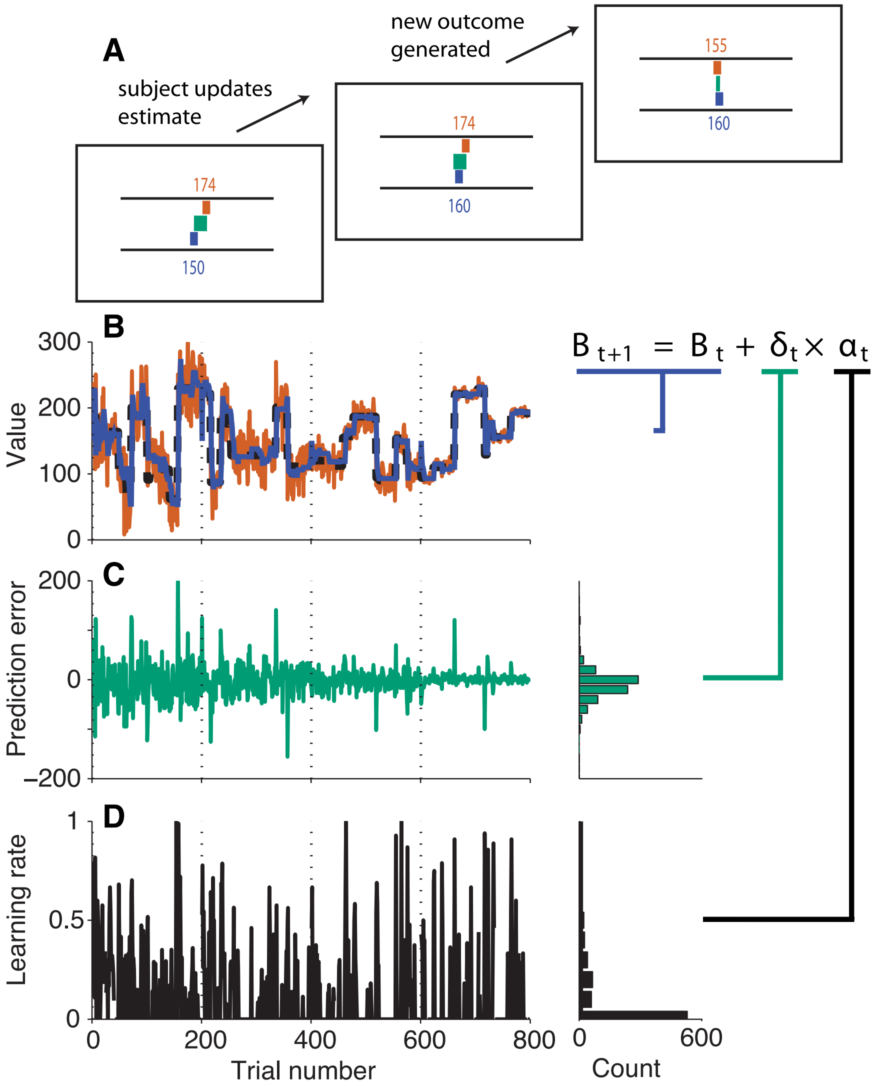

Figure 1. Estimation task and its relationship to prediction errors and learning rate. ***A***, Schematized trial of the estimation task. The subject makes a prediction (blue) and is then shown the outcome (red) and the error made in predicting the outcome (teal). After the subject updates his prediction 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.

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 before 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. 2*A*).

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 SD of the generative distribution (Fig. 2*B*,*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 SD of the generative distribution. This "*z*-scored error" was predictive of learning rate, relatively independent of the noise magnitude (Fig. 2*C*) (Spearman's ρ across all subjects was 0.15; permutation test for H_0: $\rho = 0$, $p < 0.001$). We also note that this basic trend was consistent but varied considerably in magnitude across subjects (Fig. 2*D*), 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. 3*A*, gray). In contrast, learning rates tended to decrease more gradually after a change point (Fig. 3*A*, 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. 2*D*), there were still changes in learning rate that

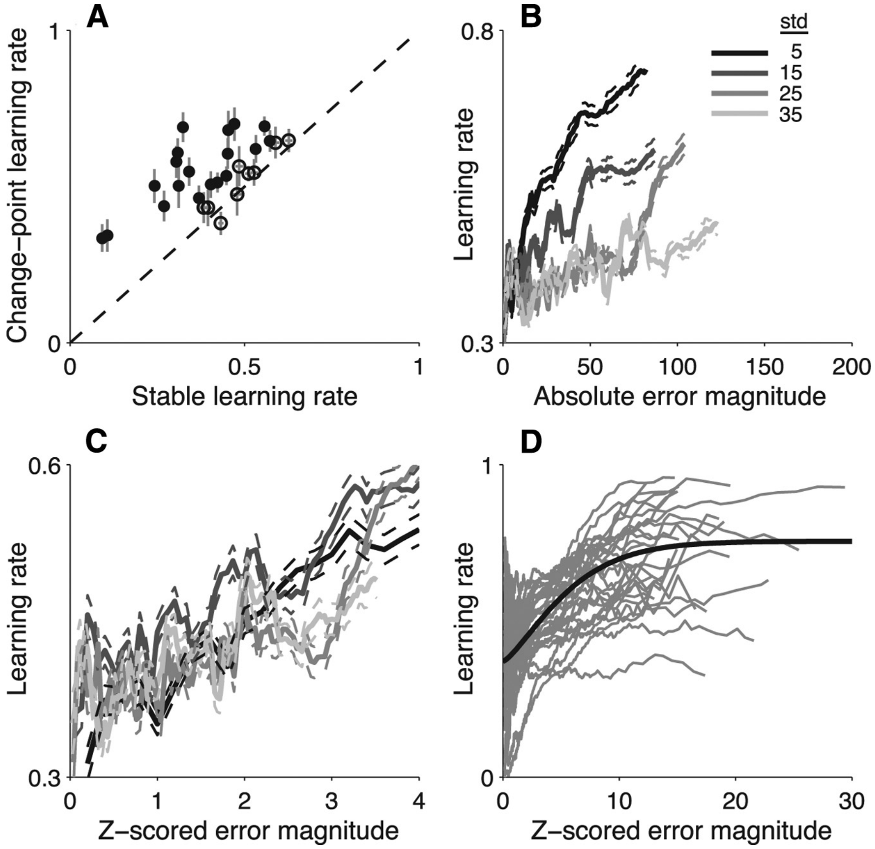

Figure 2. Learning rates increased after unexpected errors. ***A***, Mean ± SEM learning rates on trials in which the mean of the generative distribution changed (ordinate) versus on other trials (abscissa; error bars are obscured by the points). Points are data from individual subjects. Filled symbols 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 SDs of the generative distribution, as indicated. Data are 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***. The relative error magnitude was computed by dividing the absolute error magnitude by the SD of the generative distribution. ***D***, Individual subject learning rates plotted as a function of relative absolute error magnitude (gray lines). The black line indicates a cumulative Weibull function fit to data from all subjects.

persisted for many trials after a change point. The peak value in this adjusted learning rate tended to occur on the first trial after a change point and then decay gradually (Fig. 3*B*).

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 his or her current belief (Yu and Dayan, 2003; Wilson et al., 2010). 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 Figure 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 Figure 4*A* shows estimates (solid blue) and the 85% confidence windows (dashed 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 SD of the generative distribution, with greater uncertainty corresponding to higher noise (Fig. 4*B*). 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. 4*C*). 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 Figs. 4*C*, 3*B*).

In addition to these general trends across subjects, there was considerable individual variability in the choice of confidence-window size (e.g., Fig. 4*B*, whiskers) that was related to learning rate. This relationship is typified by the behavior of two example subjects, shown in Figure 5, *A* and *B*. Subject S.G. (Fig. 5*A*) used small learning rates and tended to specify large confidence windows, indicating high uncertainty (Fig. 5*A*). In contrast, subject L.Y. tended to use large learning rates and small confidence windows (Fig. 5*B*). 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 S.G., who tended to use small learning rates overall, tended to use relatively larger learning rates on trials in which she was most uncertain about her previous estimate. In contrast, subject L.Y., who tended to use large learning rates overall, 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. 5*C*). This relationship implies that subjects who tended to use large learning rates and thus were highly influenced by new information (like subject L.Y.) 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. 5*D*). Subjects who tended to use small learning rates (like subject S.G.) chose larger learning rates after 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 L.Y.) chose larger learning rates after 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.

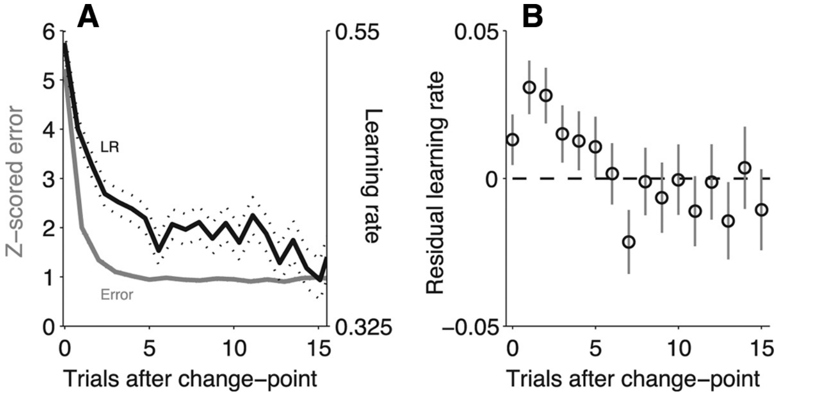

Figure 3. Learning rates decayed slowly after change points. ***A***, Prediction errors (gray, left ordinate) and learning rates (black, right ordinate) plotted as a function of trials after a change point. Solid lines indicate the mean across all subjects and all conditions; dotted lines indicate SEM. ***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 Figure 2*D* from the actual learning rates, and thus reflect the portion of learning rate that was not explained by relative error magnitude. Points and error bars are mean ± SEM across all subjects.

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. 6*A*), or as a power of the number of observations 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. 6*B*). 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. 19). The learning rate depends on both $\hat{r}$ and the probability that a change point occurred (Eq. 20). 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. 6*C*). 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. 6*D*).

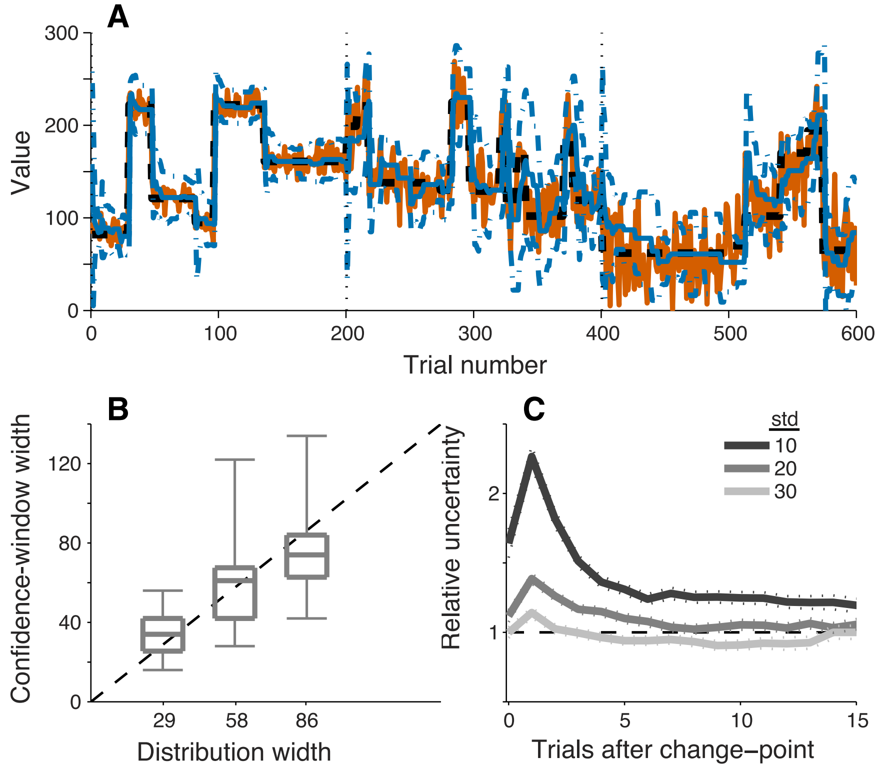

Figure 4. Subjective confidence measurements. ***A***, An example session of the confidence task. Subjects specified a symmetric window (dashed blue lines) around their estimate (solid blue line) that they were 85% certain would contain the next number (red) generated using the current mean (dashed black line) and SD (stable in blocks, indicated by the vertical, dotted lines). ***B***, Box-and-whisker plot (central line is the median, box is the 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 (***B***, *x*-axis markers). Solid and dotted lines indicate mean and SEM, respectively.

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 SD of the prior, stable distribution was small (Fig. 7*A*–*C*). Moreover, the model tended to have higher

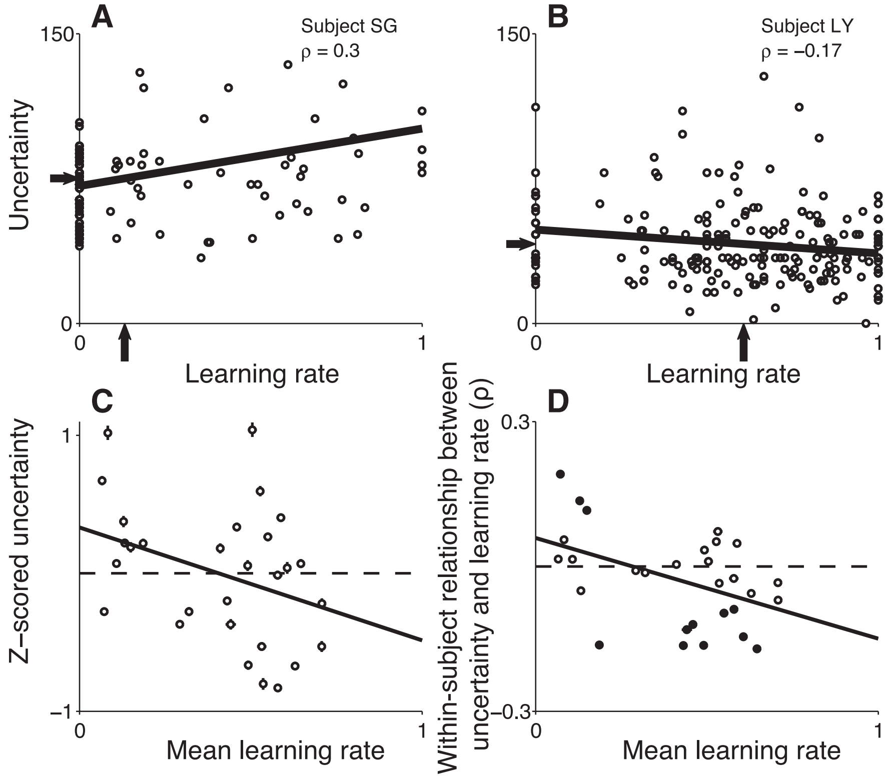

Figure 5. Relationship between confidence and learning rate. ***A***, ***B***, Trial-by-trial learning rates plotted as a function of uncertainty (confidence-window width) for an example task block (SD, 20) for two different subjects. Solid lines are linear fits. Arrows indicate the mean values of the 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. The solid line is a linear fit ($r = -0.38$; $p = 0.04$). 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***; filled symbols indicate H_0: $\rho = 0$, $p < 0.05$; a positive or negative value indicates that the subject tended to use higher or lower learning rates on trials in which they were more uncertain about their previous prediction, respectively) plotted as a function of the average learning rate used by that subject. Symbols and error bars are the mean ± SEM per subject. The solid line is a linear fit ($r = -0.44$; $p = 0.02$). 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.

learning rates on the trial after a change point, which then decayed gradually over many trials (Fig. 7*D*,*E*). In the model, this gradual decay is caused by the decay in uncertainty occurring over the same period (Fig. 7*F*). 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. 6*D*).

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 SD 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. 8*A*). This effect is attributable mostly to the fact that higher hazard-rate models tend to use higher learning rates (Fig. 8*B*) 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 (Fig. 8*A*, vertical dashed line). 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. 2*D*, 7*C*).

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. 5*C*). 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. 15). 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. 6*C*). The second type of uncertainty is related to the expected SD 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 the performance of each form of the model to the behavioral data presented in Figure 5. The simplest form used estimates of noise that were fixed within a block (Fig. 9*A*). In this case, overall uncertainty, like learning rate, declined with run length (Eq. 15). 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. 9*D*). 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. 9*G*). Thus, this model did not match the behavioral data.

The second model used a sequentially updated estimate of noise (Eq. 22). 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. 9*B*). 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 $\hat{r}$ in the prediction-error term in Eq. 22) 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. 9*E*). 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. 9*H*). Thus, this model also did not match the behavioral data.

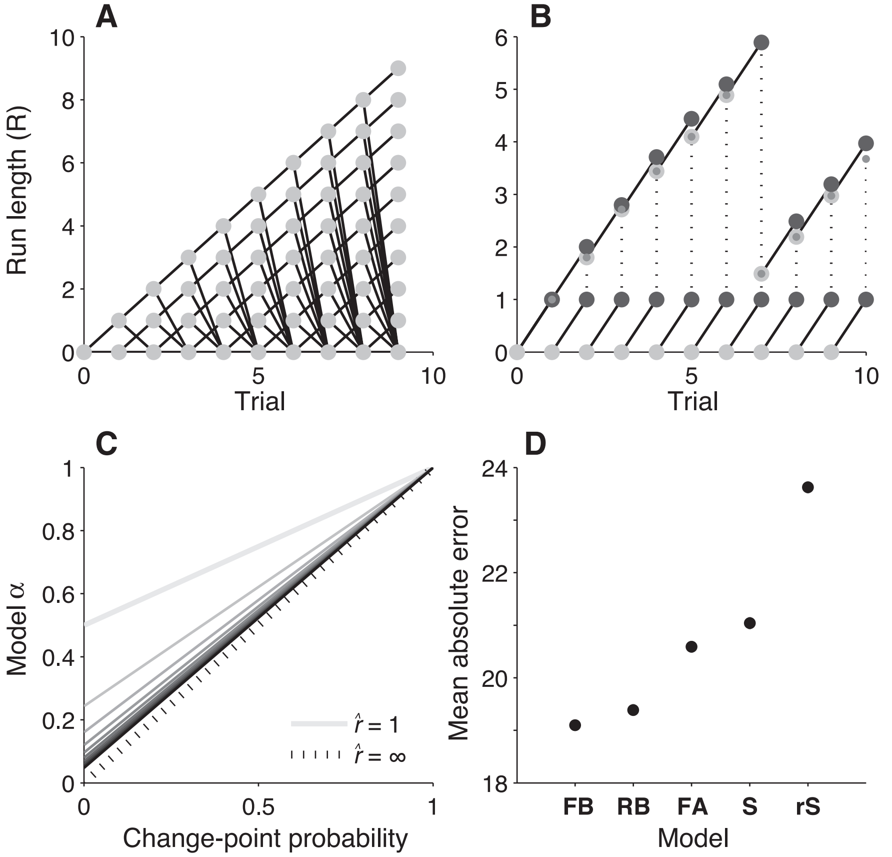

Figure 6. Bayesian model. ***A***, Message-passing algorithm for the full model. Run length (*r*) refers to the number of data points obtained previously from the current generative distribution. On each trial, the distribution either changes and *r* is set to zero, or the generative distribution does 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 Materials and 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) are shown.

The third model used a more realistic, suboptimal strategy for inferring noise (Eq. 24). 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. 9*C*) and, like the second model, a strong, negative relationship between overall uncertainty and learning rate across simulated sessions (Fig. 9*F*). 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. 9*I*). 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 intersubject 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 trade-off 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 trade-off between successful prediction amid noise and successful adaptation after change points.

Models that underweigh 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. 11*A*). This suboptimal behavior of human subjects reflected a relationship between learning rate and *z*-scored error that was too flat (Fig. 11*B*).

One explanation for this difference might be that subjects underuse likelihood information when assessing whether a change point

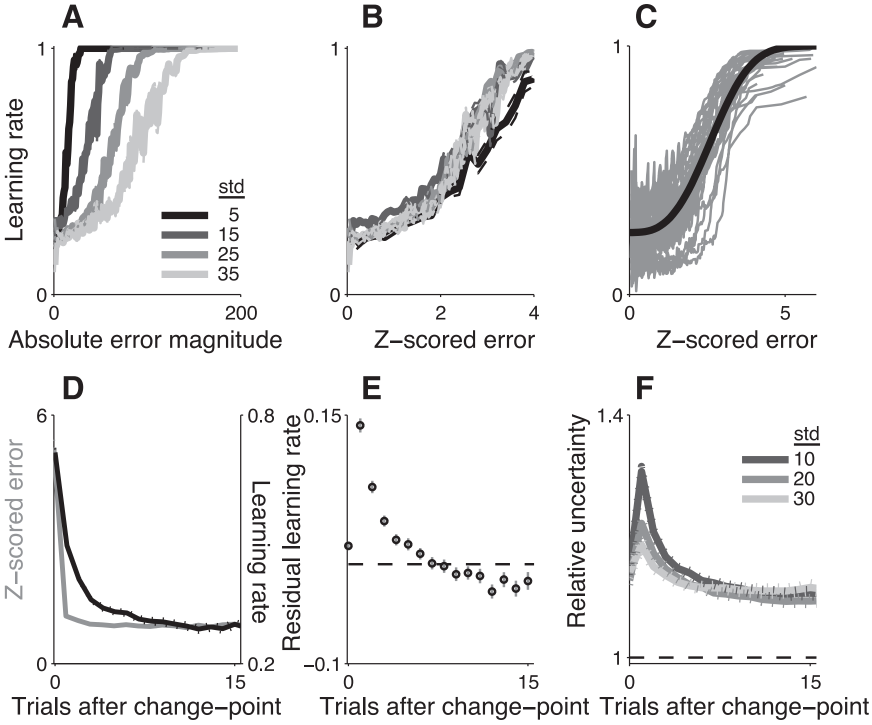

Figure 7. The reduced Bayesian model qualitatively reproduces belief-updating behavior. All plots in this figure depict simulated data using the reduced Bayesian model. One model parameter, the hazard rate, was fit for each block to minimize the difference between model and subject predictions. ***A***, Learning rate as a function of absolute error magnitude for different SDs of the generative distributions, as shown (compare Fig. 2*B*). ***B***, Learning rate as a function of *z*-scored error, plotted as in ***A*** (compare Fig. 2*C*). ***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). The black line is the cumulative Weibull fit (compare Fig. 2*D*). ***D***, *Z*-scored error (gray, left ordinate) and learning rate (black, right ordinate) plotted as a function of trials after a change point. The solid and dashed lines indicate mean ± SEM (compare Fig. 3*A*). ***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 error bars are mean ± SEM across all simulated data (compare Fig. 3*B*). ***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. The grayscale reflects the SD of the given task block, as indicated (compare Fig. 4*C*).

occurred on a given trial. Adding a parameter (Eq. 25, λ) to the reduced model that allows for such suboptimal computation lets the model range from a fixed learning rate delta-rule model ($\lambda = 0$) to the reduced-Bayesian model ($\lambda = 1$). 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. 11*C*).

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 (Eq. 26, *D*) to the variance on the predictive distribution after each time step. This model can be applied to subject data with drift (*D*), hazard rate (*H*), and expected noise (*N*) all fit as free parameters (Eqs. 26, 27), producing predictions that have a more shallow relationship between learning rate and *z*-scored error (Fig. 11*B*). This model described subject behavior better than either the reduced-Bayesian model with only the hazard rate as a free parameter (for 30 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 of 30 subjects) or a delta-rule model with a fixed learning rate (for 29 of 30 subjects) (Fig. 11*D*).

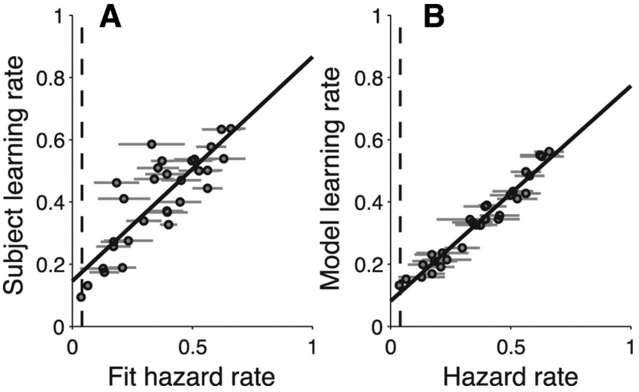

Figure 8. Relationship between learning rate and hazard rate. ***A***, Variability in subject learning rates can be described by the hazard rate in the model. Subjects that are fit best by high hazard rate versions of the reduced Bayesian model use higher learning rates, on average. The dashed line indicates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

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 (Corrado et al., 2005; Behrens et al., 2007; Krugel et al., 2009), 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, tended to rise immediately then decay slowly after 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. 2*C*, 7*B*). 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. 6*C*) and uncertainty (Eq. 15). 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. 7*E*). 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 Figs. 3*B*, 7*E*, and 4*C*, 7*F*). 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 the Rescorla–Wagner model, a straightforward form of the delta rule, and the Pearce-Hall model, 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.

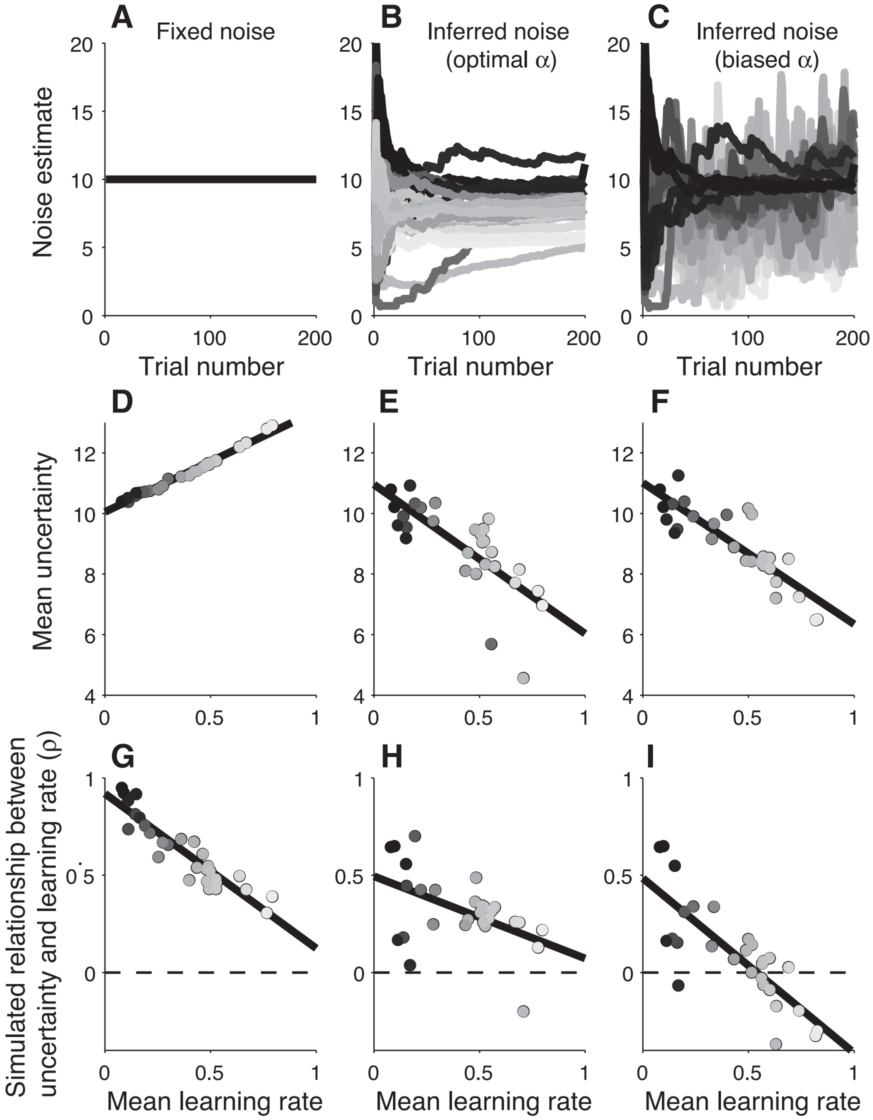

Figure 9. On-line noise inference. Individual variability was simulated by using models that employed the hazard rates fit to individual subject data (see Computing best-fitting hazard rates in Materials and Methods) (in all panels, grayscale represents the different hazard rates, with lighter shades for higher rates). Three models that differed only in their method for computing noise were used to simulate performance. The first, simplest model (left) used the actual SD 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. 22, 23). 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 (Eq. 24). ***A–C***, Noise estimates from each model over the course of each 200-trial block in which the SD of the generative distribution was equal to 10. ***D–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 infer less noise. ***G–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.

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 nonexistent 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., used learning rates near zero). In contrast, other subjects tended to adjust their beliefs dramatically in response to each new outcome (i.e., used 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

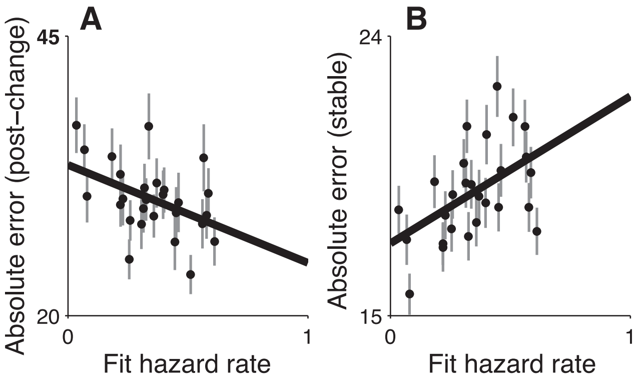

Figure 10. Hazard rate trade-off. ***A***, Average absolute errors made by subjects one to five trials after a change point plotted as a function of the fit hazard rate from the reduced Bayesian model for each subject (points). The line is a linear regression ($r = -0.43$; $p = 0.02$). The negative relationship implies that subjects who used higher hazard rates made better predictions after change points. ***B***, Average absolute errors made by subjects six or more trials after a change point plotted as a function of the fit hazard rate for each subject (points). The 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.

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 oversensitivity 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 catechol-*O*-methyltransferase (COMT) that leads to lower learning rates in reversal tasks but improved performance in working-memory tasks (Bruder et al., 2005; Krugel et al., 2009). 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

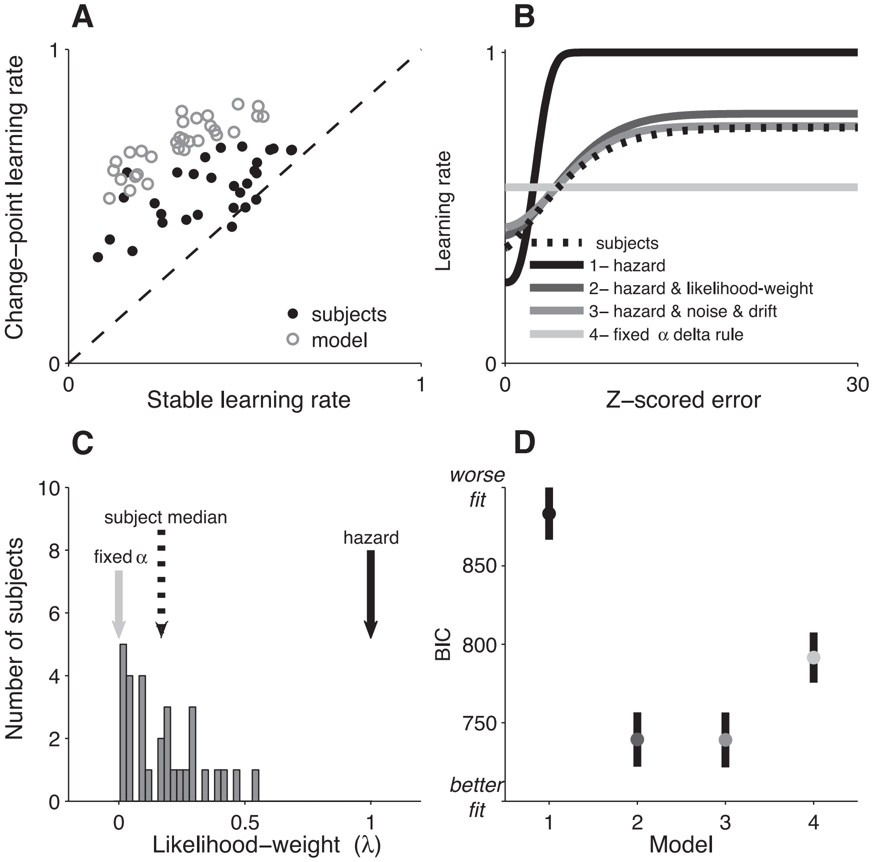

Figure 11. Better descriptive models to capture suboptimal performance. ***A***, Although subjects (filled symbols; data are plotted as in Fig. 2*A*) and the reduced Bayesian model (open symbols) both used higher learning rates after change points than during a stable period, the model tends to show 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 (Eq. 25, λ). When $\lambda = 0$, the model updates beliefs according to a fixed learning rate delta rule. When $\lambda = 1$, 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 error bars are mean ± SEM across subjects. The grayscale and model numbers are as in ***B***.

of reward-prediction error (Eq. 1, δ) (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 functional magnetic resonance imaging (fMRI) and EEG data (Debener et al., 2005; Matsumoto et al., 2007; Hayden et al., 2009). 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 (Eq. 1, δ), less is known about the learning rate (Eq. 1, α). 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 mechanisms. 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 blood oxygen level-dependent 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 ceruleus–norepinephrine system (Yu and Dayan, 2005). Our task and model provide a framework for testing this possibility.

References

Adams RP, MacKay DJ (2007) Bayesian online changepoint detection. Cambridge, UK: University of Cambridge Technical Report.

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

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.

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

Daw N, O'Doherty J, Dayan P, Seymour B, Dolan R (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879.

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

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.

Fearnhead P, Liu Z (2007) On-line inference for multiple changepoint problems. J R Stat Soc Series B Stat Methodol 69:589–605.

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

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

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.

Krugel L, Biele G, Mohr P, Li S, Heekeren HR (2009) Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. Proc Natl Acad Sci U S A 106:17591–17596.

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

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

Pearce JM, Bouton ME (2001) Theories of associative learning in animals. Annu Rev Psychol 52:111–139.

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

Rushworth MF, Behrens TE (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci 11:389–397.

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

Steyvers M, Brown S (2006) Prediction and change detection. Adv Neural Inform Processing Systems 18:1281–1288.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.

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

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

Yu A, Dayan P (2003) Expected and unexpected uncertainty: ACh and NE in the neocortex. Adv Neural Inform Processing Systems 15:173–180.

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