nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00971-z

Imprecise neural computations as a source of adaptive behaviour in volatile environments

Charles Findling[1,2], Nicolas Chopin[2] and Etienne Koechlin[1,3,4] ✉

In everyday life, humans face environments that feature uncertain and volatile or changing situations. Efficient adaptive behaviour must take into account uncertainty and volatility. Previous models of adaptive behaviour involve inferences about volatility that rely on complex and often intractable computations. Because such computations are presumably implausible biologically, it is unclear how humans develop efficient adaptive behaviours in such environments. Here, we demonstrate a counterintuitive result: simple, low-level inferences confined to uncertainty can produce near-optimal adaptive behaviour, regardless of the environmental volatility, assuming imprecisions in computation that conform to the psychophysical Weber law. We further show empirically that this Weber-imprecision model explains human behaviour in volatile environments better than optimal adaptive models that rely on high-level inferences about volatility, even when considering biologically plausible approximations of such models, as well as non-inferential models like adaptive reinforcement learning.

Everyday life features uncertain situations. In such environments, efficient adaptive behaviour requires inferring from experience the current external contingencies and, especially, stimulus–action–outcome contingencies: the agent forms posterior beliefs about external contingencies by combining their prior beliefs with current observations. Empirical studies provide ample evidence that such low-level inferential processes operate in the brain and guide human adaptive behaviour[1–10]. Everyday life also features volatile (ever-changing) situations yielding external contingencies that change over time. Optimal adaptive behaviour consequently requires making additional higher-order inferences about the environmental volatility, that is, the probability that external contingencies will change. Empirical studies show that humans consistently adjust their behaviour with respect to the environmental volatility as if they are making such higher-order inferences[1,11–14]. However, these inferences are complex and may rapidly yield intractable computations[13,15]. This computational complexity problem casts severe doubt upon the hypothesis that the brain implements such higher-order inferential processes[13]. Thus, we have a poor understanding of how humans exhibit adaptive behaviour close to optimal adaptive processes that involve biologically implausible computations.

To clarify this issue, we leveraged three facts. (1) Higher-order inferences about the environmental volatility yield posterior beliefs about external contingencies, derived from both prior beliefs and estimated volatility. Posterior beliefs thus depend on the estimated volatility: when the estimated volatility increases, posterior beliefs are less dependent on prior beliefs (and more on current observations)[1]. (2) Neural computations that derive posterior beliefs from prior beliefs are highly imprecise, which attenuates the dependence of posterior beliefs on prior ones[16]. (3) Computational imprecisions further scale with the magnitude of changes in internal representations (the so-called Weber's law in psychophysics[17–19]). We then conjectured that the corruption by imprecise computations of low-level inferences about external contingencies may counter-intuitively lead to efficient adaptive behaviour without relying on higher-order inferences about the environmental volatility[20]. The rationale is that, consistent with Weber's law, the computational imprecisions that corrupt posterior beliefs are larger in environments featuring more changes, that is, larger in high-volatility environments than in low-volatility ones. As a result, posterior beliefs in more volatile environments depend less strongly upon prior beliefs. This effect precisely corresponds to how the volatility estimated from higher-order inferences impacts the formation of posterior beliefs about external contingencies[21].

To test this conjecture, we built a model of optimal adaptive behaviour in a variety of uncertain and volatile environments, including closed (low-dimensional) and open-ended (high-dimensional) environments. Using machine learning, we investigated the model performance in these environments. We next considered a model that relies only on low-level inferences about external contingencies, without tracking volatility. We demonstrate here that when these inferences undergo computational imprecisions conforming to Weber's law, this model performs near-optimally in all tested environments. We also show that this 'Weber-imprecision model' outperforms no-inferential adaptive processes like reinforcement learning (RL). Finally, we recorded the performance of human participants in these environments and observed that the Weber-imprecision model unambiguously accounts for human performances better than the optimal adaptive models, their standard, biologically plausible algorithmic approximations and RL.

Results

Adaptive behaviour paradigm. We considered an adaptive agent that responds to successively presented stimuli. In every trial t, one among N distinct stimuli was randomly drawn and the agent responded by selecting one among M actions ($M > N$). The agent then received a positive or negative feedback. The agent thus searched for the correct responses to stimuli by trial and error. However, feedbacks were stochastic and the combination of correct responses to stimuli changed episodically. More precisely, the environment episodically switched across distinct stimulus–response combinations, and each combination specified one distinct response to each stimulus (the 'correct' response) that led to positive feedbacks with unknown, constant probability $\eta > 0.5$, while the other responses led only to positive feedbacks with probabilities

[1]Ecole Normale Supérieure, PSL Research University, Paris, France. [2]ENSAE ParisTech, Saclay, France. [3]Université Pierre et Marie Curie, Paris, France. [4]Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France. ✉e-mail: etienne.koechlin@upmc.fr

$1-\eta<0.5$. The current correct combination changed between two successive trials with a probability τ named volatility.

We investigated a range of prototypical environments by first varying number K of possible combinations (Methods and Fig. 1e). Closed environments were modelled using $K=2$ combinations corresponding to one repeated stimulus ($N=1$) and two available actions ($M=2$), that is, a two-armed bandit with potential reversals between correct actions. Open-ended environments were modelled using $K=24$ combinations corresponding to three stimuli ($N=3$) and four available actions ($M=4$), so that uncertainty also affects identification of the current correct combination. Whenever the current combination changed, a new combination k was drawn with unknown probability γ^k. Secondly, we independently varied the temporal structure of volatility τ. In stable environments, τ was set to zero: the correct combination remained unchanged across trials. We considered various stable environments by manipulating feedback sparsity, with feedbacks delivered in 100, 20, 10, 5 or 2% of trials. In changing environments, volatility τ was set to one of two non-zero constants, $\tau_{low}=0.03$ or $\tau_{high}=0.2$, which characterize rarely and frequently changing environments, respectively. Every trial delivered feedbacks. Finally, in unstable environments, τ followed a bounded, Gaussian random walk ($0.03<\tau<0.2$), such that the environmental volatility smoothly and stochastically varied across trials (Extended Data Fig. 1).

Optimal adaptive models. For each environment, we developed the optimal adaptive agent, which develops inferences from feedbacks corresponding to the environment generative process. First, in unstable environments, the optimal agent assumes the volatility of a given trial, τ_t, to vary as a bounded, Gaussian random walk with (unknown) variance ν and comprising three hierarchically organized levels of inferences[1] that relate to (1) the volatility change rate ν, (2) the successive volatility values τ_t and (3) the successive occurrences of correct combinations z_t (Fig. 1a and Extended Data Fig. 3). These inferences further combine with those about feedback probability η and, when $K>2$, about occurrence probabilities γ^k of combinations k. The agent thus forms posterior beliefs $B(t)$ about the current correct combination in trial t and, by marginalizing over these beliefs, selects the most likely correct action in response to stimuli.

This inferential agent is computationally intractable. We therefore emulated this agent using a sequential Monte Carlo method based on particle filtering that has recently been developed in machine learning to solve these models[22–25] (Methods). We refer to this agent emulation as the exact varying-volatility model.

Second, in changing environments, the optimal agent is identical to the one described for unstable environments, except that the agent now comprises only two hierarchically organized inferential levels, which relate to (1) volatility value τ, which is assumed to be constant, and as above, (2) the successive occurrences of correct combinations z_t (Fig. 1b and Extended Data Fig. 4). This inferential agent is also computationally intractable and was therefore emulated with the same particle filtering method. We refer to this emulation as the exact constant-volatility model.

Both models emulate the exact inferential process, combining offline both backward and forward inferences (Methods and Supplementary Fig. 5). Both models therefore lack biological plausibility and are computationally exorbitant. Consequently, we also investigated their online forward algorithmic approximation by restricting the particle filtering to the online, forward sampling that estimates posterior beliefs $B(t)$ from posterior beliefs $B(t-1)$ in the preceding trial[23,24] (Methods). The resulting online particle filtering is plausibly implemented in populations of cortical neurons[26–28]. We refer to these approximations as the forward varying- and constant-volatility models.

Third, in stable environments, the optimal agent is identical to those described for unstable and changing environments except that it now comprises only one inferential level related to external contingencies (Fig. 1c and Extended Data Fig. 5): namely, the identity of the correct combination z is assumed to remain unchanged over trials (no volatility). The resulting inferential complexity is considerably lower, such that this agent is computable through pure online forward closed-form computations[29] that derive posterior beliefs $B(t)$ about a combination z directly from posterior beliefs $B(t-1)$ given the feedback observed in trial t (Methods). For consistency, however, we emulated the agent using the same online, forward particle-filtering method as above, which, in this case, emulates the exact inference process. We refer to this emulation as the zero-volatility model.

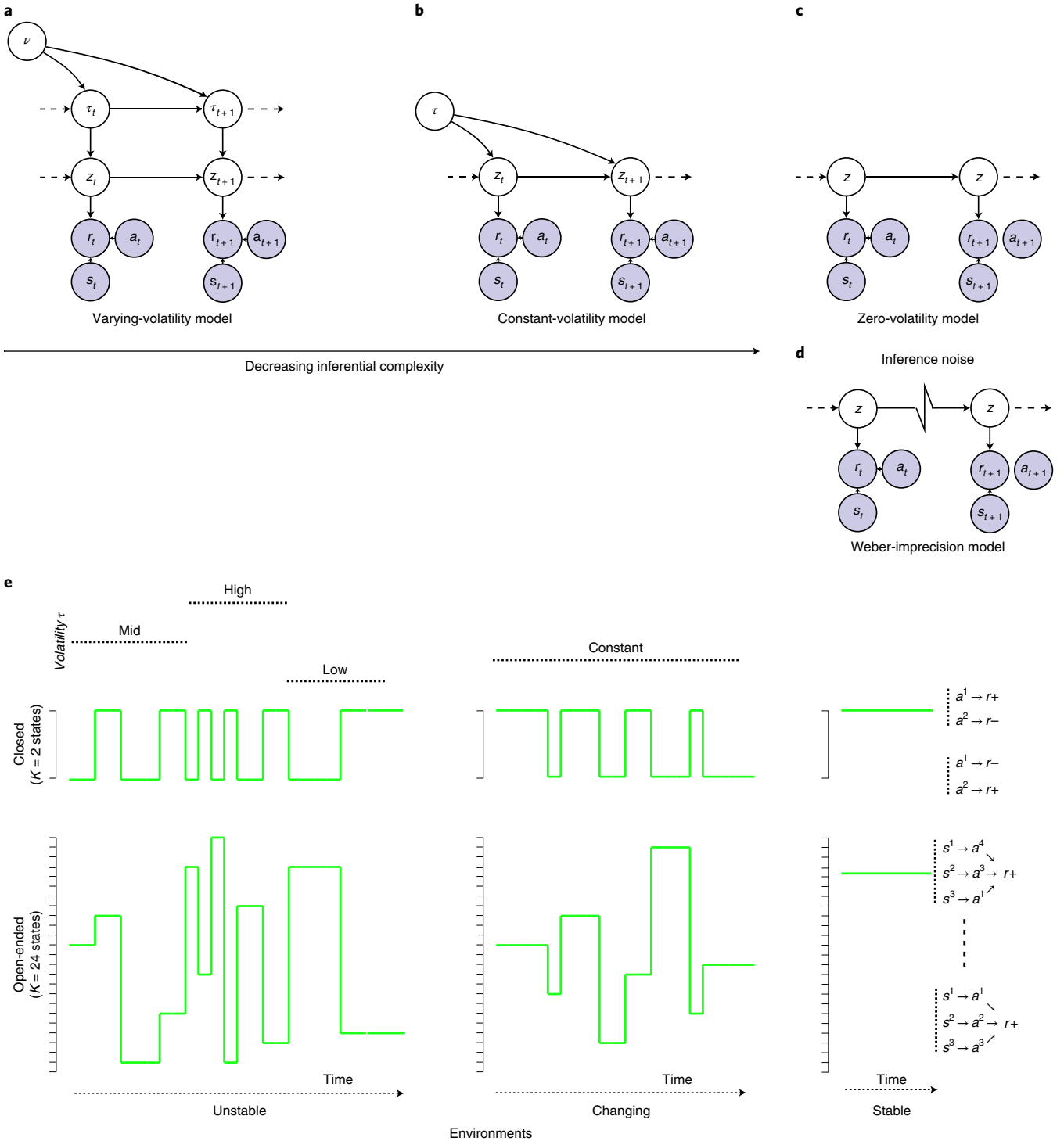
The Weber-imprecision model. The zero-volatility model evidently has poor adaptive performances in volatile environments when combinations change: posterior beliefs $B(t)$ about the correct combination move slowly from current to new correct combinations. We hypothesized that the substantial computational imprecisions previously identified in human inferential processes[16] remedy this limitation. Consistently with Weber's law[17,18], computational imprecisions presumably scale with the magnitude of belief updating and, consequently, increase whenever combinations change. Such computational imprecisions make posterior beliefs $B(t)$ less dependent upon past observations, especially when combinations change more frequently. Posterior beliefs $B(t)$ thus adjust to the environmental volatility as if they derived from higher-order inferences about that volatility. We hypothesized that with such imprecisions, the zero-volatility model approaches the optimal adaptive performance in volatile environments.

Fig. 1 | Inferential models of adaptive behaviour. a, Third-order inferential models (referred to as varying-volatility models) assume that the environment's latent state z_t in trial t changes with probability (volatility) τ_t, which in turn is assumed to vary as a Gaussian random walk with constant variance ν. Inferences relate to volatility variance ν, volatility τ_t and latent state z_t according to external feedbacks r_t observed following action a_t selected in response to stimuli s_t. Note that although variance ν is assumed to be constant, its estimates may change across trials. **b**, Second-order inferential models (referred to as constant-volatility models) assume that the environment's latent state z_t changes with constant probability (volatility) τ. Inferences relate only to volatility τ and latent state z_t. Again, although volatility τ is assumed to be constant, its estimates may change across trials. **c**, First-order inferential models (referred to as zero-volatility models) assume that the environment's latent state z remains unchanged across trials. Inferences relate to only latent state z. Again, although latent state z is assumed to be constant, its estimates may change across trials. **d**, The Weber-imprecision model is identical to the zero-volatility model, except that imprecisions occur in inferential computations. These computational imprecisions reflect neural noise in a manner consistent with Weber's law. See text for details. Extended Data Figs. 3–5 show the full generative models corresponding to the varying-volatility, constant-volatility and zero-volatility models, respectively. **e**, Examples of unstable, changing and stable environments whose generative processes correspond to the varying-, constant- and zero-volatility models, respectively. Closed environments were modelled using $K=2$ latent states (two-armed bandit potentially reversing shown on the right). Open-ended environments were modelled using $K=24$ latent states associated with distinct stimulus–response mappings combining three stimuli and four possible responses (two combination examples are shown on the right). Green lines show transitions between latent states. See text for details.

Computational imprecisions presumably derive from noisy online neural computations. We consequently assumed that in the (forward) zero-volatility model, particle filtering is noisy: every particle coding for one combination in trial t may start (mis)coding for another combination in trial $t+1$ with probability ϵ_t. Weber's law states that computational imprecisions scale with the distance d_t between posterior beliefs in trials t and $t+1$. We therefore assumed noise ϵ_t is a random variable uniformly distributed between 0 and $\mu+\lambda d_t$:

$$\epsilon_t \sim U(0, \mu + \lambda d_t)$$

where $\mu \geq 0$ and $\lambda \geq 0$ are free parameters quantifying the constant and Weber components of computational imprecision, respectively (Methods). We refer to this model as the Weber-imprecision model, which comprises the zero-volatility model as the special case (μ, λ) = (0,0) (Fig. 1d). Only when λ is non-zero are posterior beliefs $B(t)$ sensitive to the environmental volatility as if they derived from higher-order inferences about that volatility (Extended Data Fig. 2). Like the models described above, the Weber-imprecision model comprises inferential processes about (1) feedback probability η, to prevent feedback stochastic fluctuations from improperly impacting state beliefs $B(t)$ and (2) about occurrence probabilities γ^k of



combinations k (latent states), to capture their potential differential occurrences.

Alternative models. To assess the functional specificity of computational imprecisions in inferences, we investigated an alternative model whereby computational imprecisions occur in action selection rather than in belief inferences. This model, named the noisy-selection model, is identical to the zero-volatility model except that actions are probabilistically selected according to the standard softmax rule with inverse temperature β as free parameter (the zero-volatility model corresponds to $\beta \gg 1$).

Finally, to assess the role of inferential processes in optimizing adaptive behaviour, we also considered the standard, non-inferential adaptive model, namely, the Pearce–Hall RL model with computational imprecisions. The model, named the noisy-PH-RL model, combines Pearce–Hall's learning rule[30,31], noisy updates scaling with the unsigned reward prediction error and the softmax rule[32]. Free parameters included along with inverse temperature β are α and α_{PH}, scaling the constant and adjustable components of the learning rate, respectively, and ζ, scaling updating noise (Methods).

Computer simulation results. We simulated all the models in each environment described above, for a total of 1,000 trials. We set feedback probability η to 90%. In each environment, we simulated each model 50 times and computed its resulting average performance (proportion of correct responses). In all unstable environment simulations, we set volatility variance ν to 0.0001. In every open-ended environment simulation, occurrence probabilities γ^k of combinations were set randomly. We first analysed the best performance that the Weber-imprecision, noisy-selection and noisy-PH-RL models could achieve in every environment. For every environment, we thus computed the free parameters that maximized model performances. We then compared these best performances to the corresponding optimal performances. The performances of the exact zero-, constant- and varying-volatility models in stable, changing and unstable environments were virtually identical to their forward approximations (Fig. 2).

In stable environments, as expected, both the Weber-imprecision and noisy-selection models performed optimally, regardless of feedback sparsity and number K of latent states (Fig. 2a and Supplementary Fig. 1). Indeed, both models contain the zero-volatility model as a special case. By contrast, the noisy-PH-RL model reached optimal performance only when $K=2$. When $K=24$ and feedback sparsity was equal to 2%, this model exhibited a ~9% loss of performance (s.e.m. = 2%), reflecting the lack of inferential processes about latent states.

In changing environments, as hypothesized, only the Weber-imprecision model performed quasi-optimally for both $K=2$ and $K=24$ (performance losses <1% and <2% for τ_{low} and τ_{high}, respectively) (Fig. 2b,d). The noisy-PH-RL model performed moderately for $K=2$ (losses of ~3% and ~9% for τ_{low} and τ_{high}, respectively) but poorly for $K=24$ (losses of ~14% for both τ_{low} and τ_{high}). The noisy-selection model performed even more poorly for both $K=2$ and $K=24$, regardless of the environmental volatility (all losses >17%) (Fig. 2b,d).

In unstable environments, the results also confirmed our prediction (Fig. 2c). Only the Weber-imprecision model still performed quasi-optimally (losses <1% for both $K=2$ and $K=24$). Again, the noisy-PH-RL model performed moderately for $K=2$ (loss ~5%) but poorly for $K=24$ (loss ~15%), while in both cases, the noisy-selection model performed poorly (both losses >26%).

Thus, only the Weber-imprecision model reached a virtually optimal performance in a variety of stable and volatile environments. This optimal performance, however, was reached with noise parameters (μ, λ) properly adjusted to each environment. This adjustment deviates from the idea that, as they reflect neural computational imprecisions, these parameters are specifically *not* adjustable, especially through higher-order inferential processes. To address this issue, we next analysed the model versatility (or robustness) across all the environments with no parameter adjustments. We first computed for every pair of noise parameters (μ, λ) the performance loss that the Weber-imprecision model exhibited in each environment relative to the optimal performance. For each pair (μ, λ), we then computed the maximal performance loss (denoted maxloss(μ, λ)) over all the environments. Figure 3a shows that, consistent with our prediction, maxloss(μ, λ) remained below 10% when the constant noise component μ remained close to zero (<0.04) and the Weber noise component λ ranged between ~0.8 and ~1.8, while always remaining above 22% when $\lambda = 0$. Maxloss(μ, λ) was minimal and equal to 9.2% loss (s.e.m. = 0.16%) when $\mu^* = 0.01$ and $\lambda^* = 1.4$ (the asterisk (*) indicates the optimal values over all environments). In every environment, accordingly, the Weber-imprecision model with parameters (μ^*, λ^*) exhibited a performance loss relative to theoretical optima that never exceeded 9.2% (s.e.m. = 0.16%). This upper bound value, named the 'minimaxloss', defines the model versatility over environments.

Using the same minimax approach, we found that the noisy-PH-RL and noisy-selection models were dramatically less versatile than the Weber-imprecision model. The noisy-PH-RL minimaxloss was equal to 22.8% (s.e.m. = 0.61%) and corresponded to constant and adjustable learning rates $\alpha^* = 0.53$ and $\alpha^*_{PH} = 0.33$ and no noise ($\zeta^* = 0$) (Fig. 3b). The noisy-selection minimaxloss was equal to 39.6% (s.e.m. = 0.78%) and corresponded to inverse temperature $\beta^* \to +\infty$, that is, to the zero-volatility model (Fig. 3b), indicating that even considering an adaptive inverse temperature cannot improve performance. Thus, the Weber-imprecision model versatility could neither be achieved through RL processes nor noisy action selection and specifically resulted from imprecise probabilistic inferences over external contingencies.

We next compared the Weber-imprecision model versatility to the versatility of both the constant- and varying-volatility models. These models contain no free parameters, so that the minimaxloss simply reduces to the maxloss these models exhibited across the environments. We found that only the exact constant-volatility model was more versatile than the Weber-imprecision model (Fig. 3b). While the latter exhibited a 9.5% minimaxloss (s.e.m. = 0.16%) as reported above, the minimaxloss for the exact constant-volatility model was 4.5% (s.e.m. = 0.76%). Its forward approximation (that is, the forward constant-volatility model), however, has as low a versatility as the noisy-PH-RL model (minimaxloss = 20.3%, s.e.m. = 2.72%), while both the exact and forward varying-volatility models performed even more poorly (minimaxloss > 40%, s.e.m. < 2%). All these higher-order inference models exhibited the largest performance loss in environments with sparse feedbacks, indicating that (1) third-order inferences about volatility change rate ν are highly inefficient and even deleterious with sparse feedbacks, as the number of volatility trajectories between two distant informative trials are potentially infinite and (2) assuming a constant volatility overcomes this problem but requires backward processes to bring together distant feedbacks for properly inferring volatility. Finally, the Weber-imprecision model endured its maximal loss relative to optima (minimaxloss = 9.5%) in the environment where the constant-volatility model is precisely the optimal agent (changing environment), and volatility is large ($\tau_{high} \approx 0.2$) and represents the only source of uncertainty regarding correct combinations (closed environment $K=2$). Among the biologically plausible models, the Weber-imprecision model thus appeared as the most versatile adaptive model. We obtained the same results when considering the absolute rather than relative performance losses (Fig. 3c,d).

Human adaptive performances in closed environments. We next investigated how the models account for human behaviour in volatile

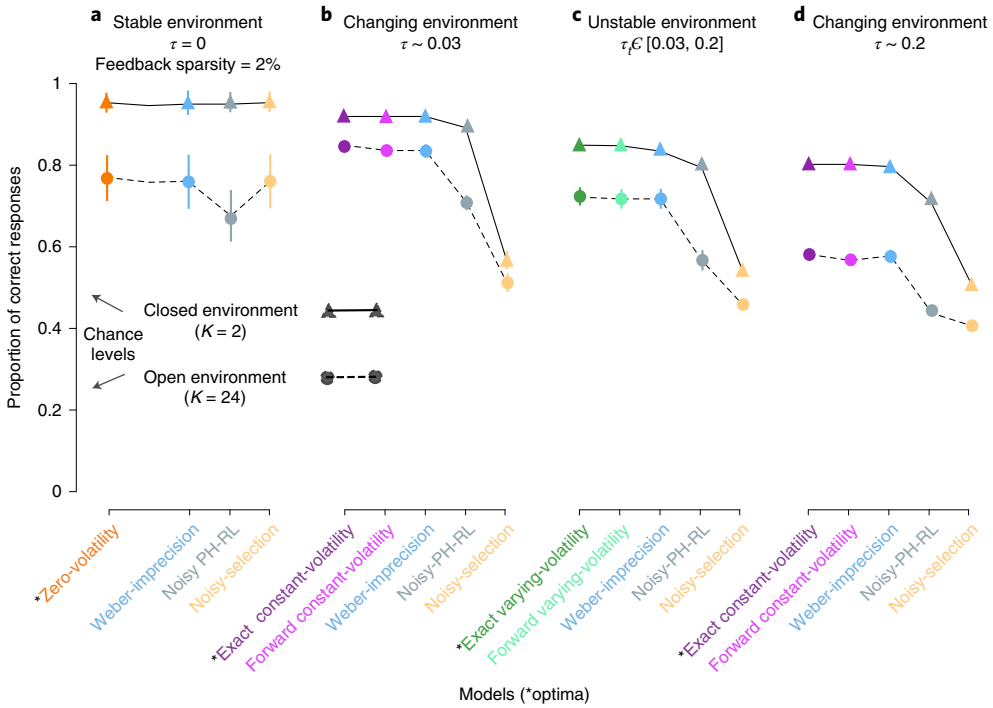


Fig. 2 | Models' maximal performances in stable, changing and unstable environments. Maximal proportions of correct responses for the Weber-imprecision, noisy-PH-RL and noisy-selection models when simulated in closed ($K=2$ latent states, triangles) and open-ended ($K=24$ latent states, circles) environments. Each model was simulated $N=50$ times in every environment (error bars are s.e.m. across simulations). * shows the theoretical optimal model for each environment. Solid and dashed lines serve as guides to the eye. **a**, Stable environments (volatility $\tau=0$) with sparse feedbacks (2% of trials). The theoretical optimal performance corresponds to the zero-volatility model. **b**, Rarely changing environments (constant volatility $\tau=0.03$). The theoretical optimal performance corresponds to the exact constant-volatility model. **c**, Unstable environments (volatility τ_t follows a bounded Gaussian random walk in the range [0.03, 0.2]). **d**, Frequently changing environments (constant volatility $\tau=0.2$). The theoretical optimal performance corresponds to the exact constant-volatility model. Note that in every environment, the forward approximation of the exact optimal model as well as the Weber-imprecision model virtually achieved the optimal performance. By contrast, the noisy-PH-RL model performed decently in closed environments but poorly in open-ended environments, while the noisy-selection model (zero-volatility model comprising a softmax affecting action selection) performed poorly in both closed and open-ended environments.

environments. We first tested $N=22$ participants in a behavioural protocol implementing the unstable, closed environment described above (that is, $K=2$, two-armed bandits with reversals) (Fig. 4a). The protocol exactly corresponded to this environment except that volatility τ_t varied as a stepwise function rather than a Gaussian random walk (Fig. 4b and Methods) and feedback probability η was set to 80% rather than 90%. The protocol was thus similar to that used in previous studies[1]. Computer simulations showed that in this protocol, the exact varying- and constant-volatility model performances were 86.9% and 86.3%, respectively (s.e.m. < 0.3%). Their forward approximation reached virtually the same performances (86.6% and 86.0%, respectively; s.e.m. < 0.3%). Participants' performances were significantly lower (79.8%, s.e.m. = 0.9%), although as previously shown[1], their behaviour was sensitive to the environmental volatility. Fitting participants' performances with the inferential model (with the distinct volatility levels corresponding to the different volatility blocks as free parameters) revealed that participants' behaviour was indeed associated with significantly larger fitted volatility values in high than in low-volatility blocks (median-split high versus low-volatility blocks; paired t-test; $t(20)=2.46$, $P=0.023$, two-tailed; Cohen's $d=0.383$; 95% confidence interval [0.0034, 0.0297]) (Fig. 4c).

To fit participants' choice data, we inserted in every model a softmax rule for action selection (inverse temperature β as the free parameter) (Methods). The Weber-imprecision model consequently included the noisy-selection model as a special case with noise parameters set to zero. Model posterior probabilities (MPPs) given the data were derived by computing model marginal likelihoods over parameter spaces using advanced Monte Carlo procedures (Methods). These MPPs are the optimal Bayesian quantification for selecting models, balancing model degrees of freedom and adequacy to data[33,34]. Among all inferential models (exact varying- or constant-volatility models, their forward approximations and the Weber-imprecision model), the Weber-imprecision model showed the largest MPPs, which further led to the rejection of the other inferential models (exceedance probability $P_{\text{exceed}}=0.973$)[33,34] (Fig. 4d). Consistently with our simulations, the best-fitting Weber-imprecision model (that is, the one with free parameters maximizing model likelihoods) relied on the dominant contribution of the Weber noise component ($\lambda_{\text{fit}}\langle d_t\rangle=0.25$; $\mu_{\text{fit}}=0.05$, s.e.m. = 0.01; $\lambda_{\text{fit}}=0.8$, s.e.m. = 0.1). The Weber-imprecision model further fit participants' performances significantly better than the reduced model with Weber noise component $\lambda=0$ ($MPP_{\lambda=0}=0.22$, $\text{MPP}_{\lambda>0}=0.78$, $P_{\text{exceed}}=0.998$) and better than the extended model with distinct Weber noise components for low and high-volatility episodes ($\text{MPP}_{\lambda=\text{constant}}=0.79$, $\text{MPP}_{\lambda\text{high}\neq\lambda\text{low}}=0.21$, $P_{\text{exceed}}>0.998$; $\lambda_{\text{high}}=0.88$, $\lambda_{\text{low}}=0.76$, s.e.m. = 0.126, paired t-test: $t(20)=1.71$, $P=0.1$).

A model recovery procedure (Methods)[35] confirmed that our behavioural protocol and fitting method properly discriminated the models. Indeed, the MPPs revealed that every model fit its own performance, as simulated from its best-fitting free parameters, better than the other models and rejected all other models (all $P_{\text{exceed}}>0.99$) (Fig. 4e). Moreover, only the Weber-imprecision model explained participants' data better than all RL models includ-

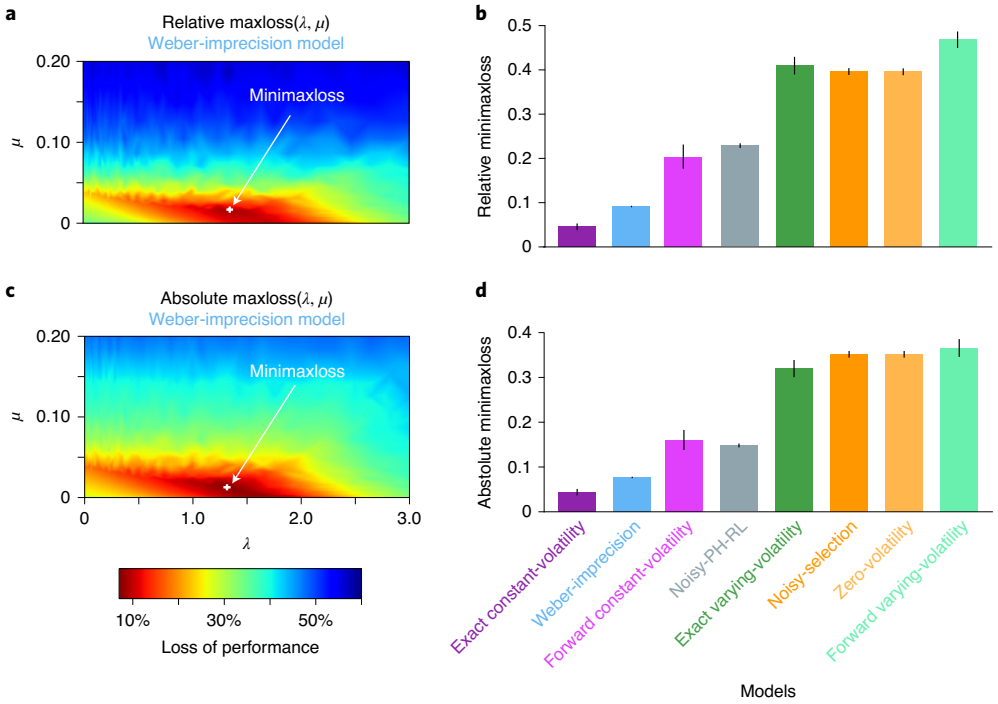


Fig. 3 | Models' versatility across environments. a, Maximal losses over all investigated environments (stable, rarely and frequently changing and unstable, either closed or open-ended) of the Weber-imprecision model relative to theoretical optimal performances according to noise free parameters μ and λ (constant and Weber noise components, respectively). These losses are normalized in each environment with respect to the related theoretical optimal performance. The arrow indicates the minimaxloss that corresponds to the parameter values minimizing these maximal losses. **b**, Minimaxlosses for all the investigated models. **c**, Same as **a**, except that absolute rather than relative losses are considered: namely, losses in every environment are not normalized with respect to the related theoretical optimal performance. **d**, Same as **b**, except that minimaxlosses are computed from absolute rather than relative losses. Each model was simulated $N = 50$ times in each environment (error bars are s.e.m. across simulations). The lower the minimaxloss, the more versatile the model is across the environments. Note that the exact constant-volatility model is the most versatile, but it is biologically implausible from a computational viewpoint. Among the biologically plausible models (Weber-imprecision, noisy-PH-RL, zero-volatility, noisy-selection and forward constant- or varying-volatility models), the Weber-imprecision model is the most versatile and approaches the versatility of the exact constant-volatility model.

ing and nested in the noisy-PH-RL model ($P_{exceed} > 0.99$), while the other models failed (all $P_{exceed} \leq 0.55$) (Fig. 4f and Supplementary Fig. 2). Consistently, the best-fitting Weber-imprecision model fully captured the time course of participants' performances following reversals, while the other best-fitting models failed (Fig. 5 and Supplementary Note 1).

Human adaptive performances in open-ended environments. We next investigated models' accounts of human behaviour in volatile, open-ended environments. We analysed the performances of 62 participants from two previous studies[5,7], who were tested in the rarely changing, open-ended environment described above (constant volatility $\tau_{low} = 0.03$, $N = 3$, $M = 4$, $K = 24$) (Fig. 6a). Unbeknownst to participants, every participant performed two distinct instantiations of the environment: in the no-recurrence condition, every combination occurred once, thus forming 24 episodes associated with the 24 distinct combinations; in the recurrence condition, the 24 episodes corresponded to distinct combinations pseudo-randomly drawn from a subset of only three combinations. Computer simulations showed that the optimal performance from the exact constant-volatility model was equal to 86.9% (s.e.m. = 0.4%) over these two conditions. The exact varying-volatility model reached virtually the same performance (86.7%, s.e.m. = 0.4%), which their forward approximation also achieved (86.1% and 85.4%, respectively; s.e.m. < 0.4%). Participants exhibited a significantly lower performance (74.2%, s.e.m. = 0.7%).

In these open-ended environments, computing model marginal likelihoods was practically intractable. To derive MPPs, we therefore used model Bayesian information criteria as standard estimates of marginal likelihoods, with model free parameters maximizing model likelihoods (Methods). We first found that in both conditions, and especially in the recurrence condition, every inferential model explained participants' data better than all RL models including and nested in the noisy-PH-RL model (all $P_{exceed} > 0.99$) (Fig. 6d and Supplementary Fig. 3). Unlike RL models, all the inferential models make inferences at the level of combinations and infer their occurrence probabilities γ_k (see model descriptions above). The result thus confirmed previous findings[5,7] that, unlike RL models, participants formed beliefs about combinations and their occurrences, and that those beliefs endowed them with the ability to capture the recurrence of combinations in the recurrence condition (Supplementary Fig. 4). Yet the same model recovery procedure as above confirmed that the protocol associated with our fitting method properly discriminated the inferential models: every model fit its own performance simulated from its best-fitting free parameters better than the other models did, leading to the rejection of those models (all $P_{exceed} > 0.99$, Fig. 6c). Critically, the Weber-imprecision model again best explained participants' data by exhibiting the largest MPP, leading to the rejection of all the other inferential models ($P_{exceed} > 0.99$) (Fig. 6b). The best-fitting Weber-imprecision model again relied on the dominant contribution of the Weber noise component ($\lambda_{fit}\langle d_t \rangle = 0.39$; $\mu_{fit} = 0.05$, s.e.m. = 0.002; $\lambda_{fit} = 1.2$, s.e.m. = 0.06) (Supplementary Note 3).

Discussion

The Weber-imprecision model consists of first-order inferential processes forming beliefs about external contingencies and under-

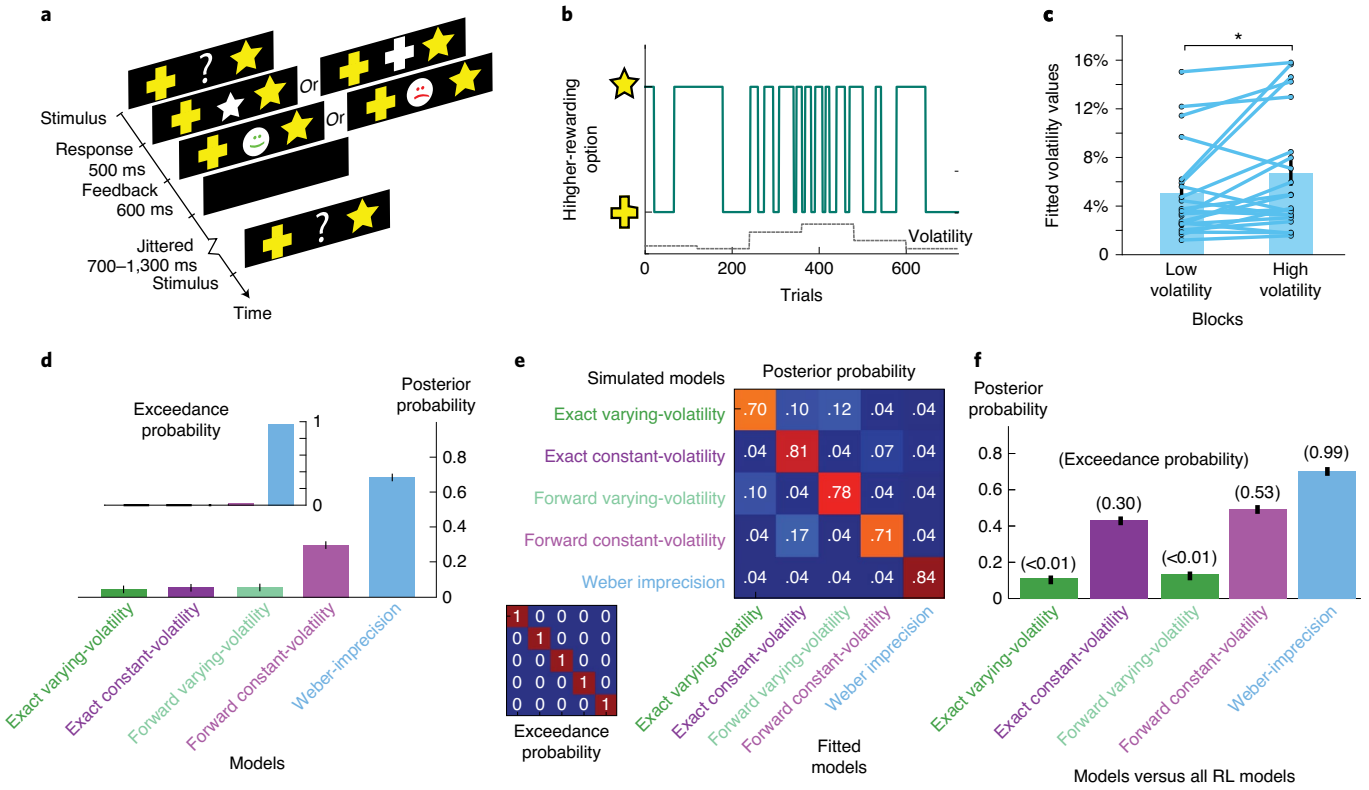


Fig. 4 | Model fits to human performances in closed, unstable environments. a, Trial structure in the two-armed bandit task. Participants chose one of two visually presented symbols by pressing a response button and received a binary feedback. One symbol led to positive feedbacks with 80% probability, the other led to negative feedbacks with 80% probability. **b**, Task temporal structure. Feedback probabilities reversed episodically between symbols. Volatility (probability of reversals) varied along the experimental session as a stepwise function comprising six distinct levels [0.01, 0.02, 0.03, 0.05, 0.08, 0.15] whose occurrence order was counterbalanced across participants. Each level comprised 120 trials. One example of an experimental session is shown. **c**, Fitted volatility values in low- and high-volatility blocks (median-split high- versus low-volatility blocks) resulting from fitting participants' performances with the inferential model taking as free parameters the distinct volatility levels corresponding to the different volatility blocks (*$P = 0.023$, two-tailed paired t-test, $t(20) = 2.46$, Cohen's $d = 0.383$, 95% confidence interval [0.0034, 0.0297]; data distribution was assumed to be normal, but this was not formally tested). **d**, MPPs given human data and related exceedance probabilities (inset) for inferential models (varying- and constant-volatility models, their forward approximations and the Weber-imprecision model). The Weber-imprecision model unambiguously best accounted for human performances ($P_{exceed} = 0.973$). Error bars are standard deviations of Bayesian estimators. **e**, Confusion matrices from the model recovery procedure assessing the protocol's ability to dissociate the models. The large matrix shows MPPs; the small matrix shows the related exceedance probabilities. Colour scale ranges from 0 (dark blue) to 1 (dark red). Each model unequivocally explains its own behaviour better than the other models. **f**, MPPs relative to the best-fitting RL model including and nested in the noisy-PH-RL model (here the noisy-RL; Supplementary Fig. 2) given human data. Related exceedance probabilities are shown in parentheses. Error bars are standard deviations of Bayesian estimators. Only the Weber-imprecision model unambiguously accounts for human data better than all RL models. In all these analyses, action selection in every model is assumed to follow a softmax probabilistic function (inverse temperature β as free parameter). See Extended Data Figs. 6–8 for model fitted parameters.

going computational imprecisions that, consistent with the Weber Law[17–19], scale with the magnitude of belief updating. The results show that the model

1. can reach the theoretically optimal adaptive behaviour in stable, changing, unstable, closed and open-ended environments;
2. is the most versatile adaptive process among biologically plausible adaptive processes tested in this study, including adaptive RL (ref. [30] and M. Lehmann, poster presented at the *Computational and System Neurosciences Conference 2015*) and forward, online approximations of higher-order inferential processes tracking volatility[1,11–14,36,37], that is, exhibits the best performance over these environments without prior knowledge about their temporal or volatility structures;
3. is the second-most versatile process approaching the maximal versatility that exact but biologically implausible second-order inferential processes tracking volatility achieves[12,36], while outperforming the versatility of higher-order inferential processes tracking volatility[1] and
4. best accounts for human performances in these environments.

These results hold neither when computational imprecisions occur in action selection nor in adaptive RL. These findings thus support the Weber-imprecision model as describing human adaptive processes in uncertain and variable environments.

Our computer simulations provide principled support to the Weber-imprecision model. In a variety of environments (closed or open-ended and with zero, constant or unstable volatility), the Weber-imprecision model can (counter-intuitively) virtually reach the adaptive optimum in every environment (adpative optima correspond to inferential processes matching the environment generative processes) and outperform RL. The Weber-imprecision model thus captures key adaptive properties of optimal inferential processes that, in previous studies, may have been (mis)interpreted as

evidence supporting the presence in humans of higher-order inferences tracking volatility.

Moreover, among the biologically plausible models involving no offline backward inferences, the Weber-imprecision model exhibits adaptive performances emerging as the most versatile and/or robust to structural uncertainty[12] (also named Knightian or Keynesian uncertainty in economics[38,39]). As is frequently the case in real life, structural uncertainty reflects a lack of knowledge about the generative structure of the environment. We investigated environments with three distinct temporal structures: stable (zero volatility), changing (constant volatility) and unstable (variable volatility) environments. As the unstable generative structure encompasses the changing one, which in turn encompasses the stable one, in principle third-order inference processes corresponding to the adaptive optimum in unstable environments should optimally unravel this structural uncertainty. However, when action feedbacks were sparse as is often the case in everyday situations, we found these processes to exhibit performances lower than RL. This finding confirms that, as previously advocated[12,15], resolving temporal structural uncertainty through higher-order inferences operating within the high-dimensional space of possible temporal structures is actually ineffective and even deleterious in view of the informative poverty of environment feedbacks.

By contrast, exact second-order inference processes corresponding to adaptive optima in changing environments were found to approach the adaptive optimum in every environment, even when feedbacks were sparse. These processes assume a constant volatility and, in principle, optimally unravel structural uncertainty only across changing and stable environments. However, these processes efficiently adapt to unstable environments, as posterior beliefs about the environmental volatility are constantly updated. Moreover, with sparse feedbacks, assuming a constant volatility rules out the considerable variety of volatility trajectories that are compatible with temporally distant feedbacks but inferentially inextricable. These processes thus represent the best compromise between the dimensionality and adequacy of the inferential space regarding temporal structures. However, their biologically plausible approximation, which excludes offline backward inferences, exhibited adaptive performances similar to RL when feedbacks were sparse. Offline back-

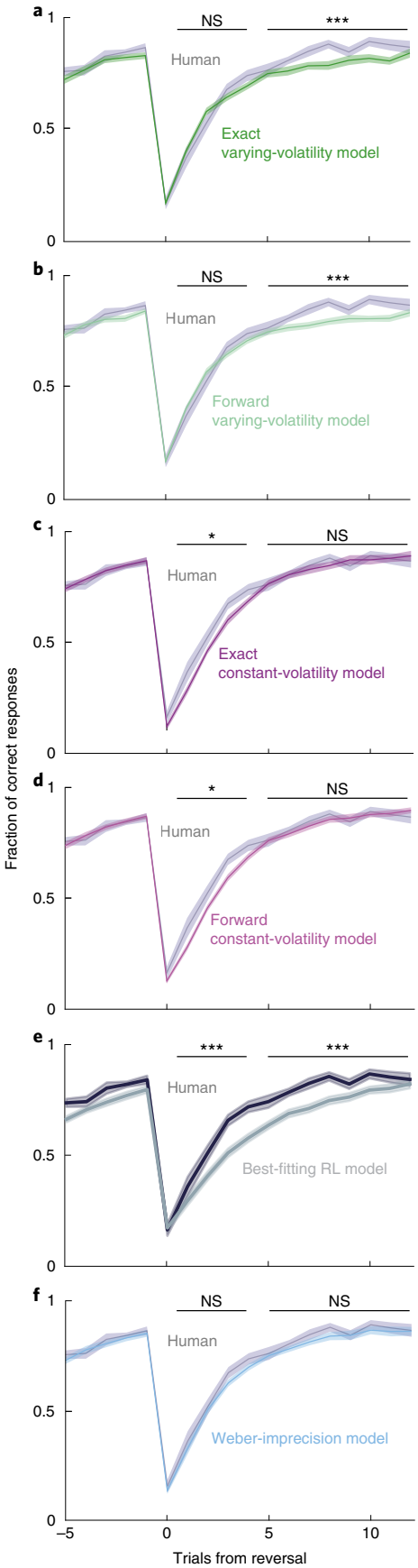


Fig. 5 | Human and model adaptive behaviour following contingency reversals. Fraction of correct responses in trials preceding and following reversals in the two-armed bandit task described in Fig. 4 (closed, volatile, unstable environment). Each graph shows human performances (same data in all graphs) overlaid with fitted model performances averaged across participants (shaded areas are s.e.m. across participants, $N = 21$). Trial 0 corresponds to the trial when, unbeknownst to participants and models, reversals occurred. Human performances showed fast early adaptive responses to reversals (trials 1 to 3) followed by a slow, late adaptation to new environment contingencies (trials 5 to 15) leading to a performance plateau. **a**, Human and exact varying-volatility model performances. Late adaptations (averaged over trials 5 to 15) were slower in the model than in participants (two-tailed paired *t*-test, $t(20) = 3.9$, $P < 0.001$). **b**, Human and forward varying-volatility model performances. Again, late adaptations were slower in the model than in participants (two-tailed paired *t*-test, $t(20) = 4.7$, $P < 0.001$). **c**, Human and exact constant-volatility model performances. Early adaptive responses were slower in the model than in participants (two-tailed paired *t*-test, $t(20) = 2.39$, $P = 0.027$). **d**, Human and forward constant-volatility model performances. Again, early adaptive responses were slower in the model than in participants (two-tailed paired *t*-test, $t(20) = 2.65$, $P = 0.015$). **e**, Human and RL model performances. Both early adaptive responses and late adaptations were slower in RL than in participants (two-tailed paired *t*-tests, both $t(20) > 2.91$, $P < 0.008$). **f**, Human and Weber-imprecision model performances. The Weber-imprecision model exhibited early adaptive responses and late adaptations similar to participants (two-tailed paired *t*-tests, both $t(20) < 1.12$, $P > 0.12$). See Supplementary Note 1 for a detailed discussion of these effects. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; NS, not significant.

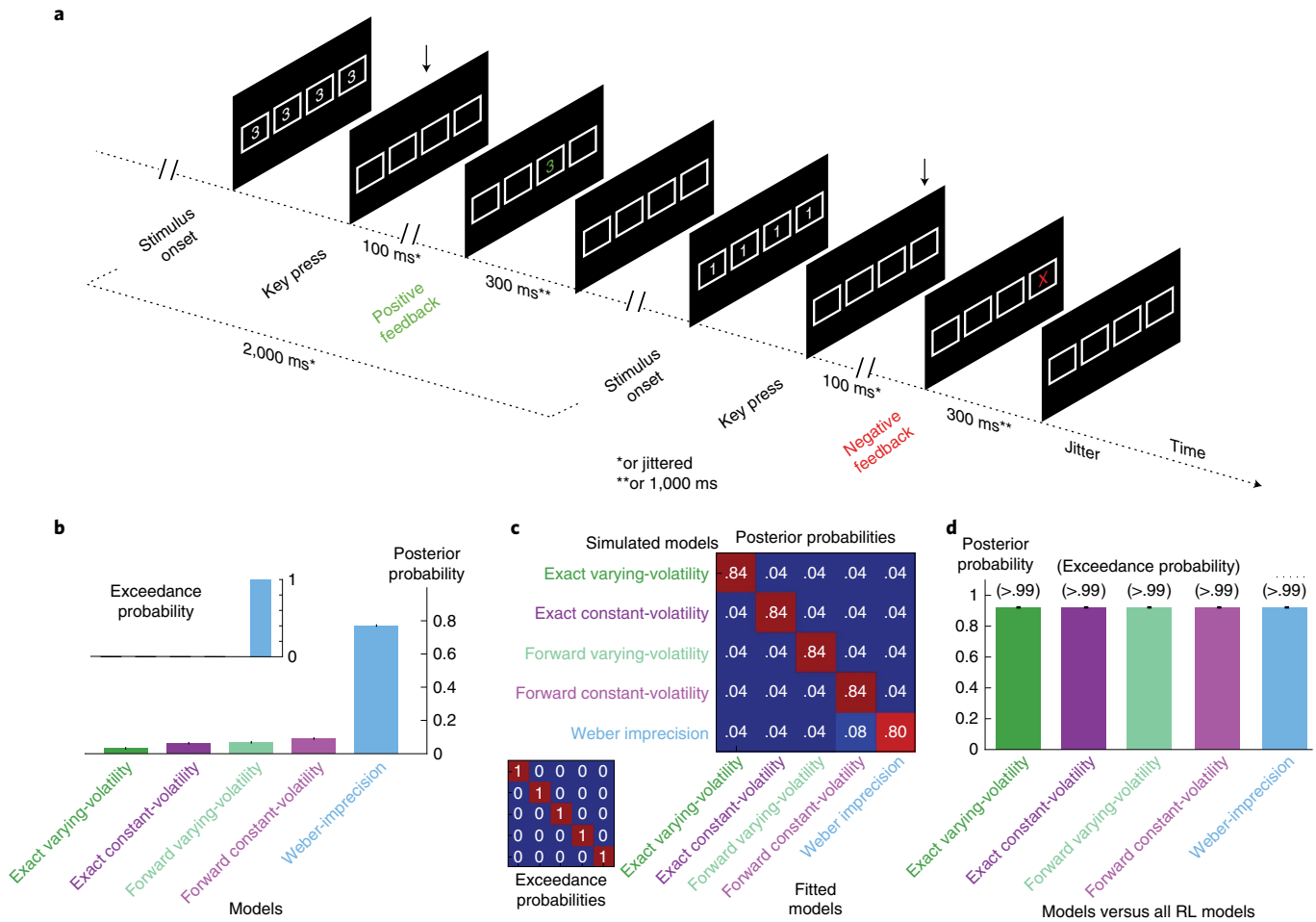


Fig. 6 | Model fits to human performances in open-ended, changing environments. a, Trial structure in the open-ended task, modelling rarely changing environments with stochastic, positive or negative feedbacks. Visual stimuli were pseudorandomly drawn from a set of three Arabic numbers, for example (1, 3, 5). Participants responded by pressing one of four possible response keys. Shortly afterparticipants' responses, stimuli were removed and a positive or negative feedback was presented. Each digit was associated with a unique correct response leading to positive feedbacks with 90% probability. The other responses led to negative feedbacks with 90% probability. Digit-response combinations changed episodically and unpredictably with constant volatility $\tau = 0.03$. Two successive combinations were orthogonal, that is, all stimulus–response pairs were distinct (see also refs. [5,7]). * and ** indicate slight timing differences in trial structures between ref. [5] and ref. [7] due to specific neuroimaging constraints in the latter study. **b**, Estimated MPPs with respect to human data and related exceedance probabilities (inset) for inferential models (varying- and constant-volatility models, their forward approximations and the Weber-imprecision model). Error bars are standard deviations of Bayesian estimators. The Weber-imprecision model unambiguously best accounted for human performances ($P_{exceed} > 0.99$). **c**, Confusion matrices from the model recovery procedure assessing the protocol's ability to dissociate the models. Colour scale ranges from 0 (dark blue) to 1 (dark red). The large matrix shows estimated MPPs; the small matrix shows the related exceedance probabilities and reveals that each model clearly explains its own behaviour better than the other models do. **d**, Estimated posterior probabilities of each model relative to the best-fitting RL model including and nested in the noisy-PH-RL model (here the noisy-PH-RL model, Supplementary Fig. 2) with respect to human data. Related exceedance probabilities are shown in parentheses. Error bars are standard deviations of Bayesian estimators. All inferential models unambiguously account for human data better than the RL model. Supplementary Fig. 3 shows the fits separately for the recurrence and no-recurrence condition. In all these analyses, action selection in every model is assumed to follow a softmax probabilistic function (inverse temperature β as the free parameter). See Extended Data Figs. 6–8 for model fit parameters.

ward inferences are thus required for bridging across no-feedback trials to properly infer the environmental volatility and exhibit efficient adaptive performances surpassing RL.

Although the Weber-imprecision model reduces to first-order inferences corresponding to the adaptive optima in stable environments, the model resolved temporal structural uncertainty close to exact second-order inference processes, even when feedbacks were sparse. The model indeed assumes external contingencies to remain constant across trials, which, whatever the environment, remains the default and 'irrefutable' hypothesis in the absence of feedbacks or when feedbacks remain consistent with posterior beliefs. Moreover, when external contingencies change and generate feedbacks inconsistent with these beliefs, Weber's computational imprecisions enable posterior beliefs to rapidly detach from priors and capture new contingencies. The Weber-imprecision model, however, is not an algorithmic approximation of higher-order inference processes tracking volatility. The model performances indeed maximally differed in environments when either the former or the latter corresponded to adaptive optima. Conceptually, the Weber-imprecision model optimizes short-term predictability by rapidly inferring locally accurate but globally inaccurate simplified world models (stable contingencies), while computational imprecisions enable rapid switching across such simplified world models. By contrast, higher-order inferential processes optimize

long-term predictability by slowly inferring more general, accurate but complex world models (volatile contingencies), detrimentally to short-term effectiveness.

The Weber-imprecision model involves online, forward, first-order inferences about external contingencies that can be expressed in closed form. We nonetheless chose to implement these inferences through an online particle filter based on forward, iterated importance sampling[23,24] for four reasons: (1) this particle filter realizes the exact first-order inference; (2) it constitutes the forward inference component in particle filtering emulating higher-order inferential processes[22], which allowed us to properly compare the models; (3) previous studies show this particle filter is plausibly implemented in populations of cortical neurons[26–28] and (4) computational imprecisions are naturally modelled in online particle filtering as sampling noise that scales with the extension of resampling occurring between successive trials. As this particle filter emulates the exact first-order inference, this implementation choice has no impact on the results. The same results would be obtained with other exact implementations, such as evidence accumulation models and their neural implementations[40,41], undergoing computational imprecisions consistent with Weber's law. We thus remain agnostic about possible neural implementations underlying the Weber-imprecision model.

In any case, we assumed that as previously proposed[19], computational imprecisions consistent with Weber's law stem from intrinsic neural noise corrupting the inferential process. Indeed, we modelled computational imprecisions as a sampling noise that arises randomly and uniformly over a range that scales with the magnitude of belief updating. An algorithmic-equivalent model assuming no neural imprecisions would compute and monitor this magnitude online to deterministically scale the sampling noise accordingly. We found, however, that the Weber-imprecision model clearly fit the human data reported here better than this algorithmic-equivalent model ($P_{\text{exceed}} > 0.99$ for both closed and open environments, obtained using the same methods as in Results). This confirms that the variability of human performances relative to first-order inferences reflects imprecise neural computations rather than additional monitoring processes. In canonical decision-making tasks featuring stable environments, such imprecise neural computations may act as a nuisance. For instance, a previous study[16] showed that these imprecisions accounts for two-thirds of human suboptimal choices in standard tasks requiring participants to infer, from a series of ambiguous cues, which of several latent states is the most likely. The present results, however, show that these imprecisions endow humans with powerful adaptive abilities in variable environments, approaching the optimal but biologically implausible adaptive processes.

In all inferential models considered here, we assumed first-order inferences about external contingencies to apply to the whole set of latent states. This is certainly an unrealistic assumption about human inferences in open-ended environments such as those investigated here ($K = 24$ potential latent states). This may explain why in the open-ended environment, but not in the closed one, participants performed below the theoretical minimaxloss computed for the Weber-imprecision model (Computer simulation results). Using the same behavioural data, we previously showed that participants indeed make first-order inferences relating to only a small subset of three or four latent states referred to as the 'inferential buffer'[5,7]. We proposed a model, named Probe, describing how this inferential buffer, through these first-order inferences, is updated with new latent states and drives adaptive behaviour[5,7]. This model, however, treats the environmental volatility as a constant free parameter that modulates first-order inferences by treating them as deriving from higher-order inferences. The present results therefore predict that replacing this free parameter in the inferential buffer with computational imprecisions consistent with Weber's law should provide an even better account of observed human performances. We tested this prediction (Methods) and found that this new model did indeed provide a decidedly better fit to human data than the original Probe model ($P_{\text{exceed}} > 0.99$). The Weber-imprecision model thus accounts for human adaptive performances without requiring first-order inferences to apply to all the latent states of the environment.

Neuroimaging studies show that brain regions exhibit activations correlating with the volatility derived from higher-order inferences[1,42], consistent with the view that higher-order inferences tracking volatility might guide human adaptive behaviour. Our results, however, suggest that these brain activations might alternatively reflect the computational imprecisions corrupting first-order beliefs about external contingencies. Indeed, in the Weber-imprecision model, computational imprecisions vary with the environmental volatility and make these beliefs directly contingent upon the environmental volatility. The Weber-imprecision model thus makes the prediction that these brain activations would, rather, reflect the processing of these imprecise beliefs guiding human behaviour. This prediction is fully testable given that, as mentioned above, computational imprecisions in the Weber-imprecision model further exhibit stochastic fluctuations independent of the volatility derived from higher-order inferences.

Additionally, previous results suggest that people's metacognitive judgments reflect both the encoded uncertainty and the coding imprecisions in internal representations[43]. Because in the Weber-imprecision model the environmental volatility is only implicitly included as computational imprecisions corrupting the beliefs about external contingencies, rather than explicitly represented, we speculate that people's metacognitive judgments would be unable to disentangle their subjective uncertainty about external contingencies and the environmental volatility.

Previous results have shown that imprecise neural computations in human inferential processes about external contingencies contribute importantly to human choice suboptimality in canonical decision-making tasks[16]. The present results show that these imprecisions in first-order inferences actually allow humans to dispense with developing higher-order inferences about the environmental volatility to reach near-optimal adaptive behaviour in volatile environments. Thus, imprecise neural computations may have been preserved throughout the evolution of computationally frugal inferential processes as efficiently contributing to near-optimal adaptive behaviour in real-life environments.

Methods

Our experiments involving human participants comply with all the French ethical regulations regarding human experiments. All participants volunteered to participate to the study and provided written informed consent. The present study was approved by the French National Ethics Committee (CPP, Inserm protocol #C15-98). Participants were paid for their participation. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[1,5,7,11–14,36,37].

Adaptive behaviour paradigm. We considered an adaptive agent responding to successively presented stimuli. In every trial, one among N distinct stimuli was randomly drawn uniformly and the agent responded by selecting one among M actions. The agent then received a positive or negative feedback. The agent thus searched for the correct responses to stimuli by trial and error. However, feedbacks were stochastic and the combination of correct responses to stimuli changed episodically. More precisely, the environment episodically switched across distinct latent states defining the current correct combination: every latent state specified one distinctive response to each stimulus (the correct response) that led to positive feedbacks with unknown, constant probability $\eta > 0.5$, while the other responses led to positive feedbacks with probability $1-\eta < 0.5$. The maximal number K of potential latent states (or correct combinations) was thus equal to $M!/(M-N)!$. Latent states changed between two successive trials with probability τ named volatility. We simulated the six environments described below.

Closed and stable environments. These environments correspond to $K = 2$ and volatility $\tau = 0$, namely, a two-armed bandit with no reversals. A unique stimulus ($N = 1$) was repeated over trials and one of two available actions ($M = 2$) had to

be selected in every trial. To make such environments non-trivial and investigate the role of feedback sparsity on adaptive behaviour, we simulated eight closed and stable environments each delivering feedbacks in 100, 50, 20, 10, 5 or 2% of trials. Feedback probability η was set to 90%.

Closed and changing environments. These environments were identical to the preceding closed and stable environments except that volatility τ was non-zero and constant. This corresponds to a two-armed bandit with reversals between two latent states. We simulated the rarely and frequently changing environments, respectively, as corresponding to constant volatility $\tau = 0.03$ and $\tau = 0.2$. As such environments are non-trivial, feedbacks were delivered in every trial (feedback sparsity was set to zero).

Closed and unstable environments. These environments were identical to the preceding closed and changing environments except that volatility τ_t varied across trials as a bounded Gaussian random walk within the range [0.03, 0.2] and with variance $\nu = 0.0001$. In every trial t, volatility τ_t is drawn from a normal distribution centered on volatility τ_{t-1} with variance ν. Whenever a value smaller than 0.03 or larger than 0.2 was drawn, volatility τ_t was set to 0.03 or 0.2, respectively (Extended Data Fig. 1).

Open-ended and stable environments. These environments were identical to the closed and stable environments described above except that number K of potential latent states was set to $K = 24$. In every trial, one of three stimuli ($N = 3$) was drawn randomly and uniformly, and one of four available actions ($M = 4$) had to be selected, which leads to $K = 24$ potential combinations. The correct combination remained unchanged across trials. Again, we simulated eight open-ended and stable environments each delivering feedbacks in 100, 50, 20, 10, 5 or 2% of trials.

Open-ended and changing environments. These environments were identical to the preceding open-ended and stable environments except that volatility τ was non-zero and constant. As above, we considered rarely and frequently changing environments corresponding to constant volatility $\tau = 0.03$ and $\tau = 0.2$, respectively. As such environments are non-trivial, feedbacks were delivered in every trial (feedback sparsity set to zero). When latent states changed, the new latent state k is drawn according to a multinomial distribution with probabilities $\gamma^1, \ldots, \gamma^k, \ldots, \gamma^K$ (excluding the preceding latent state). These probabilities were initially drawn randomly and uniformly according to a flat Dirichlet distribution (in order that $\sum_{i=1,\ldots,K} \gamma^i = 1$).

Open-ended and unstable environments. These environments were identical to the preceding open-ended and changing environments except that volatility τ_t varied across trials as a bounded Gaussian random walk within the range [0.03, 0.2] and with variance $\nu = 0.0001$, as described above for closed unstable environments.

Computational models. *Exact varying-volatility model.* The exact varying-volatility model corresponds to the optimal Bayesian adaptive process with a generative model exactly matching the generative process of unstable environments described above with K equal to either 2 (closed environments) or 24 (open environments) (Fig. 1a and Extended Data Fig. 3). This model is therefore the theoretical adaptive optimum in these environments. More precisely, the model assumes the following.

1. Volatility τ_t varies as a bounded Gaussian random walk within the range [0, 0.5] with uninformative priors about constant variance ν:

$$\nu_{\text{prior}} \sim \text{Inverse} - \text{Gamma}(3, 0.001)$$

2. The current latent state changes between trials $t-1$ and t with probability τ_t.
3. When the current latent state changes between trial $t-1$ and t, new latent state z_t is drawn from the set of potential latent states $\{1, \ldots, K\}$ according to a multinomial distribution with probabilities $\gamma = \{\gamma^1, \ldots, \gamma^K\}$:

$$z_t | z_t \neq z_{t-1} \sim \text{Multinomial}(\gamma)$$

with $\sum_{k=1,\ldots,K} \gamma^k = 1$ and excluding the preceding state.

4. Priors about probabilities γ are uninformative and follow a flat Dirichlet distribution:

$$\gamma_{\text{prior}} \sim \text{Dirichlet}(1, \ldots, 1)$$

5. Given current latent state z_t, stimulus s_t and chosen action a_t, feedbacks r_t are delivered according to a Bernoulli distribution with parameters η or $1-\eta$:

$$r_t | s_t, a_t, z_t \sim \text{Bernoulli}(\eta) \text{if } a_t \text{ is the correct response}$$

$$r_t | s_t, a_t, z_t \sim \text{Bernoulli}(1-\eta) \text{if } a_t \text{ is an incorrect response}$$

6. Priors about feedback noise $\eta > 0.5$ are uninformative and follow a flat Beta distribution:

$$\eta/2 \sim \text{Beta}(1, 1).$$

By marginalizing over posterior beliefs about latent states z_t, the model then selects in every trial the action that is most likely to lead to the positive feedback in response to the stimulus (that is, the correct action): namely, actions are selected according to an 'argmax' policy. In this model, however, computing posterior beliefs about latent states z_t and τ_t and latent parameters ν, γ and η is an intractable problem. We addressed this issue by using a sequential Monte Carlo (SMC) algorithm recently developed in machine learning to solve this class of inferential models comprising both latent states and parameters[22]. The algorithm is based on particle filtering methods and converges to the exact solution when the number of sampling particles increases to infinity[22]. The algorithm comprises two intermixed SMC procedures: (1) a particle filter[23] implementing iterated importance sampling in the space of latent states (z_t,τ_t) and (2) an iterated importance sampling combined with a particle Markov chain Monte Carlo method[25,44] in the space of parameters ν, γ and η. We implemented the algorithm using a total number of 1×10^6 particles, corresponding to 1,000 samples in the parameter space, each associated with 1,000 particles in the space of latent spaces. We verified that this number allows approaching the asymptotic convergence: we implemented the algorithm using 4×10^6 particles and obtained virtually identical posterior beliefs (Supplementary Fig. 5).

Exact constant-volatility model. The exact constant-volatility model corresponds to the optimal Bayesian adaptive process with a generative model exactly matching the generative process of changing environments described above with K equal to either 2 (closed environments) or 24 (open environments) (Fig. 1b and Extended Data Fig. 4). This model is therefore the theoretical adaptive optimum in these environments. More precisely, the exact constant-volatility model is identical to the exact varying-volatility model except that volatility τ_t is now assumed to be a constant ($\tau_t = \tau$ for all t) in the range [0, 0.5]. Priors about volatility τ were uniform over [0, 0.5] and followed a flat beta distribution:

$$2\tau \sim \text{Beta}(1, 1).$$

The exact constant-volatility model was implemented using the same SMC algorithm as the exact varying-volatility model described above.

Forward varying- and constant-volatility models. As explained above, the exact varying- and constant-volatility models both perform exact inference through the particle filtering method (when the number of particles is large enough). However, the algorithm is both computationally very costly and biologically implausible, because it constantly requires offline backward passes to resample and revise past beliefs about latent parameters according to all past observations. Resampling and revisions of beliefs are performed in a backward fashion through the particle Markov chain Monte Carlo procedure to prevent the sampling of beliefs from degenerating in local minima and to more accurately sample posterior beliefs. To overcome this limitation, we derived an online, biologically plausible, computationally frugal approximate algorithm from the exact SMC algorithm described above. This forward approximation replaces the offline, backward particle Markov chain Monte Carlo procedure with the following standard, online particle resampling of posterior beliefs: resampling occurs with respect to the current empirical estimates of parameters' distributions given parameter priors. For the varying-volatility model, accordingly, the resampling of posterior beliefs about latent parameters ν, γ and η occurs in trial t through the inverse-gamma, Dirichlet and beta distributions, respectively, and the hyper-parameters are estimated from the particle filter in trial t (Supplementary Methods). For the constant-volatility model, similarly, the resampling of posterior beliefs about latent parameters τ, γ and η occurs in trial t through the beta, Dirichlet and beta distributions, respectively, and the hyper-parameters are again estimated from the particle filter in trial t. In both cases, the resulting online particle filtering is biologically plausible and neural implementations have been previously proposed[26–28]. We referred to these approximations as the forward varying- and constant-volatility models. To make these forward models as computationally frugal as possible, we implemented them with the minimal number of particles that allows approaching the asymptotic performance in the most complex environment (that is, unstable environments): namely, $200 \times 200 = 40{,}000$ particles (Supplementary Fig. 5). Overall, the results show that these forward models provide accurate approximations of exact models, provided that the environment and model generative process are identical (Figs. 2 and 3).

Zero-volatility and noisy-selection model. The zero-volatility model corresponds to the optimal Bayesian adaptive process with a generative model exactly matching the generative process of stable environments described above, with K equal to either 2 (closed environments) or 24 (open environments) (Fig. 1c and Extended Data Fig. 5). This model is therefore the theoretical adaptive optimum in these environments. More precisely, the zero-volatility model is identical to the exact constant-volatility model except that volatility τ is now assumed to be zero. Model priors thus reduce to only combination probabilities γ and feedback noise η as described above. Posterior beliefs in this model are computable in closed form[29]. For consistency, however, we implemented this model through the same forward particle filtering method as that used for the forward varying- and constant-volatility models, knowing that in this case, the forward approximation

emulates the exact inference process. The noisy-selection model is the variant of the zero-volatility model assuming that actions are selected according to a softmax rather than argmax policy. For the sake of generality, we applied the softmax rule to the logarithm of beliefs about correct actions (Supplementary Note 2). The noisy-selection model thus has the softmax inverse temperature β as free parameter and comprises the zero-volatility model as a special case ($\beta \gg 1$).

Weber-imprecision model. The Weber-imprecision model is another variant of the zero-volatility model assuming that imprecisions stemming from online neural noise occur in computing posterior beliefs about latent states. Neural noise is naturally introduced in the zero-volatility model by implementing a noisy particle filter: every particle coding for one latent state (that is, combinations) in trial t may start (mis)coding for another latent state in trial $t+1$ with probability ϵ_t. This newly (mis)encoded state is randomly drawn according to the actual distribution of particles over latent states, which reflects probabilities γ. Consistently with Weber's law, computational imprecisions are further assumed to scale with the distance d_t between posterior beliefs in trial t and $t+1$. Probabilistic noise ϵ_t is thus itself a random variable uniformly distributed between 0 and $\mu+\lambda d_t$:

$$\epsilon_t \sim U(0, \mu + \lambda d_t),$$

where μ and λ are two free parameters quantifying the constant and Weber components of computational imprecisions, respectively. Supplementary methods show how distance d_t derives from particle filtering. Note that for $\mu=0$ and $\lambda=0$, the Weber-imprecision model simply reduces to the zero-volatility model.

Noisy-PH-RL model. This RL model combines the Pearce–Hall learning rule[30] (α and α_{PH} as free parameters) and Gaussian white noise in value updating scaling with reward prediction error (ζ as free parameter) with a softmax policy for action selection (inverse temperature β as free parameter). In response to successive stimuli s_t, accordingly, the softmax policy selects the action a_t according to action values $Q(s_t,a)$, which are updated following feedbacks r_t with respect to the following noisy updating rule:

$$\begin{aligned} Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + [\alpha + \alpha_{PH}|r_t - Q_t(s_t, a_t)|](r_t - Q_t(s_t, a_t)) \\ + N(0, \zeta)|r_t - Q_t(s_t, a_t)|, \end{aligned}$$

while $Q_t(s,a)$ remain unchanged for $s \neq s_t$ or $a \neq a_t$. $N(0,\zeta)$ denotes the zero-centered Gaussian distribution with standard deviation ζ.

Experimental procedures. *Closed environments.* Participants. We recruited 22 healthy participants (11 women and 11 men aged 18–35 years), who volunteered to participate in the study. Participants provided written informed consent. The present study was approved by the French National Ethics Committee (CPP, Inserm protocol #C15-98). Participants were paid for their participation. One participant was excluded because their overall performance was lower than two standard deviations from the mean performance.

Behavioural protocol. Each participant completed the two-armed bandit task described in Fig. 4a,b. Stimuli were presented on a computer screen and participants responded using two response buttons. The task included 720 trials broken down into six blocks of 120 trials corresponding to six distinct levels of external volatility. Stimuli remained on the screen until participants responded (self-paced task). 500 ms later, the positive or negative feedback (green happy versus red angry smiley faces) was displayed during 600 ms. One response led to the positive and negative feedback with 80% and 20% probabilities, respectively. The other responses led to positive and negative feedback with the conserve probabilities (that is, $\eta=80\%$). The inter-trial interval was jittered between 700 and 1,300 ms. Participants had short breaks, were informed that breaks had no influence on ongoing task contingencies and had to keep in mind their beliefs about these contingencies during breaks (that is, breaks occurred between reversals and within blocks of constant volatility). Participants were informed that their final pay-offs increased with the number of positive feedbacks received.

Open-ended environments. We analysed the behavioural performances of participants from two previously published studies that administered the same behavioural protocol with and without functional magnetic resonance imaging[5,7]. The behavioural protocol corresponded to the open-ended, rarely changing environment described in the main text with constant volatility $\tau=0.03$, number of distinct stimuli $N=3$, actions $M=4$ and possible combinations $K=24$ (Fig. 6a). Participants volunteered and provided written informed consent. These studies were approved by the French National Ethics Committee (CPP, Inserm protocol #C15-98). Participants were paid for their participation.

First study. 22 healthy volunteers (aged 18–35 years old, 13 of whom were women) participated in the study[5]. No participants were excluded. Stimuli were visually presented Arabic numbers. Participants responded to each stimulus by pressing one of four response keys. When key presses occurred no later than 1,500 ms after stimulus onsets, stimuli disappeared 100 ms after key presses and participants received audio-visual feedbacks (duration 300 ms). Feedbacks were positive or negative. A positive feedback consisted of an ascending sound and the apparition of the associated stimulus in a box representing the pressed key at the bottom of the screen. Negative feedbacks consisted of only a descending sound. Otherwise, stimuli were removed and no feedbacks were delivered. Stimulus onset asynchrony was 2,000 ms. Associations between actual stimuli, response fingers and feedbacks were orthorgonalized and counterbalanced across participants. Participants were instructed that feedbacks could be uncertain and variable and that pay-offs increased with the total number of positive feedbacks received. No additional instructions were provided to participants.

In every trial, a correct response was associated with each stimulus (three possible stimuli) and led to positive feedbacks with 90% probability. All other responses led to negative feedbacks with 90% probability. Distinct stimuli were associated with distinct correct responses. Digit-response combinations remained unchanged over a series of successive trials uniformly ranging from 36 to 54 trials and corresponding to an external volatility $\tau=0.03$. Two successive combinations were orthogonal, that is, all digit-response pairs differed between successive combinations.

The experiment included two behavioural sessions administered on two separate days. In each session, combinations changed 24 times. In the no-recurrence session, combinations never recurred. In the recurrence session, only three distinct combinations occurred in a pseudo-randomized order and in equal proportion. These combinations were fully orthogonal. Stimuli were pseudo-randomly chosen from the set {1,3,5} (one session) or {2,4,6} (the other session). Stimuli along with combination and session order were counterbalanced across participants. Sessions were administered according to a double-blinded procedure. Volatility of external contingencies was identical in the no-recurrence and recurrence sessions ($\tau=0.03$).

Second study. Forty healthy, right-handed volunteers (aged 18–26 years old, 20 of whom were women) participated in the fMRI study[7]. No participants were excluded. The experimental protocol was identical to the one described above, except the following differences. Positive feedbacks corresponded to stimuli presented in green in the box representing the chosen button. Negative feedbacks corresponded to a red cross shown in the box representing the chosen button. When no answer was made, red lines were shown in the four boxes. Pseudo-randomized time jittering was used between responses and feedbacks (from 400 to 3,900 ms) as well as between feedback offsets and trial onsets (from 100 to 3,600 ms). Feedbacks duration was 1,000 ms. (Fig. 6a).

Model fitting and comparison. In both closed and open-ended environments, we replaced the argmax with a softmax decision policy (inverse temperature β as the free parameter) in all models to fit models' to participants' performances. For all inferential models and the sake of generality, we applied the softmax rule to the logarithm of beliefs about correct actions (marginalized over latent states) (Supplementary Note 2). Thus, the exact/forward varying- and constant-volatility models included this unique free parameter β. The Weber-imprecision model included three free parameters, the softmax inverse temperature β along with noise parameters μ and λ. As a result, the Weber-imprecision model contained the zero-volatility and noisy-selection model as special cases. Finally, the noisy-PH-RL model included four free parameters: the softmax inverse temperature β, the noise parameter ζ and the learning rates α and α_{PH}. Model fits to human data were compared according to MPPs, that is, the model marginal likelihoods (given the data) with uniform priors over models. MPPs based on marginal likelihoods are the optimal Bayesian quantification for comparing models, balancing model degree of freedom (or complexity) and adequacy to data. Model marginal likelihoods are obtained by marginalizing model likelihoods over the whole free parameter space. This marginalization was practically tractable for closed environments and was carried out through standard importance sampling and quasi-Monte Carlo methods[45]. For open-ended environments, however, it was practically intractable (especially for the Weber-imprecision model with three free parameters) and we used the standard approximation of model marginal likelihoods, that is, the Bayesian information criterion. Computing Bayesian information criteria requires estimating maximal model likelihoods over the free parameter space, which was carried out through standard Bayesian optimization methods[46]. All the analyses were based on equal priors across models.

In any case, computing model likelihoods given free parameter values is required before marginalizing or maximizing these likelihoods. As the inferential models are particle filters, computing the likelihoods for these models may not be trivial and is explained below.

Exact varying- and constant-volatility models. For these models, computing model likelihoods given human data is straightforward. Indeed, for each model, the particle filter emulates the corresponding exact inferential process so that model posterior beliefs and, consequently, the model softmax probability of selecting the actual human choices are unrelated to particle filtering. Thus, given inverse temperature β, each model directly provides its likelihood given human data.

Forward varying- and constant-volatility models. For each of these models, the particle filter approximates and may diverge from the corresponding exact inferential process. As a result, model posterior beliefs and, consequently, the model softmax probability of selecting the actual human choices depend upon the

actual realization of the particle filter. Given inverse temperature β, computing the model likelihood given human data then requires marginalizing over the realizations of particle filtering. We solved this marginalization problem by noting that as a particle filter, each forward model actually defines a hidden Markov chain generating actions. We therefore computed the marginalization using established sequential Monte Carlo methods for hidden Markov models[24]. Note that this marginalization procedure again optimally balances the adequacy to data and the additional degree of freedom resulting from the various possible realizations of particle filtering.

Weber-imprecision model. For this model, the particle filter reflects the exact zero-volatility inferential process, but the presence of computational imprecisions makes the Weber-imprecision model diverge from the exact process. As a result, model posterior beliefs and, consequently, the model softmax probability of selecting the observed human choices again depend upon the actual realization of the noisy particle filter implementing the Weber-imprecision model. We solved this issue exactly as described above for the forward varying- and constant-volatility models. Given inverse temperature β and noise components μ and λ, we thus derived the model likelihood given human data. Note that, again, the derivation is a marginalization procedure and therefore optimally balances the adequacy to data and the additional degree of freedom resulting from the various possible realizations of the noisy particle filter.

Model recovery procedure. We implemented a model recovery procedure to assess the validity of our model fitting and selection procedure and the ability of our experimental protocols to discriminate the models[35]. The recovery procedure consists in generating synthetic data by simulating every model of interest, then applying our model fitting and selection procedure (see above) to these data. This procedure should lead to the selection of the simulated model. For each model, we generated 21 sets (62 sets) of synthetic data in the closed (open-ended) environment, each corresponding to the free parameters fitted to one participant. The results show that the recovery procedure was conclusive for both the closed and open-ended environments, thereby validating our model fitting and selection procedure along with the protocol's ability to discriminate the models.

Probe model. The Probe model is an online algorithm approximating optimal adaptive processes (that is, Dirichlet processes mixture) in environments featuring a potentially infinite number of latent states[5,7]. Probe accounts for human adaptive behaviour in open-ended, changing environments as those investigated here[5,7]. Probe assumes that first-order inferences only apply to a small subset of three or four latent states referred to as the inferential buffer[5,7] and describes how this inferential buffer is updated with new latent states through these first-order inferences and drives adaptive behaviour. The model, however, treats external volatility as a free parameter that modulates first-order inferences by treating them as deriving from higher-order inferences, which allows the expression of the inferential process in closed form. To overcome this limitation, we investigated a new model by replacing this free parameter with Weber imprecisions affecting first-order inferential computations within the inferential buffer. As Probe is expressed in closed form, these imprecisions were modelled as follows:

$$\tilde{B}_t \sim \text{Dirichlet}\left(\text{mean} = B_t, \text{concentration} = \frac{1}{\mu + \lambda d_t}\right)$$

where B_t are posterior beliefs about latent states updated according to the original Probe model with volatility parameter set to zero ($B_t = (B_t^i)_{i\in\text{buffer}}$) and $\tilde{B}_t$ are the corresponding imprecise beliefs resulting from Weber imprecisions. As in the main text, $d_t = |B_t - \tilde{B}_{t-1}|$ measures the magnitude of belief updating, while μ and λ are free parameters reflecting the constant and Weber components of neural noise. The Dirichlet distribution appropriately indicates that $\tilde{B}_t$ are randomly distributed around mean B_t with variance scaling with $\mu + \lambda d_t$. We fit and compared the original and new Probe models to human performances recorded in the open-ended environment investigated in Results, using the same model fitting and comparison method as for the Weber-imprecision model (Model fitting and comparison). MPPs revealed that the new Probe model unambiguously better accounts for human performances (exceedance probability $P_{\text{exceed}} > 0.99$).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All the data that support the findings of the present study are available from the corresponding author upon request.

Code availability

All program codes are freely available at https://github.com/csmfindling/learning_variability_and_volatility

Received: 31 October 2019; Accepted: 18 September 2020;
Published online: 9 November 2020

References

1. Behrens, T. E., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
2. Boorman, E. D., Behrens, T. E., Woolrich, M. W. & Rushworth, M. F. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* **62**, 733–743 (2009).
3. Gershman, S. J., Blei, D. M. & Niv, Y. Context learning, and extinction. *Psychological Rev.* **117**, 1997–1209 (2010).
4. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to grow a mind: statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011).
5. Collins, A. G. & Koechlin, E. Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol.* **10**, e1001293 (2012).
6. Collins, A. G. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. Rev.* **120**, 190–229 (2013).
7. Donoso, M., Collins, A. G. & Koechlin, E. Foundations of human reasoning in the prefrontal cortex. *Science* **344**, 1481–1486 (2014).
8. Kolossa, A., Kopp, B. & Fingscheidt, T. A computational analysis of the neural bases of Bayesian inference. *Neuroimage* **106**, 222–237 (2015).
9. Schuck, N. W., Cai, M. B., Wilson, R. C. & Niv, Y. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron* **91**, 1402–1412 (2016).
10. Rouault, M., Drugowitsch, J. & Koechlin, E. Prefrontal mechanisms combining rewards and beliefs in human decision-making. *Nat. Commun.* **10**, 301 (2019).
11. Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* **30**, 12366–12378 (2010).
12. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput. Biol.* **7**, e1001048 (2011).
13. Wilson, R. C., Nassar, M. R. & Gold, J. I. A mixture of delta-rules approximation to bayesian inference in change-point problems. *PLoS Comput. Biol.* **9**, e1003150 (2013).
14. McGuire, J. T., Nassar, M. R., Gold, J. I. & Kable, J. W. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870–881 (2014).
15. Bossaerts, P., Yadav, N. & Murawski, C. Uncertainty and computational complexity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20180138 (2019).
16. Drugowitsch, J., Wyart, V., Devauchelle, A. D. & Koechlin, E. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron* **92**, 1398–1411 (2016).
17. Fechner G. T. *Elemente der Psychophysik* (Breitkopf and Härtel, 1860).
18. Treisman, M. Noise and Weber's law: the discrimination of brightness and other dimensions. *Psychol. Rev.* **71**, 314–330 (1964).
19. Deco, G., Scarano, L. & Soto-Faraco, S. Weber's law in decision making: integrating behavioral data in humans with a neurophysiological model. *J. Neurosci.* **27**, 11192–11200 (2007).
20. Wyart, V. & Koechlin, E. Choice variability and suboptimality in uncertain environments. *Curr. Opin. Behav. Sci.* **11**, 109–115 (2016).
21. Faraji, M., Preuschoff, K. & Gerstner, W. Balancing new against old information: the role of puzzlement surprise in learning. *Neural Comput.* **30**, 34–83 (2018).
22. Chopin, N., Jacob, P. E. & Papaspiliopoulos, O. SMC2: an efficient algorithm for sequential analysis of state space models. *J. R. Stat. Soc. Series B Stat. Methodol.* **75**, 397–426 (2013).
23. Doucet, A., Godsill, S. & Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.* **10**, 197–208 (2000).
24. Andrieu, C., Doucet, A. & Holenstein, R. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Series B Stat. Methodol.* **72**, 269–342 (2010).
25. Chopin, N. A sequential particle filter for static models. *Biometrika* **89**, 539–552 (2002).
26. Shi, L. & Griffiths, T. L. Neural implementation of hierarchical Bayesian inference by importance sampling. *Adv. Neural Inf. Process. Syst.* **22**, 1669–1677 (2009).
27. Huang, Y. & Rao, R. P. Neurons as Monte Carlo samplers: Bayesian inference and learning in spiking networks. *Adv. Neural Inf. Process. Syst.* **27**, 1943–1951 (2014).
28. Legenstein, R. & Maass, W. Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. *PLoS Comput. Biol.* **10**, e1003859 (2014).
29. Scott, S. L. Bayesian methods for hidden Markov models. *J. Am. Stat. Assoc.* **97**, 337–351 (2002).
30. Pearce, J. M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
31. Roesch, M., Esber, G. R., Li, J., Daw, N. & Schoenbaum, G. Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. *Eur. J. Neurosci.* **35**, 1190–1200 (2012).

32. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
33. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - revisited. *Neuroimage* **84**, 971–985 (2014).
34. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
35. Palminteri, S., Wyart, V. & Koechlin, E. The importance of falsification in computational cognitive modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).
36. Payzan-LeNestour, E. *Bayesian Learning in Unstable Settings: Experimental Evidence Based on the Bandit Problem* Research Paper No. 10-28 (Swiss Finance Inst., 2010).
37. Summerfield, C., Behrens, T. E. & Koechlin, E. Perceptual classification in a rapidly changing environment. *Neuron* **71**, 725–736 (2011).
38. Knight, F. H. *Risk, Uncertainty and Profit* (Univ. Chicago Press, 1921).
39. Keynes, J. M. *A Treatise on Probability* (Macmillan, 1921).
40. Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
41. Wang, X.-J. Neural dynamics and circuit mechnaisms of decision-making. *Curr. Opin. Neurobiol.* **22**, 1039–1046 (2012).
42. Payzan-LeNestour, E., Dunne, S., Bossaerts, P. & O'Doherty, J. P. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79**, 191–201 (2013).
43. Lebreton, M., Abitbol, R., Daunizeau, J. & Pessiglione, M. Automatic integration of confidence in the brain valuation signal. *Nat. Neurosci.* **18**, 1159–1167 (2015).
44. Beaumont, M. A. Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–1160 (2003).
45. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods* (Society for Industrial and Applied Mathematics, 1992).
46. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. Preprint at *arXiv* http://arxiv.org/abs/1206.2944 (2012).

Acknowledgements

We thank J. Drevet for her help in collecting human data. Supported by a European Research Council Grant (ERC-2009-AdG #250106) to E.K. and a DGA-INSERM PhD fellowship to C.F. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

E.K. and C.F. conceived the study and designed the models. C.F. and N.C. developed the models. C.F. programmed the models, performed computer simulations and collected human data. E.K. and C.F. analysed human and simulation data. E.K. and C.F. wrote the paper.

Competing interests

The authors declare no competing interests

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-00971-z.

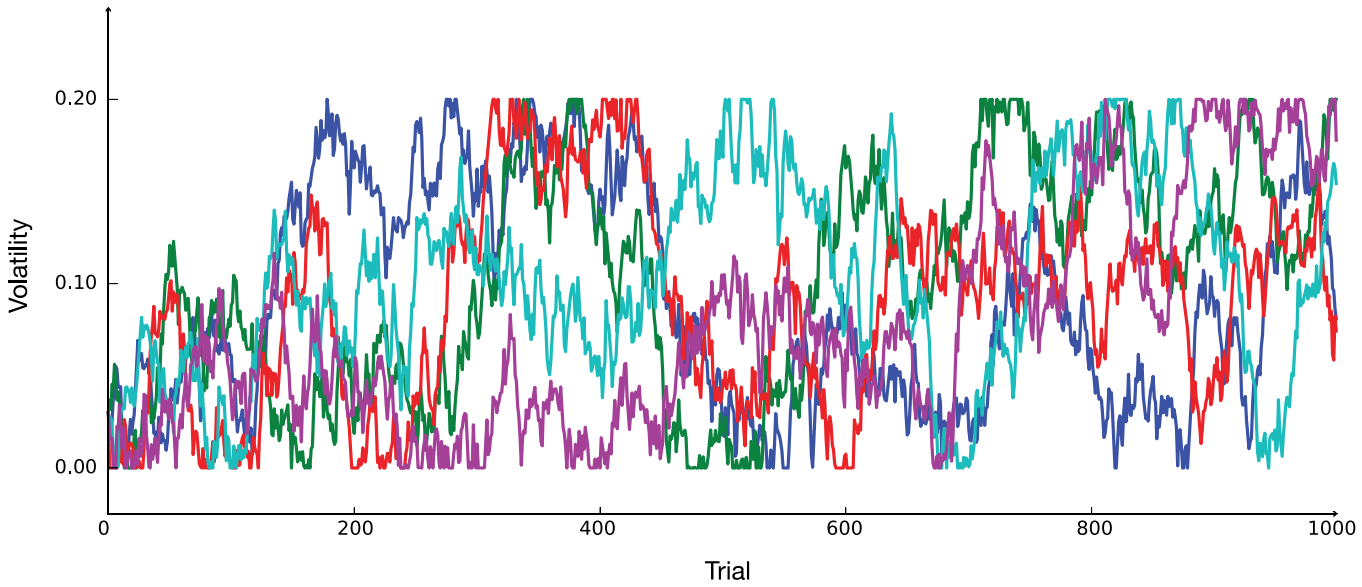
Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00971-z.

Correspondence and requests for materials should be addressed to E.K.

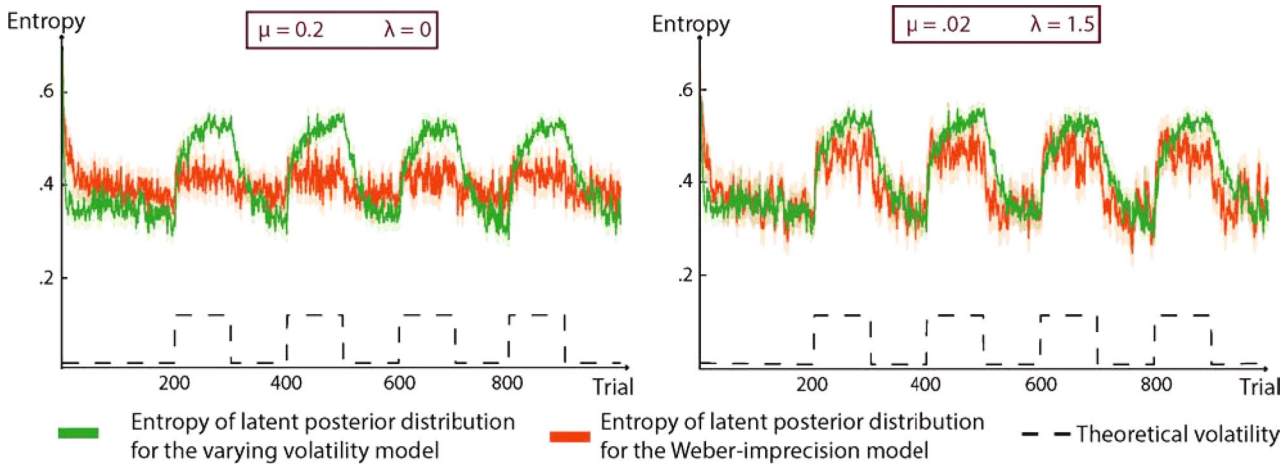
Peer review information Primary Handling Editor: Marike Schiffer

Reprints and permissions information is available at www.nature.com/reprints.

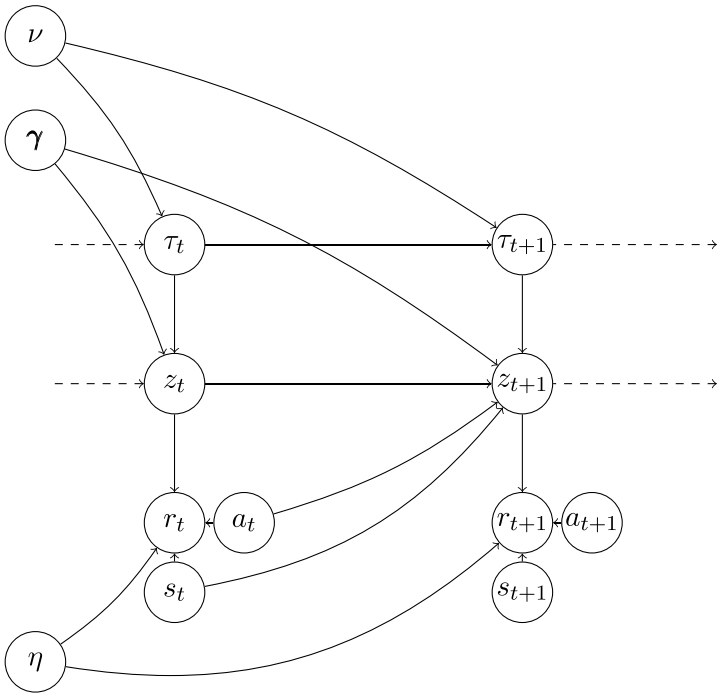
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



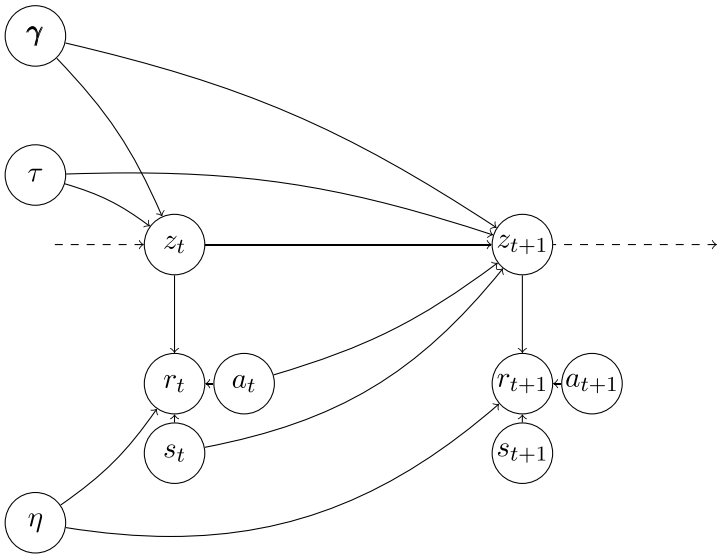
Extended Data Fig. 1 | Examples of five volatility trajectories in unstable environments. Environment volatility follows a bounded gaussian random walk between 0.03 and 0.2 with variance 0.0001 (see Methods).



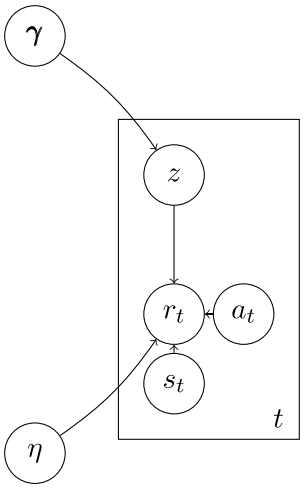
Extended Data Fig. 2 | Relation between the Weber noise component λ in the Weber imprecision model and external volatility. The figure shows the entropy of posterior beliefs about current combinations (latent state posteriors) for the exact varying-volatility and Weber imprecision model. Each model is simulated $N = 50$ times in a closed environment (K=2, two-armed bandit), which alternates between high and low-volatility periods. Left, simulations when Weber component λ is set to 0 and constant component μ is large ($\mu = 0.2$). Right, simulations when Weber component λ is large ($\lambda = 1.5$) and constant component μ is low ($\mu = 0.02$). Note that the entropies of posterior beliefs are similar between the Weber-imprecision and varying-volatility model only when the Weber component is large enough.



Extended Data Fig. 3 | Full generative model of varying-volatility models. This model is exactly the process generating unstable environments. This generative model assumes that volatility τ_t follows a bounded random walk with constant variance ν. z_t represents the current correct combination. γ represents the probabilities of combination occurrence whenever the correct combination changes. η represents feedback noise. In every trial, observables are stimuli s_t, actions a_t and binary feedback r_t. See Methods for details.



Extended Data Fig. 4 | Full generative model of constant-volatility models. This model is exactly the process generating changing environments. This generative model assumes that volatility τ is constant. z_t represents the current correct combination. γ represents the probabilities of combination occurrence whenever the correct combination changes. η represents feedback noise. In every trial, observables are stimuli s_t, actions a_t and binary feedback r_t. See Methods for details.



Extended Data Fig. 5 | Full generative model of zero-volatility models. This model is exactly the process generating stable environments. This generative model assumes that volatility is null and that observations are all equally informative. z represents the correct combination. γ represents combinations' probabilities. η represents feedback noise. In every trial, observables are stimuli s_t, actions a_t and binary feedback r_t. See Methods for details.

Volatility Models	***Softmax coefficients***	
	Closed environments	Open-ended environments
Exact Varying Volatility	1.61 (0.13)	1.13 (0.021)
Exact Constant Volatility	1.34 (0.12)	0.94 (0.017)
Forward Varying Volatility	1.63 (0.11)	1.27 (0.024)
Forward Constant Volatility	1.38 (0.11)	1.07 (0.02)

Extended Data Fig. 6 | Means (s.e.m) of parameters fitted across participants for exact and forward volatility models.

Weber-imprecision model parameters	Closed environments	Open-ended environments
Softmax coefficient β	1.77 (0.13)	1.05 (0.023)
Weber Component λ	0.8 (0.1)	1.22 (0.06)
Constant Component μ	0.05 (0.01)	0.05 (0.003)

Extended Data Fig. 7 | Means (s.e.m.) of parameters fitted across participants for the Weber-imprecision model.

Models	Parameters	Closed environment	Open-ended environment
RL			
	Softmax coefficient β	5.0 (0.46)	3.1 (0.05)
	Learning rate α	0.61 (0.02)	0.82 (0.007)
Noisy-RL		*(best-fitting RL model)**	
	Softmax coefficient β	10.3 (1.08)	9.75 (0.74)
	Learning rate α	0.50 (0.03)	0.60 (0.02)
	Learning noise ζ	0.18 (0.02)	0.35 (0.01)
Pearce-Hall-RL			
	Softmax coefficient β	4.7 (0.2)	3.16 (0.06)
	Learning rate α	0.39 (0.03)	0.77 (0.02)
	PH learning rate α_{PH}	0.23 (0.04)	0.06 (0.02)
Noisy-Pearce-Hall-RL			*(best-fitting RL model)**
	Softmax coefficient β	11.7 (1.1)	37.0 (2.2)
	Learning rate α	0.30 (0.03)	0.07 (0.007)
	PH learning rate α_{PH}	0.20 (0.03)	0.50 (0.01)
	Learning noise ζ	0.18 (0.02)	0.38 (0.008)

* See Supplementary Fig. 2.

Extended Data Fig. 8 | Means (s.e.m.) of parameters fitted across participants for Reinforcement Learning models.

nature research

Corresponding author(s): Etienne Koechlin

Last updated by author(s): Sep 17, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Matlab Psychtoolbox
Data analysis	Matlab statistical toolbox. Custom codes in Language C++ were used to run the model and fit the model to human behavioral data. All these codes are freely available at https://github.com/csmfindling/learning_variability_and_volatility, as stated in the paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability statement:
All behavioral data reported in the paper are available from the authors upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental along with model computer simulations and model fitting to experimental data.
Research sample	Sample: healthy young adult volunteers from Paris area, where the study was carried out (N total=84 participants). Aged range: 18-35 years-old. 44 females, 40 males. This is the standard practice in Cognitive Psychology. The sample is representative of healthy young adults from this area volunteering to participate to behavioral experiments.
Sampling strategy	Random sampling with sample sizes determined before carrying out the experiments. Sample sizes were determined according to the standard of field : Experiment 1: N=22 (11 females, age range: 18-35 y/o); Experiment 2: N=22 (13 females, age range: 18-35 y/o); Experiment 3: N=40 (20 females; age range: 18-35 y/o). Experiment 3 is a replication of experiment 2.
Data collection	Data were collected through computers. Noboby was present during the experiments and the persons in charge of collecting the data were blind of experimental conditions and study hypotheses.
Timing	Experiment 1: Nov to Dec 2017; Experiment 2: Jan to Feb 2010; Experiment 3: April to July 2012.
Data exclusions	Only one participant was excluded in Experiment 1, because her/is performance was beyond 2 standard deviations from the mean (standard threshold for rejecting outliers), as stated in the paper.
Non-participation	No participants declined to participate to the study.
Randomization	no distinct experimental groups were compared in the present study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	Participants were recruited through regular announcements and postings for volunteering. To minimize self-selection biases, these announcements and postings mentioned no specific features of the experiments. Participants were informed about the experimental procedures only before providing their fully informed consent. No participants declined their participation
Ethics oversight	The protocol was approved by the Comité National de Protection des Personnes (CPP) and the Institut National de la Santé et de la Recherche Médicale (biomedical protocol CPP-INSERM #15-98). This information is provided in the paper.

Note that full information on the approval of the study protocol must also be provided in the manuscript.