Inferring the causes of noise from binary outcomes: A normative theory of learning under uncertainty

Xiaotong Fang, Payam Piray*
Department of Psychology, University of Southern California
*Corresponding author: piray@usc.edu

Abstract
Inferring the true cause of noise—distinguishing between volatility (environmental change) and stochasticity (outcome randomness)—is essential for learning in noisy environments. While most studies rely on binary outcomes, previous models are designed for continuous outcome and use ad hoc approximations to handle binary data, introducing theoretical inconsistencies and interpretational issues. Here, we develop a normative framework for inferring the causes of noise from binary feedback that remains faithful to the discrete nature of the generative process and underlying statistical structure. First, we establish a generative model using a state space approach tailored for binary outcomes and derive the corresponding hidden Markov model inference procedure. Second, we introduce a computational model combining the hidden Markov model with particle filtering to simultaneously infer volatility and stochasticity from binary outcomes. Third, we validate predictions through a 2×2 probabilistic reversal learning task with human participants, systematically manipulating both noise parameters. Results show that participants adjust their learning rates consistent with model predictions, increasing learning rates under volatile conditions and decreasing them under high stochasticity. Our theoretical and experimental results offer a principled approach for dissociating volatility and stochasticity from binary outcomes, providing insights into learning processes relevant to typical cognition and psychiatric conditions.

Introduction

Learning under uncertainty is a fundamental challenge for humans and animals making decisions in complex environments. Whether foraging for food, interpreting social signals, or adapting to changing circumstances, organisms must constantly update their understanding of the world based on noisy observations. Critically, learners cannot directly observe the true state of the world; they only see outcomes and must infer the hidden causes that generated them. This inference is complicated by the fact that noise can arise from two distinct sources. First, the hidden state itself may change over time, a source of uncertainty known as volatility. Second, observations may be unreliable reflections of an otherwise stable hidden state, a source of uncertainty known as stochasticity. When an unexpected outcome occurs, the learner faces an inference problem: did the hidden state change, or was the outcome simply noisy? This distinction is critical because the two sources of noise call for opposite responses: if the hidden state changed, past information is outdated and learning should be fast; if the outcome was just noise, past information remains valid and learning should be slow. Correctly attributing noise to its source is therefore essential for adaptive behavior, yet this inference is challenging because both sources contribute to the same observable outcome, surprising events.

The idea that learning should adapt to the reliability of information has a long history in psychology. Theories of conditioning have emphasized that the rate of learning depends on the predictability of outcomes—from Mackintosh's (1975) attentional theory to Pearce and Hall's (1980) associability model. These frameworks capture the intuition that surprising or unreliable events warrant different learning responses than predictable ones. More recently, Bayesian approaches have formalized these ideas, showing that optimal learning rates depend on the noise structure of the environment (Behrens et al., 2007; Dayan et al., 2000; Nassar et al., 2010). The Kalman filter, in particular, has become influential for showing how learning rates should increase with volatility and decrease with stochasticity (Dayan & Yu, 2003; Gershman, 2015). Building on this foundation, several models have augmented the Kalman filter with inference mechanisms to estimate volatility from observations when stochasticity is assumed to be known (Behrens et al., 2007; Mathys et al., 2011). However, these models do not address the core computational problem: how learners might dissociate volatility from stochasticity when both are unknown and must be inferred from experience. We have recently developed a modeling approach that addresses this problem for continuous outcomes, allowing volatility and stochasticity to be jointly inferred from observations alone (Piray & Daw, 2021a, 2024). Normative predictions of this framework, that adaptive learners should dissociate volatility and stochasticity based on outcomes and adjust their learning rates accordingly, have been validated in our recent experimental work with human participants (Piray & Daw, 2024). However, this work focused on continuous outcomes, leaving the problem unaddressed for binary settings.

Binary outcomes are ubiquitous in both experimental paradigms and real-world settings. Probabilistic reversal learning tasks, widely used in animal and human research to study flexible learning, decision-making and their neural substrates, rely on binary feedback. Beyond the lab, many situations involve binary outcomes that may not perfectly reflect underlying states: a medical test returns positive or negative; a friendly gesture is returned or ignored; a dating request is accepted or rejected; a job application succeeds or fails. In these cases, the learner must infer a hidden state from noisy binary feedback and must determine whether unexpected outcomes reflect a change in the underlying state or noise in the observation process. Unlike continuous outcomes, which provide graded feedback, binary outcomes produce win or loss experiences that are psychologically more salient. This salience may make misattribution of noise sources particularly consequential for binary feedback, especially in psychiatric contexts. Individuals prone to self-blame, such as those with major depressive disorder, may misinterpret stochastic errors as meaningful patterns, falsely perceiving random mispredictions as evidence of failure (Abramson et al., 1978; Beck, 1970; Zahn et al., 2015). This misattribution can create a self-reinforcing cycle, where stochastic fluctuations are mistaken for systematic failures, exacerbating symptoms through increased self-blame and maladaptive belief updating. A principled framework for dissociating volatility and stochasticity from binary outcomes is therefore needed.

Existing models for binary outcomes have extended the Kalman filter framework using various approximation strategies. The most influential of these, the Hierarchical Gaussian Filter (Mathys et al., 2011), was designed to estimate volatility from binary observations when stochasticity is assumed to be known and fixed. Extending these models to the full problem, where both volatility and stochasticity must be inferred from outcomes, is not straightforward. The way these models convert continuous latent variables into binary outcome probabilities (using transformations such as sigmoid) introduces fundamental problems: stochasticity is not represented as an explicit parameter, and learning rates increase following surprising events regardless of their true source, even when volatility is zero and surprise is entirely due to stochasticity (Piray & Daw, 2020).

Here, we take a fundamentally different approach. Rather than approximating Kalman-based inference for binary data, we build on the Hidden Markov Model (HMM)—a framework specifically designed for binary latent states and binary observations. The HMM provides a tractable algorithm for binary outcomes when both volatility and stochasticity are known, analogous to the role the Kalman filter plays for continuous outcomes. Importantly, this change in perspective allows us to address the full problem: jointly inferring volatility and stochasticity from binary outcomes alone. Our contribution is three-fold: First, we establish a generative model based on the state space framework that makes explicit the latent attribution problem underlying learning and present the HMM inference procedure for this model. Second, we introduce PF-HMM, a computational model that combines HMM with particle filtering (PF)

techniques (Andrieu et al., 2010; Doucet & Johansen, 2011), enabling learning of volatility and stochasticity directly from binary outcomes. Third, we present new experimental results from human participants in two 2×2 probabilistic reversal learning tasks that systematically manipulate both noise parameters.

The paper is structured to progressively build this framework. We begin by formally presenting the generative model, then review the HMM inference procedure and demonstrate through simulations how it captures learning given known volatility and stochasticity. The core computational contribution follows: integrating HMM with particle filtering to simultaneously learn both parameters from outcomes. We validate the model with two novel experimental datasets, revealing how humans differentially respond to volatility and stochasticity when learning from binary outcomes.

Results

State space models

We begin by considering a relatively general class of probabilistic models known as state space models in control theory (Ljung, 1998) and Markov models in machine learning (Bishop, 2006). In these models, a hidden "state" variable, x_t, evolves over discrete time steps, and outcomes, o_t, are generated based on this hidden variable but corrupted by noise. Specifically, the model's structure defines two key conditional dependencies: $p(x_t|x_{t-1}, v)$, where v is the volatility parameter governing the noise in the state transition process, and $p(o_t|x_t, s)$, where s is the stochasticity parameter controlling noise in the outcome process. For inference in this model, the observer must compute the posterior probability $p(x_t|o_1, o_2, \ldots, o_t)$ given all outcomes up to the current point. Importantly, this state space modeling approach can be implemented with either continuous or binary variables. When the hidden and observed variables are continuous, inference leads to the Kalman filter, while with binary variables, it leads to the HMM. We provide a formal treatment of this framework in Appendix A; here, we focus on the key intuitions.

Kalman-based hierarchical models

When both the hidden state and outcomes are continuous, and the conditional dependencies are Gaussian, optimal inference is tractable and given by the Kalman filter. The Kalman filter provides a closed-form solution in which the posterior mean (the best estimate of the hidden state) is updated by a delta-rule with a dynamic learning rate (Appendix A.2). Crucially, this learning rate depends on both volatility and stochasticity: it increases with volatility and decreases with stochasticity, reflecting how quickly an organism should revise its expectations in the face of prediction errors. This framework has become foundational in statistical theories of conditioning, with the Kalman gain corresponding directly to the associability term in classical conditioning models (Dayan et al., 2000; Dayan & Long, 1998; Dayan & Yu, 2003). It has helped explain numerous phenomena in the conditioning literature, including blocking, trial-

order effects, auto-shaping, and latent learning (Dayan & Yu, 2003; Gershman, 2015; Kakade & Dayan, 2002). More broadly, this Bayesian approach formalizes the idea that animals infer causal relationships between variables, including unobservable ones, and estimate uncertainty surrounding these relationships, allowing them to weigh new evidence against prior knowledge based on their relative reliability (Courville et al., 2006; Gershman et al., 2010; Piray & Daw, 2021a).

Here, we will focus on hierarchical modeling approaches built on the Kalman filter framework, which extends learning to account for volatility, particularly in the context of binary outcomes. These models typically augment the standard generative model by treating volatility as a dynamic variable that evolves according to its own conditional transition probability, $p(v_t|v_{t-1})$. This extension, while theoretically elegant, introduces a significant computational challenge: the inference problem is no longer analytically tractable, necessitating various forms of approximate inference. Several notable approaches have addressed this challenge. In their seminal work, Behrens and colleagues employed a brute force numerical integration method to perform approximate inference in this hierarchical framework (Behrens et al., 2007, 2008). A more computationally sophisticated solution was subsequently introduced by Mathys and colleagues, who developed a parametric variational approach (Iglesias et al., 2013; Mathys et al., 2011). In their model, the posterior distributions over both x_t and v_t are approximated by Gaussian distributions through an explicit hierarchical update procedure. This procedure first computes the mean and variance of the approximate posterior for x_t and then calculates the mean and variance of the approximate posterior for v_t. Specifically, Mathys et al. assumed that the transition probability for v_t follows a Gaussian distribution and implemented an exponential transformation function to appropriately scale the variance of the transition probability before incorporating it into the next level of the hierarchy (Iglesias et al., 2013; Mathys et al., 2011). We have previously proposed a slightly different formulation for the transition probability of volatility, utilizing Gamma distribution instead (Piray & Daw, 2020).

Particularly relevant to our present discussion is how these models have been adapted for binary outcomes. These models modified the conditional dependency of the continuous generative model to accommodate binary outcomes, using $p(o_t|x_t) = \text{Bernoulli}\big(\sigma(x_t)\big)$, where $\sigma(x) = 1/(1 + \exp(-x))$ is the sigmoid function. This formulation has an important conceptual implication: it effectively eliminates an explicit stochasticity parameter. This does not mean the absence of stochasticity in the model; rather, stochasticity exists but becomes conflated with other parameters. For example, even with zero volatility and an initial hidden variable value of 1, the probability of observing a positive outcome (1) is 0.73. Thus, outcomes remain stochastically generated despite the lack of an explicit stochasticity parameter. To implement inference with binary outcomes, these models have employed various well-established approximation strategies, such as Taylor expansions (Mathys et al., 2011) and moment matching (Piray & Daw, 2020) techniques. Despite their technical differences, these approaches share a

common foundation: they preserve the underlying generative structure of the Kalman filter, allowing them to derive approximate inference methods with update equations that closely resemble the standard Kalman update rules. However, as we have formally shown previously (Piray & Daw, 2020), this approach introduces a significant theoretical problem. The learning rate—defined as the ratio of prediction update to prediction error—exhibits a problematic relationship with volatility: it increases dramatically following large, surprising events (i.e., large unsigned prediction errors), regardless of the true volatility level, and even when volatility is zero (Supplementary Figure 1). This spurious effect stems directly from combining the Kalman filter with the sigmoid transformation used to convert the continuous hidden variable x_t into a probability (see Appendix B for a formal treatment). This limitation is particularly problematic for experimental designs that manipulate volatility, such as reversal learning tasks with distinct low and high volatility blocks. In high-volatility blocks, where contingencies change frequently, large surprises naturally occur more often, creating an artificial correlation between volatility and the implied learning rate. But the problem runs deeper: even when volatility is fixed, the learning rate increases with unexpected outcomes that are entirely due to stochasticity. This fundamental confound underscores the need for computational models that properly disentangle these distinct sources of uncertainty.

Hidden Markov Model

In contrast to previous approaches, we develop a generative model that more faithfully represents the inherent structure of binary environments. Rather than adapting continuous-variable frameworks to binary data, we directly exploit the established architecture of state space models specifically designed for binary hidden causes—an approach with strong foundations in machine learning (Bishop, 2006; Wainwright & Jordan, 2008), for which the inference is tractable and given by the HMM (Ghahramani, 2001; Rabiner, 1989). The HMM has broad applications across the biological sciences, including areas such as genetics (Burge & Karlin, 1998; Krogh et al., 2001), and ecology (McClintock et al., 2020). HMMs have also been widely used in neuroscience for analysis of neural data (Baldassano et al., 2017; Bolkan et al., 2022; Calhoun et al., 2019; Vidaurre et al., 2018; Wiltschko et al., 2015). However, unlike the Kalman filter, HMMs have rarely been used as a computational model to explain phenomena related to conditioning and learning, though they have occasionally been used to fit human experimental data (Hampton et al., 2006; Schlagenhauf et al., 2014).

Here, we present a generative model consisting of a binary hidden state and binary outcomes, with two distinct sources of noise affecting the process at different levels. We provide the main equations below; see Appendix A.3 for the full derivation. Specifically, the model centers on a binary hidden state variable, x_t, which can take only two values (0 or 1) and evolves over discrete time steps according to probabilistic rules. From this hidden state, binary outcome, o_t, is generated but potentially corrupted by outcome noise (Figure 1). This framework allows us to

explicitly model both the volatility of the environment and the stochasticity of outcomes as independent sources of noise with separate parameters.

In particular, the hidden variable evolves according to the following rule:

$$x_t = \begin{cases} x_{t-1}, & \text{with probability } 1 - v \\ 1 - x_{t-1}, & \text{with probability } v \end{cases} \tag{1}$$

where $0 \leq v \leq 0.5$ is the volatility parameter governing the degree of noise in this process. If $v = 0$ there is no diffusion noise, and the hidden state never changes. If $v = 0.5$, there is an equal probability that the hidden state switches or stays the same on every trial, indicating that the noise is maximum. This means that the conditional transition probability is given by a Bernoulli distribution:

$$p(x_t | x_{t-1}) = \begin{cases} 1 - v, & \text{if } x_t = x_{t-1} \\ v, & \text{if } x_t = 1 - x_{t-1} \end{cases} \tag{2}$$

The binary outcome is noisily generated based on the hidden state, with the amount of noise governed by the stochasticity parameter:

$$o_t = \begin{cases} x_t, & \text{with probability } 1 - s \\ 1 - x_t, & \text{with probability } s \end{cases} \tag{3}$$

where $0 \leq s \leq 0.5$ is the stochasticity parameter governing the degree of noise. The outcome switches with respect to the hidden variable with a probability given by s. If $s = 0$, there is no outcome noise, and the outcome always follows the hidden state. If $s = 0.5$, on the other hand, there is equal probability that the outcome switches or follows the hidden state, essentially generating outputs completely randomly regardless of the hidden state. Thus, the conditional outcome probability is also given by a Bernoulli distribution:

$$P(o_t | x_t) = \begin{cases} 1 - s, & \text{if } o_t = x_t \\ s, & \text{if } o_t = 1 - x_t \end{cases} \tag{4}$$

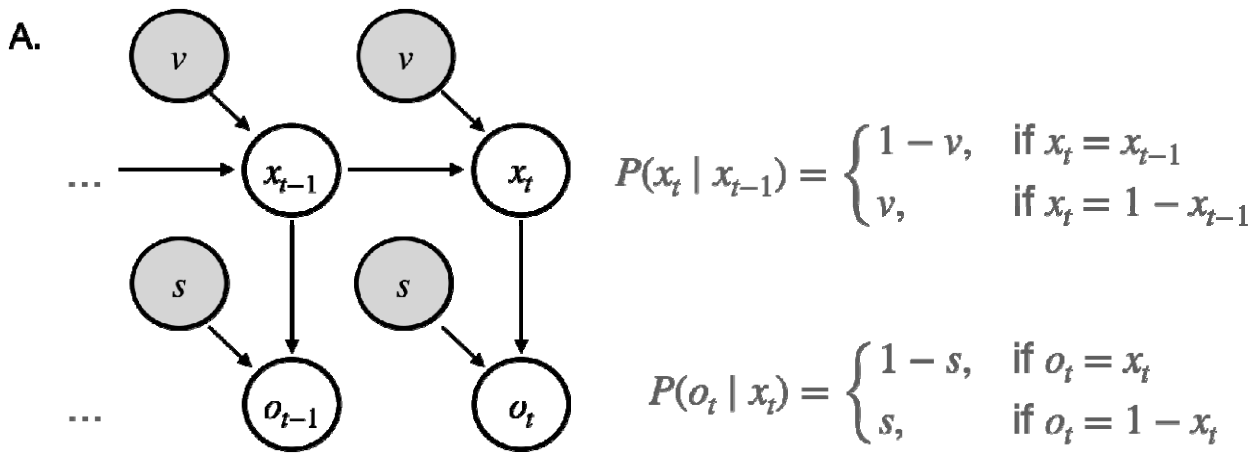

Figure 1. Graphical model of Hidden Markov Model. The model infers hidden state ( ) by updating predictions based on observed outcomes ( ), where volatility ( ) controls the probability of state transitions and stochasticity ( ) determines the noise in outcome generation. In the model, and are treated as constants, determining the volatility of the hidden state and the level of stochasticity in outcomes relative to it, respectively.

To illustrate, consider a typical probabilistic reversal learning task, where participants learn that one option is rewarded 80% of the time while the other is rewarded 20%, and these contingencies periodically reverse (Cools et al., 2002; Costa et al., 2015; Izquierdo et al., 2017). In our framework, the reversal frequency corresponds to volatility ( ), while the 80/20 reward schedule corresponds to stochasticity ( ). Critically, from the participant's perspective, both sources of noise produce the same observable result: an unexpected outcome. When a participant chooses the supposedly better option but receives no reward, this could mean the contingencies have reversed (volatility), or it could simply be the 20% chance of no reward from a still-correct option (stochasticity). Disentangling these possibilities is the core computational challenge our framework addresses.

For our binary state space model, the inference process—computing the posterior probability distribution —is analytically tractable with simple update equations within the HMM framework (Ghahramani, 2001). The HMM framework gives exact, closed-form equations for how beliefs should be optimally updated in light of new outcomes.

The posterior distribution maintains a Bernoulli form at each time step: where represents the probability that the hidden state equals 1 given all outcomes up to and including time . This parameter is updated recursively through a set of equations that incorporate the previous belief state, volatility parameter, stochasticity parameter, and current outcome. Specifically, the update process involves three steps:

(5)

$$\ell = (1 - s)o_t + s(1 - o_t) \tag{6}$$

$$r_t = \frac{q \cdot \ell}{q \cdot \ell + (1 - q)(1 - \ell)} \tag{7}$$

with $r_0 = 0.5$. These equations have clear interpretations:

- The intermediate variable q equals $p(x_t = 1|o_1, o_2, \ldots, o_{t-1})$, representing our belief about the current state before seeing the current outcome. This quantity incorporates both the previous belief (r_{t-1}) and the probability of state transitions (governed by volatility v).
- The variable ℓ equals $p(o_t|x_t = 1)$, representing the likelihood of observing the current outcome o_t if the hidden state is 1. This likelihood depends on the stochasticity parameter s, which determines how reliably the outcome reflect the state.
- The final equation for r_t is an application of Bayes' rule, combining the prior belief q with the likelihood of the outcome, ℓ, to compute the posterior probability.

This recursive Bayesian update scheme provides a computationally efficient method for tracking belief states as new outcomes arrive. The equations maintain mathematical elegance while capturing how both volatility and stochasticity affect the learning process in binary environments (for full derivations see Appendix A.3).

Unlike the Kalman filter, there is no explicit learning rate variable in the update rules of the HMM. Nevertheless, the learning rate, defined as the ratio of prediction update to prediction error serves as an informative descriptive measure of learning, even if the underlying process does not involve an explicit error-driven update. Specifically, the HMM provides a prediction on each trial (i.e., r_t in Eqs. 5-7), allowing for the computation of both a prediction error and an update. As a result, the learning rate can be quantified as $(r_t - r_{t-1})/(o_t - r_{t-1})$. This approach is similar in practice to how prior studies using binary outcomes have typically assessed learning.

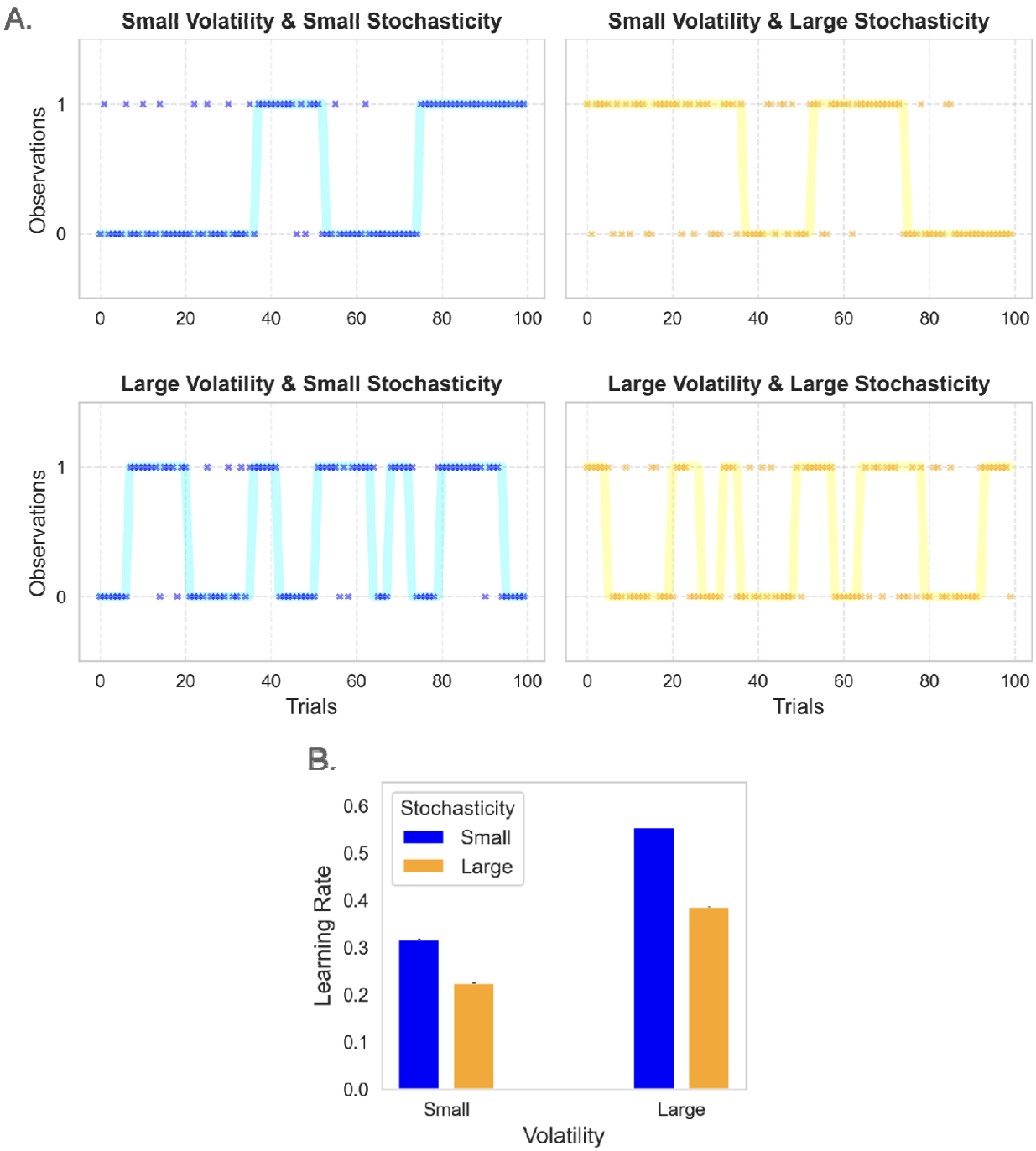

Figure 2. HMM simulation analysis. (A) Examples of generated time series with volatility levels of 0.025 and 0.1, and stochasticity levels of 0.125 and 0.25. The solid lines represent the hidden state, whose variability depends on volatility levels. Individual data points are marked separately, with their variations relative to the hidden state determined by the stochasticity level. (B) Learning rates quantified by the optimal HMM—given the true values of volatility and stochasticity—vary systematically with both factors. The simulation was repeated 1000 times and the means and standard errors are plotted. The model suggests that volatility and stochasticity exert opposing

effects: higher volatility increases the learning rate, while higher stochasticity decreases it.

We thus investigated how volatility and stochasticity influence the learning rate in the HMM. To do this, we defined a 2×2 prototypical design, which includes a systematic manipulation of both volatility and stochasticity (Figure 2A). This results in four blocks of binary time series, where values of volatility and stochasticity are either small or large. Next, we simulated the HMM for each block separately using the true values of volatility and stochasticity. This allowed us to quantify the learning rate by regressing the trial-by-trial update in predictions against prediction error across blocks. Block-specific intercepts were also included in the regression analysis. This method yields the learning rate per block. Notably, this approach reveals the theoretically optimal learning rate (i.e., ideal observer learning rate), since the HMM is the optimal inference model for this generative process and our simulations use the true parameter values for volatility and stochasticity. This simulation analysis (Figure 2B) revealed that the learning rate for binary outcomes follows the intuitive predictions: higher volatility increases the optimal learning rate, indicating rapid adaptations to new information. Conversely, higher stochasticity decreases the optimal learning rate, indicating the need for having a larger window for averaging outcomes, as each individual outcome is less reliable.

PF-HMM: a model for inferring the cause of noise based on binary outcomes

The HMM model presented above provides an optimal algorithm for recursively computing the posterior over the hidden state when new outcomes arrive. However, this algorithm, similar to the Kalman filter, relies on the critical assumption that the values of volatility and stochasticity are known and given. In real-world environments, neither of these values is known a priori, and both must be estimated from observations. This presents a fundamental challenge, especially when both volatility and stochasticity are unknown and potentially fluctuating over time.

To address this challenge, we extend the state space model presented above to include both volatility and stochasticity as dynamically changing random variables. Specifically, we expand the state space model of Figure 1 by assuming that v_t and s_t evolve according to independent diffusion processes (Figure 3B). Each variable drifts smoothly over time around its previous value, controlled by separate diffusion parameters. Importantly, this formulation allows the variables to fluctuate over time without being biased to drift in any direction. The formal mathematical definitions of these diffusion processes have been detailed in Appendix A.4.

Since this model no longer permits exact inference, we turn to approximate inference techniques. We employ a PF approach, a standard and efficient method broadly applicable to such problems (Andrieu et al., 2010; Doucet et al., 2001). The key idea is to maintain a population of candidate hypotheses (particles), each representing a possible combination of volatility and stochasticity values. As outcomes arrive, particles that better explain the data receive more weight, allowing the model to track the noise parameters over time.

Importantly, the model leverages a key insight: when provided with samples of volatility and stochasticity parameters, we can conduct tractable inference for the hidden state through the HMM framework. This combined strategy is known as Rao-Blackwellized PF (Doucet et al., 2000; Doucet & Johansen, 2011), which efficiently handles the computational challenges of our the extended generative model. In sum, the new PF-HMM model estimates v_t and s_t using PF, while inference for x_t is conducted via HMM, conditional on these estimated parameter values.

The algorithm operates in three distinct steps on each trial. The first step involves prediction, where individual particles transition forward in time according to the generative dynamics for volatility and stochasticity. This gives rise to a set of sampled values for these variables. The second step evaluates how well each particle's predictions match the observed outcome. Particles that better predict the outcome receive higher weights, while those that predict poorly are down weighted. As is customary in particle filtering methods, systematic resampling is applied during this phase if the effective number of particles drops below a certain threshold (Douc & Cappe, 2005). The third step applies the HMM to integrate the new outcome with previous estimates, updating the belief about the hidden state for each particle. The expected value for all random variables at each trial is computed as the weighted average across all particles. The complete formal algorithm is given in Appendix A.4.

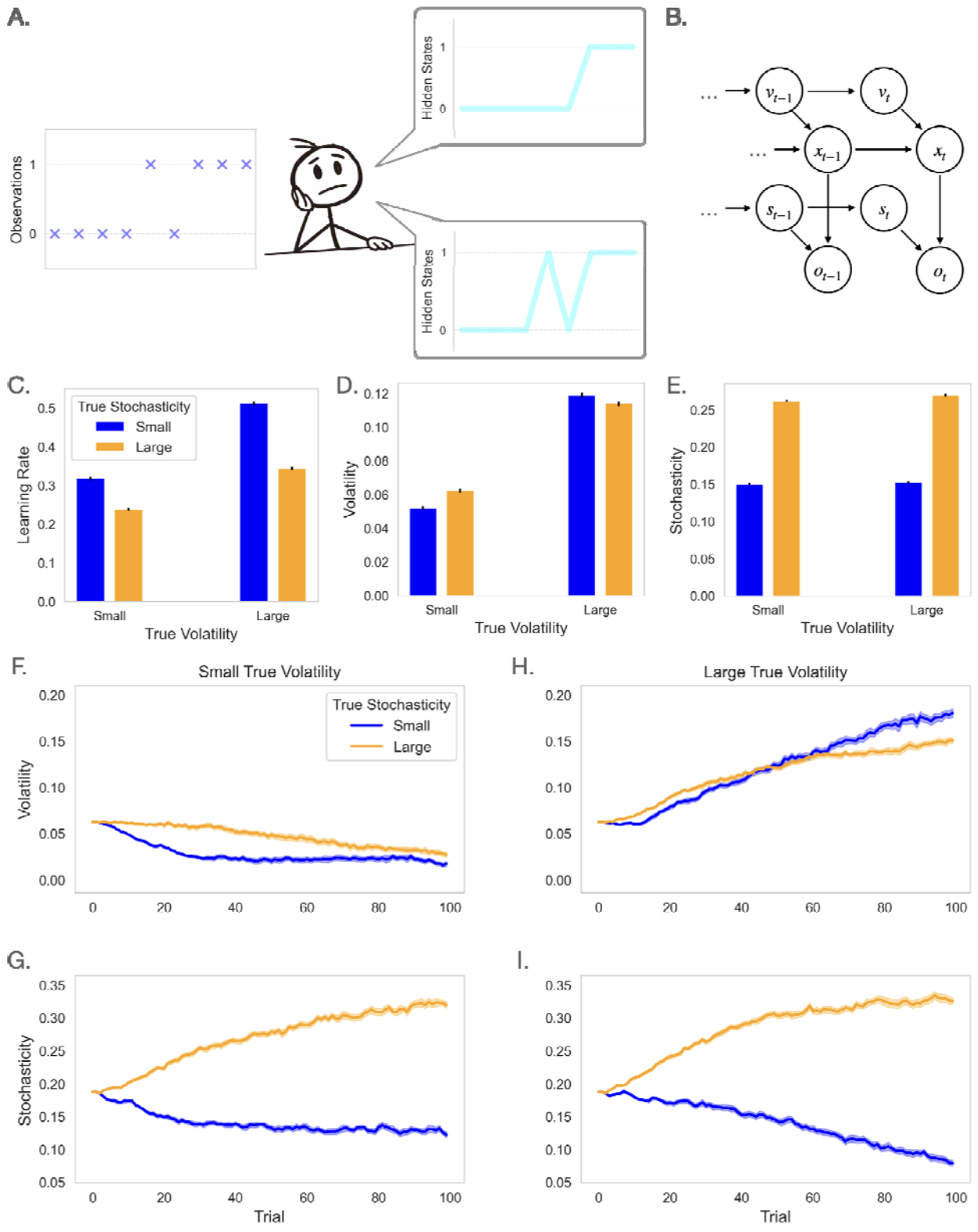

Figure 3. PF-HMM simulation analysis. (A) In the real world, one must infer the hidden variable based solely on observations, which leads to trial-by-trial attribution (or misattribution) of experienced noise to either volatility or stochasticity. The same observed outcome sequence can arise either from

frequent hidden-state changes (high volatility, low stochasticity) or from a stable state corrupted by outcome noise (high stochasticity, low volatility). Because the latent state is unobserved, learners must infer the source of variability, creating a risk of misattribution that leads to different learning-rate adjustments. (B) The structure of the (generative) model: The binary outcome on trial t, denoted o_t, is generated from a binary hidden state x_t, corrupted by independent noise whose magnitude is determined by the level of stochasticity, s_t. The hidden state itself evolves from its previous value, x_{t-1}, corrupted by another source of noise governed by the volatility parameter, v_t. Both v_t and s_t also evolve over time, changing noisily based on their values in the previous trial. The learner's task is to infer the current hidden state, as well as volatility and stochasticity, based on the sequence of observed outcomes. (C-I) PF-HMM was simulated using the same 2×2 design as in Figure 2B. The model's learning rate is influenced by both true volatility and stochasticity, which have opposite effects: higher volatility increases the learning rate, while higher stochasticity decreases it (C). Estimated volatility captures differences in true volatility (D), and estimated stochasticity captures differences in true stochasticity (E). Panels F-G show volatility and stochasticity estimates from the model for blocks with low true volatility, while panels H–I show estimates for blocks with high true volatility. The simulation was conducted on the same 1000 time series used in the analysis shown in Figure 2. Error bars in panels C–E represent the standard error of the mean, and those in panels F–I represent the standard error of the median.

Next, we tested this model using the same 2×2 paradigm described above. In this paradigm, the true values of volatility and stochasticity are fixed within each block and only change between blocks. However, the model has no prior knowledge of this structure. Instead, the initial values for volatility and stochasticity are set to the midpoint between the low and high true values used in the experiment, and the diffusion parameters for v_t and s_t are both set to 0.02. Thus, this approach allows us to examine how well the model can adapt to the different volatility and stochasticity conditions across blocks without being explicitly informed about the block structure or the true parameter values. By keeping the model's parameterization constant across all conditions, we can observe its ability to infer the underlying environmental statistics purely from outcomes.

Figure 3 shows the results of this analysis. We first quantified the learning rate per block. As shown in Figure 3C, the model's learning rate exhibits a pattern very similar to the optimal learning rate presented in Figure 2B: learning rates increase with true volatility and decrease with true stochasticity. Notably, this occurs despite the fact that the model has no prior knowledge of the true volatility and stochasticity values, requiring it to estimate these parameters solely from outcomes.

Furthermore, examining the model's estimates of volatility and stochasticity reveals it performs remarkably well in reconstructing these underlying parameters. Importantly, the model's estimate of volatility shows minimal sensitivity to manipulations of true stochasticity, while its estimate of stochasticity remains largely unchanged across different levels of true volatility. This selective sensitivity demonstrates that the model can successfully attribute experienced noise to its actual source, distinguishing between environmental change (volatility) and observational noise (stochasticity).

Ablated variants of PF-HMM

Our final set of models comprises ablated models of volatility and stochasticity. Previous simulation analyses revealed a crucial aspect of the model: volatility and stochasticity inferences are inherently interrelated. Although these are independent sources of noise from the perspective of the generative model, they become dependent for the purpose of inference (i.e., in the eyes of the observer). This is exactly why dissociating volatility and stochasticity is computationally challenging; both factors increase experienced noise (i.e., variability in outcomes; Figure 4). Separating these contributions requires a process of credit assignment, or as it is called in the Bayesian literature, "explaining away," where competing explanations for outcomes are weighed against each other. A pathological model that ignores one factor incorrectly attributes noise from that factor to the other. This dynamic may be particularly relevant to neurological and psychiatric conditions that selectively impair the processing of volatility or stochasticity. Our model predicts that such impairments would not only reduce learning rate modulation but could also lead to paradoxical reversals, where volatility and stochasticity are mistakenly substituted for one another due to a breakdown in the explaining away mechanism.

Figure 4 illustrates this effect using our 2×2 design, highlighting two distinct ablated models. In the stochasticity-ablated model, all experienced noise is attributed to volatility, causing a persistent increase in learning rate regardless of the actual source of noise (Figure 4E). Conversely, in the volatility-ablated model, all noise is attributed to stochasticity, resulting in a decrease in learning rate irrespective of the true source of noise (Figure 4F). In other words, a model lacking one noise parameter inevitably misattributes experienced noise to the remaining one, leading to systematic errors in learning. This maladaptive pattern of misattributions is also evident in the volatility and stochasticity estimates of both ablated models: In the stochasticity-ablated model, volatility estimates always increase regardless of the true source of noise (Figure 4G). Similarly, in the volatility-ablated model, stochasticity estimates always increase, irrespective of the actual source of noise (Figure 4H).

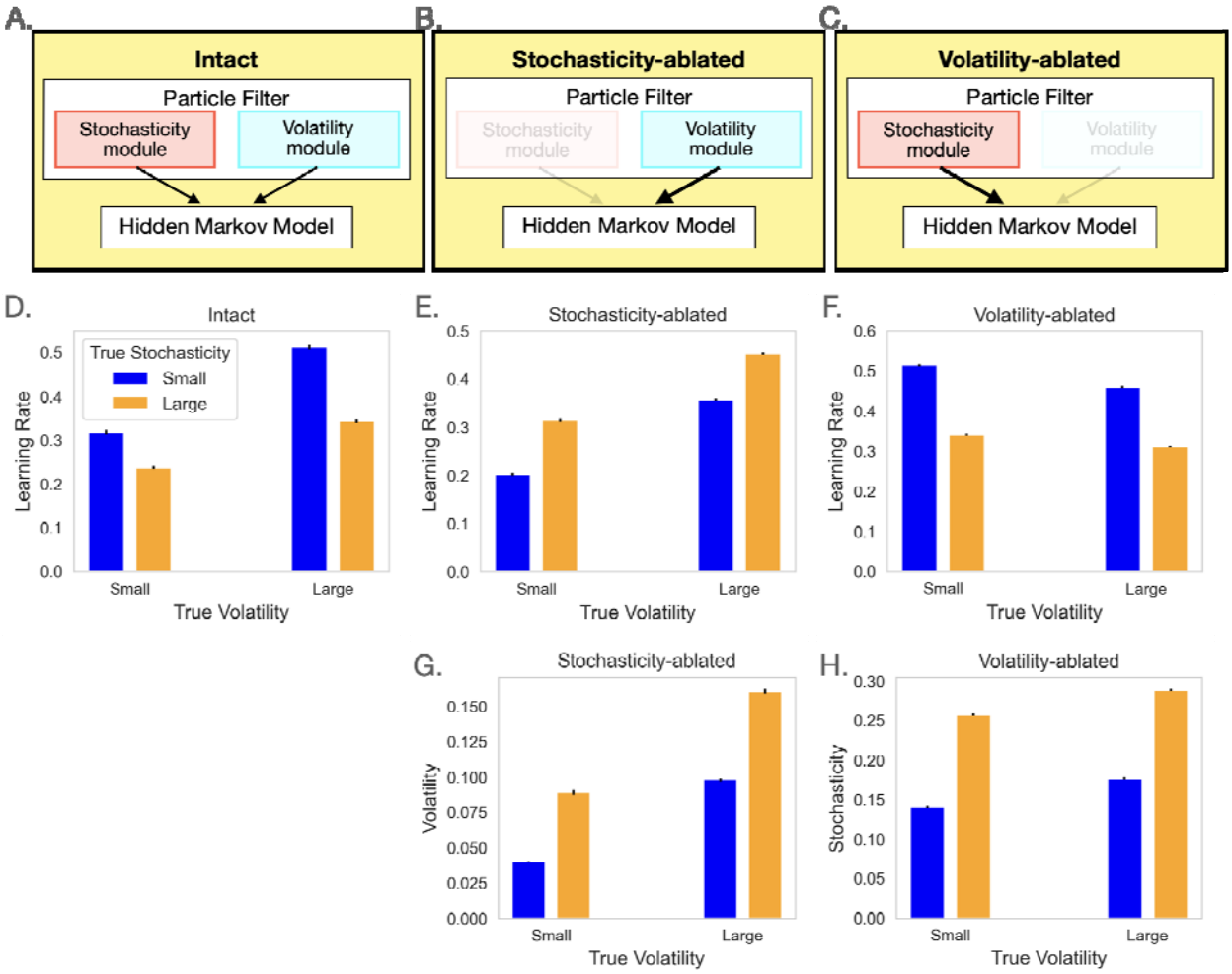

Figure 4: Ablated variants of PF-HMM simulation analysis. (A) In the intact model, stochasticity and volatility compete to explain the observed noise arising from the mismatch between the HMM component's predictions and the actual outcomes. The intact model produces behaves optimally by inferring the source of noise based on observed outcomes. (B-C) Two distinct ablated models exhibit different pathological behaviors: ablating the stochasticity module results in a fixed level of stochasticity across blocks, causing noise due to stochasticity to be misattributed to volatility. In contrast, ablating the volatility module results in a fixed level of volatility across blocks, which leads to volatility-related noise being misinterpreted as stochasticity. (D-H) The intact model's learning rate is the same as in Figure 3C and is reproduced here for comparison (D). For both ablated models, the effect of ablating does not merely eliminate the corresponding influence on the learning rate but instead causes a reversal. Specifically, the stochasticity-ablated model exhibits an elevated learning rate in response to increasing stochasticity (E). On the contrary, the volatility-ablated model shows a reduced learning rate as volatility rises (F). This reversal occurs because the noise from the ablated factor is misattributed to the remaining module. (G) The stochasticity-ablated model incorrectly infers increased volatility in highly stochastic environments. (H) Conversely, the volatility-ablated model incorrectly attributes volatility to stochasticity, resulting in an inflated stochasticity estimate in more volatile environments. The simulation was conducted on the same 1,000 time series used in the analysis shown in Figures 2–3. Error bars reflect the standard error of the mean computed over 1,000

simulations.

These ablated variants of the model offer a promising framework for understanding disrupted learning in clinical populations. This is particularly important in the context of psychiatric disorders, such as anxiety disorders, depression, impulse control disorders, and schizophrenia. Although maladaptive processing of uncertainty has been associated with all these disorders, it remains unknown how these same computational mechanisms might influence many mental illnesses transdiagnostically. Our framework can provide valuable insights into these shared computational dysfunctions across different psychiatric conditions.

Human participants infer the source of noise from binary outcomes

We conducted a behavioral experiment to examine whether humans dissociate between volatility and stochasticity from binary outcomes. Following our prototypical paradigm, we designed a task to study learning from binary outcomes in a 2×2 setting where both factors were systematically manipulated. The task consisted of four blocks, each with distinct fixed parameters for true volatility and stochasticity, though participants were not informed of these values or their order across blocks. We recruited 73 participants (34 female, age range: 18-66, median age: 35) through the online platform Prolific.

We developed an engaging cover story (Figure 5A) in which participants predicted which side of a beach a sea lion would visit, based on where treasures had been discovered on previous trials. Crucially, participants could not observe the sea lion directly; they could only use the treasure locations to guide their predictions. They were told that the sea lion tended to stay on the same side of the beach but could suddenly switch (volatility). They also learned that ocean waves could independently carry treasures to the opposite side (stochasticity). The instructions carefully emphasized that volatility and stochasticity were independent sources of uncertainty, stemming from the sea lion's behavior patterns and wave effects, respectively. Participants were also informed they would interact with four different sea lions on four different beaches. Each sea lion–beach pair constituted one block of 40 trials, although participants were not informed about the length of the blocks. This experimental design enabled us to measure participants' behavioral responses to systematic manipulations of both true volatility and true stochasticity (See Supplementary Figure 2 for the timeseries used in the task).

We first used a blockwise HMM to analyze choice data and quantify the learning rate per block. The model was fitted to participants' responses, quantifying two parameters per block: a volatility parameter, v, and a stochasticity parameter, s. We used maximum a posteriori estimation with wide priors to fit these three parameters to choice data (Piray, Dezfouli, et al., 2019); see Methods for further details. Notably, this analysis is theory-neutral with respect to inference about volatility and stochasticity, as the HMM does not impose any assumptions about how these factors should vary across blocks. Therefore, any observed differences in fitted parameters or learning rates across blocks should be attributed solely to differences in

participants' behavioral responses rather than any model-driven assumptions. We also conducted a recovery analysis in which the blockwise HMM model was simulated using different sets of known parameters. The resulting time-series data were then analyzed using the same model-fitting procedure. This analysis demonstrated that the model's parameters are almost perfectly recoverable using this approach (Supplementary Table 1 and Supplementary Figure 3).

Using the best-fitted parameters (Supplementary Table 2), we generated predictions and quantified the learning rate per block (Figure 5B). We then analyzed these learning rates across all participants. Our analysis revealed a significant positive main effect of true volatility on learning rate ($t(72) = +5.286, P < 0.001$), indicating that participants appropriately increased their learning rates in response to higher volatility. Conversely, we found a significant negative effect of true stochasticity on learning rate ($t(72) = -4.302, P < 0.001$), demonstrating that participants decreased their learning rates when true stochasticity increased, consistent with normative predictions from the model. No significant interaction between the two factors was found ($t(72) = -1.242, P = 0.22$). The regression analysis also revealed block-level effects that were independent of prediction error; see Supplementary Table 3. These results extend previous work demonstrating effects of volatility on learning from binary outcomes by additionally revealing a negative effect of stochasticity on learning.

Importantly, fitting PF-HMM to participants response exhibits a very similar pattern of effects (Figure 5C). A practical challenge in fitting PF models (and related Monte Carlo sampling approaches) is that the likelihood function is not differentiable with respect to the generative parameters, which has limited the application of these models in psychology. To address this, we followed the procedure that we have used in tasks with continuous observations (Piray & Daw, 2024), and adopted an optimization scheme based on Gaussian processes that is well suited for non-differentiable likelihoods (Gelbart et al., 2014; Snoek et al., 2012). We validated this approach through a parameter recovery analysis: the model was simulated with known parameter values, and the resulting data were subjected to the same fitting procedure. This analysis confirmed that the model's parameters are reliably recoverable (Supplementary Table 4, Supplementary Figure 4). This suggests that participants dissociate volatility and stochasticity based solely on observed binary outcomes.

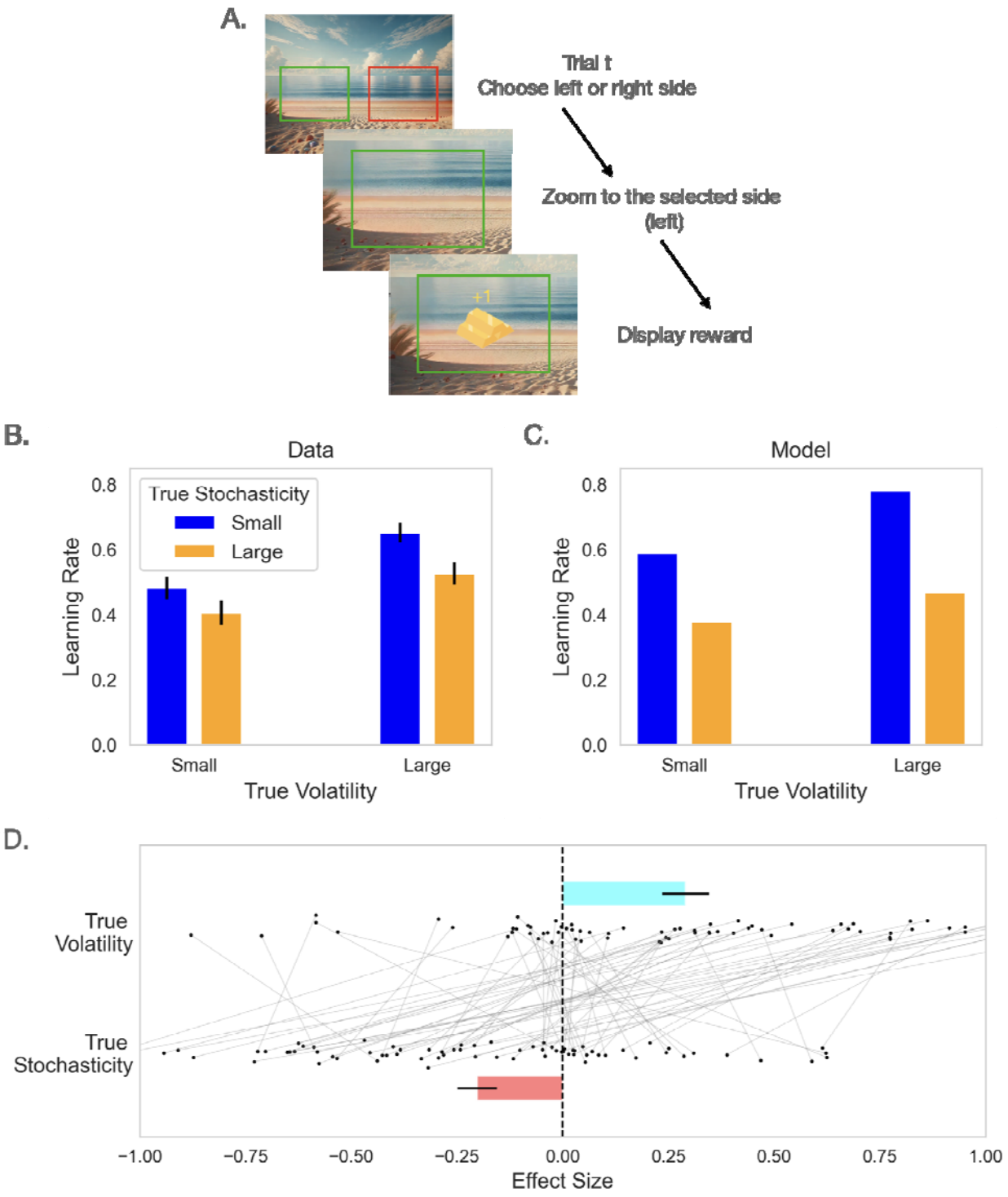

Figure 5. Experimental findings. (A) On every trial, participants choose one side of the beach to find treasures brought by a sealion not visible to them. Participants can use left or right arrow key to choose one side when they see the whole view of the beach. Then the screen will zoom in to that side and display either reward or no reward. The task has four blocks with a factorial design, manipulating both true volatility and true stochasticity. In the illustration, the left side with green

box is chosen (the boxes are not displayed in the actual task). (B) Model-neutral analysis of behavioral data. Mean learning rate per block is plotted (n = 73), which calculated by fitting the blockwise HMM model. The results show strong main effects of both factors: the learning rate increases with higher true volatility and decreases with higher true stochasticity. Note that this analysis is theory-neutral with respect to changes in learning rate across blocks, as the HMM does not impose any assumptions about how volatility and stochasticity parameters vary across blocks. (C) Learning rate from fitting the PF-HMM model on subject data shows similar effects to model neutral analysis. (D) Main effects of both factors on the learning rate coefficient have been plotted for all participants. The effect size of true volatility is calculated by learning rate of two high volatility blocks minus learning rate of two low volatility blocks. Effect size of true stochasticity is calculated similarly. Mean and standard error of the mean are plotted in B-D.

Human responses are slower when identifying the true source of noise is more difficult

The previous analysis demonstrates that human participants can dissociate volatility and stochasticity by observing outcomes alone, in line with predictions from the PF-HMM. Here, we aim to further elucidate the cognitive and computational mechanisms underlying this dissociation by examining whether the PF-HMM also accounts for a related but statistically independent aspect of behavior—namely, response times.

Our modeling approach offers insight by predicting that identifying the source of noise becomes more difficult when particles are indistinguishable. This indistinguishability arises when most or all particles assign similar likelihoods to a given outcome, resulting in low variability across per-particle likelihoods. Under these conditions, the model struggles to determine which particles best explain the observation, as all provide similarly good fits. This ambiguity makes it harder to accurately attribute the observed noise to its true source—volatility or stochasticity. In contrast, high variability across particles suggests that some particles account for the outcome much better than others, enabling more confident inference about the underlying cause of the noise.

To test these predictions, we analyzed response times, anticipating that they might reveal additional signatures of the underlying computational process. Specifically, we hypothesized that responses would be slower when the core computational task—identifying the source of noise—is more difficult, which occurs when particles are indistinguishable (i.e., when particle likelihood variability is low). We quantified trial-by-trial particle likelihood variability as the standard deviation of per-particle likelihoods across particles and performed a regression analysis examining the relationship between response times and particle likelihood variability. To control for potential confounding effects, we included additional regressors: average particle likelihood, estimated volatility, estimated stochasticity, and an accuracy regressor indicating whether the response matched the outcome, ensuring that any observed effects are not due to a response time and accuracy trade-off. Model predictions were generated using the median fitted parameters across subjects (For more details see Methods).

This analysis revealed a highly significant negative relationship between response time and particle variability (; Figure 6). These results suggest that when variability across particles is low, participants respond more slowly, likely reflecting the increased difficulty of accurately identifying the source of noise due to greater indistinguishability in how particles fit the observed outcome (See also Supplementary Table 5).

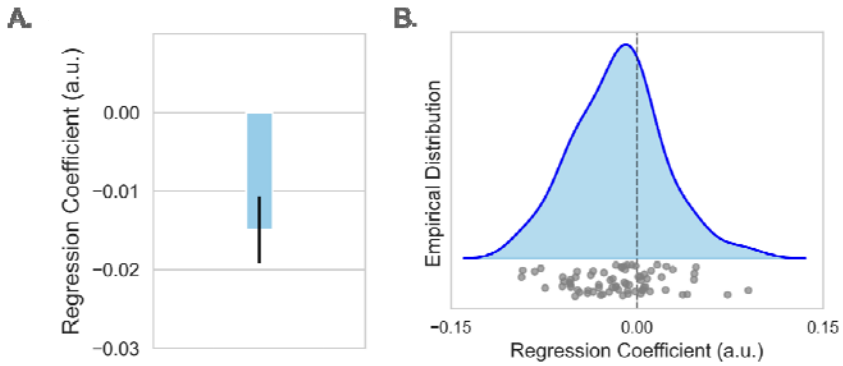

Figure 6. Computational mechanisms underlying how participants infer the source of noise. (A) The regression coefficient from an analysis of participants' response times (in seconds) against particle likelihood variability (fitted to each subject is plotted. Particle variability is quantified as the standard deviation of per-trial particle likelihoods. The significant negative relationship indicates that lower particle variability is associated with longer response times. The mean and standard error is plotted. (B) Individual data points and the empirical distribution are plotted.

Generalization to loss learning

To test whether our findings generalize beyond reward learning, we conducted an additional behavioral experiment, the turtle task, using the same prototypical 2×2 design that independently manipulated volatility and stochasticity. The turtle task is structurally identical to the sea lion task but implements loss avoidance rather than reward seeking, allowing us to test whether the dissociation between volatility and stochasticity holds across outcome valence. We recruited 30 participants for this experiment using Prolific.

Using the model-neutral blockwise HMM analysis, we first quantified learning rates for each block (Figure 7A). Replicating the results from the sea lion task, learning rates showed a significant positive main effect of true volatility ( ) and a significant negative main effect of true stochasticity ( ), with no significant interaction (; Supplementary Table 6). These results indicate that participants appropriately increased learning rate with volatility and decreased it with stochasticity, replicating our reward learning findings from the sea lion task.

We also repeated the response time analysis to test whether the PF-HMM predicts difficulty in inferring volatility versus stochasticity, as reflected in response times. Replicating the sea lion

task results, response times were significantly larger when particle likelihood variability was low, indicating greater difficulty in identifying the source of uncertainty (
; Figure 7C; Supplementary Table 7).

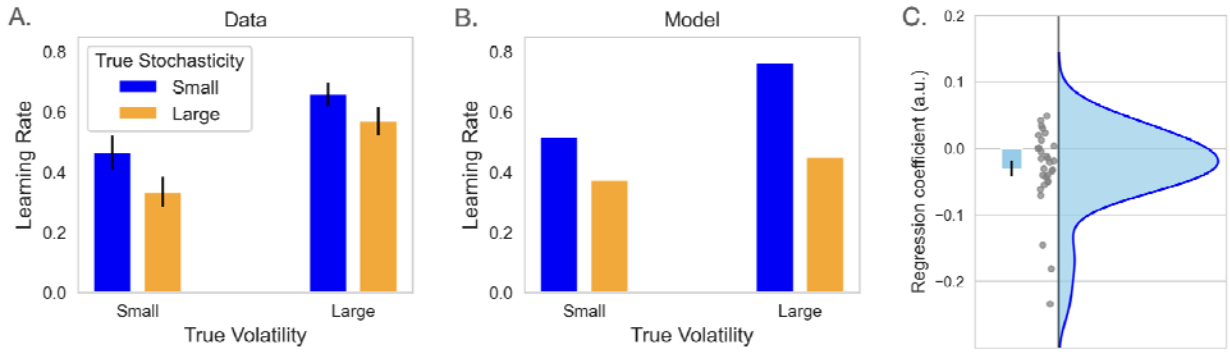

Figure 7. Generalization to an independent loss-learning task (turtle task). (A) Model-neutral analysis of an additional behavioral data with similar task setup. Mean learning rate per block is plotted (n = 30), which calculated by fitting the HMM model. The results show strong main effects of both factors: the learning rate increases with higher true volatility and decreases with higher true stochasticity. (B) Mean learning rate from fitting the PF-HMM model on subject data shows similar effects to model neutral analysis. (C) Linear regression results of participants' response time (in seconds). The regression coefficient of particle likelihood variability (is plotted. The significant negative relationship indicates that lower particle variability is associated with longer response times. The mean and standard error, along with individual data points and the empirical distribution, are plotted.

Model comparison with alternative reversal-learning models

To assess whether the PF-HMM provides a better account of participants' behavior than previous learning models, we compared it against two alternatives commonly used in reversal-learning tasks: the Pearce–Hall associability (PHA) model (Barnby et al., 2022; Li et al., 2011; Pearce & Hall, 1980) and the binary Hierarchical Gaussian Filter (HGF) (Iglesias et al., 2013; Mathys et al., 2011). All models were fitted to individual participants' data by maximum likelihood estimation. Model evidence was calculated for each model using the Bayesian information criterion (Schwarz, 1978) and were used for random effects Bayesian model selection (Piray, 2026; Rigoux et al., 2014; Stephan et al., 2009). This procedure yields the model frequency, the expected proportion of participants best described by each model, and the protected exceedance probability, which quantifies the likelihood that a given model is the most frequent, while accounting for the possibility that observed differences arise by chance.

Bayesian model selection provided strong evidence in favor of the PF-HMM across both the sea lion and turtle tasks (Table 1). In the sea lion task, the PF-HMM achieved the highest model frequency (0.757) with a protected exceedance probability of 1.00, substantially exceeding both the PHA and HGF models. A similar pattern was observed in the turtle task, where the PF-

HMM again dominated model comparison. Importantly, these alternative models fail both quantitatively and qualitatively: in addition to providing poorer fits to the data, they do not capture the model-neutral effects observed across datasets (Figures 5C and 7B), including the dissociable effects of volatility and stochasticity evident in the behavioral data (Supplementary Figure 5). In contrast, because the PF-HMM separately dissociates volatility from stochasticity, it accounts better for data, and it reproduces learning rate pattern observed across datasets.

	Model frequency			**Protected exceedance probability**		
Dataset	PHA	HGF	PF-HMM	PHA	HGF	PF-HMM
Sea lion	0.169	0.075	0.757	0.000	0.000	1.000
Turtle	0.035	0.079	0.885	0.000	0.000	1.000
Jang et al., 2019	0.186	0.223	0.591	0.000	0.000	1.000
Piray et al., 2019	0.085	0.147	0.768	0.000	0.000	1.000

Table 1: Bayesian model comparison results across four datasets.

To assess the generality of these findings, we applied the same model fitting and Bayesian model selection procedures to two publicly available datasets involving probabilistic reversal learning.

In the first dataset, 174 participants performed a reward-learning task in which they estimated the probability of reward based on binary feedback (Jang et al., 2019). On each trial, participants viewed an image and decided whether to accept or reject a gamble, with reward probability determined by the image category. Crucially, these contingencies shifted multiple times throughout the experiment, with each image category undergoing at least two reversals. The original study paired this learning task with a recognition memory test, administered entirely after the learning phase, to examine how reinforcement-learning dynamics shape episodic memory. Because the memory test did not intervene during learning, we can analyze the learning phase in isolation as an independent test of probabilistic learning under changing contingencies.

In the second dataset, 44 participants completed a probabilistic go/no-go task in which they learned to press or withhold a button in response to facial cues in order to obtain reward or avoid punishment (Piray, Ly, et al., 2019). The cues combined emotional content (happy or angry) with background color (grey or yellow for reward/punishment), and response–outcome contingencies for each cue were probabilistic and varied independently over time. The original studies examined socio-emotional modulation of learning and choice; here, we use the behavioral data from this task as a second independent test of learning under shifting

contingencies. All learning models were augmented with a response model that included valence- and emotion-specific bias parameters to account for Pavlovian approach–avoidance tendencies, so differences in model comparison are attributable to learning dynamics alone.

Across both datasets, Bayesian model comparison again favored the PF-HMM (Table 1). The model achieved the highest global model frequency in the Jang et al. dataset (0.591) and the Piray et al. dataset (0.768), with protected exceedance probabilities of 1.000 in both cases.

Discussion

Learning from binary outcomes is a fundamental challenge for organisms navigating uncertain environments. In such environments, observed outcomes provide only noisy, indirect information about an underlying hidden state that cannot be directly observed and may itself fluctuate over time. Despite the importance of understanding how organisms learn under these conditions, the literature has not yet addressed this problem from a comprehensive theoretical perspective. Instead, existing models have primarily relied on probabilistic algorithms designed for continuous outcomes (Browning et al., 2015; Iglesias et al., 2013; Mathys et al., 2011, 2014; Piray & Daw, 2020, 2021a), employing ad hoc approximations to handle binary data rather than systematically addressing the underlying statistical structure or considering how organisms might optimally learn from binary outcomes. This limitation is problematic even when stochasticity is fixed and known, as we have shown previously (Piray & Daw, 2020); see also (Ryali et al., 2018)). More importantly, it poses a fundamental challenge for understanding learning under conditions where both volatility and stochasticity are present, since continuous-based algorithms fail to distinguish between these distinct sources of noise.

We exploited the framework of probabilistic state space models, which serves as the underlying generative probabilistic graphical model for both the Kalman filter and HMM (Bishop, 2006; Ljung, 1998). While the Kalman filter is designed for continuous states and observations, the HMM framework is particularly suited to scenarios involving binary hidden states and binary outcomes (Ghahramani, 2001). Crucially, like the Kalman filter, the HMM supports exact inference, making it unnecessary to rely on Kalman-based approximate algorithms for handling binary outcomes. Additionally, while the Kalman filter recursively updates posterior statistics of the inferred hidden state—namely, its mean and variance—the HMM iteratively updates the posterior belief about the hidden state through a Bernoulli distribution, specifically estimating the probability that the state is 1 given outcomes.

Building on these parallels with the continuous case, we extended the concepts of volatility and stochasticity to the binary setting. Volatility refers to noise that disrupts the underlying binary hidden state over time, specifically the probability of the hidden state switching from one value to the other. In contrast, stochasticity reflects noise in the contingency between the hidden state and observed binary outcomes, specifically the probability that an outcome deviates from the

true hidden state. From a statistical standpoint, optimal learning should adapt to these sources of uncertainty: Higher volatility requires faster learning because old information quickly becomes irrelevant. Conversely, higher stochasticity is expected to slow down learning, as each individual outcome is less informative, necessitating the integration of more outcomes to accurately infer the hidden state. We validated these predictions through simulation analyses using an optimal HMM with true values of volatility and stochasticity. As expected, optimal learning rates increased under high-volatility conditions and decreased under high-stochasticity conditions. To further test these predictions, we designed an experimental task employing a 2x2 factorial manipulation of volatility and stochasticity. Consistent with the optimal model's behavior, human participants exhibited higher learning rates under high volatility and lower learning rates under high stochasticity.

We also developed a new model to capture learning in real-world situations where the true values of volatility and stochasticity are unknown and must be inferred from observed outcomes. To achieve this, we extended the HMM by incorporating a PF, a sequential Monte Carlo sampling method (Andrieu et al., 2010; Doucet et al., 2000). The new model, PF-HMM, allows volatility and stochasticity parameters to evolve dynamically on a trial-by-trial basis, reflecting real-time belief updates and closely mirroring participants' incremental inference processes. The model integrates exact inference from the HMM with the PF's per-trial estimates of volatility and stochasticity, enabling the model to continuously adapt to changing conditions while accurately distinguishing between different sources of noise. Simulation analyses demonstrated that the PF-HMM successfully inferred volatility and stochasticity, accurately attributing noise to its correct source. Importantly, this approach also allows us to predict two distinct patterns of impaired learning that arise from selective lesions to the model's inference processes. Specifically, when the model's ability to infer volatility is lesioned, it exhibits a maladaptive pattern characterized by a reduced learning rate, regardless of the true source of noise. Conversely, when the model's ability to infer stochasticity is lesioned, it results in an excessive learning rate, indiscriminately adjusting to all forms of noise.

Previous experimental work has primarily focused on learning from binary outcomes under conditions of volatility (Brazil et al., 2017; Browning et al., 2015; Cole et al., 2020; de Berker et al., 2016; Deserno et al., 2020; Diaconescu et al., 2014, 2020; Hauser et al., 2014; Iglesias et al., 2013; Lawson et al., 2017; Piray, Ly, et al., 2019; Reed et al., 2020). While this line of research has been influential in both neuroscience and psychiatry, it has largely overlooked a more complex and challenging problem: dissociating volatility and stochasticity when both are unknown (Piray & Daw, 2021a, 2024). This distinction is crucial because volatility and stochasticity represent fundamentally different sources of, yet most previous studies have not systematically manipulated them experimentally. A notable exception is the recent work by Pulcu and Browning, who employed a 2x2 factorial design to manipulate both volatility and stochasticity, similar to our experimental design (Pulcu & Browning, 2025). However, unlike our findings,

their study did not reveal any significant effects related to stochasticity (referred to as "noise" in their terminology). This discrepancy underscores the significance and robustness of our results, as well as the effectiveness of our experimental design. By successfully dissociating volatility and stochasticity, we established the theoretically expected opposing effects of these factors on learning rates: increased learning rates under high volatility and decreased learning rates under high stochasticity.

A related line of research has investigated learning from continuous observations under conditions of volatility and stochasticity, but until recently, it faced similar challenges in dissociating these distinct sources of uncertainty (Piray & Daw, 2021a). Our recent empirical work in that domain successfully addressed this issue by building on a well-established paradigm, demonstrating that learning rates increase with volatility and decrease with stochasticity (Piray & Daw, 2024). While both approaches employ PF models to dissociate volatility and stochasticity through dynamic tracking of these noise sources over time, the current model differs fundamentally at the level of inference. While both the previous model for continuous observations and the current model for binary observations use particle filter models to dissociate volatility and stochasticity by dynamically tracking these noise sources over time, they differ fundamentally in their inference process. The previous model relies on a Kalman filter for learning volatility and stochasticity based on continuous observations. In contrast, the model presented here is specifically designed for binary observations, employing a HMM to perform exact inference based on trial-by-trial estimates of volatility and stochasticity provided by the PF. This shift from Kalman filter-based inference to HMM-based inference is not merely a technical adjustment; it is a necessary adaptation to accurately capture the distinct statistical structure of binary outcomes. By integrating particle filtering with HMM, the current approach offers a novel and robust computational framework for dissociating volatility and stochasticity in binary learning problems, a challenge that has been unanswered by previous models focusing on continuous observations.

Beyond theoretical implications, accurately distinguishing volatility from stochasticity carries significant clinical relevance, particularly for understanding psychiatric disorders such as anxiety, schizophrenia, and impulse control disorders (Gagne et al., 2018; Horga & Abi-Dargham, 2019; Huang et al., 2017; Piray et al., 2014, 2017; Pulcu & Browning, 2019; Stephan et al., 2006; van Timmeren et al., 2023). These conditions often involve maladaptive responses to environmental uncertainty, where patients may exhibit either overly rigid or excessively flexible learning strategies (Browning et al., 2015; Lawson et al., 2017). Our model simulations, especially through targeted lesion experiments, provide valuable insights into how maladaptive estimation of volatility or stochasticity could drive clinically relevant maladaptive behaviors.

Our approach also highlights a critical issue overlooked by previous studies in computational psychiatry: the potential differences between learning from continuous and binary outcomes

with respect to uncertainty-related parameters and their relevance to psychopathology. There are several key distinctions between these types of learning. Computationally, binary and continuous outcomes require fundamentally different algorithms for exact inference—specifically, Kalman filters for continuous outcomes and HMMs for binary outcomes. This distinction reflects underlying differences in how uncertainty is represented and processed within each framework. Psychopathologically, binary outcomes may be particularly salient in conditions such as depression, where individuals are prone to self-blame or heightened sensitivity to negative feedback (Abramson et al., 1978; Beck, 1970; Peterson et al., 1981; Zahn et al., 2015). Unlike continuous feedback, binary feedback offers clear-cut evaluations of success or failure, which may exacerbate maladaptive learning patterns and may partly explain why people learn differently from positive versus negative prediction errors (Frank et al., 2004; Piray, 2011; Pulcu & Browning, 2019). Specifically, when learning from negative feedback, individuals prone to self-blame may misattribute stochasticity (bad outcomes due to bad luck) to volatility (bad outcomes perceived as resulting from their own choices). This misattribution can reinforce negative biases by systematically interpreting perceived negative outcomes as evidence of environmental change, even when no such changes occur. Consequently, these individuals may develop maladaptive learning strategies that perpetuate their symptoms.

Although we focused on the problem of learning from binary outcomes under volatility and stochasticity, the dissociation of these two forms of uncertainty has potential implications for various cognitive processes that have not yet been extensively studied within this framework. One particularly relevant area is latent cause inference and structural learning, which involves determining underlying, categorical causes that govern observable patterns, such as stimuli, rewards, or punishments (Braun et al., 2010; Collins & Frank, 2013; Courville et al., 2004, 2006; Gershman et al., 2010; Lloyd & Leslie, 2013; Mahmoodi et al., 2024; Niv, 2019; Shin & Niv, 2021; Tomov et al., 2018). In these scenarios, the challenge lies in inferring which observations should be grouped together under a common cause and when a new cause should be inferred. The computational literature on latent cause inference has largely developed around probabilistic frameworks, such as the Chinese restaurant process (Gershman & Blei, 2012), which provides a principled way of inferring the emergence of novel causes based on observed data. These approaches have been influential in understanding how the brain constructs internal models of the environment, particularly when it encounters novel or changing structures (Gershman et al., 2010; Gershman & Blei, 2012; Lloyd & Leslie, 2013; Niv, 2019; Shin & Niv, 2021). Fundamentally, however, the problem involves clustering various observables in a way that accurately reflects their underlying causal structure (Courville et al., 2004, 2006). The ability to dissociate volatility and stochasticity is likely to play a critical role in this process. Specifically, if an agent infers similar volatility parameters for different observables, it suggests that those observables are likely changing together, potentially due to a shared underlying cause. This ability to track volatility separately from stochasticity provides a valuable signal for clustering observations appropriately, which is essential for constructing accurate causal models.

Another important issue that is theoretically relevant but have not been studied in the context of volatility and stochasticity is the issue of planning and decision making from a reinforcement learning perspective. Most previous work in this area has focused on prediction rather than optimal planning or decision-making. Even studies involving binary "choices" have primarily emphasized learning, as these tasks were essentially structured as prediction problems (Behrens et al., 2007, 2008; Browning et al., 2015). In fact, in these studies, the two choices were perfectly anti-correlated, and participants were explicitly informed of this structure. Consequently, participants' learning processes could be assessed by measuring their ability to track changes in underlying contingencies, rather than by evaluating their capacity to make independent decisions aimed at maximizing reward, as is common in reinforcement learning (Sutton & Barto, 1998).

Extending the Bayesian generative processes underlying models of volatility and stochasticity to frameworks compatible with Markov decision processes (Sutton & Barto, 1998) commonly used in reinforcement learning presents substantial computational challenges. Prediction models typically focus on continuously updating beliefs about latent states, while planning and optimal control models must not only learn these parameters but also compute optimal sequences of actions that maximize future rewards over extended decision sequences (Sutton & Barto, 1998). Although some approaches have extended Kalman filtering to multi-step decision problems, resulting in Kalman-like temporal difference algorithms (Geist et al., 2009; Gershman, 2015), these methods fail to integrate the core optimal structure of either problem. Specifically, they do not combine the state space generative model characteristic of Kalman filtering with the optimal control framework essential to reinforcement learning. An alternative approach is the framework of planning as inference from control theory and machine learning (Kappen, 2005; Levine, 2018; Todorov, 2008, 2009; Toussaint & Storkey, 2006; Ziebart et al., 2008), which has recently influenced cognitive neuroscience research on planning, cognitive control, and cognitive map-making (Bazarjani & Piray, 2025; Botvinick & Toussaint, 2012; Piray & Daw, 2021b, 2025). This framework formulates optimal control as a Bayesian inference problem, providing a principled way to integrate uncertainty estimation with decision-making. By reframing planning as an inference problem, generative models of volatility and stochasticity can be incorporated directly into the generative model related to planning, rather than treating them as separate processes developed from different theoretical foundations and only superficially connected. This approach has the potential to unify prediction and control under a coherent computational framework. Overall, while significant progress has been made in understanding how agents learn to predict under conditions of volatility and stochasticity, extending these insights to reinforcement learning and planning remains a difficult and open problem. Achieving this integration requires reconciling models of uncertainty estimation with the demands of optimal control, including how agents select actions to achieve long-term goals rather than merely improving predictive accuracy.

While extending prediction models to decision-making introduces challenges related to long-term reward maximization and planning, a fundamental issue arises even in simpler multi-choice decision-making scenarios where sequential planning is unnecessary. Moving from prediction to decision-making inherently involves balancing exploration and exploitation (Amin et al., 2021; Daw et al., 2006; Song et al., 2019; Wilson et al., 2014). This trade-off requires deciding whether to exploit known options that provide reliable rewards or explore uncertain options that may yield better rewards in the future—a process strongly influenced by uncertainty estimates related to both volatility and stochasticity. Exploitation involves selecting actions based on existing knowledge to maximize immediate rewards, while exploration involves choosing actions that may be suboptimal in the short term but provide valuable information about the environment. In decision-making under uncertainty, agents must continually balance these competing demands to make adaptive choices. Multiple studies have shown that uncertainty estimates affect decision variability beyond what can be explained by learning and prediction alone (Findling et al., 2019, 2021; Lee et al., 2023). However, when both volatility and stochasticity are unknown and must be inferred from outcomes, distinguishing between prediction, behavioral variability, and information-driven exploration remains challenging.

Conclusion

Inferring the true source of noise from binary outcomes is a fundamental challenge in real-world environments, especially under conditions of volatility and stochasticity where distinguishing between these sources is critical for adaptive learning. However, existing models designed for continuous data often rely on ad hoc approximations to handle binary outcomes, resulting in theoretical inconsistencies and limited generalization. To address this, we developed a novel framework, PF-HMM, which combines HMM with PF to infer volatility and stochasticity directly from binary outcomes. The model effectively dissociates these noise sources, providing a robust approach for understanding learning under uncertainty. Furthermore, lesion analyses using the PF-HMM reveal distinct patterns of maladaptive learning, offering potentially crucial insights into psychiatric conditions. Our experimental findings show that human participants adjust their learning rates appropriately, increasing them under high volatility and decreasing them under high stochasticity, consistent with model predictions. Participants' response times are also slower when inferring the true source of noise is more difficult, as determined by the PF-HMM. Overall, this approach provides a principled framework for studying how organisms learn from binary outcomes and adapt to different sources of uncertainty in both healthy cognition and psychiatric conditions.

Methods

Experimental procedure

The study was approved by the Institutional Review Board at the University of Southern California ("Learning and decision making under uncertainty"; Protocol Number: UP-23-00359). Participants were recruited online via the Prolific platform and provided informed consent prior to beginning the experiment. This study was not preregistered. The task was implemented in JavaScript and jsPsych (Leeuw et al., 2023) and deployed using the NivTurk platform (Zorowitz & Bennett, 2022).

Participants first read instructions, completed step-by-step practice trials to familiarize themselves with the task, and then completed a comprehension quiz to ensure understanding. Those who failed the quiz twice were not allowed to proceed. A total of 73 participants were recruited for the sea lion task. The sample size was chosen based on pilot data from a separate sample, which suggested that it would yield at least 0.8 statistical power. A total of 30 participants were recruited for the turtle task, based on a power analysis using the sea lion task data. All participants passed the comprehension check and met quality control criteria, including consistent engagement with the task (i.e., not leaving the interface idle for extended periods). After completing the main behavioral task, participants also completed an additional task (not analyzed here) and the demographic questionnaire. Data from none of the participants were excluded from the analyses.

Blockwise HMM

To obtain model-neutral estimates of learning rate, we fitted a binary Hidden Markov Model (HMM) separately to each block of each participant's data. The full generative formulation and filtering equations are provided in the Appendix A.3. Briefly, the HMM assumes a latent binary state that evolves according to a first-order Markov process governed by volatility v, and generates outcomes through a stochasticity parameter s.
For each block, two parameters were fitted: v volatility (probability of hidden state switching) and s stochasticity (probability that the observed outcome deviates from the hidden state). Given these parameters, the HMM recursively computes the predicted probability that the hidden state equals 1, denoted as r_t. These predicted probabilities reflect belief updating under the assumed volatility and stochasticity levels.

Response model

Choices were modeled using a response function that maps model-derived predicted probabilities to observed responses. While the HMM provides the predicted belief about outcome probability, actual choices may also reflect motor-related perseveration. To account for this, predicted probabilities were transformed into log-odds space and augmented by a stickiness parameter ρ, capturing the tendency to repeat or switch the previous choice. This response-level parameter operates independently of belief updating and latent state inference.

Let p_t denote the final predicted probability of choosing option 1 after applying the response model. The same response model structure was applied across all fitted models to ensure comparability of likelihood evaluation.

Model fitting procedure for blockwise HMM

Parameters were estimated separately for each block using maximum a posteriori (MAP) estimation. Raw parameters were optimized in unconstrained space and transformed via a half-sigmoid function ($f(x) = \frac{0.5}{1+e^{-x}}$) to enforce valid ranges. A zero-mean Gaussian prior with fixed variance was placed on each parameter in the unconstrained space. MAP estimation was implemented using the CBM package (Piray, Dezfouli, et al., 2019), which employs a Laplace approximation around the posterior mode and provides an estimate of model evidence. We used a wide variance (100) to ensure that parameters could vary over a wide range.

For each trial, the HMM filtering equations were used to compute r_t, which was then passed through the response model to obtain p_t. The log-likelihood of observed choices was computed as:

$$\log L = \sum_t p_t c_t + (1 - p_t)(1 - c_t)$$

where c_t is the choice (0 or 1) on trial t. Trials with missing responses were excluded from the likelihood computation. Blockwise log-likelihoods were summed across blocks for subject-level model fitting.

We also performed a recovery analysis to assess how well model parameters can be recovered. We generated N=100 synthetic datasets by sampling model parameters randomly. The synthetic datasets were subsequently fitted using the same model-fitting procedure employed for empirical data. This analysis revealed that model parameters are highly recoverable (see Supplementary Table 1 and Supplementary Fig. 3).

Model fitting procedure for PF-HMM

Fitting the PF-HMM poses substantial challenges due to the stochastic nature of particle filtering and the lack of differentiability of the likelihood with respect to model parameters. We therefore employed a gradient-free optimization approach based on Bayesian optimization, implemented using MATLAB's bayesopt routine. Model parameters were fit separately for each participant by maximizing the log-likelihood of observed behavioral data under the PF-HMM. All reported PF-HMM fits include a response-level perseveration parameter, which accounts for choice repetition effects independently of latent state inference (see Response model below). For completeness, we report a comparison between PF-HMM fits with and without the perseveration parameter in the Supplementary Materials (Supplementary Table 8), demonstrating that inclusion of the response model improves model fit without altering the qualitative learning-rate effects.

Our approach is similar to our recent work for fitting PF models to human data (Piray & Daw, 2024). Specifically, for a given parameter setting, model likelihood was evaluated by running the particle filter multiple times (10 runs per evaluation) to account for stochastic variability in the inference process. The objective function minimized by the optimizer was defined as the mean negative log-likelihood across runs, augmented by a penalty proportional to the standard error of the mean across runs. This regularized objective favors parameter estimates that both achieve high likelihood and yield stable estimates across repeated stochastic simulations.

Bayesian optimization was initialized with multiple seed points and iteratively evaluated candidate parameter settings using an expected-improvement acquisition function. For each participant, the best-fitting parameters were selected based on the optimized objective. Latent trajectories—including inferred beliefs about the hidden state, volatility, and stochasticity—were obtained by averaging model outputs across repeated particle filter runs using the fitted parameters. Learning rates were computed post hoc as descriptive summaries by regressing belief updates on prediction errors derived from the model's latent state estimates (see below).

To assess parameter recoverability, we conducted a recovery analysis in which synthetic datasets were generated by simulating the PF-HMM using randomly sampled parameter values. The same fitting procedure was then applied to these synthetic datasets. This analysis demonstrated high recoverability for all model parameters (Supplementary Table 4 and Supplementary Fig. 4).

A recovery analysis was conducted for the PF-HMM with perseveration. Again, $N = 100$ synthetic datasets were generated. True parameters were sampled from the same ranges used during model fitting, including the diffusion parameters governing volatility and stochasticity and the global perseveration parameter. For each simulated subject, belief trajectories were generated using the full PF-HMM forward model with the same particle-filter configuration as used in fitting (10,000 particles, systematic resampling, identical initialization). To reduce stochastic variability, belief trajectories were averaged across multiple forward passes. Choices were then generated sequentially using the same response model.

Learning rate

For HMM and the PF-HMM models, the learning rate was defined as the ratio of update to prediction error. The belief update Δr_t measures how much the posterior belief, r_t, changes after observing the outcome:

$$\Delta r_t = r_t - r_{t-1}.$$

The prediction error δ_t quantifies the discrepancy between the observed outcome o_t and the model's prediction on trial t:

$$\delta_t = o_t - r_{t-1}$$

The trial-by-trial learning rate is defined as the ratio of the belief update to the prediction error, capturing how strongly new information influences belief revision. However, in practice, we quantified learning rates at the block level by regressing belief updates onto prediction error across trials within each block:

$$\Delta r_t = \alpha\, \delta_t + \beta_0 + \varepsilon_t,$$

where α is the regression coefficient related to prediction errors (i.e., the learning rate for that block), β_0 is learning-independent block effect, and ε_t is noise.

Response time analysis

To examine whether computational signatures of noise-source attribution are reflected in behavior beyond choice, we conducted a trial-by-trial regression analysis of response times (RTs). The analysis tested the hypothesis that responses would be slower when identifying the source of noise is more difficult, reflected as low particle variability in the PF-HMM.

For each trial, model predictions were generated using the median fitted parameter values across participants. From these trajectories, we extracted:

- Particle likelihood variability (PLV_t): the standard deviation of per-particle likelihoods at trial t.
- Average per particle likelihood (PLM_t): the average per-particle likelihood at trial t.
- Estimated volatility (V_t): the PF-HMM estimate of volatility at trial t.
- Estimated stochasticity (S_t): the PF-HMM estimate of stochasticity at trial t.
- Accuracy (Acc_t): a binary indicator (1 if the participant's choice matched the outcome; 0 otherwise).

Response times were extracted from trial t, while model-derived regressors were aligned to trial $t-1$ to ensure proper temporal ordering between predictors and behavioral response. Outlier response times were identified and removed using standard outlier detection procedures.

For each participant, we fit the following linear regression model:

$$RT_t = \beta_0 + \beta_{PLV} \cdot PLV_t + \beta_{PLM} \cdot PLM_t + \beta_V \cdot V_t + \beta_S \cdot S_t + \beta_{Acc} \cdot Acc_t + \varepsilon_t$$

Here, RT_t denotes the response time on trial t, and ε_t is noise. The regression model was fit separately for each participant separately, yielding subject-specific regression coefficients. To assess group-level effects, we performed one-sample *t*-tests on the regression coefficients across participants. Mean effects, standard errors of the mean (SEM), *t*-values, and *p*-values were computed for each regressor.

Code and data availability

Analyses were conducted in MATLAB (R2023b) and Python (3.11.5). All experimental data and analysis code are publicly available online at https://github.com/piraylab/PF-HMM. The manuscript is available on *bioRXiv*.

Acknowledgements

This work was supported by grants R21MH134217 from the National Institute of Mental Health. We thank Nathaniel Daw for helpful discussions. The authors declare no conflicts of interest.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, *87*(1), 49–74.

Amin, S., Gomrokchi, M., Satija, H., van Hoof, H., & Precup, D. (2021). *A Survey of Exploration Methods in Reinforcement Learning* (arXiv:2109.00157). arXiv. https://doi.org/10.48550/arXiv.2109.00157

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269–342. https://doi.org/10.1111/j.1467-9868.2009.00736.x

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

Barnby, J. M., Mehta, M. A., & Moutoussis, M. (2022). The computational relationship between reinforcement learning, social inference, and paranoia. *PLOS Computational Biology*, *18*(7), e1010326. https://doi.org/10.1371/journal.pcbi.1010326

Bazarjani, A., & Piray, P. (2025). *Efficient Learning of Predictive Maps for Flexible Planning*. OSF. https://doi.org/10.31234/osf.io/ak57f

Beck, A. T. (1970). *Depression: Causes and Treatment*. University of Pennsylvania Press.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

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

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bolkan, S. S., Stone, I. R., Pinto, L., Ashwood, Z. C., Iravedra Garcia, J. M., Herman, A. L., Singh, P., Bandi, A., Cox, J., Zimmerman, C. A., Cho, J. R., Engelhard, B., Pillow, J. W., & Witten, I. B. (2022). Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nature Neuroscience*, *25*(3), 345–357. https://doi.org/10.1038/s41593-022-01021-9

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*(2), 157–165. https://doi.org/10.1016/j.bbr.2009.08.031

Brazil, I. A., Mathys, C. D., Popma, A., Hoppenbrouwers, S. S., & Cohn, M. D. (2017). Representational Uncertainty in the Brain During Threat Conditioning and the Link With Psychopathic Traits. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *2*(8), 689–695. https://doi.org/10.1016/j.bpsc.2017.04.005

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

Burge, C. B., & Karlin, S. (1998). Finding the genes in genomic DNA. *Current Opinion in Structural Biology*, *8*(3), 346–354. https://doi.org/10.1016/s0959-440x(98)80069-9

Calhoun, A. J., Pillow, J. W., & Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nature Neuroscience*, 22(12), 2040–2049. https://doi.org/10.1038/s41593-019-0533-x

Cole, D. M., Diaconescu, A. O., Pfeiffer, U. J., Brodersen, K. H., Mathys, C. D., Julkowski, D., Ruhrmann, S., Schilbach, L., Tittgemeyer, M., Vogeley, K., & Stephan, K. E. (2020). Atypical processing of uncertainty in individuals at risk for psychosis. *NeuroImage: Clinical*, 26, 102239. https://doi.org/10.1016/j.nicl.2020.102239

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

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22(11), 4563–4567. https://doi.org/10.1523/JNEUROSCI.22-11-04563.2002

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: A bayesian perspective. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(6), 2407–2416. https://doi.org/10.1523/JNEUROSCI.1989-14.2015

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300. https://doi.org/10.1016/j.tics.2006.05.004

Courville, A. C., Daw, N., & Touretzky, D. (2004). Similarity and Discrimination in Classical Conditioning: A Latent Variable Account. *Advances in Neural Information Processing Systems*, *17*. https://papers.nips.cc/paper_files/paper/2004/hash/65fc9fb4897a89789352e211ca2d398f-Abstract.html

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3 Suppl*, 1218–1223. https://doi.org/10.1038/81504

Dayan, P., & Long, T. (1998). Statistical Models of Conditioning. *Advances in Neural Information Processing Systems*, 117–123.

Dayan, P., & Yu, A. (2003). Uncertainty and Learning. *IETE Journal of Research*, *49*(2–3), 171–181. https://doi.org/10.1080/03772063.2003.11416335

de Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*, 10996. https://doi.org/10.1038/ncomms10996

Deserno, L., Boehme, R., Mathys, C., Katthagen, T., Kaminski, J., Stephan, K. E., Heinz, A., & Schlagenhauf, F. (2020). Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(2), 173–183. https://doi.org/10.1016/j.bpsc.2019.10.007

Diaconescu, A. O., Mathys, C., Weber, L. A. E., Daunizeau, J., Kasper, L., Lomakina, E. I., Fehr, E., & Stephan, K. E. (2014). Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Computational Biology*, *10*(9). https://doi.org/10.1371/journal.pcbi.1003810

Diaconescu, A. O., Wellstein, K. V., Kasper, L., Mathys, C., & Stephan, K. E. (2020). Hierarchical Bayesian models of social inference for probing persecutory delusional ideation. *Journal of Abnormal Psychology*, *129*(6), 556–569. https://doi.org/10.1037/abn0000500

Douc, R., & Cappe, O. (2005). Comparison of resampling schemes for particle filtering. *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, 64–69. https://doi.org/10.1109/ISPA.2005.195385

Doucet, A., de Freitas, N., Murphy, K. P., & Russell, S. J. (2000). Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks. *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, UAI '00*, 176–183.

Doucet, A., Freitas, N. de, & Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer Science & Business Media.

Doucet, A., & Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In D. Crisan & B. Rozovskii (Eds.), *Handbook of Nonlinear Filtering*. Oxford University Press.

Findling, C., Chopin, N., & Koechlin, E. (2021). Imprecise neural computations as a source of adaptive behaviour in volatile environments. *Nature Human Behaviour*, *5*(1), 99–112. https://doi.org/10.1038/s41562-020-00971-z

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, *22*(12), 2066–2077. https://doi.org/10.1038/s41593-019-0518-9

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science (New York, N.Y.)*, *306*(5703), Article 5703. https://doi.org/10.1126/science.1102941

Gagne, C., Dayan, P., & Bishop, S. J. (2018). When planning to survive goes wrong: Predicting the future and replaying the past in anxiety and PTSD. *Current Opinion in Behavioral Sciences, Survival Circuits*, *24*, 89–95. https://doi.org/10.1016/j.cobeha.2018.03.013

Geist, M., Pietquin, O., & Fricout, G. (2009). Kalman Temporal Differences: The deterministic case. *Proceedings of the 2009 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL-09)*, 185–192.

Gelbart, M. A., Snoek, J., & Adams, R. P. (2014). Bayesian optimization with unknown constraints. *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, UAI'14*, 250–259.

Gershman, S. J. (2015). A Unifying Probabilistic View of Associative Learning. *PLoS Computational Biology*, *11*(11). https://doi.org/10.1371/journal.pcbi.1004567

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12. https://doi.org/10.1016/j.jmp.2011.08.004

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197–209. https://doi.org/10.1037/a0017808

Ghahramani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. In *Hidden Markov models: Applications in computer vision* (pp. 9–42). World Scientific Publishing Co., Inc.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(32), 8360–8367. https://doi.org/10.1523/JNEUROSCI.1010-06.2006

Hauser, T. U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S., & Brem, S. (2014). Role of the Medial Prefrontal Cortex in Impaired Decision Making in Juvenile Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry*, *71*(10), 1165–1173. https://doi.org/10.1001/jamapsychiatry.2014.1093

Horga, G., & Abi-Dargham, A. (2019). An integrative framework for perceptual disturbances in psychosis. *Nature Reviews Neuroscience*, *20*(12), 763–778. https://doi.org/10.1038/s41583-019-0234-1

Huang, H., Thompson, W., & Paulus, M. P. (2017). Computational Dysfunctions in Anxiety: Failure to Differentiate Signal From Noise. *Biological Psychiatry*, *82*(6), 440–446. https://doi.org/10.1016/j.biopsych.2017.07.007

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E. M., & Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, *80*(2), 519–530. https://doi.org/10.1016/j.neuron.2013.09.009

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, *345*, 12–26. https://doi.org/10.1016/j.neuroscience.2016.03.021

Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, *3*(7), Article 7. https://doi.org/10.1038/s41562-019-0597-3

Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review*, *109*(3), Article 3. https://doi.org/10.1037/0033-295X.109.3.533

Kappen, H. J. (2005). Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, *95*(20), 200201. https://doi.org/10.1103/PhysRevLett.95.200201

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Lee, J. K., Rouault, M., & Wyart, V. (2023). Adaptive tuning of human learning and choice variability to unexpected uncertainty. *Science Advances*, *9*(13), eadd0501. https://doi.org/10.1126/sciadv.add0501

Leeuw, J. R. de, Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351

Levine, S. (2018). Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review. *arXiv:1805.00909 [Cs, Stat]*. http://arxiv.org/abs/1805.00909

Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., & Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, *14*(10), Article 10. https://doi.org/10.1038/nn.2904

Ljung, L. (1998). System Identification. In A. Procházka, J. Uhlíř, P. W. J. Rayner, & N. G. Kingsbury (Eds.), *Signal Analysis and Prediction* (pp. 163–173). Birkhäuser. https://doi.org/10.1007/978-1-4612-1768-8_11

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

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), Article 4. https://doi.org/10.1037/h0076778

Mahmoodi, A., Luo, S., Harbison, C., Piray, P., & Rushworth, M. F. S. (2024). Human hippocampus and dorsomedial prefrontal cortex infer and update latent causes during social interaction. *Neuron*, *112*(0), 1–14. https://doi.org/10.1016/j.neuron.2024.09.001

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

Mathys, C., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*, 825. https://doi.org/10.3389/fnhum.2014.00825

McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., & Patterson, T. A. (2020). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, *23*(12), 1878–1903. https://doi.org/10.1111/ele.13610

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

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), Article 6.

Peterson, C., Schwartz, S. M., & Seligman, M. E. (1981). Self-blame and depressive symptoms. *Journal of Personality and Social Psychology*, *41*(2), 253–259. https://doi.org/10.1037//0022-3514.41.2.253

Piray, P. (2011). The role of dorsal striatal D2-like receptors in reversal learning: A reinforcement learning viewpoint. *The Journal of Neuroscience*, *31*(40), Article 40. https://doi.org/10.1523/JNEUROSCI.3008-11.2011

Piray, P. (2026). Addressing low statistical power in computational modelling studies in psychology and neuroscience. *Nature Human Behaviour*, *10*(2), 347–356. https://doi.org/10.1038/s41562-025-02348-6

Piray, P., & Daw, N. D. (2020). A simple model for learning in volatile environments. *PLoS Computational Biology*, *16*(7), e1007963. https://doi.org/10.1371/journal.pcbi.1007963

Piray, P., & Daw, N. D. (2021a). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587. https://doi.org/10.1038/s41467-021-26731-9

Piray, P., & Daw, N. D. (2021b). Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature Communications*, *12*(1), 4942. https://doi.org/10.1038/s41467-021-25123-3

Piray, P., & Daw, N. D. (2024). Computational processes of simultaneous learning of stochasticity and volatility in humans. *Nature Communications*, *15*(1), 9073. https://doi.org/10.1038/s41467-024-53459-z

Piray, P., & Daw, N. D. (2025). Reconciling flexibility and efficiency: Medial entorhinal cortex represents a compositional cognitive map. *Nature Communications*, *16*(1), 7444. https://doi.org/10.1038/s41467-025-62733-7

Piray, P., den Ouden, H. E. M., van der Schaaf, M. E., Toni, I., & Cools, R. (2017). Dopaminergic Modulation of the Functional Ventrodorsal Architecture of the Human Striatum. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(1), 485–495. https://doi.org/10.1093/cercor/bhv243

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLOS Computational Biology*, *15*(6), e1007043. https://doi.org/10.1371/journal.pcbi.1007043

Piray, P., Ly, V., Roelofs, K., Cools, R., & Toni, I. (2019). Emotionally aversive cues suppress neural systems underlying optimal learning in socially anxious individuals. *Journal of Neuroscience*, 1394–18. https://doi.org/10.1523/JNEUROSCI.1394-18.2018

Piray, P., Zeighami, Y., Bahrami, F., Eissa, A. M., Hewedi, D. H., & Moustafa, A. A. (2014). Impulse control disorders in Parkinson's disease are associated with dysfunction in stimulus valuation but not action valuation. *The Journal of Neuroscience*, *34*(23), 7814–7824. https://doi.org/10.1523/JNEUROSCI.4063-13.2014

Pulcu, E., & Browning, M. (2019). The Misestimation of Uncertainty in Affective Disorders. *Trends in Cognitive Sciences*, *23*(10), 865–875. https://doi.org/10.1016/j.tics.2019.07.007

Pulcu, E., & Browning, M. (2025). Humans adapt rationally to approximate estimates of uncertainty. *eLife*, *14*. https://doi.org/10.7554/eLife.103734.1

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. Proceedings of the IEEE. https://doi.org/10.1109/5.18626

Reed, E. J., Uddenberg, S., Suthaharan, P., Mathys, C. D., Taylor, J. R., Groman, S. M., & Corlett, P. R. (2020). Paranoia as a deficit in non-social belief updating. *eLife*, *9*. https://doi.org/10.7554/eLife.56345

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies—Revisited. *NeuroImage*, *84*, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Ryali, C., Reddy, G., & Yu, A. J. (2018). Demystifying excessively volatile human learning: A Bayesian persistent prior and a neural approximation. *Advances in Neural Information Processing Systems*, *31*. https://papers.nips.cc/paper_files/paper/2018/hash/78b9cab19959e4af8ff46156ee460c74-Abstract.html

Schlagenhauf, F., Huys, Q. J. M., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., Dolan, R., & Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage*, *89*(100), 171–180. https://doi.org/10.1016/j.neuroimage.2013.11.034

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), Article 2. https://www.jstor.org/stable/2958889

Shin, Y. S., & Niv, Y. (2021). Biased evaluations emerge from inferring hidden causes. *Nature Human Behaviour*, *5*(9), 1180–1189. https://doi.org/10.1038/s41562-021-01065-0

Snoek, J., Larochelle, H., & Adams, R. P. (2012). *Practical Bayesian Optimization of Machine Learning Algorithms* (arXiv:1206.2944; Issue arXiv:1206.2944). arXiv. https://doi.org/10.48550/arXiv.1206.2944

Song, M., Bnaya, Z., & Ma, W. J. (2019). Sources of suboptimality in a minimalistic explore-exploit task. *Nature Human Behaviour*, *3*(4), 361–368. https://doi.org/10.1038/s41562-018-0526-x

Stephan, K. E., Baldeweg, T., & Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia. *Biological Psychiatry*, *59*(10), Article 10. https://doi.org/10.1016/j.biopsych.2005.10.005

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), Article 4. https://doi.org/10.1016/j.neuroimage.2009.03.025

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Todorov, E. (2008). General duality between optimal control and estimation. *2008 47th IEEE Conference on Decision and Control*, 4286–4292. https://doi.org/10.1109/CDC.2008.4739438

Todorov, E. (2009). Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(28), 11478–11483. https://doi.org/10.1073/pnas.0710743106

Tomov, M. S., Dorfman, H. M., & Gershman, S. J. (2018). Neural Computations Underlying Causal Structure Learning. *Journal of Neuroscience*, *38*(32), 7143–7157. https://doi.org/10.1523/JNEUROSCI.3336-17.2018

Toussaint, M., & Storkey, A. (2006). Probabilistic inference for solving discrete and continuous state Markov Decision Processes. *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, 945–952. https://doi.org/10.1145/1143844.1143963

van Timmeren, T., Piray, P., Goudriaan, A. E., & van Holst, R. J. (2023). Goal-directed and habitual decision making under stress in gambling disorder: An fMRI study. *Addictive Behaviors*, *140*, 107628. https://doi.org/10.1016/j.addbeh.2023.107628

Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., & Woolrich, M. W. (2018). Discovering dynamic brain networks from big data in rest and task. *Neuroimage*, *180*(Pt B), 646–656. https://doi.org/10.1016/j.neuroimage.2017.06.077

Wainwright, M. J., & Jordan, M. I. (2008). Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, *1*(1–2), 1–305. https://doi.org/10.1561/2200000001

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology. General*, *143*(6), 2074–2081. https://doi.org/10.1037/a0038199

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P., & Datta, S. R. (2015). Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, *88*(6), 1121–1135. https://doi.org/10.1016/j.neuron.2015.11.031

Zahn, R., Lythe, K. E., Gethin, J. A., Green, S., Deakin, J. F. W., Young, A. H., & Moll, J. (2015). The role of self-blame and worthlessness in the psychopathology of major depressive disorder. *Journal of Affective Disorders*, *186*, 337–341. https://doi.org/10.1016/j.jad.2015.08.001

Ziebart, B. D., Maas, A., Bagnell, J. A., & Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 3, AAAI'08*, 1433–1438.

Zorowitz, S., & Bennett, D. (2022). *nivlab/nivturk: Prolific v1.2* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6609218

Appendix A

A.1 State space models

We begin by describing state space models (SSMs), which provide a general probabilistic framework for modeling dynamical systems in which a sequence of observations is generated from an underlying sequence of latent states. Let $\{o_t\}_{t=1}^{T}$denote the observed outcomes and $\{x_t\}_{t=1}^{T}$denote the corresponding latent states. In general, SSMs might have optional inputs, and might also have

Note that the term state space model is used broadly to refer to any recurrent process with a latent state. In general, SSMs may be noisy or noiseless, may be driven by external inputs, and may involve latent states that are one- or multi-dimensional. Here, we focus on noisy SSMs with a one-dimensional latent state and observation.

SSMs assume the following conditional independence structure:

1. First-order Markov process of the latent state:

$$p(x_t \mid x_{1:t-1}) = p(x_t \mid x_{t-1}) \tag{A1.1}$$

2. Conditional independence of observations given the latent state:

$$p(o_t \mid x_{1:t}, o_{1:t-1}) = p(o_t \mid x_t) \tag{A1.2}$$

The primary inference goal (filtering) in an SSM is to compute the posterior over the current latent state given observations up to time t, i.e. $p(x_t \mid o_{1:t})$. Importantly, this posterior can be computed recursively, regardless of the specific functional forms of the transition and observation models. This typically is done in two steps:

- Prediction: Given the posterior at time $t-1$, $p(x_{t-1} \mid o_{1:t-1})$, the predicted distribution at time t (prior to observing o_t) is obtained by marginalizing over the previous state:

$$p(x_t \mid o_{1:t-1}) = \int p(x_t \mid x_{t-1})\, p(x_{t-1} \mid o_{1:t-1})\, dx_{t-1} \tag{A1.3}$$

- Update: After observing o_t, the estimate is adjusted by actual measurement using Bayes' rule, giving rise to the posterior given all observations:

$$p(x_t \mid o_{1:t}) = \frac{1}{c_t} p(o_t \mid x_t) p(x_t \mid o_{1:t-1}) \tag{A1.4}$$

where c_t is the normalization constant with respect to x_t:

$$c_t = \int p(o_t \mid x_t) p(x_t \mid o_{1:t-1}) dx_t = p(\, o_t \mid o_{1:t-1}\,) \tag{A1.5}$$

Equations (A1.3 – 1.5) hold for any model satisfying the conditional independence structure above, irrespective of the functional forms of $p(x_t \mid x_{t-1})$and $p(o_t \mid x_t)$. However, the tractability of the integral in Eq. (A1.3) depends on those forms. Two important special cases admit exact recursive solutions. In linear–Gaussian SSMs, both the transition and observation models are Gaussian, and the filtering distribution $p(x_t \mid o_{1:t})$ remains Gaussian at every step. Exact inference in this setting is provided by the Kalman filter. Although the Kalman filter applies more generally to multidimensional systems, we briefly review the one-dimensional case below. In discrete-state SSMs, where latent states take values in a finite set, the filtering distribution is categorical, and the recursion reduces to finite summation rather than integration. Exact inference in this setting is provided by the hidden Markov model (HMM) filtering equations. In the present work, we focus on the binary system, in which $x_t \in \{0,1\}$ and $o_t \in \{0,1\}$.

A.2 Kalman filter

The Kalman filter arises as a special case of the SSM under linear transition dynamics and Gaussian noise. It provides an exact recursive solution when both the hidden state and the observations are continuous-valued, and all noise sources are Gaussian:

$$x_t = \mathrm{A}x_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathrm{V}), \tag{A2.1}$$

$$o_t = \mathrm{C}x_t + e_t, \quad e_t \sim \mathcal{N}(0, \mathrm{S}), \tag{A2.2}$$

where x_t denotes the hidden state, o_t denotes the observation, and ϵ_t and e_t are independent Gaussian noise whose covariances is given by V and S, respectively. We focus on the special scalar case with $\mathrm{A} = \mathrm{C} = 1$, $\mathrm{V} = v$ (volatility), and $\mathrm{S} = s$ (stochasticity). In probabilistic form, these become:

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t|x_{t-1}, v) \tag{A2.3}$$

$$p(o_t|x_t) = \mathcal{N}(o_t|x_t, s) \tag{A2.4}$$

Under these assumptions, it can be shown that the posterior remains Gaussian at every time step:

$$p(x_t \mid o_{1:t}) = \mathcal{N}(x_t \mid m_t, w_t). \tag{A2.5}$$

Thus, inference reduces to recursively updating the posterior mean m_t and variance w_t. Specifically, inference proceeds in two steps: prediction and update, corresponding to Eqs. (A1.3 – 1.4) in the general SSM formulation. The predicted mean and covariance prior to observing o_t, $p(x_t \mid o_{1:t-1}) = \mathcal{N}(x_t \mid m_t^-, w_t^-)$, are obtained by propagating the previous posterior through the linear dynamics. It can be shown that:

$$m_t^- = m_{t-1} \tag{A2.6}$$

$$w_t^- = w_{t-1} + v \tag{A2.7}$$

After observing o_t, the posterior is given by Equation A1.4. Defining the Kalman gain as:

$$k_t = \frac{w_t^-}{w_t^- + s} = \frac{w_{t-1} + v}{w_{t-1} + v + s} \tag{A2.8}$$

the posterior mean and variance are then given by:

$$m_t = m_t^- + \alpha_t(o_t - m_t^-) = m_t + k_t(o_t - m_t) \tag{A2.9}$$

$$w_t = (1 - \alpha_t)w_t^- = (1 - k_t)(w_{t-1} + v) \tag{A2.10}$$

Thus, the Kalman gain k_t determines how strongly new observations influence the posterior estimate. It increases with v (volatility) and decreases with s stochasticity. In this setting, learning rate, defined as update divided by prediction error, $(m_t - m_{t-1})/(o_t - m_{t-1})$, is equivalent to the Kalman gain.

A.3 Hidden Markov model

We consider a binary HMM, in which both hidden states and observations take values of either 0 or 1. The hidden state evolves according to a binary diffusion process governed by volatility parameter v. Specifically,

$$x_t = \begin{cases} x_{t-1}, & \text{with probability } 1 - v \\ 1 - x_{t-1}, & \text{with probability } v \end{cases} \tag{A3.1}$$

where $v \in [0, 0.5]$ denotes the probability that the hidden state switches on trial t. When $v = 0$, the state is perfectly stable; when $v = 0.5$, the state is completely noisy. The corresponding conditional distribution of x_t can be written compactly as:

$$p(x_t \mid x_{t-1}) = \begin{cases} 1 - v, & \text{if } x_t = x_{t-1} \\ v, & \text{if } x_t = 1 - x_{t-1} \end{cases} \tag{A3.2}$$

Observations are generated by corrupting the hidden state with observation noise governed by the stochasticity parameter s:

$$o_t = \begin{cases} x_t, & \text{with probability } 1 - s \\ 1 - x_t, & \text{with probability } s \end{cases} \tag{A3.3}$$

where $s \in [0, 0.5]$ is the stochasticity parameter indicating how noisy the observation process is. Larger values of s indicate noisier observations, with $s = 0.5$ indicating that the observation is equally likely to match or differ from the latent state x_t. In contrast, when $s = 0$, the observation matches the state. Thus, the conditional distribution of o_t can be written as:

$$p(o_t \mid x_t) = \begin{cases} 1 - s, & \text{if } o_t = x_t \\ s, & \text{if } o_t = 1 - x_t \end{cases} \tag{A3.4}$$

Under these assumptions, it can be shown that the posterior at every time step takes the form of a Bernoulli distribution. In other words, for binary SSMs, the Bernoulli distribution plays the role that the Gaussian plays in linear-Gaussian SSMs. Recall that for a binary random variable $z \in \{0,1\}$, the Bernoulli distribution with parameter $\mu \in [0, 1]$ is given by:

$$\mathrm{Bern}(z \mid \mu) = \begin{cases} \mu, & \text{if } z = 1 \\ 1 - \mu, & \text{if } z = 0 \end{cases} \tag{A3.5}$$

Thus, parameter μ represents probability of $z = 1$.

Now suppose the posterior belief at time $t - 1$ is Bernoulli with parameter r_{t-1},

$$p(x_{t-1} \mid o_{1:t-1}) = \mathrm{Bern}(x_{t-1} \mid r_{t-1}) \tag{A3.6}$$

where $r_{t-1} = p(x_{t-1} = 1 \mid o_{1:t-1})$ represents the probability that the hidden state equals 1 given all outcomes up to and including time $t - 1$. We now show that the posterior at time t is also Bernoulli.

Since the binary HMM satisfies the conditional independence structure of the general state space model (Section A.1), the recursive update equations (A1.3–1.4) apply directly. The prediction step computes the belief about the current state before observing o_t, that is, $p(x_t = 1 \mid o_{1:t-1})$. Using the general SSM recursion (Eq. A1.3) with integration replaced by summation, the predicted distribution is obtained by marginalizing over the previous hidden state:

$$p(x_t = 1 \mid o_{1:t-1}) = \sum_{x_{t-1} \in \{0,1\}} p(x_t = 1 \mid x_{t-1}) p(x_{t-1} \mid o_{1:t-1}) \tag{A3.7}$$

Substituting the transition model (Eq. A3.2) and definition of r_{t-1}, we obtain:

$$q = p(x_t = 1 \mid o_{1:t-1}) = (1 - v) r_{t-1} + v(1 - r_{t-1}) \tag{A3.8}$$

This yields Eq. (5) in the main text. Thus, q represents belief about that $x_t = 1$ before incorporating the current observation. Similarly, we have $p(x_t = 0 \mid o_{1:t-1}) = 1 - q$, and consequently $p(x_t \mid o_{1:t-1}) = \mathrm{Bern}(x_t \mid q)$. While not needed for the HMM update, it is useful for the next section to also obtain $p(o_t | o_{1:t-1})$ before observing o_t:

$$p(o_t = 1 | o_{1:t-1}) = \sum_{x_t \in \{0,1\}} p(o_t = 1 \mid x_t) p(x_t | o_{1:t-1}) = (1 - s)q + s(1 - q) \tag{A3.9}$$

Next, we use Eq (A1.4) to calculate the posterior after observing o_t. First, the likelihood of the observed outcome o_t under the assumption that $x_t = 1$, can be written as:

$$\ell = p(o_t \mid x_t = 1) = (1 - s) o_t + s(1 - o_t) \tag{A3.10}$$

This corresponds to Eq. (6) in the main text. Note that ℓ depends only on the stochasticity parameter s and the observed outcome o_t. Similarly, from the observation model, we obtain $p(\, o_t \mid x_t = 0\,) = so_t + (1-s)(1-o_t) = 1 - \ell$.

Second, applying Eq. A1.4 yields the posterior belief after observing the current outcome:

$$p(\, x_t = 1 \mid o_{1:t}\,) = \frac{1}{c_t} p(\, o_t \mid x_t = 1\,)\, p(\, x_t = 1 \mid o_{1:t-1}\,) = \frac{1}{c_t}\ell q \tag{A3.11}$$

and similarly:

$$p(\, x_t = 0 \mid o_{1:t}\,) = \frac{1}{c_t} p(\, o_t \mid x_t = 0\,)\, p(\, x_t = 0 \mid o_{1:t-1}\,) = \frac{1}{c_t}(1-\ell)(1-q) \tag{A3.12}$$

Since, $p(\, x_t \mid o_{1:t}\,)$ is a probability distribution, it must sum to one, which yields the normalizing constant, c_t:

$$c_t = \ell q + (1-\ell)(1-q) \tag{A3.13}$$

Thus, the Bernoulli form is preserved under recursive Bayesian updating, $p(x_t \mid o_{1:t}) = \mathrm{Bern}(x_t \mid r_t)$, with

$$r_t = \frac{\ell q}{\ell q + (1-\ell)(1-q)} \tag{A3.14}$$

This corresponds to Eq. (7) in the main text.

A.4 Particle filter-HMM

Now we extend the binary SSM by treating both volatility and stochasticity as dynamically changing random variables. Since their values are bounded between 0 and 0.5, we model these processes using the Beta distribution. Specifically, we define a reparametrized version of the Beta distribution, $\mathcal{B}(z|\mu,\sigma)$, as follows:

$$\mathcal{B}(z|\mu,\sigma) = \mathrm{Beta}\left(z \,\middle|\, \frac{\mu}{\sigma}, \frac{1-\mu}{\sigma}\right) \tag{A4.1}$$

Here, Beta denotes the standard Beta distribution. This reparameterization ensures that the expected value of z is μ, while σ controls the dispersion of the distribution. Using this formulation, we define the volatility and stochasticity processes as:

$$v_t = 0.5u, \qquad u \sim \mathcal{B}(2v_{t-1}, \sigma_v) \tag{A4.2}$$

$$s_t = 0.5w, \qquad w \sim \mathcal{B}(2s_{t-1}, \sigma_s) \tag{A4.3}$$

where σ_v and σ_s determine the stepwise diffusion for volatility and stochasticity, respectively. Importantly, the conditional expected values of v_t and s_t are v_{t-1} and s_{t-1}, respectively, thereby maintaining the martingale property of the process.

The hidden state and outcomes are generated through the same generative process underlying the HMM on each trial:

$$p(x_t|x_{t-1}, v_t) = \begin{cases} 1 - v_t, & \text{if } x_t = x_{t-1} \\ v_t, & \text{if } x_t = 1 - x_{t-1} \end{cases} \tag{A4.4}$$

$$P(o_t|x_t, s_t) = \begin{cases} 1 - s_t, & \text{if } o_t = x_t \\ s_t, & \text{if } o_t = 1 - x_t \end{cases} \tag{A4.5}$$

Since exact inference is no longer tractable, we use Rao-Blackwellized particle filtering (Doucet et al., 2000; Doucet & Johansen, 2011): each particle i maintains estimates v_t^i, s_t^i, and r_t^i, updated in three steps per trial.

First, for each particle i, volatility and stochasticity transition forward in time according to their generative dynamics. This gives rise to sampled values v_t^i and s_t^i. Specifically, we sample these from the generative distributions defined in Eqs. (A4.2–A4.3):

$$v_t^i \leftarrow 0.5u^i, \qquad u^i \sim \mathcal{B}\left(2v_{t-1}^i, \sigma_v\right) \tag{A4.6}$$

$$s_t^i \leftarrow 0.5w^i, \qquad w^i \sim \mathcal{B}\left(2s_{t-1}^i, \sigma_s\right) \tag{A4.7}$$

The second step updates particle weights based on outcome likelihood:

$$q^i = \left(1 - v_t^i\right)r_{t-1}^i + v_t^i\left(1 - r_{t-1}^i\right) \tag{A4.8}$$

$$\mathscr{p}^i = \left(1 - s_t^i\right)q^i + s_t^i\left(1 - q^i\right) \tag{A4.9}$$

$$y^i = \mathrm{Bern}\left(o_t \middle| \mathscr{p}^i\right) \tag{A4.10}$$

$$w_t^i = \frac{w_{t-1}^i y^i}{\sum_j w_{t-1}^j y^j} \tag{A4.11}$$

where w_t^i represents particle i's weight on trial t. The intermediate variable q^i represents the probability of the current state being 1 conditional on v_t^i and s_t^i before observing the current outcome. The intermediate variable $\mathscr{p}^i$ is the predictive probability of the outcome o_t being 1, conditional on v_t^i and s_t^i. As is customary in PF methods, systematic resampling is applied during this step if the effective number of particles drops below a certain threshold.

The third step applies the HMM to integrate the new outcome with previous estimates. For each particle, we update r_t^i using HMM update equations.

$$\ell^i = \left(1 - s_t^i\right) o_t + s_t^i (1 - o_t) \tag{A4.12}$$

$$r_t^i = \frac{q^i \ell^i}{q^i \ell^i + (1 - q^i)(1 - \ell^i)} \tag{A4.13}$$

The expected value for all random variables at trial t is computed as the weighted average across all particles, with w_t^i serving as the weighting coefficients.

Appendix B

In the main text, we noted that binary models built on the Kalman filter exhibit a problematic relationship with the learning rate. Here, we formalize this claim. This argument appeared in our earlier work (Piray & Daw, 2020); here, we present it more slightly more generally and include it for completeness. Specifically, we show that a class of generative models based on sigmoid transformations of linear–Gaussian state-space models, and the resulting Kalman-filter-based algorithms, such as hierarchical Gaussian filters (HGF) and the volatile Kalman filter, do not yield the correct relationship between learning rate and volatility.

These models assume the following conditional dependencies:

$$p(x_t \mid x_{t-1}, v_t) = \mathcal{N}(x_t \mid x_{t-1}, v_t) \tag{B1}$$

$$p(o_t \mid x_t) = \mathrm{Bern}(o_t \mid \sigma(x_t)) \tag{B2}$$

where $\sigma(x) = 1/(1 + \exp(-x))$is the sigmoid function. Here, v_t is volatility, which is itself dynamically changing. While different models make different assumptions about the dynamics of v_t, our analysis is independent of these assumptions. In fact, our argument holds even if volatility is fixed, i.e., $v_t = v$. Under these models, the posterior over x_t is approximated as Gaussian. Its mean, m_t, takes the form:

$$m_t = m_{t-1} + \kappa_t \delta_t \tag{B3}$$

where κ_t is the modified Kalman gain, which typically depends on inferred volatility, and $\delta_t = o_t - \sigma(m_{t-1})$ is the prediction error. Since $o_t \in \{0,1\}$ and the sigmoid is bounded between 0 and 1, it follows that $-1 < \delta_t < 1$.

We now show that the learning rate is directly contaminated by large prediction errors, regardless of κ_t and regardless of whether the prediction error is driven by volatility. The learning rate is defined as:

$$\alpha_t = \frac{\sigma(m_t) - \sigma(m_{t-1})}{\delta_t} \tag{B4}$$

Substituting the update equation for m_t:

$$\alpha_t = \frac{\sigma(m_{t-1} + \kappa_t \delta_t) - \sigma(m_{t-1})}{\delta_t} \tag{B5}$$

The key point is that the numerator of α_t depends on second-order effects of δ_t. To see this, we expand $\sigma(m_{t-1} + \kappa_t \delta_t)$ using a Taylor series around m_{t-1}:

$$\sigma(m_{t-1} + \kappa_t \delta_t) = \sigma(m_{t-1}) + \kappa_t \delta_t \, \sigma'(m_{t-1}) + 0.5\kappa_t^2 \delta_t^2 \, \sigma''(m_{t-1}) + O(\delta_t^3) \tag{B6}$$

where $\sigma'$and σ'' denote the first and second derivatives of the sigmoid, and $O(\delta_t^3)$ represents terms of order δ_t^3 and higher (that are small given that $|\delta_t| < 1$. Substituting into the expression for α_t:

$$\alpha_t = \kappa_t \, \sigma'(m_{t-1}) + 0.5\kappa_t^2 \, \delta_t \, \sigma''(m_{t-1}) + O(\delta_t^2) \tag{B7}$$

For the sigmoid function:

$$\sigma'(x) = \sigma(x)\big(1 - \sigma(x)\big), \tag{B8}$$

$$\sigma''(x) = \sigma'(x)(1 - 2\sigma(x)) \tag{B9}$$

Note that $\sigma'(x) \in [0,0.25]$ and $(1 - 2\sigma(x)) \in [-1,1]$. Substituting:

$$\alpha_t = \kappa_t \, \sigma'(m_{t-1}) + 0.5\kappa_t^2 \, \sigma'(m_{t-1})(1 - 2\sigma(m_{t-1})) \, \delta_t + O(\delta_t^2) \tag{B10}$$

We are interested in the dependence of α_t on the prediction error δ_t. The key term is $(1 - 2\sigma(m_{t-1}))\, \delta_t$, which is large in magnitude whenever $|\delta_t|$ is large, regardless of volatility and its effects on κ_t. In particular, there are two cases corresponding to large prediction errors:

- $o_t = 1$ and m_{t-1} is large and negative: both $(1 - 2\sigma(m_{t-1}))$ and δ_t are positive and large.
- $o_t = 0$ and m_{t-1} is large and positive: both $(1 - 2\sigma(m_{t-1}))$ and δ_t are negative and large.

Thus, in both cases, α_t is inflated by $|\delta_t|$. In other words, the learning rate is inherently contaminated by the magnitude of the prediction error, regardless of the source of noise, and independent of volatility estimate affecting κ_t.