A simple model for learning in volatile environments

Payam Piray* and Nathaniel D. Daw
Princeton Neuroscience Institute, Princeton University
* corresponding author: ppiray@princeton.edu

Abstract

Sound principles of statistical inference dictate that uncertainty shapes learning. In this work, we revisit the question of learning in volatile environments, in which both the first and second-order statistics of observations dynamically evolve over time. We propose a new model, the volatile Kalman filter (VKF), which is based on a tractable state-space model of uncertainty and extends the Kalman filter algorithm to volatile environments. The proposed model is algorithmically simple and encompasses the Kalman filter as a special case. Specifically, in addition to the error-correcting rule of Kalman filter for learning observations, the VKF learns volatility according to a second error-correcting rule. These dual updates echo and contextualize classical psychological models of learning, in particular hybrid accounts of Pearce-Hall and Rescorla-Wagner. At the computational level, compared with existing models such as hierarchical Gaussian filter, the VKF gives up some flexibility in the generative model to enable a more faithful approximation to exact inference. Importantly, as expected based on theory and empirical observations, this results in a positive relationship between volatility and learning rate signals, which does not hold for the binary version of the hierarchical Gaussian filter. When fit to empirical data, the VKF is better behaved than alternatives and better captures human choice data in two independent datasets of probabilistic learning tasks. The proposed model provides a coherent account of learning in stable or volatile environments and has implications for decision neuroscience research.

Author Summary

Sound principles of statistical learning dictate that uncertainty influences behavior. However, despite the success of statistically founded algorithms for learning in stable environments, in which uncertainty behaves in simple and predictable ways, it is challenging to develop a simple yet efficient algorithm for learning in volatile environments, in which uncertainty dynamically changes over time. In this article, we develop a model for learning in volatile environments. The proposed model is consistent with key concepts of classical learning theories from behavioral psychology. Furthermore, our model is algorithmically simpler, theoretically more accurate, and empirically more parsimonious than the state-of-the-art models of learning in volatile environments. The proposed model provides a coherent theory of learning under uncertainty.

Introduction

Our decisions are guided by our ability to associate environmental cues with the outcomes of our chosen actions. Accordingly, a central theoretical and empirical question in behavioral psychology and neuroscience has long been how humans and other animals learn associations between cues and outcomes. According to both psychological and normative models of learning [1–4], when animals observe pairings between cues and outcomes, they update their belief about the value of the cues in proportion to prediction errors, the difference between the expected and observed outcomes. Importantly, the degree of this updating depends on a stepsize or learning rate parameter. Although some accounts take this as a free parameter, analyses based on statistical inference, such as the Kalman filter [5], instead demonstrate that the learning rate should in principle depend on the learner's uncertainty. The dynamics of uncertainty – and hence, of learning rates – then depend on the assumed or learned dynamics of the environment. For instance, the Kalman filter is derived assuming that the true associations fluctuate randomly, but at a known, constant speed. In this case the asymptotic uncertainty, and learning rate, are determined by how quickly the associations fluctuate and how noisily they are observed. However, in volatile environments, in which the speed by which true associations change might itself be changing, uncertainty (and learning rates) should fluctuate up and down according to how quickly the environment is changing [6,7]. This normative analysis parallels classical psychological theories, such as the Pearce-Hall model [3], which posit that surprising outcomes increase the learning rate while expected ones decrease it. Those models measure surprise by the absolute value of the discrepancy between the actual outcome and the expected value, i.e. the unsigned prediction error [3].

Behavioral studies have also shown that human learning in volatile environments is consistent with the predictions of this class of models [6,8]: learning rates fluctuate with the volatility of the environment. There is also evidence for a neural substrate for such dynamic learning rates: for instance, neuroimaging studies have shown that activity in the dorsal anterior cingulate cortex covaries with the optimal learning rate [6] and recent work suggests a mechanistic model of adaptive learning rate [9]. Theoretical and empirical work also suggests that neuromodulatory systems, particularly acetylcholine, norepinephrine and serotonin, might be involved in encoding uncertainty signals necessary for computing learning rate in stable and volatile environments, respectively [7,8,10–12]. Finally, various approximate learning algorithms have been fruitful for studying individual differences in learning [13–16].

The theoretical studies establish the question of *why* the learning rate should be dynamically adjusted, and the empirical studies provide evidence that it does so. However, a complete understanding of a learning system requires understanding of *how* these theories could be realized at the process or

algorithmic level [17]. This is as yet much less clear: as discussed below, statistically grounded theories of learning under volatility tend to be somewhat impractical and opaque. Furthermore, while their general analogy with the more rough and ready psychological theories like Pearce-Hall seems clear, there is not a direct mapping comparable to the way for example the Kalman filter encompasses and rationalizes the classical Rescorla-Wagner theory.

A fully Bayesian account of learning turns on two different aspects. The first is a generative model, that is a set of explicit probabilistic assumptions about how the environment evolves and generates outcomes. Inverting the generative model with Bayes' rule gives rise to an optimal inference algorithm for estimating the latent environmental variables from observable outcomes. In situations more complicated than the smooth world of the Kalman filter, however, exact inference is generally intractable and the second aspect comes into play: additional approximations are required to achieve a practical, algorithmic- or process-level inference model. An influential model, proposed by Behrens and colleagues [6] addressed the first but not the second of these points. It comprises a two-level generative model for learning, in which a variable governing the observed associations fluctuates according to a Gaussian random walk with a speed (i.e., diffusion variance) determined, at each timepoint by a second higher-level, variable, itself fluctuating with analogous dynamics. Although this model has been conceptually influential, it lacked any tractable or biologically-plausible inference model at the algorithmic level: instead, its predictions were simulated by brute-force integration. Another model, called the hierarchical Gaussian filter (HGF) [8,18] extended Behrens' generative model to a more general one consisting of multiple levels of hierarchy, in which the extent of diffusion noise at each level is determined by the preceding level [19]. Importantly, it also addressed the second issue by offering a tractable approximate inference rule for the process, based on a variational approach. This filter provides a biologically-plausible model for hierarchical learning and is able to capture dynamics over an arbitrary number of cascading layers with an algorithm that is easily generalizable.

Here we revisit the question of approximate inference in a two-level model of learning in volatile environments. A key theoretical complication faced by the aforementioned models is that the variables at each level of the hierarchy, above the first, represent variances for the corresponding random walks at the next level down. Since variances must be positive, if the hierarchy is taken as a cascade of analogous unbounded Gaussian random walks, nonlinear transformations must be introduced at each stage to ensure positivity. This, in turn, complicates inference: in particular, even after employing a variational approximation (as the HGF does) to decouple inference about each variable from the others, solving the resulting subproblems still requires further approximation to accommodate the nonlinearity at each

stage. Informed by this reasoning, we propose a novel model for learning in volatile environments that is conceptually derived from that of Behrens et al. [6] and the 2-level case of the HGF. However, we introduce a distinct diffusion dynamics for the upper level, which ensures positivity and also an exact, conjugate solution to the variational maximization. The resulting model thus gives up some of the elegant flexibility of the HGF (its ability recursively to chain through an arbitrarily deep hierarchy) in return for a simpler inference rule requiring fewer approximations for the most widely used, 2-level, case. We separately encounter, and take a distinct approach to, a second issue of nonlinearity that arises in these models, which arises when (as has almost always been the case in empirical studies of human learning in this area) the observable outcomes like rewards are binary-valued instead of continuous. The resulting algorithm, called VKF, is a generalization of Kalman filter algorithm to volatile environments and resembles models that hybridize the error-driven learning from the Rescorla-Wagner model and the Kalman filter with Pearce-Hall's dynamic learning rate (as proposed by different authors, for example by Li et al. and Le Pelley [20,21]). Notably, in volatile environments, the learning rate fluctuates with larger and smaller than expected prediction errors, as suggested by models such as Pearce-Hall.

In the next section, we review the Kalman filter algorithm and present the generative model underlying the VKF and the resulting learning algorithm. The full formal treatment is given in the S1 Appendix. Next, we show that the proposed model outperforms existing models in predicting empirical data.

Results

Theoretical results

Kalman filter for tracking in environments with stable dynamics

The Kalman filter is the cornerstone of statistical tracking theories, with widespread applications in many technological and scientific domains including psychology and neuroscience [4,7,22,23]. The Kalman filter corresponds to optimal statistical inference for a particular class of linear state space environments with Gaussian dynamics. In those environments, the hidden state of the environment is gradually changing across time (according to Gaussian diffusion) and the learner receives an outcome on each time depending on the current value of the state (plus Gaussian noise). In these circumstances, the posterior distribution over the hidden state is itself Gaussian, thus tracking it amounts to maintaining two summary statistics: a mean and a variance.

Formally, consider the simplest case of prediction: that of tracking a noisy, fluctuating reward (e.g. that associated with a particular cue or action), whose magnitude o_t is observed on each trial t. Assume a state space model in which, on trial t, the hidden state of the environment, x_t (the true mean reward), is equal to its previous state x_{t-1} plus some process noise

$$x_t = x_{t-1} + e_t, \qquad \text{Equation 1}$$

where the process noise $e_t \sim Normal(0, v)$ has a variance given by v. A critical assumption of the Kalman filter is that the process uncertainty, v, is constant and known. The outcome then noisily reflects the hidden state, $o_t \sim Normal(x_t, \sigma^2)$. Here the observation noise is again Gaussian with known, fixed variance, σ^2. The Kalman filtering theory indicates that the posterior state at time t, given all previous observations, o_1, …, o_t, will itself be Gaussian. Its mean, m_t, is updated at each step according to the prediction error:

$$m_t = m_{t-1} + k_t(o_t - m_{t-1}), \qquad \text{Equation 2}$$

where k_t is the learning rate (or Kalman gain)

$$k_t = (w_{t-1} + v)/(w_{t-1} + v+, \sigma^2), \qquad \text{Equation 3}$$

which depends on the noise parameters and the posterior variance w_{t-1} on the previous trial. Note that $k_t < 1$ in all trials and it is larger for larger values of v. On every trial, the posterior variance also gets updated:

$$w_t = (1 - k_t)(w_{t-1} + v). \qquad \text{Equation 4}$$

Note that although it is not required for prediction (that is for updating m_t), it is also possible to compute the autocovariance (given observations o_1,…, o_t), defined as the covariance between consecutive states, $w_{t-1,t} = \mathrm{cov}[x_{t-1}, x_t]$, which is given by

$$w_{t-1,t} = (1 - k_t)w_{t-1}. \qquad \text{Equation 5}$$

This equation indicates that when the Kalman gain is relatively small, the autocovariance is large, which means that information transmitted by observing a new outcome is expected to be quite small. We will see in the next section that this autocovariance plays an important role in inference in volatile environments.

VKF: A novel algorithm for tracking in volatile environments

We next consider a volatile environment in which the dynamics of the environment might themselves change. In the language of the state space model presented above, the process noise dynamically changes. Thus, the variance of the process noise (e_t in Equation 1) is a stochastic variable changing with time (Figure 1A). To build a generative model, we need to make some assumptions about the dynamics of this variable. Our approach here is essentially the same as that taken by Smith and Miller

[24] and by Gamerman et al. [25] (see also West et al. [26]). Consider a problem in which the process variance dynamically changes. In the previous section, we saw that the state x_t diffused according to additive noise. Because variances are constrained to be positive, it makes sense to instead assume their diffusion noise is multiplicative to preserve this invariant. Therefore, we assume that the current value of precision (inverse variance), z_t, is given by its previous value multiplied by some independent noise. Formally, the current state of z_t is given by

$$z_t = z_{t-1}\epsilon_t, \qquad \text{Equation 6}$$

where ϵ_t is an independent random variable on trial t, which is distributed according to a rescaled beta distribution (as detailed in S1 Appendix), such that the mean of ϵ_t is 1 (that is, conditional expectation of z_t is equal to z_{t-1}) but has spread controlled by a free parameter $0 < \lambda < 1$. The value of noise, ϵ_t, is always positive and is smaller than $(1-\lambda)^{-1}$. Therefore, while "on average", z_t is equal to z_{t-1}, z_t can be smaller than z_{t-1}, or larger by a factor up to $(1-\lambda)^{-1}$. Thus, the higher the parameter λ, the faster the diffusion.

To build up the full model, first consider a simplified case in which the latent state $x_t \sim Normal(\mu_t, z_t^{-1})$ is directly observed at each step and, further, has some known mean μ_t. Although this sub-problem greatly simplifies the main problem by isolating x_t from x_{t-1}, it provides the foundation for inference in the full model. This is because it is similar to the problem that remains once a variational approximation is introduced. It can be formally shown (see S1 Appendix) that in the simplified case, estimating the process variance, i.e. the posterior over z_t at each step given all previous observations, is tractable. Specifically, the posterior distribution over z_t takes the form of a Gamma distribution, whose inverse mean, v_t, is updated according to the observed sample variance:

$$v_t = v_{t-1} + \lambda((x_t - \mu_t)^2 - v_{t-1}), \qquad \text{Equation 7}$$

where $v_t = E[z_t]^{-1}$. Note that this equation amounts to an error-correcting rule for updating v_t, in which the error is given by $(x_t - \mu_t)^2 - v_{t-1}$: the difference between the observed and expected squared prediction errors.

Now we are ready to build the full generative model in which both z_t and x_t dynamically evolve over time (Fig 1). In this model, as before, the observation, $o_t \sim Normal(x_t, \sigma^2)$, follows a Gaussian distribution with the mean given by the hidden variable x_t, which itself is given by $x_t = Normal(x_{t-1}, z_t^{-1})$: a diffusion whose precision (i.e. inverse variance) is given by another dynamic random variable, z_t. The value of z_t in turn depends on its previous value multiplied by some noise that finally depends on parameter λ according to Equation 6. Therefore, this generative model consists of two chains of random variables, which are hierarchically connected to each other. Unlike each of those two

chains separately, however, when they are conjoined in this hierarchical model, inference is not tractable and therefore we need some approximation. We use structured variational inference for approximate inference in this model [27,28]. This technique assumes a factorized approximate posterior distribution and minimizes the mismatch between this approximate posterior and the true posterior using the principle of variational inference.

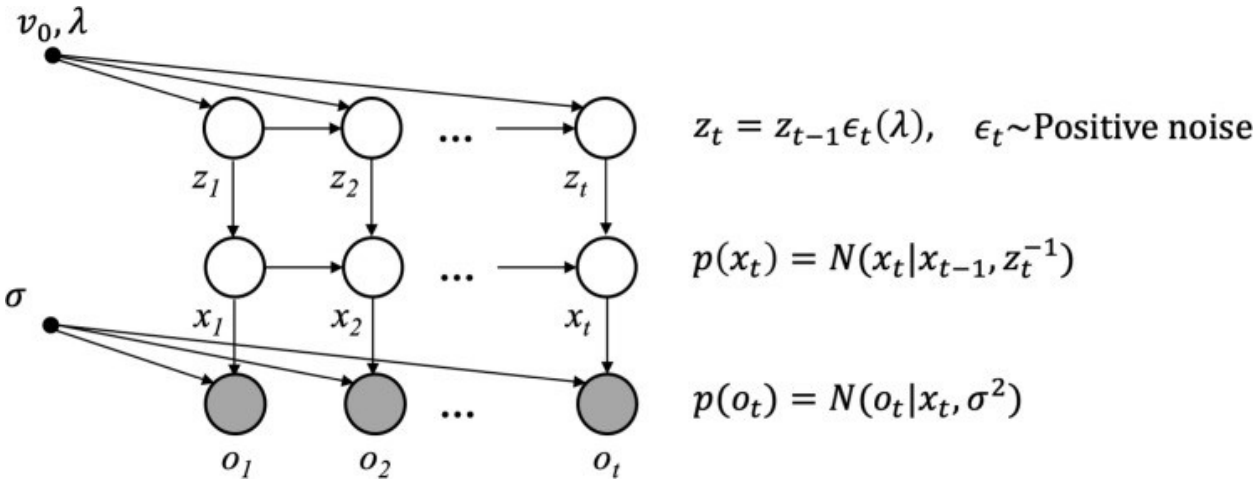

Fig 1. The generative model of VKF. The generative model consists of two interconnected hidden temporal chains, x_t and z_t, governing observed outcomes, o_t. Arrows indicate the direction of influence. On each trial, t, outcome, o_t, is generated based on a Gaussian distribution, its mean is given by the hidden random variable x_t, and its variance is given by the constant parameter ω. This variable is itself generated according to another Gaussian distribution, which its mean is given by x_{t-1}, and its variance is given by another hidden random variable, z_t^{-1}. This variable is itself generated based on its value on the previous trial, z_{t-1}, multiplied by some positive noise distributed according to a scaled Beta distribution governed by the parameter λ. The inverse mean of this variable on the first trial is assumed to be given by another constant parameter, v_0. See Equations 1-2 for further explanation.

The resulting learning algorithm is very similar to the Kalman filtering algorithm and encompasses Kalman filter as a special case. Importantly, the new algorithm also tracks volatility on every trial, denoted by v_t, which is defined as the inverse of expected value of z_t. Therefore, we call the new algorithm the "volatile Kalman filter". In this algorithm, the update rule for the posterior mean m_t and variance w_t over x_t (Equations 9-13 below) is exactly the same as the Kalman filtering algorithm, but in which the constant process variance is replaced by the estimated volatility on the previous trial, v_{t-1}. The volatility also gets updated on every trial according to expected value of $(x_t - x_{t-1})^2$

$$v_t = v_{t-1} + \lambda(E[(x_t - x_{t-1})^2] - v_{t-1}), \qquad \text{Equation 8}$$

where the expectation should be taken under the approximate posterior over x_{t-1} and x_t. Therefore, the volatility update rule takes a form of error correcting, in which the error is given by $E[(x_t - x_{t-1})^2] -$

v_{t-1} with the noise parameter, λ, as the step size. Thus, the higher the noise parameter, the higher the speed of volatility update is. Therefore, we call λ the "volatility update rate". Also, note that the expectation in this equation depends on the autocovariance.

It is then possible to write $E[(x_t - x_{t-1})^2]$ in terms of the variance and covariance of x_{t-1} and x_t to obtain Equation 8 below and complete the VKF learning algorithm:

$$k_t = (w_{t-1} + v_{t-1})/(w_{t-1} + v_{t-1} + \sigma^2), \quad \text{Equation 9}$$

$$m_t = m_{t-1} + k_t(o_t - m_{t-1}), \quad \text{Equation 10}$$

$$w_t = (1 - k_t)(w_{t-1} + v_{t-1}), \quad \text{Equation 11}$$

$$w_{t-1,t} = (1 - k_t)w_{t-1}, \quad \text{Equation 12}$$

$$v_t = v_{t-1} + \lambda\left((m_t - m_{t-1})^2 + w_{t-1} + w_t - 2w_{t-1,t} - v_{t-1}\right), \quad \text{Equation 13}$$

where σ is the constant variance of observation noise. In addition to the volatility update parameter, λ, which is constrained in the unit range, this algorithm depends on another parameter, $v_0 > 0$, which is the initial value of volatility. Notably, the Kalman filter algorithm is a special case of the VKF in which $\lambda = 0$ and the process variance is equal to v_0 on all trials. In the next section, we test this model with synthetic and empirical datasets.

Binary VKF

We have also developed a binary version of VKF for situations in which observations are binary. The generative model of the binary VKF is the same as that of VKF with the only difference that binary outcomes are generated according to Bernoulli distribution with the parameter given by $s(x_t) = 1/(1 + \exp(-x_t))$, where $s(x_t)$ is the sigmoid function, which maps the normally distributed variable x_t to the unit range. For this generative model, the inference is more difficult because the relationship between hidden states and observations is nonlinear. Therefore, further approximation is required to perform inference here, because observations are not normally distributed and Equation 1 does not hold. For the binary VKF, we assumed a constant posterior variance, ω, and employed moment matching (which is sometimes called assumed density filtering [29,30]) to obtain the posterior mean. The resulting algorithm is then very similar to the original VKF with the only difference that the update rule for the mean (i.e. Equation 10) is slightly different:

$$k_t = (w_{t-1} + v_{t-1})/(w_{t-1} + v_{t-1} + \omega), \quad \text{Equation 14}$$

$$\alpha_t = \sqrt{w_{t-1} + v_{t-1}}\,, \quad \text{Equation 15}$$

$$m_t = m_{t-1} + \alpha_t\big(o_t - s(m_{t-1})\big), \quad \text{Equation 16}$$

$$w_t = (1 - k_t)(w_{t-1} + v_{t-1}), \quad \text{Equation 17}$$

$$w_{t-1,t} = (1 - k_t)w_{t-1}, \quad \text{Equation 18}$$

$$v_t = v_{t-1} + \lambda\left((m_t - m_{t-1})^2 + w_{t-1} + w_t - 2w_{t-1,t} - v_{t-1}\right). \quad \text{Equation 19}$$

Note that the learning rate for the binary VKF is α_t defined by Equation 15. Furthermore, we have introduced a parameter, $\omega > 0$, specifically for inference (i.e. does not exist in the generative model). We call this parameter the noise parameter, because its effects on volatility are similar to the noise parameter, σ for linear observations (through k_t, Equation 14). However, it is important to note that unlike σ, this parameter does not have an inverse relationship with the learning rate, α_t.

Simulation analyses

Comparing VKF with known ground truth

First, we study the performance of VKF on simulated data with known ground truth. We applied the VKF to a typical volatility tracking task (similar to [6,8]). In this task, observations are drawn from a normal distribution whose mean is given by the hidden state of the environment. The hidden state is constant +1 or –1. Critically, the hidden state was reversed occasionally. The frequency of such reversals itself changes over time; therefore, the task consists of blocks of stable and volatile conditions. We investigate performance of VKF in this task because its variants have been used for studying volatility learning in humans [6,8]. Note that, as here, it is common to study these models' learning performance when applied to situations that do not exactly correspond to the generative statistical model for which they were designed. In particular, it has been shown using this type of task that humans' learning rates are higher in the volatile condition [6,8]. Fig 2 shows the performance of VKF in this task. As this figure shows, the VKF tracks the hidden state very well and its learning rate is higher in the volatile condition. Furthermore, the volatility signal increases when there is a dramatic change in the environment. Note that as for other models previously fit to tasks of this sort, the switching dynamics at both levels of hidden state here are not the same as the random walk generative dynamics assumed by our model. These substitutions demonstrate that the principles relating volatility to learning are general, and so the resulting models are robust to this sort of variation in the true generative dynamics.

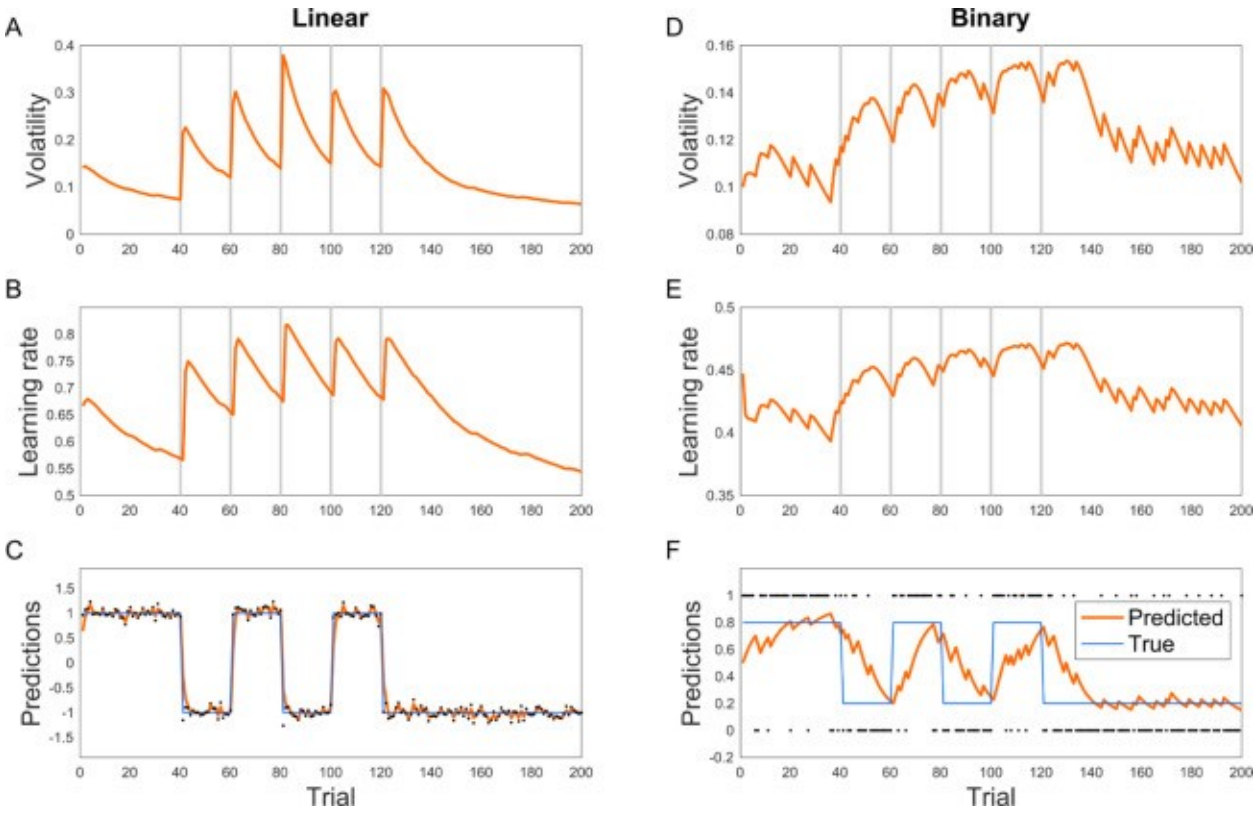

Fig 2. Behavior of the VKF. A-C) A switching probabilistic learning task in which observations are randomly drawn from a hidden state (with variance 0.01) that switches several times. The relationship between the hidden state and observations are either linear (A-C) or binary (D-F). The volatility signal with the VKF increases after a switch in the underlying hidden state and the learning rate signal closely follows the volatility. The grey lines show the switching time. Dots are actual observations. These parameters used for simulating VKF: $\lambda = 0.1$, $v_0 = 0.1$, $\sigma^2 = 0.1$, $\omega = 0.1$.

In another simulation analysis, we studied the performance of the binary VKF in a similar task, but now with binary observations, which were used in previous studies [6,8]. Here, observations are drawn from a Bernoulli distribution whose mean is given by the hidden state of the environment. The hidden state is a constant probability 0.8 or 0.2, except that it is reversed occasionally. As Fig 2 shows, predictions of the binary VKF match with the hidden state. Furthermore, the volatility signal increases when there is a dramatic change in the environment, and the learning rate is higher in the volatile condition. Similar simulation analyses with higher levels of volatility show the same behavior: the volatility signal and the learning rate are generally larger in volatile blocks, in which the environment frequently switches (Fig 3).

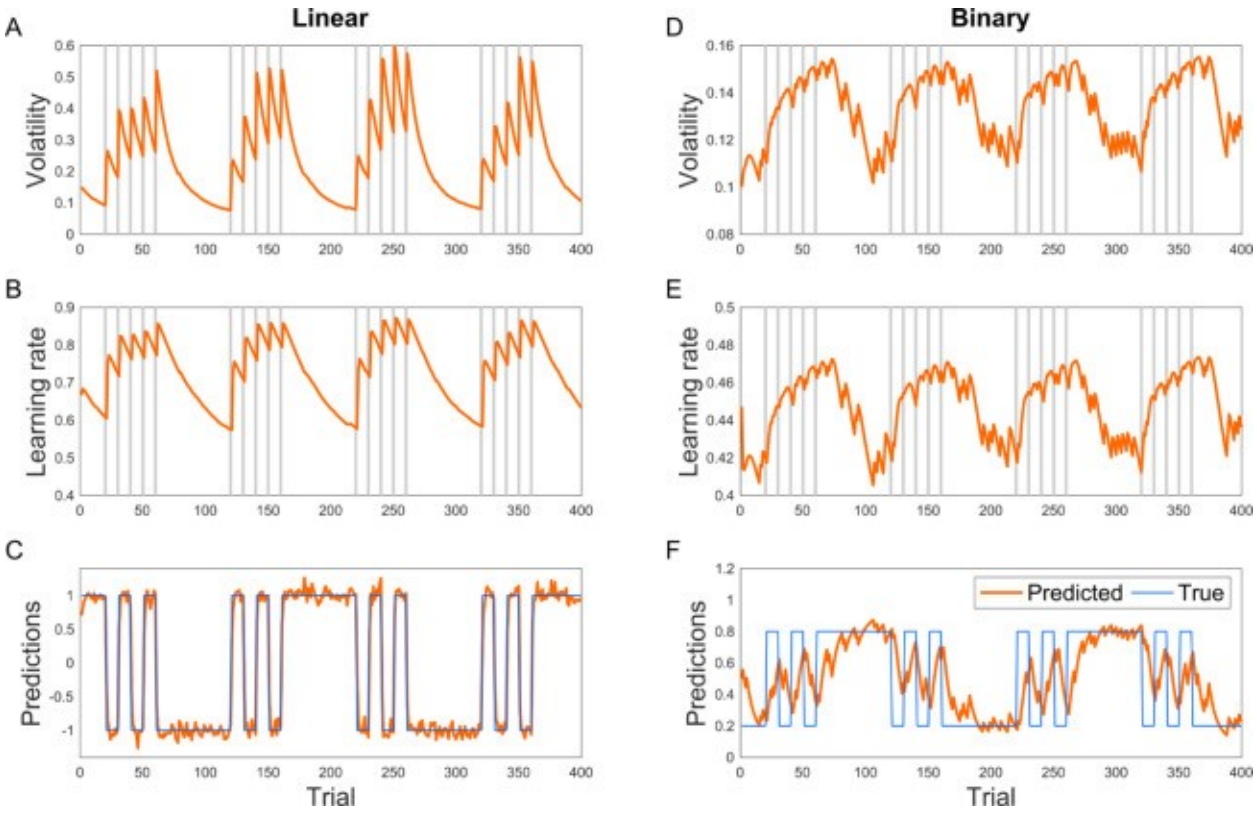

Fig 3. Behavior of the VKF in highly volatile condition for linear. (A-C) and binary (D-F) observations. The volatility and learning rate signal are larger in volatile conditions in which the hidden state is frequently changing. The parameters used were the same as Fig 2.

The VKF is a relatively simple model that approximates exact solution to an intractable inference problem. To further address the question how accurate is this approximation, next, we compared the performance of the VKF to other approaches: first, one representing (as close as was feasible) exact Bayesian inference, and, second, a different approach to approximate inference, the HGF.

Comparing VKF with the particle filter

We first compared the behavior of the VKF with a computationally expensive but near-exact method as the benchmark. For sequential data, the particle filter is a well-known Monte Carlo sequential sampling method, which approximates the posterior at every timepoint with an ensemble of samples; it approaches optimality in the limit of infinite samples. We used a Rao-Blackwellized particle filter (RBPF) [31] for this analysis, which combines sampling of some variables with analytical marginalization of others, conditional on these samples. In this way, it exploits the fact that inference on the lower level variable (i.e. x_t in Fig 1) is tractable using a Kalman filter, given samples from the upper level variable (i.e. z_t in Fig 1). Therefore, this approach essentially combines particle filter with Kalman filter. We used the same sequence

generated in previous analyses (Fig 2) and compared the RBPF algorithm with 10,000 samples as the benchmark, assuming the same parameters for both algorithms. As shown in Fig 4, the behavior of VKF is very well matched with that of this benchmark for Gaussian observations. In particular, predictions and volatility estimates of the two algorithms were highly correlated, with correlation coefficients of 1.00 and 0.95, respectively, in this task. We also quantified the relative error, defined as the average mismatch between predictions of VKF and ground truth lower-level states, x_t, measured in units of increased error relative to the benchmark error of RBPF (see Methods), as 22.9%.

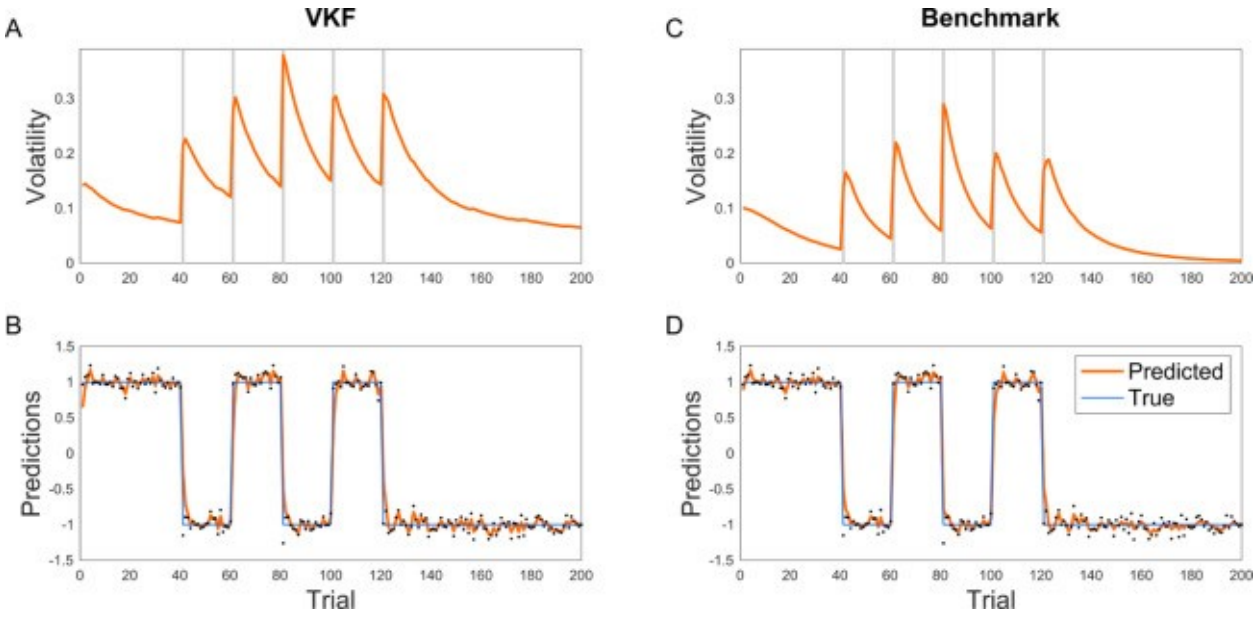

Fig 4. Comparison of VKF with the benchmark sampling methods. RBPF was used as the benchmark (with 10000 particles). The behavior of the VKF closely follows that of the benchmark. The parameters used were the same as Fig 2.

We next compared the performance of the binary VKF with that of the particle filter benchmark. For the binary VKF, the inference is more difficult because the relation between the state and the observation is also nonlinear. For the same reason, it is not possible to use RBPF to marginalize part of the computation here because the submodel conditional on the variance level is also not analytically tractable; instead we used a conventional particle filter sampling across both levels. We ran the particle filter on this problem with 10,000 samples and the same parameters as that of the binary VKF. As Fig 5 shows, the latent variables estimated by the VKF and the benchmark particle filter are again quite well matched, although the particle filter responds more sharply following contingency switches. Similar to the

previous analysis, predictions and volatility estimates were well correlated with correlation coefficients given by 0.91 and 0.68, respectively. The relative error in state estimation for binary VKF was 5.4%.

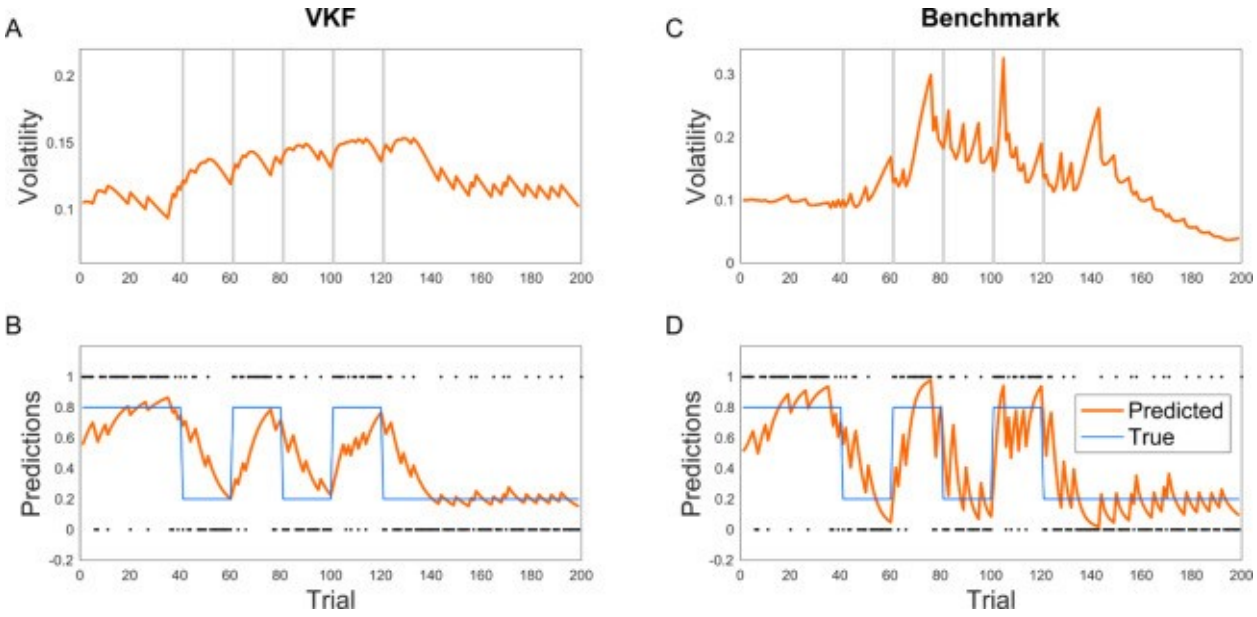

Fig 5. Comparison of binary VKF with the benchmark sampling methods. Particle filter was used as the benchmark (with 10000 particles). Both the VKF and the benchmark show higher learning rate in the volatile condition and estimated volatility signals by the two models are highly correlated. The parameters used were the same as Fig 2.

Comparing VKF with HGF

We next compared VKF to another approximate approach with more comparable computational complexity. In particular, we compared it to the HGF (with two latent levels), as the latter is the most commonly used algorithm for learning in volatile environments. Similar to our model, the generative model of the HGF assumes that there are two chains of state variables, x_t and z_t, in which the distribution over x_t is given by a normal distribution with the mean and variance depending on x_{t-1} and z_t, respectively. These two algorithms differ in two ways, however. First, for the generative models, the form of the process noise controlling the diffusion of z_t is different. In our model, the noise is a positive variable with a beta distribution. The HGF, however, assumes that the process noise for z_t is additive and normally-distributed, and hence exponential transformation is needed to ensure that the variance of x_t is non-negative. These differences, in turn, give rise to differences in the resulting approximate inference rules, particularly for inference over the volatility level, z_t. Our model has specific conjugacy properties, which

make the inference for z_t tractable if x_t is isolated from x_{t-1} (e.g. if x_{t-1} is fixed and under a variational approximation; see theoretical explanation above and S1 Appendix for mathematical proofs). However, even in this case when x_t is isolated from x_{t-1}, the inference over z_t in the HGF is not tractable and therefore requires further approximation. The approximate inference model of the HGF thus relies on a second-order Taylor approximation to deal with this issue, in addition to the variational approximation used by both the VKF and the HGF. By eliminating this additional approximation, the VKF avoids one source of potential inaccuracy. Also, the VKF's conjugacy ensures a simple (one-parameter Gamma) distribution for the posterior estimate at the top (volatility) level, which may also contribute to stability, vs. a two-parameter Gaussian approximation, in which numerical problems (e.g., negative values for the approximated posterior uncertainty) can arise.

However, directly comparing the effect of these two approaches to inference is difficult because they are closely linked with two different sets of generative assumptions. To focus our comparison on the approximations, we studied the ability of each algorithm to reproduce data generated from its own, corresponding generative model, in roughly comparable parameter regimes. Then to ensure any differences were related to the approximate inference itself, rather than to the different generative statistics themselves (e.g., supposing one problem is just harder), we scored each model in terms of error relative to optimal inference as proxied by the particle filter. This represents the best achievable performance for each specific generative process. We also performed a number of followup analyses to further pursue this question, in S1 Text.

Thus, we generated a sequence of observations using the generative model of the HGF (see Methods). We then used the generated sequence as the input to the HGF inference algorithm, using true generative parameters. We repeated this process 1000 times. In 82 of these simulations, the HGF encountered numerical problems: its inferred trajectory encountered numerical problems, i.e. negative estimates of the posterior variance over the top (volatility) level. This problem is due to the Taylor approximation used to extrapolate the variational posterior in the approximate inference model of the HGF. Only for the remaining 918 simulations were we able to quantify the error of HGF, defined as the mismatch between predicted and true lower level states x_t (see Methods). We also performed inference on the same generated data using a RBPF (derived under the HGF generative assumptions) [31], as a proxy of exact inference. We compared these two results to obtain a measure of fractional relative error in the HGF over and above the (unavoidable) error from the RBPF. Similarly, we quantified the error of the VKF relative to the RBPF: data were simulated under the VKF generative model (see Methods), which were then used as the input to the VKF inference model and its associated RBPF. This analysis revealed that the

relative error of the HGF and VKF is 21.4% (SE=6.8%) and 2.7% (SE=0.3%), respectively, indicating that the VKF performance is closer to the particle filter than the HGF. We performed a number of control analyses confirming these results with different sets of parameters, and using an alternative baseline independent of the RBPF (S1 Text).

Effects of volatility on learning rate for binary outcomes

Following Behrens' seminal study [6], the majority of previous work studying volatility estimation used binary outcomes. This is also important from a psychological perspective, because seminal models such as Pearce-Hall indicate that learning rate, or as it is called in that literature "associability", reflects the extent that the cue has been surprising in the past. In fact, modern accounts of learning in decision neuroscience are partly built on this classic psychological idea, and there is evidence that volatility-induced surprising events increase the learning rate in probabilistic learning tasks in humans [6,32–34]. Here, we highlight a crucial difference between the binary versions of the HGF and VKF regarding the relationship between volatility and learning rate.

We compared performance of HGF and VKF for binary observations, mimicking the sorts of experiments typically used to study volatility learning in the lab [6]. In this case, the generative dynamics of the latent variables did not match either model, but instead used a discrete change-point dynamics again derived from these empirical studies. The binary versions of the HGF and VKF use different approximations to deal with the nonlinear mapping between hidden states and binary observations, on top of the ones discussed before. Whereas the HGF employs another Taylor approximation to deal with binary observations, the binary VKF is based on a moment-matching approximation. Importantly, these different treatments of binary observations result in qualitative differences in the relationship between the learning rate and volatility. Accordingly, in the binary VKF, the learning rate closely follows the volatility estimate, similar to the VKF for continuous-valued observations and echoing the intuition from the psychological and decision neuroscience literatures. In the binary version of HGF, however, the relationship between the learning rate and volatility is more involved because the learning rate is also affected by the sigmoid transformation of the estimated mean on the previous trial. Fig 6 illustrates these signals in an example probabilistic switching task for both models. To demonstrate this point quantitatively, we generated 500 time-series and fitted both models to these time-series. The correlation between the learning rate and volatility estimates for these time-series were then calculated using the fitted parameters of each model. As expected, the learning rate and volatility signals of the binary VKF were positively correlated in all simulations (average correlation coefficient about 1.00), whereas those

of binary HGF were, unexpectedly, negatively correlated in all simulations (average correlation coefficient = –0.74). Note that this qualitative difference is a general behavior of these models and is not due to a particular setting of parameters (S1 Text).

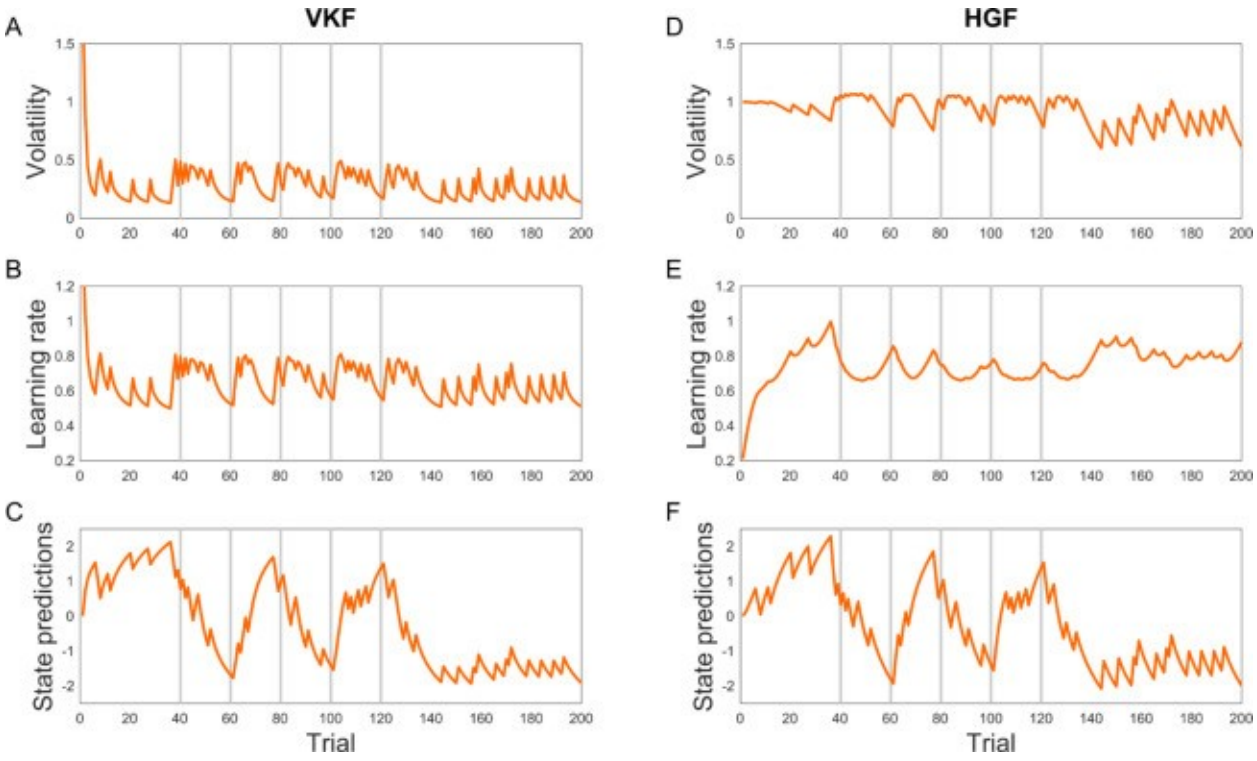

Fig 6. Behavior of the binary VKF and HGF. Predictions of the two models about the relationship between volatility and learning rate are different. Fitted parameters for each model were used for obtaining volatility, learning rate and state prediction signals here.

Testing VKF using empirical data

We then tested the explanatory power of the VKF to account for human data using two experimental datasets. In both experiments, human subjects performed a decision-making task, repeatedly choosing between two options, where only one of them was correct. Participants received binary feedback on every trial indicating which option was the correct one on that trial.

Such a test requires fitting the binary VKF to choice data by estimating its free parameters. The binary VKF relies on three free parameters: volatility learning rate, λ, the initial volatility value, v_0, and the noise parameter for binary outcomes, ω. Fig 7 shows the behavior of the binary VKF as a function of these parameters. As Fig 7 shows, the volatility learning rate parameter determines the degree by which

the volatility signal is updated and its effects are particularly salient following a contingency change. The effects of initial volatility are more prominent in earlier trials. The noise parameter indicates the scale of volatility throughout the task. Note that the noise parameter also has a net positive relationship with the learning rate in the binary VKF.

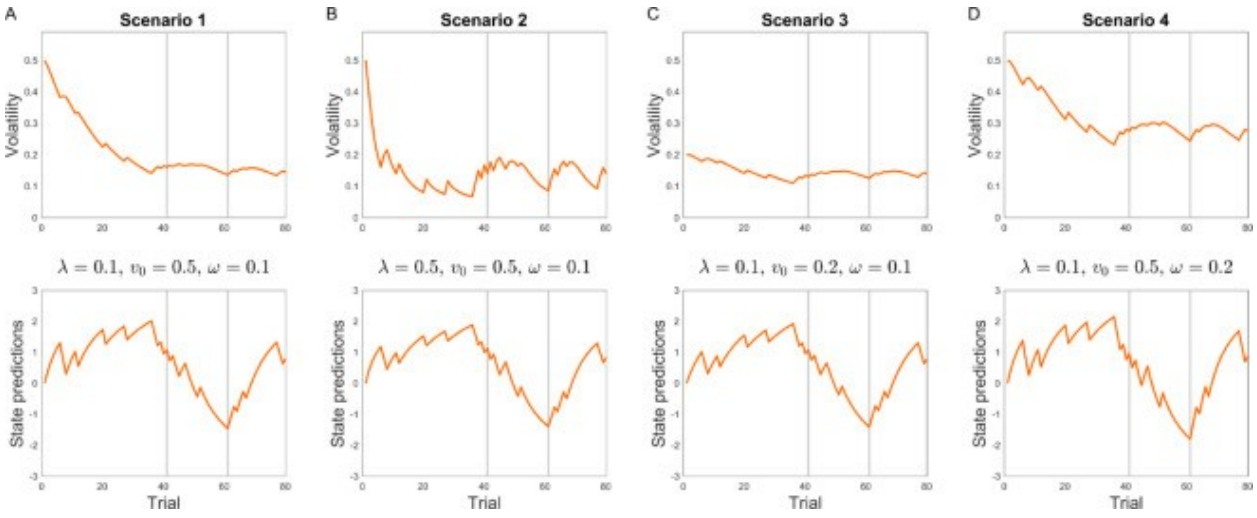

Fig 7. Effects of parameters on the VKF. A) The volatility and state prediction signals within the binary VKF are shown for a baseline scenario. B-D) The impact of changing parameters with respect to the baseline scenario is shown for (B) the volatility learning rate, λ, (C) the initial volatility v_0, and (D) the noise parameter, ω.

To model choice data, the predictions of the binary VKF were fed to a softmax function, which has a decision noise parameter, β. We first verified that these parameters can be reliably recovered using simulation analyses with synthetic datasets, in which observations and choices of 50 artificial subjects were generated based on the binary VKF and the softmax (see Methods for details of this analysis). We then fitted the parameters of the VKF to each dataset using hierarchical Bayesian inference (HBI) procedure [35], an empirical Bayes approach with the advantage that fits to individual subjects are constrained according to the group-level statistics. We repeated this simulation analysis 500 times. As reported in Table 1, this analysis revealed that the model parameters were fairly well recoverable, although estimation of the volatility learning rate and initial volatility were more prone to error.

Parameters	True parameters	Estimated parameters		
		25% quantile	Median	75% quantile
λ	0.2	0.20	0.23	0.27
v_0	5.0	4.80	5.24	5.64

ω	1.0	0.96	1.06	1.15
β	1.0	1.12	1.14	1.17

Table 1. Recovery analysis for parameters of the VKF. A dataset including 50 artificial subjects were generated based on the binary VKF and a softmax choice model. The same procedure used in analysis of empirical data (HBI) was used then to estimate the parameters. We have reported the mean of parameters across all 50 artificial subjects. This procedure was repeated 500 times.

In the first experiment (Fig 8), 44 participants carried out a probabilistic learning task (originally published in [36]), in which they were presented with facial cues and were asked to make either a go- or a no-go-response (i.e., press a button, or withhold a press, respectively) for each of four facial cues in order to obtain monetary reward or avoid monetary punishment. The cues were four combinations of the emotional content of the face image (happy or angry) and its background color (grey or yellow) representing valence of the outcome. Participants were instructed that the correct response is contingent on these cues. The response-outcome contingencies for the cues were probabilistic and manipulated independently, and reversed after a number of trials, varying between 5 and 15 trials, so that the experiment consisted of a number of blocks with varying trial length (Figure 4B). Within each block, the probability of a win was fixed. The task was performed in the scanner, but here we only focus on behavioral data. 120 trials were presented for each cue (480 trials in total). Participants learned the task effectively: performance, quantified as the number of correct decisions given the true underlying probability, was significantly higher than chance across the group ($t(43)=14.68$, $p<0.001$; mean: 0.68, standard error: 0.01). The focus of the original studies using this dataset was on socio-emotional modulation of learning and choice [34,36]. Here, however, we use this dataset because the task is a probabilistic learning tasks in which the contingencies change throughout the experiment. We considered a model space including the binary VKF, binary HGF, a Kalman filter (KL) that quantifies uncertainty but not volatility, and also a Rescorla-Wagner (RW) model that does not take into account uncertainty or vary its learning rate. For all these learning models, we used a softmax rule as the response model to generate probability of choice data according to action values derived for each model as well as value-independent biases in making a go response for all these models. Note that the response model also contained value-independent biases in making or avoiding a go response due to the emotional or reinforcing content of the cues (see Methods).

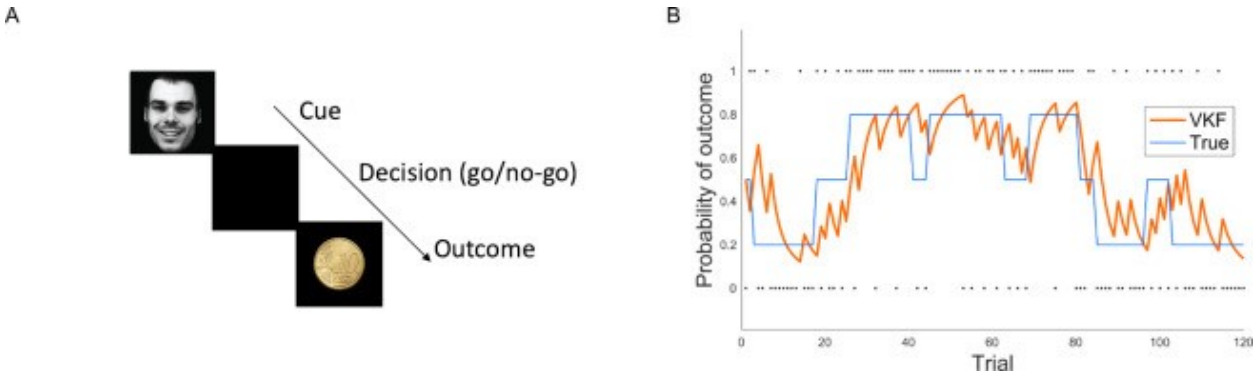

Fig 8. The probabilistic learning task in Experiment 1 used for testing the VKF. A) Participants had to respond (either go or no-go) after a face cue was presented. A probabilistic outcome was presented following a delay. B) An example of the probability sequence of outcome (of a go response) for one of the four trial-types and predictions of the (binary) VKF model. The response-outcome contingencies for the cues were probabilistic and manipulated independently. The dots show actual outcomes seen by the model.

This model space was then fit to choice data using HBI [35]. Importantly, the HBI combines advantages of hierarchical model for parameter estimation [37] with those of approaches that treat the model identity as a random effect [38], because it assumes that different subjects might express different models and estimates both the mixture of models and their parameters, in a single analysis. Thus, the HBI performs random effects model comparison by quantifying model evidence across the group (goodness of fit penalized by the complexity of the model [39]). This analysis revealed that, across participants, VKF was the superior model in 37 out of 44 participants (Fig 9). Furthermore, the protected exceedance probability in favor of VKF (i.e. the probability that a model is more commonly expressed than any other model in the model space [38,40] taking into account the null possibility that differences in model evidence might be due to chance [40]) was indistinguishable from 1. In a supplementary analysis, we also considered the particle filter model. Since that model is a Monte Carlo sampling method and too computationally intensive to embed within a further hierarchical estimation over subjects and other models, we fitted the model separately to each subject and compared it with the other models (see S1 Text for details). This analysis revealed that the VKF is the most parsimonious even compared to the particle filter model.

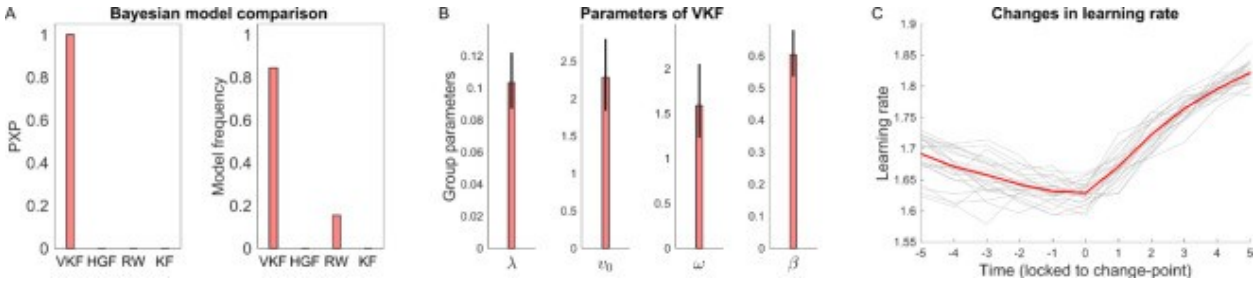

Fig 9. Bayesian analysis of VKF in the first experiment. A) Bayesian model comparison results comparing VKF with HGF, Rescorla-Wagner (RW) and Kalman filter (KF). Protected exceedance probabilities (PXPs) and model frequencies are reported. The PXP of the VKF is indistinguishable from 1 (and from 0 for other models) indicating that VKF is the most likely model at the group level. The model frequency metric indicates the ratio of subjects explained by each model. B) Estimated parameters of the VKF at the group-level. C) Learning rate signals, time-locked to the change points, estimated by the VKF for all participants (gray) and the mean across participants (red). The x-axis indicates trials relative to change points. The error-bars in B are obtained by applying the corresponding transformation function on the group-level error-bars obtained by the HBI [35] and, therefore, are not necessarily symmetric.

To examine the detailed behavior of the model, we also analyzed learning rate signals estimated by the VKF at the time of changes in action-outcome contingencies. For this analysis, we fitted each subject's choice data individually to the binary VKF model to generate learning rate signals independently. Fig 9C shows variations in the learning rate signal time-locked to the change points. Across 44 participants, 40 (i.e. 91%) showed a positive change in learning rate following change points. There was a significant difference between learning rate before (obtained by averaging over 5 trials prior to change) and after change points (obtained by averaging over 5 trials) (mean increase = 0.10, $P<0.001$), similar to previous studies with change-point dynamics [41].

In the second experiment, 174 participants performed a learning task (originally published in [42]), in which they chose to accept or reject an opportunity to gamble on the basis of their estimation of potential reward. Thus, subjects were asked to estimate the probability of reward based on binary feedback given on every trial. The reward probability was contingent on the category of the image presented during each trial and its time-course has been manipulated throughout the experiment. During each trial, participants were also presented with the value of a successful gamble. In the original study [42], Jang et al. used this learning task, in combination with a follow-up recognition memory test, to study the influences of computational factors governing reinforcement learning on episodic memory. Here, however, we only analyze the learning part of this dataset because the task is a probabilistic learning task with switching contingencies. For every image category, the reward probability has switched at least twice during the task. Most participants learned the task effectively, as their gambling decisions varied in

accordance with the probability of reward. In particular, they accepted to gamble more often in trials with high reward probability than those with low reward probability ($t(173)=18.5$, $p<0.001$; mean: 0.25, standard error: 0.01). Thirteen subjects were excluded from the analysis because a logistic regression (with intercept and two regressors: reward probability and trial value) showed a negative correlation between reward probability and their decisions to gamble, suggesting that they had not understood the task instructions. We analyzed data of the remaining 161 subjects.

We fitted the same model space used above, in which the learning models were combined with the softmax as the response model to generate probability of choices. This model space was then fit to choice data using the HBI (Fig 10). This analysis revealed that the binary VKF outperformed other models in 102 out of 161 participants (with 0.62 model frequency). Across all participants, the protected exceedance probability in favor of the VKF was indistinguishable from 1. Note that the second-best model was the Kalman filter model with model frequency of 0.3, suggesting that about 30% of subjects reduced their learning rate over time but were not sensitive to volatility. This may be because there are overall fewer switches in this task compared to the previous one. Nevertheless, across all subjects, analysis of learning rate signals time-locked to the change point shows a significant increase in learning rate (median increase = 0.03, Wilcoxon sign rank test because samples were not normally distributed, $p<0.001$). This effect was positive for 64% of participants.

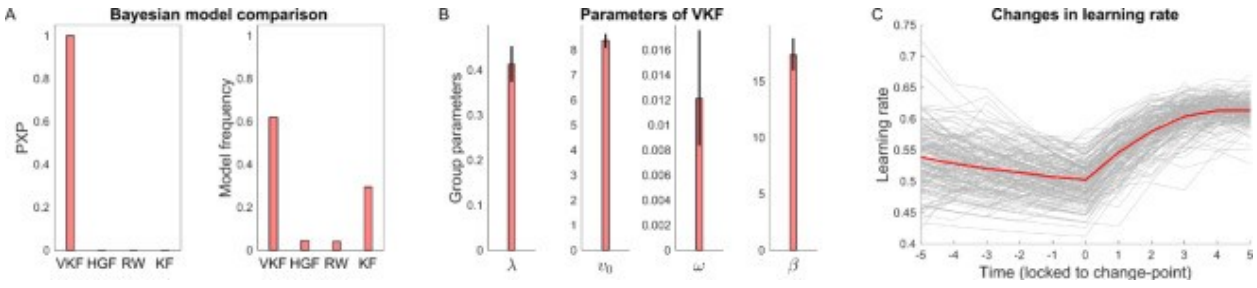

Fig 10. Bayesian analysis of VKF in the second experiment. A) Bayesian model comparison results comparing VKF with HGF, Rescorla-Wagner (RW) and Kalman filter (KF). Similar to the previous dataset, the PXP of the VKF is indistinguishable from 1 (and from 0 for other models) indicating that VKF is the most likely model at the group level. B) Estimated parameters of the VKF at the group-level. C) Learning rate signals, time-locked to the change points, estimated by the VKF for all participants (gray) and the mean across participants (red). The x-axis indicates trials relative to change points. The error-bars in B are obtained by applying the corresponding transformation function on the group-level error-bars obtained by the HBI [35] and, therefore, are not necessarily symmetric.

Discussion

In this work, we have introduced a novel model for learning in volatile environments. The proposed model has theoretical advantages over existing models of learning in volatile environments, because it is based on a novel generative model of volatility that makes it possible to have a simple approximate inference model, which is also very loyal to the exact inference. Using empirical choice data in a probabilistic learning task, we showed that this model captures human behavior better than the state-of-the-art HGF.

The Kalman filter is the cornerstone of tracking theories, with widespread applications in many technological and scientific domains including psychology and neuroscience [4,7,22,23]. For example, in movement neuroscience, the Kalman filter has been used as a model of how the brain tracks sensory consequences caused by a motor command. In learning theories, the Kalman filter provides a normative foundation for selective attention among multiple conditioned stimuli predicting a target stimulus, such as food. Nevertheless, the Kalman filter is limited to environments in which the structure of process noise is constant and known. Like other models such as the HGF, the VKF fills this gap by extending the Kalman filter to inferring the process variance in volatile environments in which the variance is itself dynamically changing. In particular, VKF contains two free parameters, a volatility update rate (i.e. λ) indicating the extent of noise in process variance dynamics and the initial value of volatility (v_0). The Kalman filter is a special case of VKF, in which the volatility update rate is zero and the constant process noise is equal to v_0.

A complete understanding of a learning system requires understanding of how computational theories should be realized at the algorithmic level [17]. Previous works have shown that at this level, the normative perspective might shed light on or even encompass related psychological models, as for example temporal difference models of reinforcement learning encompass the classical Rescorla-Wagner theory. The proposed model builds a normative foundation for the intuitive hybrid models combining critical features of Rescorla-Wagner and Pearce-Hall theories for conditioning [20,21,34]. Specifically, the learning process of VKF depends on two components. The first component is the classical prediction error signal, which is defined as the difference between the observed and expected outcome, similar to Rescorla-Wagner error signal. The second component is the learning rate modulated by the volatility estimate, which is itself a function of surprise (i.e. the squared prediction error) according to Equation 13. Therefore, although the detailed algebraic computation of the surprise term slightly differs, the structure of the model is consistent with the classical Pearce-Hall model, and like it the rate of learning in VKF depends on surprise. This construction clarifies the relationship between the Pearce-Hall associability, its

update, and volatility inference in hierarchical learning models such as the HGF and that of Behrens et al. [6].

The generative model of VKF is based on a novel state-space model of variance dynamics for Gaussian-distributed observations with known mean, in which the inference (for the volatility level considered in isolation) is tractable. This particular generative model leads to exact inference without resorting to any approximation. To build a fully generative model for volatile environments, we then combined this state-space model with the state-space model of the Kalman filter. Therefore, the full generative model of VKF contains two temporal chains (Fig 1), one for generating the mean and the other one for generating the variance. Although the inference is tractable within each chain, the combination of both chains makes the exact inference intractable. Therefore, we used structured variational techniques for making approximate inference, which isolates the tractable submodel as part of the variational maximization.

The state-of-the-art algorithm for learning in volatile environments is HGF [8,18], which is a flexible filter which can be extended iteratively to hierarchies of arbitrary depth. The VKF has the same conditional dependencies as the HGF with two latent levels. There are, however, two critical differences between the VKF and the HGF. First, the generative process underlying the variance is different between the two models. The HGF assumes additive Gaussian diffusion at each level, transformed by an exponential to allow it to serve as a variance for the next level down. In contrast, the generative model of process variance in the VKF uses a form of multiplicative diffusion, which guarantees non-negativity. Secondly, these generative differences lead to algorithmic ones. Although both models conduct inference under a variational approximation, due to the nonlinearity in the HGF, it is not possible to analytically maximize the variational family, and additional approximations are required. As mentioned, the generative process of the VKF is tailored to permit exact maximization of the variational distribution. Thus altogether, the VKF trades away the more general generative structure that underlies the HGF, to achieve a simpler and more accurate approximation to a more specific (two latent chains) case. We compared the performance of VKF and HGF in predicting human choice data in a probabilistic learning task, similar to those tasks that have been modeled with HGF in the recent past [8]. Bayesian model comparison showed that the VKF predicts choice data better in the majority of participants.

There is an additional difference between these models for binary outcomes, which results in a qualitative difference in the relationship of the volatility signal and learning rate (Fig 6). Consistent with classical Pearce-Hall models, surprising events increase learning rate in our model, a quality that is expected based on normative considerations [6] and empirical observations [6,32–34]. As we have shown

using simulation analyses, volatility and learning rate signals are negatively correlated for binary HGF, which is a consequence of using a Taylor approximation to account for binary data. We employed a different approximation strategy to account for binary outcomes, called moment matching, which preserves the positive relationship between volatility and learning rate.

Recent work highlight the importance of uncertainty processing and its effects on learning rate for understanding a number of psychiatric disorders [33,34,43–45]. For example, Browning et al. [33] found that anxiety reduces people's ability to adjust their learning rate according to volatility. In a recent study [34], we also found that in threatening contexts evoked by angry face images, socially anxious individuals show disruptions in benefiting from stability in the action-outcome contingency. Their dorsal anterior cingulate cortex, a region previously shown to reflect volatility estimates [6,32], also failed to reflect volatility in those contexts. This is because anxious individuals updated their expectations too rapidly in the stable conditions, possibly because they perceive any uncertainty as a signal of contingency change (i.e. volatility). Process-level models, such as VKF, can play important role in this line of research and we hope that this work be useful for quantifying critical computations underlying learning in the future.

In the last decade, scholars in the field of computational psychiatry have started to map deficits in decision making to parameters of computational models. In fact, these parameters could serve as computationally-interpretable summary statistics of decision making and learning processes. Parameters of the VKF are also particularly useful for such a mapping. There are three parameters that influence different aspects of learning dynamics in the VKF. The volatility update parameter captures the degree that the individual updates its estimate of volatility. Given the generative model of the VKF, this parameter also determines the subjective feeling of noise at the volatility level. Another parameter of the VKF is the initial volatility, which influences the learning process on early trials. These two parameters have similar effects for linear and binary observations. For the binary VKF, there is another parameter that is only relevant to the inference model (not the generative process), ω, which we called it the noise parameter. As shown in Fig 8, this parameter governs the scale of volatility and learning rate throughout the learning process. Notably, our simulation analyses showed that these parameters are fairly identifiable from choice data.

In this study, we assumed that the observation noise in the generative process of the VKF is a free parameter, σ; the HGF and traditional Kalman filters also have analogous parameters. In many situations, however, humans and other animals might have to learn the value of this noise. In particular, in addition to volatility, trial-by-trial estimation of this noise is relevant for optimal learning in situations in which

observation noise might dynamically change. Deriving an efficient inference model for simultaneous tracking of both signals is substantially more difficult due to dependencies arising between variables. Our generative model of the variance and the corresponding tractable inference can be helpful for that purpose, which should be further explored in future studies.

The goal of the current study was to further process-level models of volatility by proposing a model that closely match optimal statistical inference, building on a number of studies that have been proposed in the past 15 years. Since exact inference is not possible, these studies have relied on different approximate inference approaches, such as sampling, Taylor approximation, variational inference or message-passing algorithms. Our approach for treatment of binary observations using moment matching is similar to the message-passing approach taken recently for studying dynamical systems [46,47]. We chose this approach rather than the Taylor approximation used previously [18] because it has been shown that methods based on moment matching perform better than derivative-based methods in approximating exact inference for binary Gaussian process models [48].

An important concern about Bayesian process-level models is whether their computations are biologically plausible. Similar to any other model that extends the Kalman filter, normalization is required for computing the Kalman gain in the VKF. Furthermore, our model requires the squared prediction error for updating volatility. Although performing these computations might not be straightforward with current neural network models, they are not biologically implausible. Another crucial question is how these approximate Bayesian models could be realized at the mechanistic level [49]. Recently, metaplasticity has been proposed as a mechanistic principle for learning under uncertainty [9,12,50]. Metaplasticity allows synaptic states to change without substantial changes in synaptic efficacy [51] and therefore provides a mechanism for reinforcement learning under volatility [9].

Bayesian models have recently been used for online inference of sudden changes in the environment [41,52]. Although those situations can be modeled with generative processes with discrete change-point detection, Behrens et al. [6] showed that models with volatility estimate might be as good as or even better than models with specific discrete change-points. Our simulations also showed that VKF can be successfully applied to those situations. In such situations, the volatility signal plays the role of a continuous change-point estimator, which substantially increases after a major change. This is because those sudden changes in the environments cause a large “unexpected uncertainty” signal [7,11], which substantially increases the volatility.

In this article, we introduced a novel model for learning under uncertainty. The VKF is more accurate than existing models and explains human choice data better than alternatives. This work

provides new opportunities to characterize neural processes underlying decision making in uncertain environments in healthy and psychiatric conditions.

Methods

Simulation analysis for comparing VKF with benchmark

We implemented particle filter models using MATLAB Control Systems Toolbox. The particle filter is a sequential Monte Carlo method, which draws samples (i.e. particles) from the generative process and sequentially updates them. We implemented this model separately for the linear and binary observations, with 10000 particles and generative parameters. Note that for linear outcomes, inference on the lower level chain is analytically tractable given samples from the upper level chain. Therefore, we used RBPF [31] for the linear problem, which combines Monte Carlo sampling with analytical marginalization. We used Spearman rank correlation because signals were not normally distributed.

Simulation analysis for comparing accuracy of VKF and HGF

The generative model of the HGF with 2 levels is based on a probabilistic model with the same dependencies as those in our generative model (Fig 1). Under this generative model, 3 chains of random variables are hierarchically organized to generate observation, o_t:

$$x_3^{(t)} \sim Normal\left(x_3^{(t-1)}, \nu\right),$$

$$x_2^{(t)} \sim Normal(x_2^{(t-1)}, \exp(\kappa x_3^{(t)} + \omega)),$$

$$x_1^{(t)} \sim Normal\left(x_2^{(t-1)}, \sigma^2\right),$$

$$o_t = x_1^{(t)}$$

where $\nu > 0$ is the variance at the third level, $\kappa > 0$ determines the extent to which the third level affects the second level, ω indicates the tonic level variance at the second level, and σ is the observation noise. For the simulation analysis of the accuracy VKF and HGF, we generated data according to the HGF generative model (with normal observations) according to these parameters: $\nu = 0.5$, $\kappa = 1$, $\omega = -3$, and $\sigma = 1$. The initial mean at the second and third levels were 0 and 1, respectively. We generated 1000 time-series (100 trials) using these parameters. These parameters were also then used for inference based on the HGF algorithm. However, in 82 simulations, the HGF encountered numerical problems (negative posterior variance), because its inferred trajectory conflicted with the assumptions of the inference model. For the remaining time-series, we also performed inference using an RBPF under the generative model of the HGF and the true parameters (10000 particles). The RBPF exploits the fact that inference on

the lower level chain (i.e. x_2) is tractable given samples from the upper level. Specifically, this approach samples x_3 and marginalizes out x_2 using the Kalman filter. The relative error of the HGF with respect to RBPF was defined as $E_{HGF}/E_{RBPF-HGF} - 1$, in which E_{HGF} and $E_{RBPF-HGF}$ are the median of absolute differences between the ground truth generated latent variable, x_2, and the predictions of the HGF and the particle filter, respectively. A similar analysis was performed by generating 1000 time-series (100 trials) using the generative model of the VKF with parameters $\lambda = 0.15$, $v_0 = 1$, and $\sigma = 1$. We also performed inference using the RBPF based on VKF generative assumptions and the true parameters and computed the relative error for VKF, $E_{VKF}/E_{RBPF-VKF} - 1$. For both algorithms, we computed correlation coefficient between their estimated signals at both levels and those from the corresponding RBPF. To compute correlation coefficient across simulations, correlation coefficients were Fisher-transformed, averaged, and transformed back to correlation space by inverse Fisher transform [53,54].

For comparing accuracy of the VKF and HGF for binary observations, we generated 500 time-series in the probabilistic switching task with binary outcomes (Fig 2). Parameters of the HGF and VKF were fitted to these time-series using a maximum-a-posteriori procedure, in which the variance over all parameters was assumed to be 15.23. The prior mean for all parameters (except ω in HGF) was assumed to be 0. We followed the suggestions made by Mathys et al. [55] and assumed an upper bound of 1 for both ν and κ. We particularly chose –1 as the initial prior mean for ω to ensure that the HGF is well defined at the prior mean.

Recovery analysis of parameters

For this analysis, data were generated based on the binary VKF (Equations 14-19). In particular, the observation on trial t, o_t, was randomly drawn based on the sigmoid-transformation of m_{t-1}. The choice data were also generated randomly by applying the softmax as the response model with parameter β. Similar to experiment 1, for each artificial subject, we assumed 4 sequences of observations and actions (i.e. 4 cues) with 120 trials. These values were used as the group parameters: $\lambda = 0.2$, $v_0 = 5$, $\omega = 1$, and $\beta = 1$. For generating synthetic datasets for simulations, the parameters of the group of subjects (50 subjects) assigned to each model were drawn from a normal distribution with the standard deviation of 0.5.

Implementation of models for analysis of choice data

We considered a 3-level (binary) HGF [18] for analysis of choice data, with parameters, $0 < \nu < 1$, $0 < \kappa < 1$, and ω. We also considered a constant parameter for ω at –4, as Iglesias et al. [8]. However, since

the original HGF with a free ω outperformed this model (using maximum-a-posteriori estimation and random effects model comparison [38,40]), we included the original HGF in the model comparison with binary VKF. Similar to the HGF toolbox, we assumed that the initial mean at the second and third levels are 0 and 1, respectively, and the initial variance of the second and third levels are 0.1 and 1, respectively. For implementing the binary VKF, we assumed an upper bound of 10 for the initial volatility parameter, v_0.

Ethics statement

All human subjects data used here are reanalyses of anonymized data from previously published studies. Data from human subjects in experiment 1 are from a study [36] that was approved by the local ethical committee ("Comissie Mensgebonden Onderzoek" Arnhem-Nijmegen, Netherlands). Data from human subjects in experiment 2 are reported by Jang et al. [42] and the study was approved by the Brown University Institutional Review Board.

Experiment 1

Each of the learning models was combined with a choice model to generate probabilistic predictions of choice data. Expected values were used to calculate the probability of actions, a_1 (go response) and a_2 (no-go response), according to a sigmoid (softmax) function:

$$p_t(a_1) = \frac{1}{1 + e^{-\beta \bar{m}_t(s_t, a_1) - b(s_t)}}$$

$$p_t(a_2) = 1 - p_t(a_1)$$

where $\bar{m}_t$ was equal to m_t for the VKF, and it was equal to $2\sigma(\hat{\mu}_1^t) - 1$ for the HGF (as implemented in the HGF toolbox). Moreover, β is the decision noise parameter encoding the extent to which learned contingencies affect choice (constrained to be positive) and $b(s_t)$ is the bias towards a_1 due to the stimulus presented independent from learned values. The bias is defined based on three free parameters, representing bias due to the emotional content (happy or angry), b_e, bias due to the anticipated outcome valence (reward or punishment) cued by the stimulus, b_v, and bias due to the interaction of emotional content and outcome, b_i. No constraint was assumed for the three bias parameters. For example, a positive value of b_e represents tendencies towards a go response for happy stimuli and for avoiding a go response for angry stimuli (regardless of the expected values). Similarly, a positive value of b_v represents a tendency towards a go-response for rewarding stimuli regardless of the expected value of the go response. Critically, we also considered the possibility of an interaction effect in bias encoded by b_i. Therefore, the bias, $b(s_t)$, for the happy and rewarding stimulus is $b_e + b_v + b_i$, the bias for the angry

and punishing stimulus is $-b_e - b_v + b_i$, the bias for the happy and punishing stimulus is $b_e - b_v - b_i$ and the bias for the angry and rewarding stimulus is $-b_e + b_v - b_i$.

Experiment 2

This experiment was conducted by Jang et al. [42] to test the effects of computational signals governing reinforcement learning on episodic memory. The learning task consisted of 160 trials. On each trial, the trial value was first presented, followed by the image (from one of the animate or inanimate categories), response (play or pass) and the feedback. The feedback was contingent on the response made by the participant and given the probability of reward given the image category. Thus, if the participant chose play and the trial was rewarding, they were rewarded the amount shown as the trial value. If they chose to play and the trial was not rewarding, they lost 10 points. If the choice was pass, the participant did not earn any reward (i.e. 0 point), but was shown the hypothetical reward of choosing play. Data were collected using Amazon Mechanical Turk.

For modeling choice data, the softmax function with parameter β was used as the response model, in which the expected value of play was calculated based on the probability of reward estimated by the learning models and the value of trial shown at the beginning of each trial. The expected values were divided by 100 (the maximum trial value in the task) before being fed to the softmax to avoid numerical problems.

Model fitting and comparison

We used a hierarchical and Bayesian inference method, HBI [35], to fit models to choice data. The HBI performs concurrent model fitting and model comparison with the advantage that fits to individual subjects are constrained according to hierarchical priors based on the group-level statistics (i.e. empirical priors). Furthermore, the HBI takes a random effects approach to both parameter estimation and model comparison and calculates both the group-level statistics and model evidence of a model based on responsibility parameters (i.e. the posterior probability that the model explains each subject's choice data). The HBI quantifies the protected exceedance probability and model frequency of each model, as well as the group-level mean parameters and corresponding hierarchical errors. This method fits parameters in the infinite real-space and transformed them to obtain *actual* parameters fed to the models. Appropriate transform functions were used for this purpose: the sigmoid function to transform parameters bounded in the unit range or with an upper bound and the exponential function to transform

parameters bounded to positive value. To ensure that the HGF is well defined at the initial prior mean of fitted parameters (i.e. zero), we assumed $\omega = \omega_f - 1$, where ω_f is the fitted parameter.

The initial parameters of all models were obtained using maximum-a-posteriori procedure, in which the initial prior mean and variance for all parameters were assumed to be 0 and 6.25, respectively. This initial variance is chosen to ensure that the parameters could vary in a wide range with no substantial effect of prior. These parameters were then used to initialize the HBI. The HBI algorithm is available online at https://github.com/payampiray/cbm.

Supporting information

S1 Text. Supplementary results.

S1 Appendix. Formal treatment of the VKF model.

Acknowledgment

We are grateful to Matthew Nassar for sharing the empirical dataset used here.

References

1. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press; 1998.
2. Rescorla RA, Wagner AR, Black AH, Prokasy WF. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement. Classical Conditioning II: Current Research and Theory. New York: Appleton Century-Crofts; 1972. pp. 64–69.
3. Pearce JM, Hall G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev. 1980;87: 532–552.
4. Dayan P, Kakade S, Montague PR. Learning and selective attention. Nat Neurosci. 2000;3 Suppl: 1218–1223. doi:10.1038/81504
5. Kalman RE. A New Approach to Linear Filtering and Prediction Problems. Trans ASME--J Basic Eng. 1960;82: 35–45.
6. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. Nat Neurosci. 2007;10: 1214–1221. doi:10.1038/nn1954
7. Dayan P, Yu A. Uncertainty and Learning. IETE J Res. 2003;49: 171–181. doi:10.1080/03772063.2003.11416335
8. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron. 2013;80: 519–530. doi:10.1016/j.neuron.2013.09.009
9. Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, Soltani A. Metaplasticity as a Neural Substrate for Adaptive Learning and Choice under Uncertainty. Neuron. 2017;94: 401-414.e6. doi:10.1016/j.neuron.2017.03.044
10. Payzan-LeNestour E, Bossaerts P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput Biol. 2011;7: e1001048. doi:10.1371/journal.pcbi.1001048
11. Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. Neuron. 2005;46: 681–692. doi:10.1016/j.neuron.2005.04.026
12. Iigaya K, Fonseca MS, Murakami M, Mainen ZF, Dayan P. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. Nat Commun. 2018;9. doi:10.1038/s41467-018-04840-2
13. de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, et al. Computations of uncertainty mediate acute stress responses in humans. Nat Commun. 2016;7: 10996. doi:10.1038/ncomms10996
14. Katthagen T, Mathys C, Deserno L, Walter H, Kathmann N, Heinz A, et al. Modeling subjective relevance in schizophrenia and its relation to aberrant salience. PLoS Comput Biol. 2018;14: e1006319. doi:10.1371/journal.pcbi.1006319
15. Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment. Nat Neurosci. 2017;20: 1293–1299. doi:10.1038/nn.4615
16. Powers AR, Mathys C, Corlett PR. Pavlovian Conditioning-Induced Hallucinations Result from Overweighting of Perceptual Priors. Science. 2017;357: 596–600. doi:10.1126/science.aan3458
17. Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. 1st ed. San Francisco (CA): W. H. Freeman and Company; 1982.

18. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. Front Hum Neurosci. 2011;5: 39. doi:10.3389/fnhum.2011.00039
19. Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK. An Introduction to Variational Methods for Graphical Models. Mach Learn. 1999;37: 183–233. doi:10.1023/A:1007665907178
20. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND. Differential roles of human striatum and amygdala in associative learning. Nat Neurosci. 2011;14: 1250–1252. doi:10.1038/nn.2904
21. Le Pelley ME. The role of associative history in models of associative learning: a selective review and a hybrid model. Q J Exp Psychol B. 2004;57: 193–243. doi:10.1080/02724990344000141
22. Wolpert DM. Probabilistic models in human sensorimotor control. Hum Mov Sci. 2007;26: 511–524. doi:10.1016/j.humov.2007.05.005
23. Wolpert DM, Ghahramani Z. Computational principles of movement neuroscience. Nat Neurosci. 2000;3 Suppl: 1212–1217. doi:10.1038/81497
24. Smith RL, Miller JE. A Non-Gaussian State Space Model and Application to Prediction of Records. J R Stat Soc Ser B Methodol. 1986;48: 79–88. doi:10.2307/2345640
25. Gamerman D, dos Santos TR, Franco GC. A Non-Gaussian Family of State-Space Models with Exact Marginal Likelihood. J Time Ser Anal. 2013;34: 625–645. doi:10.1111/jtsa.12039
26. West M. On Scale Mixtures of Normal Distributions. Biometrika. 1987;74: 646–648.
27. Saul L, Jordan MI. Exploiting Tractable Substructures in Intractable Networks. Advances in Neural Information Processing Systems 8. MIT Press; 1995. pp. 486–492.
28. Ghahramani Z, Jordan MI. Factorial Hidden Markov Models | SpringerLink. Mach Learn. 1997;29: 245–273.
29. Minka TP. Expectation Propagation for Approximate Bayesian Inference. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2001. pp. 362–369. Available: http://dl.acm.org/citation.cfm?id=2074022.2074067
30. Boyen X, Koller D. Tractable Inference for Complex Stochastic Processes. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1998. pp. 33–42. Available: http://dl.acm.org/citation.cfm?id=2074094.2074099
31. Doucet A, Freitas N de, Murphy KP, Russell SJ. Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks. Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2000. pp. 176–183. Available: http://dl.acm.org/citation.cfm?id=647234.720075
32. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Associative learning of social value. Nature. 2008;456: 245–249. doi:10.1038/nature07538
33. Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ. Anxious individuals have difficulty learning the causal statistics of aversive environments. Nat Neurosci. 2015;18: 590–596. doi:10.1038/nn.3961
34. Piray P, Ly V, Roelofs K, Cools R, Toni I. Emotionally Aversive Cues Suppress Neural Systems Underlying Optimal Learning in Socially Anxious Individuals. J Neurosci. 2019;39: 1445–1456. doi:10.1523/JNEUROSCI.1394-18.2018

35. Piray P, Dezfouli A, Heskes T, Frank MJ, Daw ND. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. PLOS Comput Biol. 2019;15: e1007043. doi:10.1371/journal.pcbi.1007043
36. Ly V, Cools R, Roelofs K. Aversive disinhibition of behavior and striatal signaling in social avoidance. Soc Cogn Affect Neurosci. 2014;9: 1530–1536. doi:10.1093/scan/nst145
37. Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. PLoS Comput Biol. 2011;7: e1002028. doi:10.1371/journal.pcbi.1002028
38. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. NeuroImage. 2009;46: 1004–1017. doi:10.1016/j.neuroimage.2009.03.025
39. MacKay DJC. Information Theory, Inference, and Learning Algorithms. Cambridge University Press; 2003.
40. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies - revisited. NeuroImage. 2014;84: 971–985. doi:10.1016/j.neuroimage.2013.08.065
41. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci. 2010;30: 12366–12378. doi:10.1523/JNEUROSCI.0822-10.2010
42. Jang AI, Nassar MR, Dillon DG, Frank MJ. Positive reward prediction errors during decision-making strengthen memory encoding. Nat Hum Behav. 2019;3: 719–732. doi:10.1038/s41562-019-0597-3
43. Piray P, Zeighami Y, Bahrami F, Eissa AM, Hewedi DH, Moustafa AA. Impulse control disorders in Parkinson's disease are associated with dysfunction in stimulus valuation but not action valuation. J Neurosci. 2014;34: 7814–7824. doi:10.1523/JNEUROSCI.4063-13.2014
44. Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci. 2011;14: 154–162. doi:10.1038/nn.2723
45. Eldar E, Rutledge RB, Dolan RJ, Niv Y. Mood as Representation of Momentum. Trends Cogn Sci. 2016;20: 15–24. doi:10.1016/j.tics.2015.07.010
46. Wadehn F, Weber T, Loeliger H-A. State Space Models with Dynamical and Sparse Variances, and Inference by EM Message Passing. 27th European Signal Processing Conference (EUSIPCO). 2019.
47. Şenöz İ, de Vries B. ONLINE VARIATIONAL MESSAGE PASSING IN THE HIERARCHICAL GAUSSIAN FILTER. 28th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE; 2018. pp. 1–6. doi:10.1109/MLSP.2018.8517019
48. Kuss M, Rasmussen CE. Assessing Approximate Inference for Binary Gaussian Process Classification. J Mach Learn Res. 2005;6: 1679–1704.
49. Soltani A, Izquierdo A. Adaptive learning under expected and unexpected uncertainty. Nat Rev Neurosci. 2019;20: 635–644. doi:10.1038/s41583-019-0180-y
50. Khorsand P, Soltani A. Optimal structure of metaplasticity for adaptive learning. PLoS Comput Biol. 2017;13: e1005630. doi:10.1371/journal.pcbi.1005630
51. Abraham WC. Metaplasticity: tuning synapses and networks for plasticity. Nat Rev Neurosci. 2008;9: 387. doi:10.1038/nrn2356
52. Wilson RC, Nassar MR, Gold JI. Bayesian On-line Learning of the Hazard Rate in Change-Point Problems. Neural Comput. 2010;22: 2452–2476. doi:10.1162/NECO_a_00007

53. Cohen MX, Schoene-Bake J-C, Elger CE, Weber B. Connectivity-based segregation of the human striatum predicts personality characteristics. Nat Neurosci. 2009;12: 32–34. doi:10.1038/nn.2228
54. Piray P, den Ouden HEM, van der Schaaf ME, Toni I, Cools R. Dopaminergic Modulation of the Functional Ventrodorsal Architecture of the Human Striatum. Cereb Cortex N Y N 1991. 2017;27: 485–495. doi:10.1093/cercor/bhv243
55. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. Uncertainty in perception and the Hierarchical Gaussian Filter. Front Hum Neurosci. 2014;8: 825. doi:10.3389/fnhum.2014.00825