A transparent 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 environments 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. Algorithmically, the proposed model is simpler and more transparent than existing models, 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, the VKF is more accurate, particularly in estimating volatility, as it is based on more faithful approximations to the exact inference. Accordingly, when fit to empirical data, the VKF is better behaved than alternatives and better captures human choice data in a probabilistic learning task. The proposed model provides a transparent and coherent account of learning in stable or volatile environments and has implications for decision neuroscience research.

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 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. In situations more complicated than the smooth world of the Kalman filter, however, exact inference is generally intractable and additional approximations are required to achieve a practical, algorithmic- or process-level inference model. An influential model, proposed by Behrens and colleagues [6], comprises a hierarchical generative model for learning, in which a variable governing the observed associations fluctuates with a speed determined by a second higher-level, fluctuating variable. Although this model has been conceptually influential, it lacks a tractable and 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], proposed a related generative model also consisting of multiple levels of hierarchy, in which the speed of change at each level is determined by the preceding level. An approximate inference rule was also proposed, based on the so-called mean-field approximation [19], that the posterior at each time point is fully independent of all other time points. Although this model provides a biologically-plausible algorithm for learning, the assumption that the probability distribution at each time point is fully independent of all the other time points poses important theoretical and practical challenges, particularly in estimating volatility and learning rate. Furthermore, the resulting learning algorithm is much more convoluted than classical models, such as the Pearce-Hall model or the Kalman filter.

Here we revisit the question of approximate inference in a hierarchical model of learning in volatile environments. Rather than assuming a mean-field approximation, we show the importance of preserving the key statistics reflecting volatility of the environment, in particular temporal auto-correlations. Informed by this reasoning, we propose a novel model for learning in volatile environments that is conceptually aligned with that of Behrens et al. [6] and the HGF but more clearly related to the Kalman filter and to psychological models such as Pearce-Hall. In particular, the proposed model is based on the hierarchical structure of Behrens et al. [6] and principles of probabilistic graphical models [20]. 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 [21,22]). 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.

Theory

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 [4,7,23,24]. 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, $o_t \sim Normal(x_t, \omega)$. Here the observation noise is again Gaussian with known, fixed variance ω. 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:

$$m_t = m_{t-1} + k_t(o_t - m_{t-1}),$$

where k_t is the learning rate (or Kalman gain)

$$k_t = (w_{t-1} + v)/(w_{t-1} + v + \omega),$$

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:

$$w_t = (1 - k_t)(w_{t-1} + v).$$

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

$$w_{t-1,t} = (1 - k_t)w_{t-1}.$$

This equation indicates that when the Kalman gain is relatively small, the autocorrelation 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 autocorrelation 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 [25] and by Gamerman et al. [26] (see also West et al. [27]). 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 like 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

$$z_t = z_{t-1}\epsilon_t, \qquad \text{Equation 2}$$

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. This sub-problem provides the foundation for the inference in the full model, using estimates of these quantities. It can be formally shown (see S1 Appendix) that in this case, estimating the process variance, i.e. the posterior over z_t at each step given all previous observations, is tractable. Specifically, it can be shown that 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:

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

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 error.

Now we are ready to build the full generative model in which both z_t and x_t dynamically evolve over time (Figure 1). In this model, as before, the observation, $o_t \sim Normal(x_t, \omega)$, 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 2. 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 structural variational inference for approximate inference in this model [28,29]. This technique assumes a factorized approximate posterior distribution and minimizes the mismatch between this approximate posterior and the true posterior using principle of variational inference.

The resulting learning algorithm is very similar to the Kalman filtering algorithm and encompass 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 4-6 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 $(x_t - x_{t-1})^2$

$$v_t = v_{t-1} + \lambda(E[(x_t - x_{t-1})^2] - v_{t-1}), \qquad \text{Equation 3}$$

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 autocorrelation.

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

$$k_t = (w_{t-1} + v_{t-1})/(w_{t-1} + v_{t-1} + \omega), \qquad \text{Equation 4}$$

$$m_t = m_{t-1} + k_t(o_t - m_{t-1}), \qquad \text{Equation 5}$$

$$w_t = (1 - k_t)(w_{t-1} + v_{t-1}), \qquad \text{Equation 6}$$

$$w_{t-1,t} = (1 - k_t)w_{t-1}, \qquad \text{Equation 7}$$

$$v_t = v_{t-1} + \lambda\left((m_t - m_{t-1})^2 + w_{t-1} + w_t - 2w_{t-1,t} - v_{t-1}\right), \qquad \text{Equation 8}$$

where ω 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.

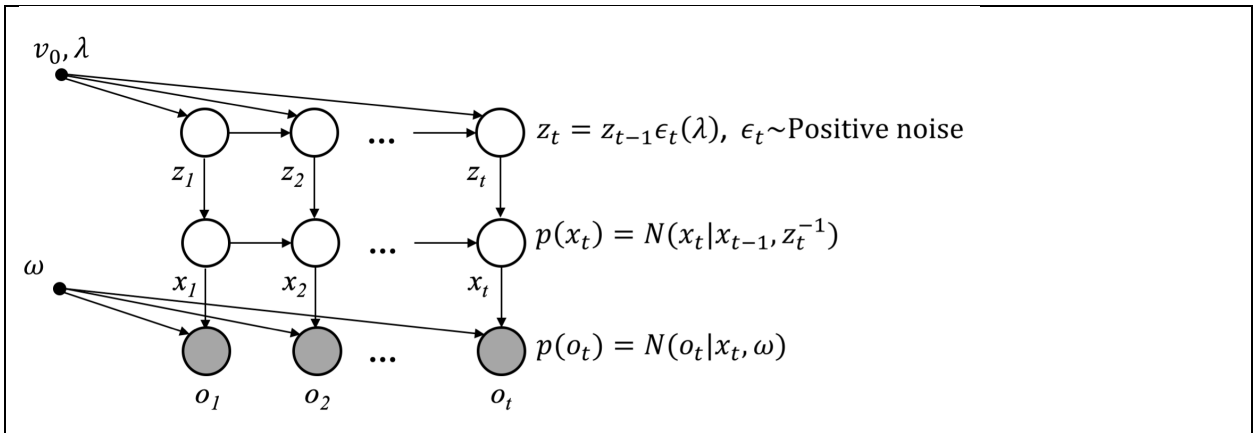

Figure 1. The generative model of VKF. The generative model consists of two interconnected hidden temporal chains, x_t and z_t, governing observed outcomes, o_t. Arrows indicate the direction of influence. On each trial, t, outcome, o_t, is generated based on a Gaussian distribution, its mean is given 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.

Binary VKF

We have also developed a binary 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

$$\sigma(x_t) = 1/(1 + \exp(-x_t))$$

where $\sigma(x_t)$ is the sigmoid function, which maps the normally distributed variable x_t to the unit range. Further approximation is required to make inference here, because observations are not normally distributed and Equation 1 does not hold. For binary VKF, we assumed a constant variance and employed moment matching (which is sometimes called assumed density filtering) to obtain the mean. The resulting algorithm is then very similar to the original VKF with the only difference that the update rule for the mean (replacing Equation 5) is slightly different

$$m_t = m_{t-1} + \sqrt{w_{t-1} + v_{t-1}}(o_t - \sigma(m_{t-1})). \qquad \text{Equation 9}$$

The update rules for the Kalman gain, variance, covariance and the volatility are the same as those for the VKF (Equations 4,6-8). Note that the outcome o_t is binary here and therefore $o_t - \sigma(m_{t-1})$ is always between –1 and +1.

Results

Simulation analyses

We first 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]. In particular, it has been shown that humans learning rate is higher in the volatile condition. Figure 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, k_t, 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.

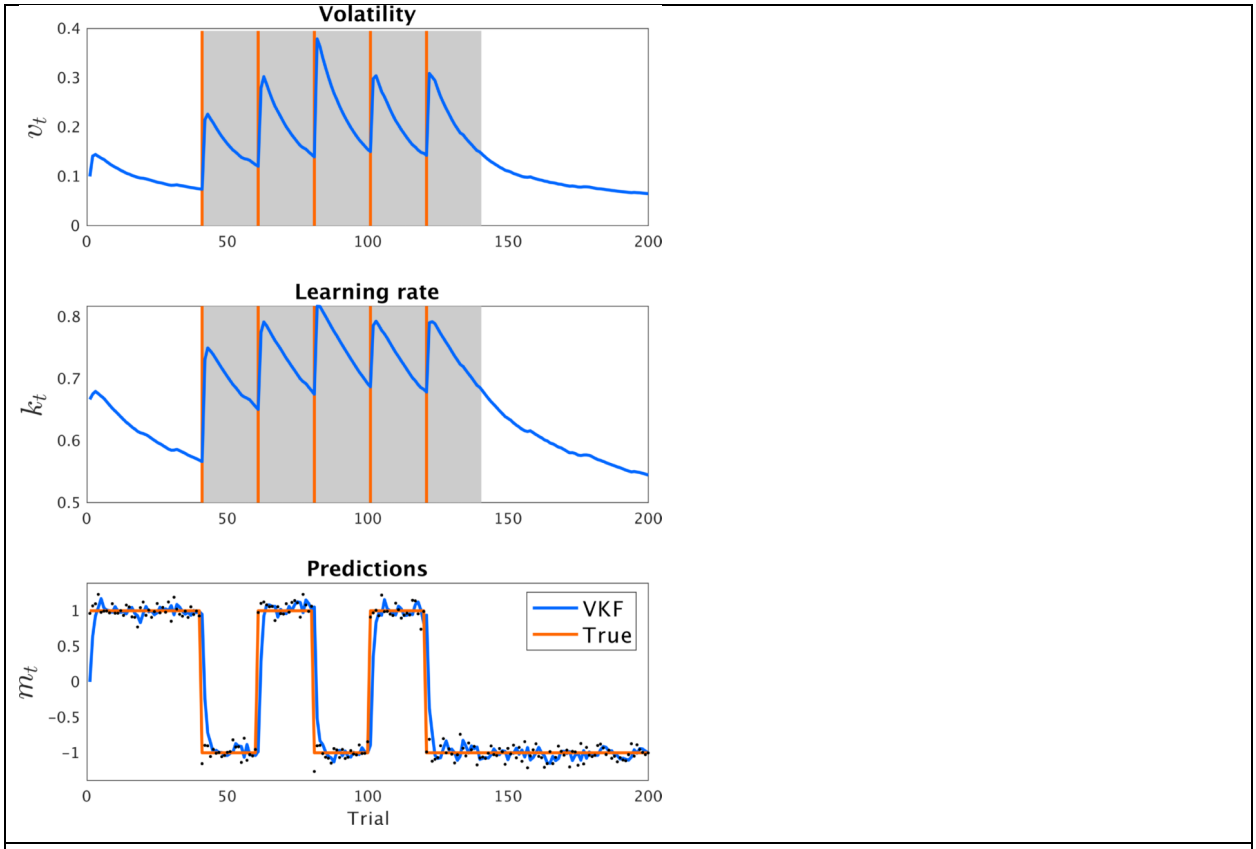

Figure 2. Behavior of the VKF in a volatility learning task, in which observations are randomly drawn from a hidden state (with variance 0.01). In the stable condition, the hidden state is constant. In the volatile condition (shaded area), however, the state of the environment is frequently changing. Top: Volatility estimates by the VKF, which increases after true change points. Middle: the learning rate (Kalman gain) in the VKF, which is higher in the volatile condition. Bottom: Predicted state by the VKF reflects well the actual state and observations (dots). These parameters used for simulating VKF: $\lambda = 0.1$, $v_0 = 0.1$, $\omega = 0.1$.

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. Figure 3 shows the performance of the binary VKF model in this task. As this figure 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, α_t, is higher in the volatile condition.

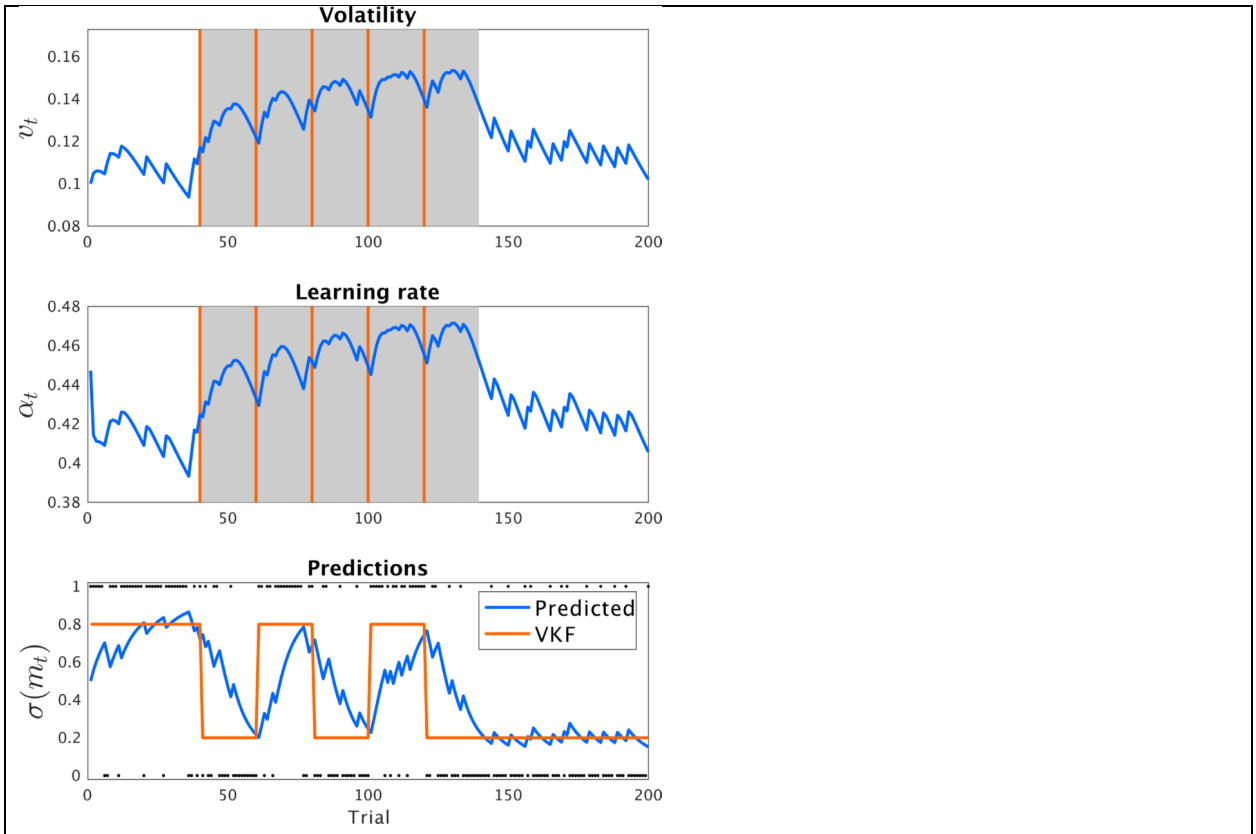

Figure 3. Behavior of the binary VKF model in a volatility learning task, in which observations are binary and are drawn from a hidden state with probability p. In the stable condition, the hidden state is constant with probability $p = 0.8$. In the volatile condition (shaded area), however, the state of the environment is frequently reversing between $p = 0.8$ and $p = 0.2$. Top: Volatility estimates by the VKF, which increases after true change points. Middle: the learning rate, which is higher in the volatile condition. Bottom: Predicted state by the VKF reflects well the actual state and observations (dots). These parameters used for simulating the binary VKF: $\lambda = 0.1$, $v_0 = 0.1$, $\omega = 0.1$.

We then compared VKF and HGF (with three levels), as the latter is the most commonly used algorithm for learning in volatile environments. These two models differ in two ways. First, although both models employ the same conditional dependency assumptions in their generative models, the distribution over the volatility variable is different. The HGF assumes normal distribution at both levels,

and therefore exponential transformations are needed to ensure variances are positive, which requires further approximation for inference. Our model of dynamic variance (with multiplicative rather than additive Gaussian diffusion) is tractable. Another important difference between the two models is in the assumptions about the approximate posterior. The approximate inference model of HGF assumes that the posterior at each time point is fully independent of all the other time points. We argue that this assumption poses critical challenges, particularly because the autocorrelation between hidden states x_t and x_{t-1} is neglected. Note that when the process noise variance, v, is constant, the Kalman filter indicates that the posterior covariance is nonzero and has an inverse relation with the learning rate. In the VKF model, unlike the HGF, this covariance plays a key role as it influences volatility update because $E[(x_t - x_{t-1})^2]$ depends on $w_{t-1,t}$ as can be seen in Equation 8. Moreover, the inference approximate model of the HGF relies on Taylor approximation up to the first or second degree, which can cause further problems if data violate (relatively strong) assumptions of this approximation.

To demonstrate these issues, 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 also performed inference using VKF on these generated time-series, by fitting its free parameter, volatility learning rate, λ, to data. Note that we have set up this experiment to favor the HGF, in that any performance differences due to the models assuming different generative dynamics will work against the VKF. We repeated this process 500 times. Note that we have set up this experiment to favor the HGF, in that any performance differences due to the models assuming different generative dynamics will work against the VKF. Across all simulations, in 51 simulations, the HGF encountered numerical problems: its inferred trajectory conflicted with the assumptions of the inference model, resulting in negative posterior variance at the third level. Only for the remaining 449 simulations were we able to calculate the median error of predicted outcomes (defined as the absolute normalized difference between predicted and observed outcome) across all time points. We found that in 325 of simulations (i.e. 72%), the error of the VKF was smaller than the error of the HGF, despite the simulations being conducted under the HGF's own generative assumptions. That is, the VKF's more accurate approximate resulted in overall better performance even given the mismatch between the assumed and true generative models.

Testing VKF using Empirical data

We tested VKF using an empirical dataset from a probabilistic learning task (previously published in [30]). In the experiment (Figure 4A), 44 participants 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 different combinations of the emotional content of the face image and its background color; together, these indicated the chance, at the end of each trial, that a go or no-go response would earn the better of two outcomes (monetary reward vs nothing, or nothing vs loss) would occur. 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 trial-type (480 trials in total).

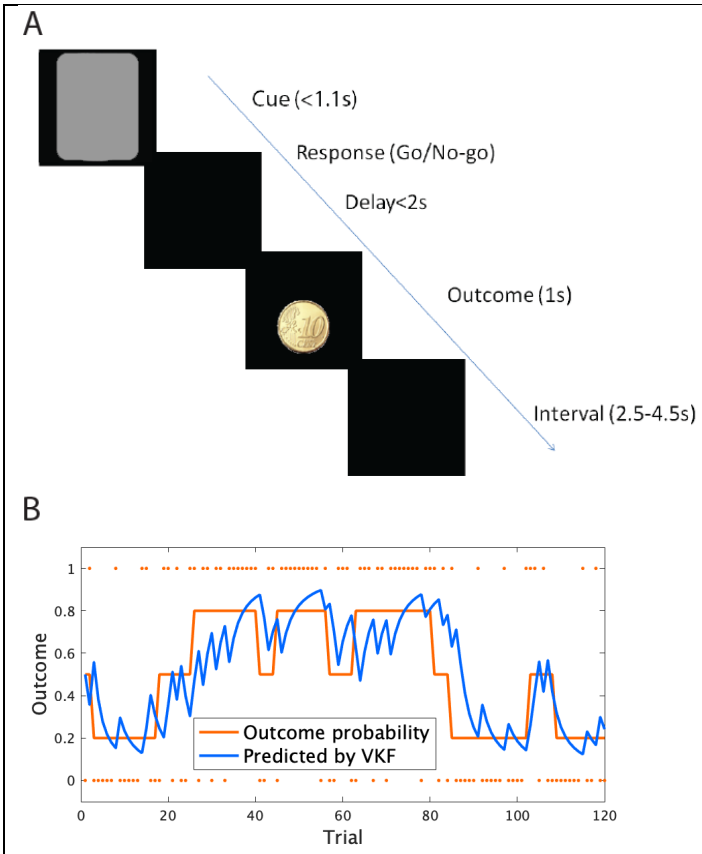

Figure 4. The probabilistic learning task used for testing. A) Participants had to respond (either go or no-go) after a cue was presented. A probabilistic outcome was presented following a delay. B) An example of the probability sequence of outcome (of a go response) for one of the four trial-types and 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.

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 ($t(43)=14.68$, $p<0.001$). The focus of the original studies using this dataset was on socio-emotional modulation of learning and choice [30,31]. Here, however, we use this dataset because the task is a volatile environment in which the contingencies change frequently. We considered a model space including both VKF and HGF, and also a Rescorla-Wagner (RW) model that does not take into account uncertainty or vary its learning rate. We used the softmax rule as a choice 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. Note that the choice model controlled value-independent biases in making or avoiding a go response due to the emotional or reinforcing content of the cues (see Methods).

This model space was then fit to choice data using hierarchical Bayesian inference (HBI) procedure [32], a Bayesian algorithm to perform concurrent model fitting and model comparison with the advantage that fits to individual subjects are constrained according to the group-level statistics (i.e. empirical priors) taking into account the possibility that different subjects might express different models. The HBI also performs random effects model comparison by quantifying model evidence across the group (goodness of fit penalized by the complexity of the model [33]). Across participants, VKF was the superior model in 37 out of 44 participants. 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 [34,35] taking into account the null possibility that differences in model evidence might be due to chance [35]) was indistinguishable from 1 (Figure 5). These results provide evidence that volatility affects choice and VKF is more parsimonious than HGF in tracking volatility. Note that since the outcome is binary here, we used the binary version of HGF, as described in [18] (considering constraints described by Mathys et al. [36], see Methods), and the binary version of VKF.

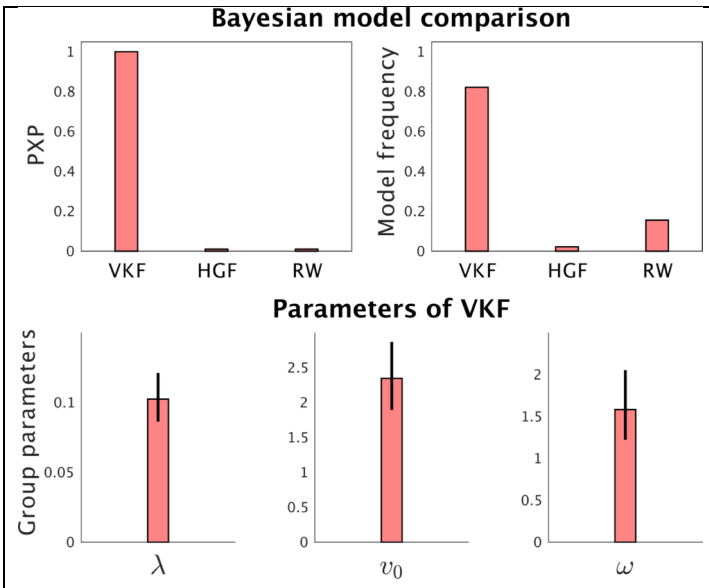

Figure 5. Comparison of VKF with alternative models using empirical data. Top: Bayesian model comparison results using protected exceedance probabilities (PXP) and model frequencies. The PXP of the VKF is indistinguishable from 1. The model frequency metric indicates the ratio of subjects explained by each model, which is more than 0.8 for the VKF. Bottom: Estimated parameters of the VKF at the group-level. The error-bars are obtained by applying the corresponding transformation function on the group-level error-bars obtained by the HBI [32] 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, mainly because it better captures the temporal dependencies and it is based on a novel generative model of uncertainty. 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 [4,7,23,24]. 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. The VKF fills this gap by extending the Kalman filter to inferring the process

variance in volatile environments in which that 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 indeed 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 [21,22,31]. 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 learning rate modulated by volatility estimate, which is itself a function of surprise (i.e. the squared prediction error) according to Equation 3. 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 of 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 Kalman filter. Therefore, the full generative model of VKF contains two temporal chains (Figure 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. This can be seen as a minimal mean-field assumption in which, unlike the HGF, all temporal dependencies are persevered.

The state-of-the-art algorithm for learning in volatile environments is HGF [8,18]. There are two critical differences between the VKF and the HGF. First, although the generative conditional dependencies of our model and HGF are the same, there is a critical difference between the generative process underlying uncertainty in the two models. The HGF assumes a log-normal distribution to maintain

positivity constraint on variance, and uses variational approximations to update both its posterior mean and variance on every trial. In contrast, the generative model of process variance in the VKF makes it possible to make exact inference by tracking the expected value of process variance (i.e. volatility). This feature of the VKF not only makes it an extension of Kalman filter, but also makes it algorithmically simpler than HGF. Secondly, the HGF is based on more severe approximations to the exact inference and therefore is less accurate. Specifically, while the VKF is based on a structured variational approximation that keeps all temporal dependencies, the HGF is based on a full mean-field approximation and therefore neglects covariance between consecutive states. Note that the mean-filed assumption across timepoints can be less accurate in online algorithms like HGF that do not iterate in updating posteriors. 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 is more parsimonious than the HGF and predicts choice data better in vast majority of participants.

Bayesian models have recently been used for online inference of sudden changes in the environment [37,38]. 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 unifying model for learning under uncertainty. The VKF has theoretical and practical advantages over existing models and explains human choice data better than alternatives.

Methods

Ethics statement

Data from human subjects are from a study [30] that is approved by the local ethical committee (“Comissie Mensgebonden Onderzoek” Arnhem-Nijmegen, Netherlands).

Implementation of models for analysis of human choice data

We have fitted a 3-level (binary) HGF [18] to choice data, which contains three free parameters: the variance of the third level, $\nu > 0$, $\kappa > 0$ determining the extent that the third level affects the second level uncertainty and ω, which indicates the tonic level variance at the second level. Formally, if x_3^t is the third level random variable on trial t, the (generative) variance of the second level is given by $\exp(\kappa x_3^t + \omega)$. As suggested by Mathys et al. [36], we assumed an upper bound at 1 for both ν and κ. 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 [34,35]), 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. The initial value of m_t and w_t were assumed to be 0 and equal to ω, respectively.

Each of the learning models was then 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 $-b_e + b_v - b_i$.

Simulation analysis using HGF generative model

We generated data according to the HGF generative model (with Gaussian observations) according to these parameters: $\nu = 0.5$, $\kappa = 1$ and $\omega = -3$. The observational noise was assumed to be 1. The initial mean at the second and third levels were 0 and 1, respectively. These parameters were also used for inference using the HGF algorithm. For inference using the VKF algorithm, we used the generative initial volatility (i.e. $\exp(\kappa\mu_3^0 + \omega) = 0.135$) and generative observational noise (i.e. 1). We fitted the VKF volatility learning rate parameter to data for each time-series using a maximum-a-posteriori procedure, in which the prior mean and variance were assumed to be 0 and 15.23, respectively.

Model fitting and comparison

We used a hierarchical and Bayesian inference method, HBI [32], 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 (Figure 5). 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. The initial parameters of all models were obtained using maximum-a-posteriori procedure, in which the 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. Specifically, for any parameter bounded in the unit range, the log-effect of this prior is less than one chance-level choice (i.e log0.5) for any value of that parameter between 0.05 and 0.95. The HBI algorithm is available online at https://github.com/payampiray/cbm.

Supporting information

S1 Appendix. Formal treatment of the VKF model.

Reference

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. Jordan MI. An Introdcution to Probabilistic Graphical Models [Internet]. In preparation. Available: http://people.eecs.berkeley.edu/~jordan/prelims/

21. 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

22. 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

23. Wolpert DM. Probabilistic models in human sensorimotor control. Hum Mov Sci. 2007;26: 511–524. doi:10.1016/j.humov.2007.05.005

24. Wolpert DM, Ghahramani Z. Computational principles of movement neuroscience. Nat Neurosci. 2000;3 Suppl: 1212–1217. doi:10.1038/81497

25. 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

26. 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

27. West M. On Scale Mixtures of Normal Distributions. Biometrika. 1987;74: 646–648.

28. Saul L, Jordan MI. Exploiting Tractable Substructures in Intractable Networks. Advances in Neural Information Processing Systems 8. MIT Press; 1995. pp. 486–492.

29. Ghahramani Z, Jordan MI. Factorial Hidden Markov Models | SpringerLink. Mach Learn. 1997;29: 245–273.

30. 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

31. 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

32. Piray P, Dezfouli A, Heskes T, Frank MJ, Daw ND. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. bioRxiv. 2018; 393561. doi:10.1101/393561

33. MacKay DJC. Information Theory, Inference, and Learning Algorithms. Cambridge University Press; 2003.

34. 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

35. 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

36. 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

37. 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

38. 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

Supplementary Appendix

A transparent model for learning in volatile environments

Payam Piray and Nathaniel D. Daw

Appendix A

In this appendix, we give a formal treatment of the VKF. We first review state space models and principles of making inference in those models. Next, we explain two specific state space models with different types of noise. The first one is Kalman filter, a well-known model in which the noise is additive and Gaussian. Next, we present another space model, based on a class of non-Gaussian state space models, in which the noise is multiplicative. Inference for these models is tractable given some assumptions about the noise. Building on these two tractable models, we are then in the position to provide an approximate solution for inference in volatile environments by using the powerful technique of structured variational inference. This technique makes use of exact inference algorithms on tractable substructures of an intractable model and provides an approximate variational solution.

A.1 State Space Models

We begin by describing state space model, in which a sequence of observations, $\{u_i\}$ where $i = 1, ..., t$, is modeled by specifying a probabilistic relation between the observations and a sequence of latent states, $\{s_i\}$ and a Markov structure linking the latent states. A state space model assumes that i) u_t is independent of all other observations and states given s_t, and ii) s_t is independent of all other states given s_{t-1}. Thus, the joint distribution for the state space model is given

by

$$p(s_1, ..., s_t, u_1, ..., u_t) = p(s_1) \prod_{i=2}^{t} p(s_i|s_{i-1}) p(u_i|s_i) \tag{A.1}$$

The inference goal in the state space model is to obtain the posterior over s_t given observations. It can be shown then that regardless of the form of transition probabilities and the probabilistic relation between s_t and u_t, the posterior can be obtained recursively. First, the current state estimate is projected ahead in time. In other words, the posterior of s_t given all observations prior to t is calculated based on the previous posterior $q(s_{t-1}) = p(s_{t-1}|u_1, ...u_{t-1})$:

$$\begin{aligned} \hat{q}(s_t) = p(s_t|u_1, ..., u_{t-1}) &= \int q(s_{t-1}) p(s_t|s_{t-1}) \, \mathrm{d}s_{t-1} \quad \text{for } t > 1 \\ \hat{q}(s_1) &= p(s_1) \end{aligned} \tag{A.2}$$

Next, the projected estimate is adjusted by actual measurement, giving rise to the posterior given all observations:

$$q(s_t) = \frac{1}{c_t} p(u_t|s_t) \hat{q}(s_t) \tag{A.3}$$

where $c_t = p(u_t|u_1, ..., u_{t-1})$ is constant with respect to s_t. Note that since c_t is independent of s_t, it is only needed to be calculated if one is interested in the likelihood of data, which is easy to see to be given by

$$p(u_1, ..., u_t) = \prod_{i=1}^{t} c_i \tag{A.4}$$

Furthermore, the posterior over two consecutive states s_{t-1} and s_t is given by

$$q(s_{t-1}, s_t) = p(s_{t-1}, s_t|u_1, ..., u_t) = C q(s_{t-1}) p(s_t|s_{t-1}) p(u_t|s_t) \tag{A.5}$$

where C is a normalization constant independent of s_{t-1} and s_t.

As noted above, equations (A.2-A.5) are valid for any model with conditional dependencies as given in equation (A.1) regardless of the form of $p(s_t|s_{t-1})$ and $p(u_t|s_t)$. However, the form of $p(s_t|s_{t-1})$ and $p(u_t|s_t)$ determine whether the integral in equation (A.2) is tractable and whether the process can be summarized in an efficient and simple algorithm by ensuring that $q(s_t)$ has the same functional form as $q(s_{t-1})$. Kalman filter provides a tractable solution for a

state space model in which transition probabilities are Gaussian with a fixed variance, which is suitable for prediction in stable environments. We discuss Kalman filter in section A.2. In section A.3, we build another state space model in which observations are Gaussian with a stochastic variance determined by the states. Furthermore, while the source of stochasticity in state transitions of Kalman filter is additive noise, the source of stochasticity in the second state space model is a multiplicative noise, which makes it appealing as a model of uncertainty. Then, in section A.4, we combine the two models to address prediction problems in volatile environments.

A.2 Kalman filter

Consider a state space model in which, on trial n, hidden state of the environment, x_t, is equal to its state on trial $n-1$ plus some Gaussian noise.

$$x_t = x_{t-1} + e_t \tag{A.6}$$

where e_t is a zero-mean Gaussian random variable with variance v. Therefore, the generative probabilistic model of x_t is

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t|x_{t-1}, v) \tag{A.7}$$

The initial state is also assumed to be Gaussian

$$p(x_1) = \mathcal{N}(x_1|0, v) \tag{A.8}$$

Observations are given by a Gaussian distribution:

$$p(o_t|x_t) = \mathcal{N}(o_t|x_t, \omega) \tag{A.9}$$

where ω is the observation variance. Using equations (A.2-A.3), we obtain:

$$q(x_t) = \mathcal{N}(x_t|m_t, w_t) \tag{A.10}$$

where

$$m_t = m_{t-1} + k_t(o_t - m_{t-1}) \tag{A.11}$$

$$w_t = (1 - k_t)(w_{t-1} + v) \tag{A.12}$$

and k_t is the learning rate (also called Kalman gain) and is given by:

$$k_t = \frac{w_{t-1} + v}{w_{t-1} + v + w} \tag{A.13}$$

Furthermore, we can make use of (A.5) to obtain the covariance between consecutive states, x_{t-1} and x_t:

$$w_{t-1,t} = (1 - k_t)w_{t-1} \tag{A.14}$$

A.3 A tractable filter for variance

Let's now consider another state space model in which hidden state on trial t, z_t, depends on its value on previous trial, z_{t-1} multiplied by some (independent) noise, ϵ_t. Then, a Gaussian random variable, x_t, is drawn with a known mean, μ_t, and a precision given by z_t,

$$p(x_t|z_t) = h_t(z_t) = \mathcal{N}(x_t|\mu_t, z_t^{-1}) \tag{A.15}$$

Therefore, conditional dependencies indicate that equations (A.2-A.3) hold for the posterior over z_t.

Our goal in this section is to build a rich model of dynamic (inverse) variance with tractable and simple update equations. First, we inspect the functional form of $h_t(z_t)$ with respect to z_t:

$$h_t(z_t) \propto z_t^{\frac{1}{2}} e^{-\frac{1}{2}z_t(x_t-\mu_t)^2} \tag{A.16}$$

Therefore, $h_t(z_t)$ takes a form of gamma distribution in which information about new observations is always coming through the rate and the shape is constant. Moreover, equation (A.3) indicates that if $\hat{q}(z_t)$ is a gamma distribution, then $q(z_t)$ takes the form of gamma distribution. Thus, for the moment, we assume that $q(z_t)$ is a gamma distribution with a constant shape a and a dynamic rate b_t:

$$q(z_t) = \mathcal{G}(z_t|a, b_t)$$

In the previous section, we saw that the transition noise was additive. For variance, it makes sense to assume a multiplicative noise. Thus, we suppose

that the value of z_t is equal to z_{t-1} multiplied by some noise

$$z_t = z_{t-1}\epsilon_t$$

Assuming that the noise is bounded between zero and $R \geq 1$, we can write this equation as

$$z_t = Rz_{t-1}\varepsilon_t \tag{A.17}$$

where $\epsilon_t = R\varepsilon_t$, in which ε_t is constrained in the unit range and has a beta distribution. Assuming that conditional expectation of z_t is z_{t-1}, we obtain $\mathbb{E}[\varepsilon_t] = R^{-1}$. A general beta distribution with this property is given by:

$$p(\varepsilon_t) = \mathcal{B}(\varepsilon_t|\eta\nu, (1-\eta)\nu) \tag{A.18}$$

where $\eta = R^{-1}$ and ν is a constant positive parameter. We show below that there is a specific value for ν, which makes the inference tractable.

Using transform theorem for random variables, the conditional distribution of z_t is given by

$$p(z_t|z_{t-1}) = \mathcal{B}(\eta z_t/z_{t-1}|\eta\nu, (1-\eta)\nu)\eta z_{t-1}^{-1} \qquad 0 < z_t < \eta^{-1}z_{t-1}$$

Applying equation (A.2) to $\hat{q}(z_t)$, we obtain

$$\hat{q}(z_t) \propto \int z_{t-1}^{a-1}e^{-b_{t-1}z_{t-1}}(\frac{\eta z_t}{z_{t-1}})^{\eta\nu-1}(1-\frac{\eta z_t}{z_{t-1}})^{(1-\eta)\nu-1}\frac{1}{z_{t-1}}\,\mathrm{d}z_{t-1}$$

If we take a to be equal to ν, the integral is tractable resulting in a gamma distribution for $\hat{q}(z_t)$. Note that evaluating those terms that are independent of z_t is not even necessary because $\hat{q}(z_t)$ is a probability distribution. Therefore, we have

$$\hat{q}(z_t) = \mathcal{G}(z_t|\eta\nu, \eta b_{t-1}) \tag{A.19}$$

Combining $\hat{q}(z_t)$ with $h(z_t)$, we obtain $q(z_t)$:

$$q(z_t) = \mathcal{G}(z_t|\eta\nu + \frac{1}{2}, b_t)$$

where

$$b_t = \eta b_{t-1} + \frac{1}{2}(x_t - \mu_t)^2 \tag{A.20}$$

If we now take the shape of $q(z_t)$ to be the same as the shape of $q(z_{t-1})$, the

inference is similarly tractable for all next trials. Consequently, ν is given by

$$\nu = \frac{1}{2(1-\eta)} \tag{A.21}$$

Therefore, the posterior distribution is a gamma distribution with a fixed shape and an evolving rate

$$q(z_t) = \mathcal{G}(z_t|\nu, b_t) \tag{A.22}$$

We can also write b_t in terms of $v_t = \mathbb{E}[z_t]^{-1}$

$$b_t = 2(1-\eta)\mathbb{E}[z_t]^{-1}$$

Now we can also write (A.20) in terms of $v_t = \mathbb{E}[z_t]^{-1}$

$$v_t = \eta v_{t-1} + (1-\eta)(x_t - \mu_t)^2$$

If we define $\lambda = 1 - \eta$, we can write this equation in the form of an error-correcting update rule:

$$v_t = v_{t-1} + \lambda((x_t - \mu_t)^2 - v_{t-1}) \tag{A.23}$$

It is also important to note that $\mathbb{E}[z_t]$ under $\hat{q}(z_t)$ is given by v_{t-1} according to equation (A.19). Furthermore, as equation (A.2) indicates $\hat{q}(z_1) = p(z_1)$, we take the probability of initial state to be given by

$$p(z_1) = \mathcal{G}(z_1|\eta\nu, \eta b_0) \tag{A.24}$$

where $b_0 > 0$ is a free parameter, which we can equivalently write it according to the initial volatility value v_0. Note that as a consequence of making the inference tractable, the form of multiplicative transition noise is restricted according to equation (A.21). However, this is not a major restriction, as the mean of noise is still free to be chosen. Note that it is also possible to assume other forms for ν, although those assumptions might make the inference slightly more complicated.

A.4 Volatile Kalman filter

In this section, we provide a solution to the general prediction problem in volatile environments. Specifically, consider a problem in which on trial t a latent ran-

dom variable, x_t, is given by its previous value, x_{t-1}, plus some Gaussian noise with a precision given by another dynamic random variable, z_t:

$$x_t = x_{t-1} + e_t \tag{A.25}$$

$$p(e_t) = \mathcal{N}(e_t|0, z_t^{-1}) \tag{A.26}$$

The dynamic of z_t is given by equations (A.17-A.18). The observation on trial t, o_t, is then drawn based on the value of x_t according to a normal distribution:

$$p(o_t|x_t) = \mathcal{N}(o_t|x_t, \omega) \tag{A.27}$$

Therefore, the graphical model corresponding to this problem consists of two coupled chains of latent variables, x_t and z_t. Exact inference for this problem is not possible due to coupling between the two chains. However, it is possible to derive a variational approximation by inducing factorization between the two chains, while preserving all dependencies within each chain. Here, we assume

$$p(x_1, ..., x_t, z_1, ..., z_t|o_1, ..., o_t) \simeq q(x_1, ..., x_t)q(z_1, ..., z_t) \tag{A.28}$$

Using variational calculations, we obtain

$$\begin{aligned} q(x_1, ..., x_t) &\propto \prod_{i=1}^{t} \exp\Big(\mathbb{E}[\log p(z_i|z_{i-1})p(x_i|x_{i-1}, z_i)p(o_i|x_i)]\Big) \\ &= C \prod_{i=1}^{t} f_i(x_{i-1}, x_i)g_i(x_i) \end{aligned} \tag{A.29}$$

where all terms independent of $x_1, ..., x_t$ are absorbed into C and

$$\begin{aligned} f_i(x_{i-1}, x_i) &\triangleq \exp\big(\mathbb{E}[\log p(z_i|z_{i-1})p(x_i|x_{i-1}, z_i)]\big) \\ &\propto \mathcal{N}(x_i|x_{i-1}, \mathbb{E}[z_i]^{-1}) \\ &= \mathcal{N}(x_i|x_{i-1}, v_i) \end{aligned} \tag{A.30}$$

$$g_i(x_i) \triangleq p(o_i|x_i) \tag{A.31}$$

Since the conditional dependencies for x_{i-1} and x_i in (A.29) is the same as those in (A.1), exact inference for $q(x_t)$ follows (A.2-A.3). Moreover, since dependencies are in Gaussian forms, we obtain $q(x_t) = \mathcal{N}(x_t|m_t, w_t)$, we can use Kalman equations (A.10-A.13) to update m_t and w_t in which we replace v

in those equations with v_t.

Next, we use variational techniques to obtain $q(z_t)$:

$$\begin{aligned} q(z_1, ..., z_t) &\propto \prod_{i=1}^{t} \exp\Big(\mathbb{E}[\log p(z_i|z_{i-1})p(x_i|x_{i-1}, z_i)p(o_i|x_i)]\Big) \\ &= C \prod_{i=1}^{t} p(z_i|z_{i-1})h_i(z_i) \end{aligned} \tag{A.32}$$

where all terms independent of $z_1, ..., z_t$ are absorbed into C and

$$\begin{aligned} h_i(z_i) &\triangleq \exp\Big(\log \mathbb{E}[p(x_i|x_{i-1}, z_i)\Big) \\ &= \frac{1}{\sqrt{2\pi}} z_i^{\frac{1}{2}} \exp(-\frac{1}{2} z_i \mathbb{E}\big[(x_i - x_{i-1})^2\big]) \end{aligned} \tag{A.33}$$

Since the conditional dependencies for z_{i-1} and z_i in (A.32) is the same as those in (A.1), inference for $q(z_t)$ follows (A.2-A.3). Moreover, since dependencies between z_{i-1} and z_i have the same functional form as (A.16) and (A.17), the posterior $q(z_t)$ takes the form of Gamma distribution, in which its mean, v_t, gets updated according to $\mathbb{E}\big[(x_t - x_{t-1})^2\big]$:

$$v_t = v_{t-1} + \lambda(\mathbb{E}\big[(x_t - x_{t-1})^2\big] - v_{t-1}) \tag{A.34}$$

Since the expectation in (A.34) should be taken under the posterior $q(x_{t-1}, x_t)$, we have

$$\begin{aligned} \mathbb{E}\big[(x_t - x_{t-1})^2\big] =& \mathbb{E}\big[x_{t-1}^2\big] + \mathbb{E}\big[x_t^2\big] - 2\mathbb{E}\big[x_{t-1}x_t\big] \\ =& m_{t-1}^2 + w_{t-1} + m_t^2 + w_t - 2(m_{t-1}m_t + w_{t-1,t}) \\ =& (m_t - m_{t-1})^2 + w_{t-1} + w_t - 2w_{t-1,t} \end{aligned} \tag{A.35}$$

These results give rise to an efficient and simple algorithm for learning in volatile environments. On every trial t, we first compute $\mathbb{E}[z_t]^{-1}$ under $\hat{q}(z_t)$, which is given by v_{t-1}. Next, we update m_t and w_t and v_t, according to equations (A.11-A.14) by replacing v with $\mathbb{E}[z_t]^{-1} = v_{t-1}$. Finally, making use of equations (A.35), (A.12) and (A.14), we can simplify (A.34) and obtain equation (8). Note that here, unlike typical algorithms obtained based on variational Bayes, we only update statistics once to have an efficient algorithm for learning.

A.5 Binary VKF

In this appendix, we give a formal treatment of the binary VKF, in which observations are drawn using the sigmoid function:

$$p(o_t|x_t) = \sigma(x_t)^{o_t}(1 - \sigma(x_t))^{1-o_t} \tag{A.36}$$

where $\sigma(.)$ is the sigmoid function:

$$\sigma(x_t) = \frac{1}{1 + \exp(-x_t)} \tag{A.37}$$

Note that for binary observations, (A.3) indicates that $q(x_t)$ is not normally distributed even if we assume that $\hat{q}(x_t)$ is normal. Therefore, some approximate strategies are required for binary observations.

A well-known method approximates $q(x_t)$ by matching its moments to those of $p(o_t|x_t)\hat{q}(x_t)$. For a Gaussian distribution, this means that first and second order moments should be matched. However, since matching variance might lead to negative values for the variance, we assume that the variance is constant (similar to Kalman filter) and match only the mean of distributions. Formally, we take the "message" from o_t to x_t to be an unnormalized Gaussian, $g_t(x_t) = c_t\mathcal{N}(x_t|\tilde{m}_t, \omega)$, in which the variance ω is a constant and the mean will be obtained by matching the mean of $q(x_t) = \mathcal{N}(x_t|m_t, w_t)$ with that of $p(o_t|x_t)\hat{q}(x_t)$, in which $\hat{q}(x_t) = p(x_t|o_1, ..., o_{t-1}) = \mathcal{N}(x_t|m_{t-1}, w_{t-1} + v_{t-1})$. We used a simple approximation for the arising integral, which is inspired by Mackay (1992):

$$m_t \simeq m_{t-1} + \sqrt{w_{t-1} + v_{t-1}}(o_t - \sigma(m_{t-1})) \tag{A.38}$$

It is also possible to approximate m_t using a slightly more complicated equation given by MacKay (1992). Using simulation analyses, however, we found that (A.38) results in better approximations. The variance of $q(x_t)$ is then given by

$$w_t^{-1} = \hat{w}_t^{-1} + \omega^{-1} \tag{A.39}$$

which results in equations (6-8).