

Supplementary Information for

Learning Optimal Decisions with Confidence

Jan Drugowitsch, André G. Mendonça, Zachary F. Mainen and Alexandre Pouget

Corresponding Author Name.

E-mail: jan_drugowitsch@hms.harvard.edu

This PDF file includes:

Supplementary text
Figs. S1 to S2
References for SI reference citations

Supporting Information Text

We here provide a self-consistent and extended derivation of all the results provided in the main text. We will use standard font for scalars, lower-case bold symbols for vectors, and upper-case bold symbols for matrices. All vectors are column vectors, unless transposed. $x \sim \mathcal{N}(\mu, \sigma^2)$ denotes that x is a normal random variable with mean μ and variance σ^2 .

Optimal decision making with high-dimensional momentary evidence

One-dimensional momentary evidence. Within each individual trial, we assume the latent state μ to be drawn from $\mu \sim \mathcal{N}(m, \sigma_\mu^2)$. $m = 0$ corresponds to the case of an unbiased prior for which $p(\mu \geq 0) = p(\mu < 0)$. In each small time step n of size δt from trial-onset at $t = 0$ (i.e., $n = 1$), the decision maker observes the momentary evidence $\delta z_n | \mu \sim \mathcal{N}(\mu \delta t, \delta t)$ that provides noisy information about the value of μ . The decision-maker's aim is to infer the sign of μ from the sequence $\delta z_1, \delta z_2, \dots$ of momentary evidence, to make choice $y = 1$ (for $\mu \geq 0$) or $y = -1$ (for $\mu < 0$).

For Bayes-optimal choices, we find the posterior $p(\mu \geq 0 | \delta z_1, \delta z_2, \dots)$ in two steps. First, for N pieces (i.e., $t = N\delta t$ seconds) of accumulated evidence, the posterior μ is give by Bayes' rule,

$$p(\mu | \delta z_{1:N}) \propto p(\mu) \prod_{n=1}^N p(\delta z_n | \mu) \propto e^{-\frac{\mu^2}{2} \left(\frac{1}{\sigma_\mu^2} \right) + \mu \left(\frac{m}{\sigma_\mu^2} + z \right)} \propto \mathcal{N} \left(\mu \left| \frac{\sigma_\mu^{-2} m + z}{\sigma_\mu^{-2} + t}, \frac{1}{\sigma_\mu^{-2} + t} \right. \right), \quad [1]$$

where all proportionalities are with respect to μ , and the second proportionality results from substituting the respective normal distributions, and defining $t = \sum_n \delta t$ and $z(t) = \sum_n \delta z_n$. This shows that the sufficient statistics of the posterior are $z(t)$ and t . For the second step we integrate this posterior over the non-negative half-line to find

$$p(\mu \geq 0 | z, t) = \int_0^\infty p(\mu | z, t) d\mu = \Phi \left(\frac{\sigma_\mu^{-2} m + z}{\sqrt{\sigma_\mu^{-2} + t}} \right), \quad [2]$$

where $\Phi(\cdot)$ is the normal cumulative function. This posterior is more certain (i.e., closer to zero or one) for larger $|\sigma_\mu^{-2} m + z|$ and smaller times t .

Using the correspondence between $\mu \geq 0$ (and $\mu < 0$) and $y = 1$ (and $y = -1$), the fact that $1 - \Phi(a) = \Phi(-a)$, and $p(\mu < 0 | z, t) = 1 - p(\mu \geq 0 | z, t)$, the more generic posterior over y is given by

$$p(y | z, t) = \Phi \left(y \frac{\sigma_\mu^{-2} m + z}{\sqrt{\sigma_\mu^{-2} + t}} \right). \quad [3]$$

This posterior captures both $y = 1$ and $y = -1$. If y is the made decision, then the expression is the belief that this decision was correct, and hence the *decision confidence* (1).

So far we have assumed a prior over μ with arbitrary mean m . With this prior, the a-priori belief that $y = 1$ is correct is given by $P^+ \equiv p(\mu \geq 0) = \Phi(m/\sigma_\mu)$. The prior is thus unbiased for $m = 0$, in which case $P^+ = 1/2$. In this case, the posterior Eq. [3] prefers $y = 1$ for all $z > 0$ and $y = -1$ for all $z < 0$. Therefore, we can bound evidence accumulation from above and below by the (potentially time-dependent) $\pm\theta(t)$ to make Bayes-optimal choices. In particular, once z reaches $\theta(t)$ (or $-\theta(t)$), it would trigger choice $y = 1$ (or $y = -1$). Observing that the unbounded accumulated evidence follows a Wiener process with drift μ , that is, $z(t) | \mu \sim \mathcal{N}(\mu t, t)$, supports the use of drift-diffusion models for Bayes-optimal decision making. Biased priors, which we discuss in a later section, require additional attention to achieve Bayes-optimal choices.

High-dimensional momentary evidence. To move to J -dimensional momentary evidence $\delta \mathbf{x}$ while preserving parallels to the one-dimensional case, we assume that there exist some (for now, known) combination weights \mathbf{w} such that $\delta z_n = \mathbf{w}^T \delta \mathbf{x}_n$. We achieve this by the generative model,

$$\delta \mathbf{x} | \mu \sim \mathcal{N}((\mathbf{a}\mu + \mathbf{b}) \delta t, \mathbf{\Sigma} \delta t), \quad [4]$$

for vectors \mathbf{a} and \mathbf{b} that satisfy $\mathbf{a}^T \mathbf{w} = 1$ and $\mathbf{b}^T \mathbf{w} = 0$, and a covariance matrix $\mathbf{\Sigma}$ for which $\mathbf{w}^T \mathbf{\Sigma} \mathbf{w} = 1$. With these properties it becomes easy to show that $\mathbf{w}^T \delta \mathbf{x} | \mu \sim \mathcal{N}(\mu \delta t, \delta t)$, as required. We will discuss our specific choices for \mathbf{a} , \mathbf{b} , and $\mathbf{\Sigma}$ for the simulations shown in the main text further below.

Using the same steps as before, we find the posterior μ given N steps (i.e., $t = N\delta t$ seconds) of momentary evidence to be given by

$$p(\mu | \delta \mathbf{x}_{1:N}, \mathbf{w}) = \mathcal{N} \left(\mu \left| \frac{\sigma_\mu^{-2} m + \mathbf{w}^T \mathbf{x}}{\sigma_\mu^{-2} + t}, \frac{1}{\sigma_\mu^{-2} + t} \right. \right), \quad [5]$$

where we have defined the accumulated evidence $\mathbf{x}(t) = \sum_n \delta \mathbf{x}_n$. The posterior over $\mu \geq 0$ is correspondingly given by

$$p(\mu \geq 0 | \mathbf{x}, t, \mathbf{w}) = \Phi \left(\frac{\sigma_\mu^{-2} m + \mathbf{w}^T \mathbf{x}}{\sqrt{\sigma_\mu^{-2} + t}} \right). \quad [6]$$

Both expressions differ from the one-dimensional case by replacing z by $\mathbf{w}^T \mathbf{x}$. Expressed as a posterior over y , the above turns into

$$p(y|\mathbf{x}, t, \mathbf{w}) = \Phi\left(y \frac{\sigma_\mu^{-2} m + \mathbf{w}^T \mathbf{x}}{\sqrt{\sigma_\mu^{-2} + t}}\right). \quad [7]$$

If y is the made decision, the above is again the decision confidence. Note that the unbounded accumulated evidence follows the multi-dimensional drifting Wiener process, $\mathbf{x}(t)|\mu \sim \mathcal{N}((\mathbf{a}\mu + \mathbf{b})t, \mathbf{\Sigma}t)$, whose \mathbf{w} -weighted linear combination reduces to the same one-dimensional process $z(t) = \mathbf{w}^T \mathbf{x}(t) \sim \mathcal{N}(\mu t, t)$ as before.

Assuming again an unbiased prior, $m = 0$, Bayes-optimal decisions are by the same logic as for the one-dimensional case cast by the boundaries $\pm\theta(t)$ on $\mathbf{w}^T \mathbf{x}$. Here, the positive (negative) boundary correspond to choice $y = 1$ ($y = -1$). We will discuss Bayes-optimal choices for biased priors in a later section.

If difficulty $|\mu|$ varies across trials, the decision confidence at a constant decision boundary drops over time. As the previous sections have shown, the decision confidence is the same for one- and high-dimensional momentary evidence as long as the decision boundary is on $z(t)$ and $\mathbf{w}^T \mathbf{x}(t)$, respectively. Furthermore, for time-invariant decision boundaries, $\pm\theta(t) = \pm\theta$, this decision confidence drops as a function of time. Here we show that this drop is a general property of symmetric priors over μ for which the difficulty $|\mu|$ can vary across trials, that extends beyond the Gaussian $p(\mu)$ we assume in other parts of this supplement. To show this, let us redefine $p(\mu)$ — in this section only — to be (as in (2)) given by

$$p(\mu) = \sum_{i=1}^L \frac{p_i}{2} (\delta(\mu - \mu_i) + \delta(\mu + \mu_i)), \quad [8]$$

which features L point masses at $\pm\mu_1, \pm\mu_2, \dots, \pm\mu_L$, each weighted by $p_i/2$, and where we have assumed positive p_i that satisfy $\sum_i p_i = 1$. Furthermore, without loss of generality, we assume the μ_i 's to be positive, ordered, and unique, that is $0 < \mu_1 < \mu_2 < \dots < \mu_L$. Here, we disallow $\mu_1 = 0$ for notational convenience, but our argument can be easily extended to include this possibility. Assuming the same one-dimensional momentary evidence as before, $\delta z_n | \mu \sim \mathcal{N}(\mu \delta t, \delta t)$, it follows from Bayes' rule that

$$p(\mu = \mu_i | z, t) = \frac{p_i e^{-\frac{t}{2}\mu_i^2 + z\mu_i}}{\sum_j p_j e^{-\frac{t}{2}\mu_j^2} (e^{z\mu_j} + e^{-z\mu_j})}. \quad [9]$$

Therefore, the belief that $\mu \geq 0$ ($y = 1$) at the upper boundary $z = \theta$ is given by

$$p(y = 1 | z = \theta, t) = \sum_i p(\mu = \mu_i | z = \theta, t) = \frac{\sum_i p_i e^{-\frac{t}{2}\mu_i^2 + \theta\mu_i}}{\sum_j p_j e^{-\frac{t}{2}\mu_j^2} (e^{\theta\mu_j} + e^{-\theta\mu_j})}. \quad [10]$$

For our symmetric prior, this belief at the upper boundary equals the decision confidence at both boundaries. Therefore, we will use it as a proxy for decision confidence.

In what follows, we will show that this belief is a mixture of two components. The first is the belief that $\mu = \mu_i$ given some fixed difficulty, $\mu \in \{-\mu_i, \mu_i\}$, and the second is the probability that this is indeed the current difficulty. The first part turns out to be independent of time, whereas the second changes. In particular, as we will show, the probability that the difficulty is high (i.e., that $|\mu|$ is small) increases over time, resulting in a re-weighting of the per-difficulty beliefs. This re-weighting causes the overall belief to drop, as we argue in the main text.

Mathematically, this mixture can be written as

$$p(y = 1 | z = \theta, t) = \sum_i p(\mu = \mu_i | z = \theta, t) = \sum_i p(\mu = \mu_i | z = \theta, t, \mu = \pm\mu_i) p(\mu = \pm\mu_i | z = \theta, t). \quad [11]$$

In the right-most sum, the first probability is the per-difficulty belief g_i for assumed difficulty μ_i , and the second is the probability that μ_i is indeed the current difficulty. Both follow from [9], and are given by

$$g_i \equiv p(\mu = \mu_i | z = \theta, t, \mu = \pm\mu_i) = \frac{e^{\theta\mu_i}}{e^{\theta\mu_i} + e^{-\theta\mu_i}} = \frac{1}{1 + e^{-2\theta\mu_i}}, \quad [12]$$

$$p(\mu = \pm\mu_i | z = \theta, t) = \frac{p_i e^{-\frac{t}{2}\mu_i^2} (e^{\theta\mu_i} + e^{-\theta\mu_i})}{\sum_j p_j e^{-\frac{t}{2}\mu_j^2} (e^{\theta\mu_j} + e^{-\theta\mu_j})} = \frac{w_i(t)}{\sum_j w_j(t)}, \quad [13]$$

where we have defined $w_i(t) = p_i e^{-\frac{t}{2}\mu_i^2} (e^{\theta\mu_i} + e^{-\theta\mu_i})$ as the unnormalized, time-dependent, per-difficulty weights. This allows us to write the overall belief as the weighted mixture

$$p(y = 1 | z = \theta, t) = \sum_i \frac{w_i(t)}{\sum_j w_j(t)} g_i, \quad [14]$$

which is a weighted mixture of per-difficulty beliefs, g_i , in which only the mixture weights are time-dependent. Note that the per-difficulty beliefs are strictly increasing in μ_i , such that are also ordered, that is, $g_1 < g_2 < \dots < g_L$. Furthermore,

increasing time by $\delta t > 0$ results in a drop in unnormalized weights, $w_i(t + \delta t) = a_i(\delta t)w_i(t)$ with $a_i(\delta t) = e^{-\frac{\delta t}{2}\mu_i^2} \in [0, 1]$. This drop is larger for larger μ_i , that is $a_1(\delta t) > a_2(\delta t) > \dots > a_L(\delta t)$. Therefore, increasing time results in putting proportionally more weight on per-difficulty beliefs associated with lower μ_i 's with lower associated g_i , such that the overall belief, and equally the decision confidence, drops.

Once we stop varying the difficulty (i.e., $L = 1$), the belief reduces to the per-difficulty belief g_1 , which does not drop over time. Therefore, the only condition required for the decision confidence to drop at a time-invariant boundary is for the difficulty $|\mu|$ to vary across trials.

Learning the input combination weight w from choice feedback

So far we have assumed w to be known. Here we derive learning rules for w based on feedback on the correctness of a choice. Specifically, we assume that the decision maker accumulated evidence \mathbf{x} for some time t and (potentially, but not necessarily) made decision y , after which feedback about the correct choice y^* is provided. Before evidence accumulation we assume the decision maker to hold belief $p(w)$ about the input combination weights w . Our aim is to find the posterior $p(w|\mathbf{x}, t, y^*)$ given all the available evidence. We focus here on a feedback after a single choice. The same principles apply to choice sequences, by turning the posterior after a choice into the prior for the subsequent choice.

The desired posterior can be found by Bayes' rule

$$p(w|\mathbf{x}, t, y^*) \propto p(y^*|\mathbf{x}, t, w)p(w), \quad [15]$$

where the likelihood $p(y^*|\mathbf{x}, t, w)$ is conditional on all observed quantities, \mathbf{x} and t , and some hypothetical weights w , and specifies the probability that y^* is the correct choice given these weights. This likelihood turns out to correspond to the previously derived decision-making posterior, Eq. [7], which is a normal cumulative function with argument linear in w . In general, problems with such a likelihood function structure are known as *Probit regression*. Such problems don't yield solutions for which the posterior has the same functional form as the prior — which is a desirable property to support efficient Bayesian input weight learning across longer sequences of choices, and to gain insight into the learning rule. Therefore, we derive below different approximations to such Bayes-optimal learning.

All of the below assumes an unbiased prior over μ by fixing m to $m = 0$. We can extend the below rules to also learn the prior bias m by extending the accumulated evidence vector \mathbf{x} by one element fixed to σ_μ^{-2} , and the weight vector w by one element containing m . Learning this extended weight vector then correspond to simultaneously learning the input weight and the prior bias.

The marginal decision confidence. Before deriving approximate weight learning rules, let us consider the consequences of uncertain weights on the decision confidence $p(y|\mathbf{x}, t)$ with these weights marginalized out. To do so, we assume our prior weight belief to be normal, $w \sim \mathcal{N}(\mu_w, \Sigma_w)$ with mean μ_w and covariance Σ_w . Then, we find this marginal decision confidence by first finding the marginal posterior over μ , which is given by

$$p(\mu|\mathbf{x}, t) = \int p(\mu|\mathbf{x}, t, w)p(w)dw = \mathcal{N}\left(\frac{\mu_w^T \mathbf{x}}{\sigma_\mu^{-2} + t}, \frac{1}{\sigma_\mu^{-2} + t} + \frac{\mathbf{x}^T \Sigma_w \mathbf{x}}{(\sigma_\mu^{-2} + t)^2}\right), \quad [16]$$

where we have used Eq. [5] with $m = 0$. We find the marginal decision confidence $p(y|\mathbf{x}, t)$ by integrating the above over the non-negative halfline, which results after some simplification in

$$p(y|\mathbf{x}, t) = \Phi\left(y \frac{\mu_w^T \mathbf{x}}{\sqrt{\sigma_\mu^{-2} + t + \mathbf{x}^T \Sigma_w \mathbf{x}}}\right) = \Phi\left(y \frac{\mu_w^T \tilde{\mathbf{x}}}{\sqrt{1 + \tilde{\mathbf{x}}^T \Sigma_w \tilde{\mathbf{x}}}}\right), \quad [17]$$

where we have defined $\tilde{\mathbf{x}} \equiv \mathbf{x}/\sqrt{\sigma_\mu^{-2} + t}$ for the second equality. Comparing this expression to Eq. [7] reveals the additional term $\mathbf{x}^T \Sigma_w \mathbf{x}$ that lowers the overall posterior confidence (i.e., moving it towards 1/2) due to uncertainty in w . If y is the made choice, the above is the decision confidence that takes into account weight uncertainty.

Weight learning by Assumed Density Filtering. Assumed Density Filtering (ADF; (3–6)) approximates the posterior by assuming a particular functional form of the approximate posterior $q(w|\mathbf{x}, t, y^*)$ and finding the parameters of this functional form by minimizing the Kullback-Leiber divergence $\text{KL}(p(w|\mathbf{x}, t, y^*) || q(w|\mathbf{x}, t, y^*))$ between the true posterior and its approximation. To minimize this divergence we again assume a normally distributed prior $w \sim \mathcal{N}(\mu_w, \Sigma_w)$ with mean μ_w and covariance Σ_w . To support sequential choices, we assume the approximate posterior to also be normal, $q(w|\mathbf{x}, t, y^*) = \mathcal{N}(w|\mu_w^*, \Sigma_w^*)$, with updated moments μ_w^* and Σ_w^* .

To find these updated moments, we use the fact that the KL-divergence is in our case minimized by matching the moments of the Gaussian sufficient statistics w and $w w^T$ (7). For the source distribution, $p(w|\mathbf{x}, t, y^*)$, these moments can be found by the gradients of the log-normalizing constant of this source distribution, $\nabla \log p(y^*|\mathbf{x}, t)$ (7, 8), where we will use the already derived marginal likelihood $p(y|\mathbf{x}, t)$ in Eq. [17]. Using these principles, the updated moments of the approximate posterior can be found by

$$\mu_w^* = \mu_w + \Sigma_w \nabla_{\mu_w} \log p(y^*|\mathbf{x}, t), \quad [18]$$

$$\Sigma_w^* = \Sigma_w - \Sigma_w \left(\nabla_{\mu_w} \log p(y^*|\mathbf{x}, t) \left(\nabla_{\mu_w} \log p(y^*|\mathbf{x}, t) \right)^T - 2 \nabla_{\Sigma_w} \log p(y^*|\mathbf{x}, t) \right) \Sigma_w. \quad [19]$$

The required gradients are given by

$$\nabla_{\mu_w} \log p(y^* | \mathbf{x}, t) = \xi_w(\gamma) y^* \frac{\tilde{\mathbf{x}}}{\sqrt{1 + \tilde{\mathbf{x}}^T \Sigma_w \tilde{\mathbf{x}}}}, \quad [20]$$

$$\nabla_{\Sigma_w} \log p(y^* | \mathbf{x}, t) = -\xi_w(\gamma) y^* \frac{\boldsymbol{\mu}_w^T \tilde{\mathbf{x}}}{2(1 + \tilde{\mathbf{x}}^T \Sigma_w \tilde{\mathbf{x}})^{3/2}} \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T, \quad [21]$$

$$[22]$$

where $\xi_w(\gamma)$ is given by

$$\xi_w(\gamma) = \frac{\partial}{\partial \gamma} \log \Phi(\gamma) = \frac{\mathcal{N}(\gamma | 0, 1)}{\Phi(\gamma)}, \quad [23]$$

and we have defined γ as

$$\gamma \equiv y^* \frac{\boldsymbol{\mu}_w^T \tilde{\mathbf{x}}}{\sqrt{1 + \tilde{\mathbf{x}}^T \Sigma_w \tilde{\mathbf{x}}}}. \quad [24]$$

Overall, this leads to the moments update equations,

$$\boldsymbol{\mu}_w^* = \boldsymbol{\mu}_w + y^* \frac{\xi_w(\gamma)}{\sqrt{1 + \tilde{\mathbf{x}}^T \Sigma_w \tilde{\mathbf{x}}}} \Sigma_w \tilde{\mathbf{x}}, \quad [25]$$

$$\Sigma_w^* = \Sigma_w + \xi_{cov}(\gamma) \left((\Sigma_w^{-1} + \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T)^{-1} - \Sigma_w \right), \quad [26]$$

where the covariance learning rate is given by

$$\xi_{cov}(\gamma) = \xi_w(\gamma)^2 + \xi_w(\gamma) \gamma. \quad [27]$$

As illustrated in Fig. S1, both mean and covariance updates are modulated by the marginal decision confidence in the feedback, y^* , given by Eq. [17]. To see how ξ_w and ξ_{cov} are a function of the marginal decision confidence about the actual choice y (rather than the feedback y^*) let us first focus on correct choices. For correct choices, $y = y^*$, such that the marginal decision confidence about y equals that of y^* , that is, $p(y | \mathbf{x}, t) = p(y^* | \mathbf{x}, t)$. Furthermore, by the definition of $p(y^* | \mathbf{x}, t)$, it can be written as $p(y^* | \mathbf{x}, t) = \Phi(\gamma)$ (where γ is defined in Eq. [24]). This function is strictly increasing in γ , such that small/large γ 's corresponds to low/high confidence. Therefore, as $\xi_w(\gamma)$ and $\xi_{cov}(\gamma)$ are functions of only γ , they are in turn functions of the marginal decision confidence $p(y | \mathbf{x}, t)$.

For incorrect choices we have $y \neq y^*$, such that $p(y | \mathbf{x}, t) = 1 - p(y^* | \mathbf{x}, t) = 1 - \Phi(\gamma)$, which is strictly decreasing in γ . Therefore, we can again assign a unique decision confidence $p(y | \mathbf{x}, t)$ to each γ , such that $\xi_w(\gamma)$ and $\xi_{cov}(\gamma)$ are again functions of the decision confidence about the made decision y .

Assumed Density Filtering with a diagonal covariance matrix. The above update equations require tracking of the full covariance matrix, making these updates scale badly with the size of the input space, J , and require non-local interactions. To find alternative, local update equations, we here assume that both the prior covariance, as well as the approximate posterior covariance are given by diagonal matrices, given by $\Sigma_w = \text{diag}(\sigma_{w,1}^2, \dots, \sigma_{w,k}^2)$ and $\Sigma_w^* = \text{diag}(\sigma_{w,1}^{2*}, \dots, \sigma_{w,k}^{2*})$. Following the same derivation as before, this leads to the update equations

$$\mu_{w,i}^* = \mu_{w,i} + y^* \frac{\xi_w(\gamma)}{\sqrt{1 + \sum_j \sigma_{w,j}^2 \tilde{x}_j^2}} \sigma_{w,i}^2 \tilde{x}_i, \quad [28]$$

$$\sigma_{w,i}^{2*} = \sigma_{w,i}^2 - \xi_{cov}(\gamma) \frac{\sigma_{w,i}^4 \tilde{x}_i^2}{\sqrt{1 + \sum_j \sigma_{w,j}^2 \tilde{x}_j^2}} \quad [29]$$

where $\mu_{w,i}$ and $\mu_{w,i}^*$ are the i th element of $\boldsymbol{\mu}_w$ and $\boldsymbol{\mu}_w^*$, respectively, and γ is given by

$$\gamma = y^* \frac{\boldsymbol{\mu}_w^T \tilde{\mathbf{x}}}{\sqrt{1 + \sum_j \sigma_{w,j}^2 \tilde{x}_j^2}}. \quad [30]$$

Thus, other than a global divisive normalization and the marginal decision confidence-related term γ , all updates are local.

Approximating the weight posterior by a second-order Taylor series. A simpler alternative to ADF that also yields a normally distributed approximate posterior is to approximate the true log-posterior, $\log p(\mathbf{w}|\mathbf{x}, t, y^*)$ by a second-order Taylor series in \mathbf{w} around $\mathbf{w} = \boldsymbol{\mu}_w$. The relevant terms in this log-posterior are

$$\log p(\mathbf{w}|\mathbf{x}, t, y^*) = \Phi(y^* \mathbf{w}^T \tilde{\mathbf{x}}) - \frac{1}{2} \mathbf{w}^T \boldsymbol{\Sigma}_w^{-1} \mathbf{w} + \mathbf{w}^T \boldsymbol{\Sigma}_w^{-1} \boldsymbol{\mu}_w + \text{const.} \quad [31]$$

The required gradient and Hessian are

$$\nabla_w \log p(\mathbf{w}|\mathbf{x}, t, y^*) \Big|_{\mathbf{w}=\boldsymbol{\mu}_w} = \xi_w(\gamma) y^* \tilde{\mathbf{x}}, \quad [32]$$

$$\nabla \nabla_w \log p(\mathbf{w}|\mathbf{x}, t, y^*) \Big|_{\mathbf{w}=\boldsymbol{\mu}_w} = -\boldsymbol{\Sigma}_w^{-1} - \xi_{cov}(\gamma) \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T, \quad [33]$$

where $\xi_w(\cdot)$ and $\xi_{cov}(\cdot)$ are defined as for ADF, but γ changes to $\gamma = y^* \boldsymbol{\mu}_w^T \tilde{\mathbf{x}}$. Using the above to find the second-order Taylor series and reading off the resulting posterior moments yields the moment updates

$$\boldsymbol{\mu}_w^* = \boldsymbol{\mu}_w + y^* \xi_w(\gamma) \boldsymbol{\Sigma}_w^* \tilde{\mathbf{x}}, \quad [34]$$

$$\boldsymbol{\Sigma}_w^* = \left(\boldsymbol{\Sigma}_w^{-1} + \xi_{cov}(\gamma) \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T \right)^{-1}. \quad [35]$$

These have a similar form as for ADF, Eqs. [25] and [26], with the main difference that they are missing the normalization by $\sqrt{1 + \tilde{\mathbf{x}}^T \boldsymbol{\Sigma}_w \tilde{\mathbf{x}}}$. Given that this normalization modulates the moment update strength by the weight uncertainty, this implies that the update equations based on the second-order Taylor series will be less influenced by this uncertainty.

Assumed density filtering with noisy feedback. So far we have assumed the feedback y^* to always be correct. We will now consider how ADF changes when the feedback itself is noisy. In particular, we assume that feedback is inverted with probability β , such that the weight likelihood given feedback y^* becomes

$$p(y^*|\mathbf{x}, t, \mathbf{w}) = \beta \Phi(-y^* \mathbf{w}^T \tilde{\mathbf{x}}) + (1 - \beta) \Phi(y^* \mathbf{w}^T \tilde{\mathbf{x}}) = \beta + (1 - 2\beta) \Phi(y^* \mathbf{w}^T \tilde{\mathbf{x}}). \quad [36]$$

In this case, the marginal decision confidence about feedback y^* becomes

$$p(y^*|\mathbf{x}, t, \beta) = \beta + (1 - 2\beta) \Phi\left(y^* \frac{\boldsymbol{\mu}_w^T \tilde{\mathbf{x}}}{\sqrt{1 + \tilde{\mathbf{x}}^T \boldsymbol{\Sigma}_w \tilde{\mathbf{x}}}}\right). \quad [37]$$

The gradients of the log marginal decision confidence thus become

$$\nabla_{\boldsymbol{\mu}_w} \log p(y^*|\mathbf{x}, t, \beta) = \xi_{\beta,w}(\gamma) y^* \frac{\tilde{\mathbf{x}}}{\sqrt{1 + \tilde{\mathbf{x}}^T \boldsymbol{\Sigma}_w \tilde{\mathbf{x}}}}, \quad [38]$$

$$\nabla_{\boldsymbol{\Sigma}_w} \log p(y^*|\mathbf{x}, t, \beta) = -\xi_{\beta,w}(\gamma) y^* \frac{\boldsymbol{\mu}_w^T \tilde{\mathbf{x}}}{2(1 + \tilde{\mathbf{x}}^T \boldsymbol{\Sigma}_w \tilde{\mathbf{x}})^{3/2}} \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T \quad [39]$$

with $\xi_{\beta,w}(\gamma)$ given by

$$\xi_{\beta,w}(\gamma) = \frac{\partial}{\partial \gamma} \log(\beta + (1 - 2\beta) \Phi(\gamma)) = \frac{(1 - 2\beta) \mathcal{N}(\gamma|0, 1)}{\beta + (1 - 2\beta) \Phi(\gamma)} \quad [40]$$

and where γ is, as for vanilla ADF, given by Eq. [24]. Using again Eqs. [18] and [19] results in the update equations

$$\boldsymbol{\mu}_w^* = \boldsymbol{\mu}_w + y^* \frac{\xi_{\beta,w}(\gamma)}{\sqrt{1 + \tilde{\mathbf{x}}^T \boldsymbol{\Sigma}_w \tilde{\mathbf{x}}}} \boldsymbol{\Sigma}_w \tilde{\mathbf{x}}, \quad [41]$$

$$\boldsymbol{\Sigma}_w^* = \boldsymbol{\Sigma}_w + \xi_{\beta,cov}(\gamma) \left(\left(\boldsymbol{\Sigma}_w^{-1} + \tilde{\mathbf{x}} \tilde{\mathbf{x}}^T \right)^{-1} - \boldsymbol{\Sigma}_w \right), \quad [42]$$

with covariance learning rate

$$\xi_{\beta,cov}(\gamma) = \xi_{\beta,w}(\gamma)^2 + \xi_{\beta,w}(\gamma) \gamma. \quad [43]$$

This illustrates that the only impact of noisy feedback is on the update strength modulators, $\xi_{\beta,w}(\cdot)$ and $\xi_{\beta,cov}$. As shown in Fig. S1, these modulators become smaller for larger feedback noise. For high-confidence choices that the feedback flags as incorrect, $\xi_{\beta,cov}$ even becomes negative, indicating that uncertainty in \mathbf{w} increases. This increase arises from approximate inference, as additional information in strictly Bayes-optimal inference should not increase uncertainty, even if this information is (knowingly) noisy.

Alternative learning heuristics. Let us now discuss alternative heuristics that do not track a belief over \mathbf{w} , but instead update a point estimate. The first alternative is the delta rule that performs stochastic gradient descent on the sum of squared distances between the chosen decision boundary and the correct decision boundary. For the current choice, this squared distance is $(\mathbf{w}^T \mathbf{x}(t) - y^* \theta(t))^2$, where $\mathbf{w}^T \mathbf{x}(t) \in \{-\theta(t), \theta(t)\}$ at decision time t equals the chosen bound, and $y^* \theta(t) \in \{-\theta(t), \theta(t)\}$ is the boundary that would have led to the correct choice. Thus, the delta rule update is given by

$$\mathbf{w}^* = \mathbf{w} + \frac{\alpha}{2\theta(0)} (y^* \theta(t) - \mathbf{w}^T \mathbf{x}) \mathbf{x}, \quad [44]$$

where we have chosen to normalize the learning α by $2\theta(0)$ to make the update magnitude less dependent of the bound height. The residual in the above is either zero or $\pm 2\theta(t)$, such that the learning rule only makes adjustments to the weight estimate in case of incorrect choices.

The delta rule aims to minimize the probability that incorrect choices are made. In diffusion models this can be achieved by accumulating more evidence before reaching the decision boundary. This, in turn, can be accomplished by reducing the overall magnitude of \mathbf{w} . In particular for small learning rates, this is exactly what the delta rule does, leading to progressively smaller $\|\mathbf{w}\|$, and weight learning that does not converge in expectation. To work around this degeneracy, we introduced the normalized delta rule. This rule performs the update exactly like the standard delta rule, but subsequently adjusts the weight magnitude to match that of the true weights. It therefore needs access to the true weight's magnitude in each trial, making it a rule that has access to an oracle that other rules don't. Thus, it uses strictly more information than other rules, which needs to be kept in mind when comparing its performance to that of other rules.

As a last heuristic we considered performing stochastic gradient ascent on the log-likelihood of the feedback, $\log p(y^* | \mathbf{x}, \mathbf{w}, t) = \log \Phi(y^* \mathbf{w}^T \tilde{\mathbf{x}})$. Taking the gradient of this log-likelihood results in the learning rule

$$\mathbf{w}^* = \mathbf{w} - \alpha y^* \xi_w (y^* \mathbf{w}^T \tilde{\mathbf{x}}) \tilde{\mathbf{x}}, \quad [45]$$

where $\xi_w(\cdot)$ is defined as for ADF. Due to the inclusion of $\xi_w(\cdot)$, this rule modulates the update strength by decision confidence, unlike the normalized delta rule above. It differs from probabilistic learning rules in that it uses a fixed learning rate α , instead of a learning rate modulation by a current estimate of the certainty about \mathbf{w} .

Tracking non-stationary combination weights

So far we have assumed the true weights, underlying the generation of the momentary evidences, $\delta \mathbf{x}$, to be stationary, allowing us to use a sequence \mathbf{x} 's, t 's, and y^* 's to learn successively better posteriors over \mathbf{w} . In the ideal case (i.e., if we wouldn't use approximate inference), this would — after enough observations — lead to a very good approximation of the true \mathbf{w} . We now change this setup to assume that the true weights change slightly across successive trials, and the learner's task is to track these changes as well as possible. This implies that, as the weights are now a moving target, they can never be learned perfectly.

We model the non-stationary of the weights by a first-order autoregressive process. That is, we assume that the true weights \mathbf{w}_n in trial n depend on the true weights \mathbf{w}_{n-1} in trial n by

$$\mathbf{w}_n | \mathbf{w}_{n-1} \sim \mathcal{N}(\mathbf{A} \mathbf{w}_{n-1} + \mathbf{b}, \Sigma_d), \quad [46]$$

where \mathbf{A} , \mathbf{b} and Σ_d are parameters of the process.

Let us now consider a probabilistic learner that maintains belief $\mathbf{w}_n \sim \mathcal{N}(\boldsymbol{\mu}_{w,n}, \Sigma_{w,n})$ before observing \mathbf{x} , t , and y^* in the n th trial. Despite the successive weight change across trials, the learner would first follow its standard learning rule (discussed above different approximations) to compute posterior parameters $\boldsymbol{\mu}_{w,n}^*$ and $\Sigma_{w,n}^*$. This is followed by taking account of the weight change by updating its parameters according to

$$\boldsymbol{\mu}_{w,n+1} = \mathbf{A} \boldsymbol{\mu}_{w,n}^* + \mathbf{b}, \quad \Sigma_{w,n+1} = \mathbf{A} \Sigma_{w,n}^* \mathbf{A}^T + \Sigma_d. \quad [47]$$

These weights then act as a starting point, i.e., prior, for learning in the next trial. No other changes to the learning rules are required to take the non-stationarity of the combination weights into account.

Sampling the Bayes-optimal posterior

Finding a tractable closed-form expression for the Bayes-optimal posterior over \mathbf{w} is unfortunately impossible. However, we can approximate this posterior to almost arbitrary precision by drawing samples from this posterior. We will first discuss such sampling for stationary combination weights, in which case we can use Gibbs sampling.

Gibbs sampling for stationary weights. For Gibbs sampling, we assume prior $\mathbf{w} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$, and observations \mathbf{x}_n , t_n , and y_n^* in the n th trial. The aim is to, after N trials, draw samples from $p(\mathbf{w} | \mathbf{x}_{1:N}, t_{1:N}, y_{1:N}^*)$. With the per-trial likelihood $p(y_n^* | \mathbf{x}_n, t_n, \mathbf{w}) = \Phi(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n)$, this posterior is given by

$$p(\mathbf{w} | \mathbf{x}_{1:N}, t_{1:N}, y_{1:N}^*) \propto \mathcal{N}(\mathbf{w} | \boldsymbol{\mu}_0, \Sigma_0) \prod_{n=1}^N \Phi(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n). \quad [48]$$

The covariance of this posterior is given by

$$\Sigma_w = \left(\Sigma_0^{-1} + \sum_{n=1}^N \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^T \right)^{-1} \quad [49]$$

which can be efficiently updated with each successive trial by the Sherman-Morrison update. To sample from the posterior \mathbf{w} , we introduce the auxiliary variables $a_n \sim \mathcal{N}(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n, 1)$ for each n , such that $a_n \geq 0$ for a choice consistent with y_n^* . Thus, for a fixed \mathbf{w} , we can draw a_n according to

$$a_n | \tilde{\mathbf{x}}_n, \mathbf{w}, y_n^* \sim \mathcal{N}_{\geq 0}(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n, 1), \quad [50]$$

where $\mathcal{N}_{\geq 0}$ denotes a draw from a truncated normal distribution, guaranteeing $a_n \geq 0$. With these samples, the posterior \mathbf{w} is given by

$$\mathbf{w} | \mathbf{x}_{1:N}, t_{1:N}, y_{1:N}^*, a_{1:N} \sim \mathcal{N} \left(\Sigma_w \left(\Sigma_0^{-1} \boldsymbol{\mu}_0 + \sum_{n=1}^N y_n^* \tilde{\mathbf{x}}_n a_n \right), \Sigma_w \right). \quad [51]$$

Overall, Gibbs sampling consists in alternating between sampling $a_{1:N}$ and \mathbf{w} until a sufficient number of \mathbf{w} -samples are drawn.

Particle filtering for non-stationary weights. Once the weights become non-stationary, particle filtering turns out to be a more efficient approach to posterior sampling. The aim is to approximate the sequential weight update

$$p(\mathbf{w}_n | \mathbf{x}_{1:n}, t_{1:n}, y_{1:n}^*) \propto p(y_n^* | \mathbf{x}_n, t_n, \mathbf{w}_n) \int p(\mathbf{w}_n | \mathbf{w}_{n-1}) p(\mathbf{w}_{n-1} | \mathbf{x}_{1:n-1}, t_{1:n-1}, y_{1:n-1}^*) d\mathbf{w}_{n-1}, \quad [52]$$

by using the particle approximation

$$p(\mathbf{w}_n | \mathbf{x}_{1:n}, t_{1:n}, y_{1:n}^*) \approx \frac{1}{K} \sum_{k=1}^K \delta_{\mathbf{w}_n^{(k)}}, \quad [53]$$

consisting of the K particles $\{\mathbf{w}_n^{(1)}, \dots, \mathbf{w}_n^{(K)}\}$. With this approximation, the above sequential update becomes

$$p(\mathbf{w}_n | \mathbf{x}_{1:n}, t_{1:n}, y_{1:n}^*) \propto \sum_k p(y_n^* | \mathbf{x}_n, t_n, \mathbf{w}_n) p(\mathbf{w}_n | \mathbf{w}_{n-1}^{(k)}). \quad [54]$$

We can sample from this posterior by an importance sampling re-sampling scheme in three steps. First, we draw K samples $\tilde{\mathbf{w}}_n^{(k)}$ from a Gaussian proposal density

$$\tilde{\mathbf{w}}_n^{(k)} \sim \mathcal{N} \left(\boldsymbol{\mu}_w \left(\mathbf{w}_{n-1}^{(k)} \right), \Sigma_w \left(\mathbf{w}_{n-1}^{(k)} \right) \right). \quad [55]$$

Second, we compute the importance sampling weights,

$$\lambda_n^{(k)} = \frac{p(y_n^* | \mathbf{x}_n, t_n, \tilde{\mathbf{w}}_n^{(k)}) p(\tilde{\mathbf{w}}_n^{(k)} | \mathbf{w}_{n-1}^{(k)})}{\mathcal{N} \left(\boldsymbol{\mu}_w \left(\mathbf{w}_{n-1}^{(k)} \right), \Sigma_w \left(\mathbf{w}_{n-1}^{(k)} \right) \right)}. \quad [56]$$

Third, we re-sample the $\mathbf{w}_n^{(k)}$'s from the $\tilde{\mathbf{w}}_n^{(k)}$'s with probabilities proportional to their respective weights, $\lambda_n^{(k)}$. To ensure efficiency of the procedure, the proposal density for each weight should be close to $p(y_n^* | \mathbf{x}_n, t_n, \mathbf{w}_n) p(\mathbf{w}_n | \mathbf{w}_{n-1})$, appropriately normalized, which we achieve by computing the proposal moments $\boldsymbol{\mu}_w \left(\mathbf{w}_{n-1}^{(k)} \right)$ and $\Sigma_w \left(\mathbf{w}_{n-1}^{(k)} \right)$ according to the ADF variant that assumes non-stationary combination weights.

Relating learning through inference to learning through optimization

In all of the above we have treated learning as an inference problem, where we want to find the posterior weights given all of the observed evidence. Here, we address the parallels between inference and optimization in two ways. First, we will describe more general decision theoretical principles that highlight these parallels. Second, we will show explicitly how our learning problem can be formulated as an optimization problem.

Decision theoretic perspective. In decision theory, the Bayes-optimal decision rule is the rule that minimizes some expected loss (9). In our case, we have defined the loss as the negative reward rate, which is the negative average number of correct decisions per unit time, across a long sequence of such decisions. Furthermore, we have tuned the diffusion model boundaries such that there exists an optimal set of weights \mathbf{w}^* that maximize the reward rate, and thus minimize the loss. Formally, the loss function $L(\mathbf{w}^*, \mathbf{w})$ returns this loss for a given action \mathbf{w} (in our case a particular set of chosen weights) given some unobserved state of nature, \mathbf{w}^* (in our case the set of weights that maximize the reward rate).

Given observations X (here, all information gathered from past trials), the Bayes-optimal action is the one minimizing the posterior loss, that is

$$\underset{\mathbf{w}}{\operatorname{argmin}} \langle L(\tilde{\mathbf{w}}, \mathbf{w}) \rangle_{p(\tilde{\mathbf{w}}|X)}, \quad [57]$$

where $p(\tilde{\mathbf{w}}|X)$ are the posterior weights given all past information. If we assume the loss to be approximately quadratic around $\tilde{\mathbf{w}}$, then it is (approximately) minimized by $\langle \tilde{\mathbf{w}}|X \rangle$ (9). This justifies computing the posterior to perform learning through inference, and the use of the posterior mean for decision-making, as used in the main text.

Using Bayes-optimal decision rules for decision making has several appealing properties. One of particular interest in relation to learning through optimization is that it is an admissible rule (9, Ch. 4, Th. 9). Here, admissibility is a concept from the frequentist school of decision theory, and specifies a (not necessarily unique) decision rule $\delta(\cdot)$ whose associated risk function $R(\mathbf{w}^*, \delta)$ is smallest among all possible decision rules and all possible states of nature \mathbf{w}^* . Here, the risk function is the expected loss for a given \mathbf{w}^* , with the expectation taken over possible observations X given \mathbf{w}^* , that is $R(\mathbf{w}^*, \delta) = \langle L(\mathbf{w}^*, \delta(X)) \rangle_{p(X|\mathbf{w}^*)}$. Therefore, the Bayes-optimal decision rule doesn't only minimize the expected loss under the posterior, but also the expected loss across different (frequentist) repetitions of the same "experiment", that is, different observations for the same state of nature \mathbf{w}^* , and does so across all possible states of nature. As a consequence, finding the posterior \mathbf{w} through inference allows us to make decisions that (approximately) minimize the loss in multiple senses, which, in our case, maximizes the reward rate.

Explicit learning through optimization. Here we demonstrate for the stationary-weight case that our inference problem can be formulated as an optimization problem that aims at maximizing performance — here for simplicity measured as the probability of making correct choices. To do so, assume that, in each trial, the decision maker observes some J -dimensional momentary evidence $\delta \mathbf{x}$ that relates to the underlying latent state μ by Eq. [4], as before. They accumulate this evidence into $\mathbf{x}(t)$, and at some point (e.g., when a decision boundary is reached) decide according to $y = \operatorname{sign}(\mathbf{w}^T \mathbf{x}(t))$, using some decision strategy weight parameters \mathbf{w} . Their aim is to optimize these weight parameters to maximize their probability of making correct choices.

To find the solution to this maximization problem, let us establish which weight parameters maximize the probability of making correct choices. For this, note that by Eq. [4], the accumulated evidence is distributed as

$$\mathbf{x}(t)|\mu^* \sim \mathcal{N}((\mathbf{a}\mu^* + \mathbf{b})t, \Sigma t), \quad [58]$$

where μ^* is the (unobserved) latent state that determines the correct choice by $y^* = \operatorname{sign}(\mu^*)$. As a consequence, $\mathbf{w}^T \mathbf{x}(t)/t$ is distributed as

$$\frac{\mathbf{w}^T \mathbf{x}(t)}{t} | \mu^* \sim \mathcal{N}\left(\mathbf{w}^T \mathbf{a} \mu^* + \mathbf{w}^T \mathbf{b}, \frac{1}{t} \mathbf{w}^T \Sigma \mathbf{w}\right). \quad [59]$$

Recall that \mathbf{a} , \mathbf{b} and Σ in Eq. [4] have been defined to satisfy $\mathbf{w}^{*T} \mathbf{a} = 1$, $\mathbf{w}^{*T} \mathbf{b} = 0$, and $\mathbf{w}^{*T} \Sigma \mathbf{w}^* = 1$ for some particular \mathbf{w}^* . For these parameters, we thus have $\mathbf{w}^{*T} \mathbf{x}(t)/t \sim \mathcal{N}(\mu^*, t^{-1})$, which provides the best estimate of μ^* (in the mean squared error sense (9)), that can in turn be used as a basis for decision-making.

To find \mathbf{w}^* from an observed sequence of $(\mathbf{x}_1(t_1), t_1, y_1^*), (\mathbf{x}_2(t_2), t_2, y_2^*), \dots, (\mathbf{x}_N(t_N), t_N, y_N^*)$, we can use maximum likelihood, which is consistent and asymptotically efficient. For the diffusion model, the likelihood of \mathbf{w} for a particular choice y is by Eq. [7] (using $m = 0$) given by $p(y|\mathbf{x}, t, \mathbf{w}) = \Phi(y\mathbf{w}^T \tilde{\mathbf{x}})$, where $\tilde{\mathbf{x}} \equiv \mathbf{x}/\sqrt{t + \sigma_\mu^{-2}}$, as previously defined. Therefore, the maximum (log-)likelihood estimate for the observed sequence is given by

$$\hat{\mathbf{w}}_{ML} = \operatorname{argmax}_{\mathbf{w}} \sum_{n=1}^N \log \Phi(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n). \quad [60]$$

Finding this estimate is an optimization problem. For a small number of observations N , this optimization problem might be underdetermined. To avoid instabilities, we can additionally add a regularization term that penalizes too large $\|\mathbf{w}\|^2$, leading to

$$\hat{\mathbf{w}}_{ML,reg} = \operatorname{argmax}_{\mathbf{w}} \left(-\lambda \mathbf{w}^T \mathbf{w} + \sum_{n=1}^N \log \Phi(y_n^* \mathbf{w}^T \tilde{\mathbf{x}}_n) \right), \quad [61]$$

where $\lambda > 0$ is some regularization parameter. Overall, this demonstrates how to formulate our learning problem as an optimization problem. We have used this approach to formulate one of our heuristics, resulting in Eq. [45].

To see how this approach relates to learning through inference, compare the expression for $\hat{\mathbf{w}}_{ML,reg}$ to Eq. [48]. As can be seen, with prior parameters $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\Sigma_0 = \lambda^{-1} \mathbf{I}$, $\hat{\mathbf{w}}_{ML,reg}$ finds the maximum of the Bayesian parameter posterior over \mathbf{w} , given by Eq. [48]. In other words, it equals the maximum a-posteriori estimate. However, the optimization approach does not directly provide an estimate of the uncertainty in $\hat{\mathbf{w}}_{ML,reg}$. This makes it hard to form consistent sequential updates, in which uncertain weights should be updated more strongly than certain weights. More generally, formulating the learning problem as

an optimization problem reduces our ability to interpret the resulting expressions. For example, we might not have been able to identify that the learning rate is modulated by decision confidence without the inference formulation. All of these points made us follow the learning-as-inference route instead.

Implementing prior biases

So far we have assumed $P^+ \equiv p(\mu \geq 0) = 1/2$, making both $\mu \geq 0$ and $\mu < 0$ equally likely. Let us now consider how to consistently implement prior biases for which $P^+ \neq 1/2$. To do so, we will restrict our discussion to the one-dimensional momentary evidence δz . The high-dimensional momentary evidence case follows the same principles, and yields the same conclusions, but it notationally more burdensome.

With a *consistent* implementation of a prior bias we mean that we want to be able to choose a pair of arbitrary, potentially time-changing boundaries* $\pm\theta(t)$, each of which triggers a different Bayes-optimal choice. This requirement turns out to become critical.

Let us discuss two ways to implement biased priors in turn. The first corresponds to a shift in the mean of $p(\mu)$, while the second modulates the mass of $\mu \geq 0$ while keeping the shape of $p(\mu)$ otherwise unchanged (as in (10)).

Shifting the prior mean. If we assume prior $\mu \sim \mathcal{N}(m, \sigma_\mu)$, then $P^+ = \Phi(m/\sigma_\mu)$, such that $P^+ \neq 1/2$ if and only if $m \neq 0$. This is the case we have discussed further above (see Sec. "One-dimensional momentary evidence"), where we have found the posterior

$$p(\mu > 0|z, t) = \Phi\left(\frac{\sigma_\mu^{-2}m + z}{\sqrt{\sigma_\mu^{-2} + t}}\right). \quad [62]$$

Thus, the posterior is $p(\mu \geq 0|z, t) \geq 1/2$ if and only if $\sigma_\mu^{-2}m + z \geq 0$. This implies that Bayes-optimal decisions are determined by the sign of $\sigma_\mu^{-2}m + z$. As a result, we cannot simply bound the accumulated evidence z , as this might not guarantee a unique association between boundaries and Bayes-optimal choices. For example, consider some negative $m < 0$ and a positive $z < \sigma_\mu^{-2}|m|$ that has just reached the upper boundary $z = \theta(t)$. At this point we would intuitively make choice $y = 1$, corresponding to $\mu \geq 0$. However, as $\sigma_\mu^{-2}m + z < 0$, our expression for the posterior shows that $p(\mu \geq 0|z, t) < 1/2$, such that $y = -1$ would be the Bayes-optimal choice. This shows that bounding z directly can in some cases violate the boundary - choice correspondence.

We can regain this correspondence by instead bounding $\tilde{z}(t) = \sigma_\mu^{-2}m + z(t)$, which, by definition, starts at $\tilde{z}(0) = \sigma_\mu^{-2}m$. For this new accumulation variable it is easy to see that $p(\mu \geq 0|\tilde{z}, t) \geq 1/2$ if and only if $\tilde{z} \geq 0$, thus restoring the boundary - choice correspondence.

Directly modulating $p(\mu \geq 0)$. An alternative approach to introducing a biased prior, which was taken in (10), is to boost one half of $p(\mu)$, while modulating down the other half,

$$p(\mu) = 2\mathcal{N}(\mu|0, \sigma_\mu^2) \begin{cases} P^+ & \text{if } \mu \geq 0, \\ 1 - P^+ & \text{otherwise,} \end{cases} \quad [63]$$

ensuring again $p(\mu \geq 0) = P^+$. This prior, and the corresponding solution, has previously been investigated by (10).

This choice of prior results in the posterior over μ ,

$$p(\mu|z, t) \propto \mathcal{N}\left(\mu \mid \frac{z}{\sigma_\mu^{-2} + t}, \frac{1}{\sigma_\mu^{-2} + t}\right) \begin{cases} P^+ & \text{if } \mu \geq 0, \\ 1 - P^+ & \text{otherwise.} \end{cases} \quad [64]$$

Adding the normalization constant and integrating the above over all $\mu \geq 0$ results in the posterior

$$p(\mu \geq 0|z, t) = \frac{P^+ \Phi\left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}}\right)}{P^+ \Phi\left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}}\right) + (1 - P^+) \left(1 - \Phi\left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}}\right)\right)}. \quad [65]$$

This posterior is $p(\mu \geq 0|z, t) \geq 1/2$, and thus promotes choice $y = 1$, if

$$\log \frac{\Phi\left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}}\right)}{1 - \Phi\left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}}\right)} \geq \log \frac{1 - P^+}{P^+}, \quad [66]$$

that is, if the log-odds provided by the accumulated evidence exceeds that of the prior log-odds for $\mu < 0$. For the same accumulator value z , the evidence log-odds drops to zero over time. As a result, it might be that the Bayes-optimal choice at the same boundary changes over time, thus violating the boundary - decision correspondence.

*They might even follow different time-courses, without changing any of the discussed concepts. To keep notation simple, we won't consider this case.

When compared to the previous section, the way the prior impacts the posterior is more complex. This makes recovering the boundary - decision correspondence more complex. The aim is to find a $C_2(P^+, t)$ such that $\tilde{z}(t) = z(t) + C_2(P^+, t)$ determines Bayes-optimal decisions by its sign alone. This can be achieved by

$$C_2(P^+, t) = \sqrt{\sigma_\mu^{-2} + t} \Phi^{-1} \left(\frac{P^+ \Phi \left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}} \right)}{P^+ \Phi \left(\frac{z}{\sqrt{\sigma_\mu^{-2} + t}} \right) + (1 - P^+) \Phi \left(\frac{-z}{\sqrt{\sigma_\mu^{-2} + t}} \right)} \right) - z(t), \quad [67]$$

which unfortunately doesn't yield a closed-form expression. To gain further insight, we approximate the cumulative Gaussian function by the logistic sigmoid $\Phi(z) \approx (1 + \exp(-C_\sigma z))^{-1}$ with $C_\sigma = \pi^2/6$ to have matching slope at $z = 0$. After some algebra, this results in

$$C_2(P^+, t) \approx \sqrt{\sigma_\mu^{-2} + t} \frac{1}{C_\sigma} \log \frac{P^+}{1 - P^+}, \quad [68]$$

showing that it becomes insufficient to use a shift of the accumulation starting point, as for the previous prior. Instead, we require both a shifted starting point, as well as an additional shift in the accumulated evidence that varies over time.

The relation between decision confidence and choice accuracy for biased priors. For either choice of the prior, the solutions that regains the boundary - decision correspondence result in a decision confidence that is the same at both boundaries, as long as these boundaries are symmetric around zero. For example, for the posterior Eq. [62] this is easy to see by replacing $\sigma_\mu^{-2}m + z$ by \tilde{z} , and, at the decision by $\tilde{z} = \pm\theta(t)$, depending on which choice has been made. This is seemingly at odds with expecting a different choice accuracy at either boundary, imposed by the biased prior.

To show that no overall inconsistency between choice accuracy and choice confidence exists, let us consider the simpler case of a prior with a single "difficulty" μ_0 , which is given by

$$p(\mu) = \frac{P^+}{2} \delta(\mu - \mu_0) + \frac{1 - P^+}{2} \delta(\mu + \mu_0), \quad [69]$$

where $\delta(\cdot)$ is the Dirac delta function. That is, $\mu = \mu_0$ with probability P^+ , and $\mu = -\mu_0$ with probability $1 - P^+$. With this prior, it is easy to show that the posterior becomes

$$p(\mu = \mu_0 | z, t) = p(\mu \geq 0 | z, t) = \frac{1}{1 + e^{-2\mu_0 \left(z - \frac{1}{2\mu_0} \log \frac{P^+}{1 - P^+} \right)}}. \quad [70]$$

For symmetric boundaries at $\pm\theta$, rather than shifting the accumulation starting point, we can equivalently shift the boundaries by the same amount to

$$\theta^+ = \theta - \frac{1}{2\mu_0} \log \frac{P^+}{1 - P^+}, \quad \theta^- = -\theta - \frac{1}{2\mu_0} \log \frac{P^+}{1 - P^+}, \quad [71]$$

again leading to a constant decision confidence $(1 + \exp(-2\mu_0\theta))^{-1}$ at either boundary.

To show that this decision confidence equals the probability of making the correct choice on average, we find this probability for each possible latent state value, using known expression for boundary hitting probabilities for diffusion models with asymmetric boundaries, as given in (11, 12). For $\mu = \mu_0$, the upper boundary θ^+ leads to the correct choice. This boundary is reached with probability

$$p(z = \theta^+ | z \in \{\theta^+, \theta^-\}, \mu = \mu_0) = \frac{e^{2\mu_0\theta} - \frac{1 - P^+}{P^+}}{e^{2\mu_0\theta} - e^{-2\mu_0\theta}}, \quad [72]$$

which is the probability of making correct choices if $\mu = \mu_0$. Note that, unlike the confidence, this probability is modulated by P^+ . In particular, it grows with an increase in P^+ . In other words, the larger the a-priori probability that the upper boundary leads to the correct choice, the larger the probability that the decision maker chooses correctly in trials in which the upper boundary is indeed the correct choice.

For $\mu = -\mu_0$, the lower boundary θ^- leads to correct choices, which happens with probability

$$p(z = \theta^- | z \in \{\theta^+, \theta^-\}, \mu = -\mu_0) = \frac{e^{2\mu_0\theta} - \frac{P^+}{1 - P^+}}{e^{2\mu_0\theta} - e^{-2\mu_0\theta}}, \quad [73]$$

where the only difference to the previous expression is the impact of the prior. Specifically, this probability shrinks with an increasing P^+ .

The average probability of choosing correctly is a combination of both bound-hitting probabilities, weighted by the latent state probabilities, which, after some algebra, results in

$$p(\text{correct}) = p(z = \theta^+ | z \in \{\theta^+, \theta^-\}, \mu = \mu_0) p(\mu = \mu_0) + p(z = \theta^- | z \in \{\theta^+, \theta^-\}, \mu = -\mu_0) p(\mu = -\mu_0) = \frac{1}{1 + e^{-2\mu_0\theta}}, \quad [74]$$

where we have used $p(\mu = \mu_0) = P^+$ and $p(\mu = -\mu_0) = 1 - P^+$. This demonstrates that, even though the decision confidence differs from the probability of making the correct choices for individual choices, it equals the average probability of making

correct choices. This unintuitive result follows from conditioning the choice probabilities on the latent state, which is unknown to the decision maker, and thus cannot be reflected in their decision confidence. Once this latent state is marginalized out (by averaging over it in Eq. [74]), consistency with the decision confidence is restored (13). The same principle applies to the more complex priors used further above, but for those, it becomes hard to establish the equivalence between choice probability and decision confidence analytically.

Generating correlated momentary evidence

Recall that, for a given latent state μ , the multi-dimensional momentary evidence is drawn according to

$$\delta\mathbf{x}|\mu \sim \mathcal{N}((\mathbf{a}\mu + \mathbf{b})\delta t, \mathbf{\Sigma}\delta t), \quad [75]$$

where the parameters \mathbf{a} , \mathbf{b} and $\mathbf{\Sigma}$ satisfy $\mathbf{a}^T\mathbf{w} = 1$, $\mathbf{b}^T\mathbf{w} = 0$, and $\mathbf{w}^T\mathbf{\Sigma}\mathbf{w} = 1$ (see Sec. "High-dimensional momentary evidence").

We satisfy the requirement on \mathbf{a} and \mathbf{b} by choosing

$$\mathbf{a} = \frac{\mathbf{w}}{\mathbf{w}^T\mathbf{w}}, \quad \mathbf{b} = f_0 \left(\mathbf{1} - \frac{\mathbf{1}^T\mathbf{w}}{\mathbf{w}^T\mathbf{w}}\mathbf{w} \right), \quad [76]$$

where f_0 is a parameter. The expression for \mathbf{b} minimizes $\|\mathbf{b} - f_0\mathbf{1}\|$ under the constraint $\mathbf{b}^T\mathbf{w} = 0$, effectively introducing an approximate baseline at f_0 .

For our choice for the covariance we were guided by observations that the noise covariance spectrum in neural population recordings has few dominant components, and otherwise rapidly drops towards small values. We achieve this while satisfying $\mathbf{w}^T\mathbf{\Sigma}\mathbf{w} = 1$ by designing a $\mathbf{\Sigma}$ that has one eigenvector $\mathbf{w}/\|\mathbf{w}\|$ with associated eigenvalue $1/\mathbf{w}^T\mathbf{w}$, and otherwise the desired eigenspectrum. To do so, we fill a $J \times J$ matrix \mathbf{B} (J is the size of $\delta\mathbf{x}$) with zero mean unit variance Gaussian random numbers, except for the first row, which we set to \mathbf{w} . This is followed by Gram-Schmidt orthonormalization of \mathbf{B} , such that the first row becomes $\mathbf{w}/\|\mathbf{w}\|$, while all other rows unit vectors, orthogonal to \mathbf{w} . We then choose a diagonal \mathbf{D} with the first diagonal element $d_{11} = 1/\mathbf{w}^T\mathbf{w}$, and all other diagonal elements $d_{jj} = \max\{\sigma_x^2 e^{-j+1}, \sigma_0^2\} / \mathbf{w}^T\mathbf{w}$, with parameters σ_x^2 and σ_0^2 . The final covariance matrix is then given by $\mathbf{\Sigma} = \mathbf{B}\mathbf{D}\mathbf{B}^T$.

If the weights change across consecutive trials $n-1$ and n , the momentary evidence needs to satisfy $\mathbf{a}_n^T\mathbf{w}_n = 1$, $\mathbf{b}_n^T\mathbf{w}_n = 0$, and $\mathbf{w}_n^T\mathbf{\Sigma}_n\mathbf{w}_n = 1$ in each trial. For \mathbf{a}_n and \mathbf{b}_n this is easily achieved by re-computing them in each trial according to the above expressions.

The generation of $\mathbf{\Sigma}_n$ relies on a stochastic process, such that re-generating a new $\mathbf{\Sigma}_n$ in each trial might lead $\mathbf{\Sigma}$ to change significantly across trials despite only small changes in \mathbf{w} . To avoid this, we instead modify $\mathbf{\Sigma}_{n-1}$ by finding the smallest rotation \mathbf{U} of $\mathbf{\Sigma}_{n-1}$ that satisfies $\mathbf{w}_n^T\mathbf{\Sigma}_n\mathbf{w}_n = 1$. To do so, we aim at finding \mathbf{U} that satisfies $\mathbf{w}_n \propto \mathbf{U}\mathbf{w}_{n-1}$. This leaves \mathbf{U} underconstrained. To introduce additional constraints, we would like to restrict the rotation imposed by \mathbf{U} to the $(\mathbf{w}_{n-1}, \mathbf{w}_n)$ plane. We express this by using ψ_3, \dots, ψ_J that are orthonormal unit vectors that are also orthogonal to \mathbf{w}_{n-1} and \mathbf{w}_n , which we can find by Gram-Schmidt orthonormalization. For those vectors, we desire $\psi_n = \mathbf{U}\psi_n$. Overall, this leads to the linear equation

$$\mathbf{U} \begin{pmatrix} \frac{\mathbf{w}_{n-1}}{\|\mathbf{w}_{n-1}\|} & \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|} & \psi_3 & \dots & \psi_J \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|} & \frac{\tilde{\mathbf{w}}}{\|\tilde{\mathbf{w}}\|} & \psi_3 & \dots & \psi_J \end{pmatrix}, \quad [77]$$

where $\tilde{\mathbf{w}}$ is given by

$$\tilde{\mathbf{w}} = 2 \frac{\mathbf{w}_{n-1}^T\mathbf{w}_n}{\mathbf{w}_n^T\mathbf{w}_n} \mathbf{w}_n - \mathbf{w}_{n-1}, \quad [78]$$

and which we can easily solve for[†] \mathbf{U} . With this rotation matrix, $\mathbf{\Sigma}_n$ is given by

$$\mathbf{\Sigma}_n = \frac{\mathbf{w}_{n-1}^T\mathbf{w}_{n-1}}{\mathbf{w}_n^T\mathbf{w}_n} \mathbf{U}\mathbf{\Sigma}_{n-1}\mathbf{U}^T, \quad [79]$$

where the re-scaling by the fraction ensures the correct scaling of the eigenvalues.

Simulation details

We used parameters $\sigma_0^2 = 0.001$, $\sigma_x^2 = 2$ and $f_0 = 0$ to generate the momentary evidence $\delta\mathbf{x}|\mu$, as described in the previous section. At the beginning of each trial sequence we drew the true weights according to $\mathbf{w} \sim \mathcal{N}(\mathbf{m}_w, \mathbf{S}_w)$, with unit mean $\mathbf{m}_w = \mathbf{1}$ and identity covariance $\mathbf{S}_w = \mathbf{I}$. For that sequence, the diffusion model bounds $\pm\theta$ were time-invariant, and tuned to maximize the reward rate if the true weights were used to combined the inputs. We used $\sigma_\mu^2 = 3^2$ to draw μ in each trial. This μ determined the correct choice by $y^* = 1$ if $\mu \geq 0$, and $y^* = -1$ otherwise. The reward rate was given by

$$RR = \frac{p(\text{correct}) - c_{\text{accum}} \langle t \rangle}{\langle t \rangle + t_{\text{iti}}}, \quad [80]$$

where the average was across trials, and we set the evidence accumulation cost to $c_{\text{accum}} = 0.01$ and the inter-trial interval to $t_{\text{iti}} = 2s$. For non-stationary weights, we re-drew the weights after each choice according to Eq. [46], with $\mathbf{A} = \lambda\mathbf{I}$,

[†]Most likely there exists a closed-form expression for \mathbf{U} . We found it by solving the above expression numerically in each trial.

$\mathbf{b} = (1 - \lambda)\mathbf{m}_w$, and $\Sigma_d = (1 - \lambda^2)\mathbf{S}_w$, and set the decay factor to $\lambda = 1 - 0.01$. This yields a weight diffusion that follows a first-order autoregressive process with steady-state mean \mathbf{m}_w and covariance \mathbf{S}_m .

To compare the weight learning performance of ADF to alternative models, we simulated 1,000 learning trials 5,000 times, and reported the reward rate per trial averaged across these 5,000 repetitions. To assess steady-state performance, we performed the same procedure with non-stationary weights, and reported reward rate averaged over the last 100 trials, and over 5,000 repetitions. The sequential choice dependencies in Fig. 4A/B were also computed from these last 100 trials. The learning rate in Fig. 1D in the main text shows the pre-factor to $\Sigma_w \tilde{\mathbf{x}}$ in Eq. [41] over decision confidence for a subsample of the last 10,000 trials of a single 15,000 trial simulation with non-stationary weights. For the Gibbs sampler, we drew 10 burn-in samples, followed by 200 samples in each trial. For the particle filter we simulated 1,000 particles.

We sped up the diffusion model simulations by simulating the diffusion directly in the one-dimensional $\mathbf{w}^T \mathbf{x}_n(t)$ space. This resulted in a one-dimensional diffusion model whose first-passage time distribution is known and can be efficiently drawn from (14). The final $\mathbf{x}_n(t_n)$ was recovered by drawing it from

$$\mathbf{x}_n(t_n) \sim \mathcal{N}\left(\frac{\mu_n t_n}{\mathbf{w}^{*T} \mathbf{w}^*} \mathbf{w}^*, \frac{t_n}{\mathbf{w}^{*T} \mathbf{w}^*} \mathbf{I}\right), \quad [81]$$

subject to the constraint $\mathbf{w}^T \mathbf{x}_n(t_n) = y_n \theta$, and where \mathbf{w}^* and \mathbf{w} denote the true weights, and the weights used for evidence accumulation, respectively.

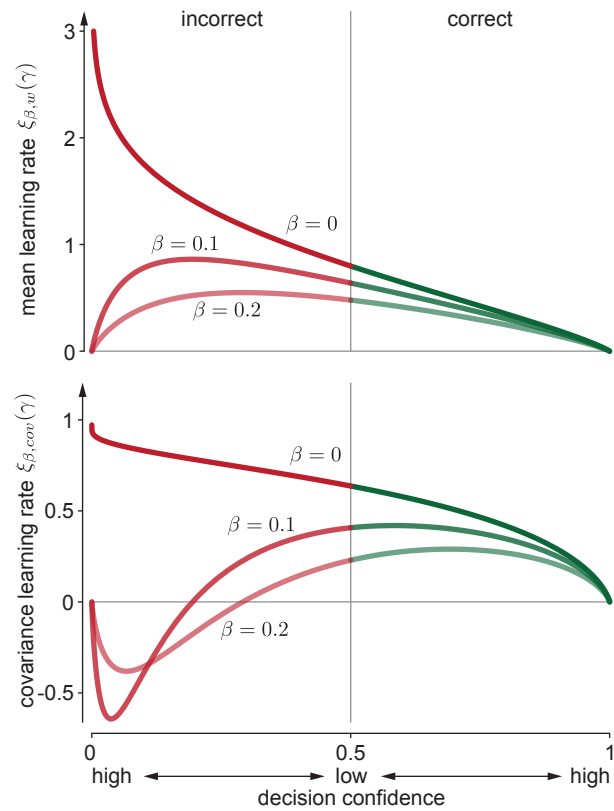


Fig. S1. Assumed density filtering learning rate modulators for noise-free and noisy feedback y^* . The top panel shows the learning rate modulator $\xi_{\beta,w}(\gamma)$ of the mean update for different levels of feedback noise, β . The bottom panel shows the same for the learning rate modulator $\xi_{\beta,cov}(\gamma)$ of the covariance update. In both cases, the marginal decision confidence associated with the feedback $p(y^*|\mathbf{x}, t) = \Phi(\gamma)$ is varied along the horizontal axis. This marginal decision confidence is $> 1/2$ for correct (green), and $< 1/2$ for incorrect (red) choices. $\beta = 0$ corresponds to the noise-free case, for which $\xi_w(\gamma) = \xi_{\beta,w}(\gamma)$ and $\xi_{cov}(\gamma) = \xi_{\beta,cov}(\gamma)$.

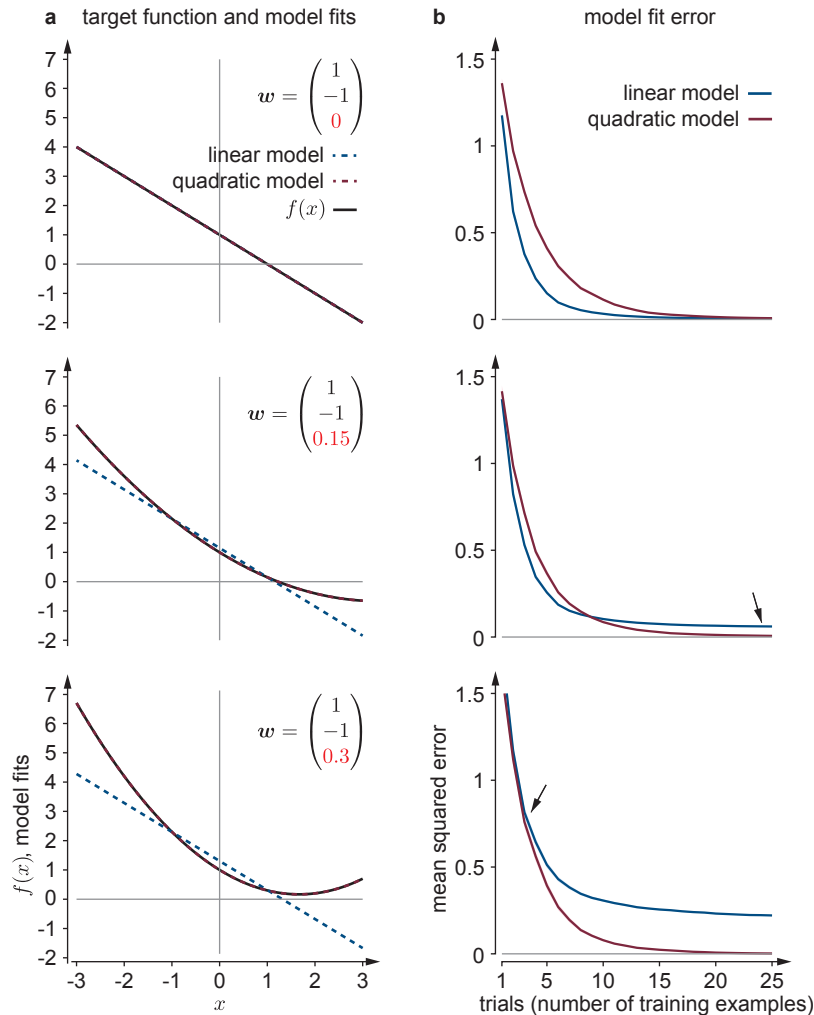


Fig. S2. Simpler models can learn more rapidly than more complex models, even if they are unable to provide perfect fits. We show model fits and model error for a linear (blue) and quadratic (red) model when fitting the target function $f(x) = w_1 + w_2x + w_3x^2$ for different w_3 's (red number in the w 's in (a)) for the top, middle, and bottom row. The quadratic model has the same functional form as $f(x)$ and learns all w . The linear model fixes $w_3 = 0$, and only learns w_1 and w_2 . Both models are fitted to training data consisting of $(x_n, f(x_n))$ -pairs, by finding the model weights that minimize the mean squared error between model predictions and given $f(x_n)$'s across all observed x_n 's. (a) With 10^4 training examples, both the linear and the quadratic model can fit a linear function (top; model fits and target function plotted on top of each other). As soon as the target function becomes quadratic (middle & bottom), the linear model fails to perfectly fit this function. (b) The mean squared error, here shown as an average across 500 repetitions across different training sets, drops more rapidly for the linear model than for the quadratic model if the target function is linear (top). This is because the linear model needs to learn fewer parameters for the same training set size. The error of both models goes to zero once the training set size increases. Even if the target function becomes quadratic (middle), the linear model can still learn more rapidly than the quadratic model (blue initially drops faster than red), even if it can't reduce its error to zero (arrow). This is only possible if the target function is still close-to-linear over the range of interest. Once it becomes too non-linear (bottom), the linear model learns slower than the quadratic model (arrow), and features a significantly worse asymptotic error. In (b), the mean squared error was in each repetition and for each training set size computed over 1000 new x 's that were not part of the training set. For all simulations, the x 's were drawn from $x \sim \mathcal{N}(0, 1)$. All learning was performed through optimization, by minimizing the mean squared error. We could have equally used learning by inference (using Bayesian linear regression with sufficiently uninformative priors), without affecting the results. Therefore, the shown effects are independent of the chosen learning formalism.

References

1. Pouget A, Drugowitsch J, Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience* 19(3):366–374.
2. Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience* 32(11):3612–3628.
3. Opper M (1998) A Bayesian approach to on-line learning in *On-line Learning in Neural Networks*, ed. Saad D. (Cambridge University Press, New York, NY, USA), pp. 363–378.
4. Minka TP (2001) Expectation propagation for approximate Bayesian inference in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01. (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA), pp. 362–369.
5. Graepel T, Candela JQ, Borchert T, Herbrich R (2010) Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, eds. Fürnkranz J, Joachims T. (Omnipress, USA), pp. 13–20.
6. Chu W, Zinkevich M, Li L, Thomas A, Tseng B (2011) Unbiased online active learning in data streams in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, eds. Apte C, Gosh J, Smyth P. (ACM, New York, NY, USA), pp. 195–203.
7. Bishop CM (2006) *Pattern Recognition and Machine Learning*. (Springer-Verlag).
8. Murphy KP (2012) *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning Series. (The MIT Press).
9. Berger JO (1993) *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics. (Springer-Verlag), 2nd edition.
10. Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience* 31(17):6339—6352.
11. Cox DR, Miller HD (1965) *The Theory of Stochastic Processes*. (Chapman and Hall).
12. Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision* 5:376–404.
13. Drugowitsch J, Moreno-Bote R, Pouget A (2014) Relation between belief and performance in perceptual decision making. *PLoS ONE* 9(5):e96511.
14. Drugowitsch J (2016) Fast and accurate Monte Carlo sampling of first-passage times from Wiener diffusion models. *Scientific Reports* 6:20490.