

Reconciling time and prediction error theories of associative learning

Received: 25 January 2025

Accepted: 2 October 2025

Published online: 21 November 2025

 Check for updatesNoé Hamou¹, Samuel J. Gershman^{2,3} & Gautam Reddy^{3,4,5} 

Learning involves forming associations between sensory events that have a consistent temporal relationship. Influential theories based on prediction errors explain numerous behavioral and neurobiological observations but do not account for how animals measure the passage of time. Here, we propose a theory for temporal causal learning, where the structure of inter-stimulus intervals is used to infer the singular cause of a rewarding stimulus. We show that a single assumption of timescale invariance, formulated as an hierarchical generative model, is sufficient to explain a puzzling set of learning phenomena, including the power-law dependence of acquisition on inter-trial intervals and timescale invariance in response profiles. A biologically plausible algorithm for inference recapitulates salient aspects of both timing and prediction error theories. The theory predicts neural signals with distinct dynamics that encode causal associations and temporal structure.

Animals learn the structure of a novel environment by forming associations between events that share consistent spatial and temporal relationships. Many principles of associative learning have been discovered within the classical conditioning paradigm, where learning is typically measured by an animal's anticipatory response to a rewarding stimulus (US) that consistently follows a cue (CS)¹. Classical conditioning experiments reveal a rich set of behavioral phenomena, including contingency degradation, blocking, and conditioned inhibition, which can be explained by reward prediction error (RPE) models. Notable examples include the Rescorla-Wagner model² and its temporal-difference (TD) generalizations^{3,4}. Neuroscientific studies provide strong support for RPE models, demonstrating that the dynamics of mesolimbic dopamine during learning and extinction match those of an RPE signal^{5–11}.

Classical RPE models do not easily explain how animals form associations across events separated by timescales spanning many orders of magnitude^{12,13}. TD models typically discretize time into states that tile the interval between the cue and reward. A TD learning rule sequentially propagates prediction errors backward in time along those discrete states, explaining how associations between distal cues and rewards could be learned^{4,14,15}. The choice of

discretization fixes an intrinsic timescale that governs the rate at which an association is acquired and the temporal precision available when anticipating reward.

However, a puzzling empirical observation is the absence of an intrinsic timescale that sets the rate of learning. Instead, the number of trials (n_{acq}) required for an animal to exhibit an anticipatory response is primarily determined by the ratio of the cue-reward interval (T) to the reward-reward interval (C) (Fig. 1a). Specifically, n_{acq} has an approximate power-law relationship with C/T ^{16,17} (Fig. 1b). Several additional phenomena challenge conventional reward prediction error (RPE) models. These include discontinuous learning curves¹⁸ (Fig. 1c) and Weber-law-like scaling of anticipatory responses¹⁹, where response profiles across experiments collapse when time is rescaled by the cue-reward interval (Fig. 1d). Furthermore, prior exposure to reward in the absence of cues impacts the number of trials to acquisition (Fig. 1e). Collectively, these behavioral results, along with broader timing-related evidence²⁰ suggesting animals explicitly encode temporal intervals, highlight key limitations of standard RPE models. Recent neurobiological findings have further questioned whether mesolimbic dopamine encodes a pure RPE signal²¹,

¹Sainsbury Wellcome Centre for Neural Circuits and Behaviour and Gatsby Computational Neuroscience Unit, University College London, London, UK.

²Department of Psychology, Harvard University, Cambridge, MA, USA. ³Center for Brain Science, Harvard University, Cambridge, MA, USA. ⁴Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ, USA. ⁵Physics & Informatics Laboratories, NTT Research, Inc., Sunnyvale, CA, USA.

 e-mail: greddy@princeton.edu

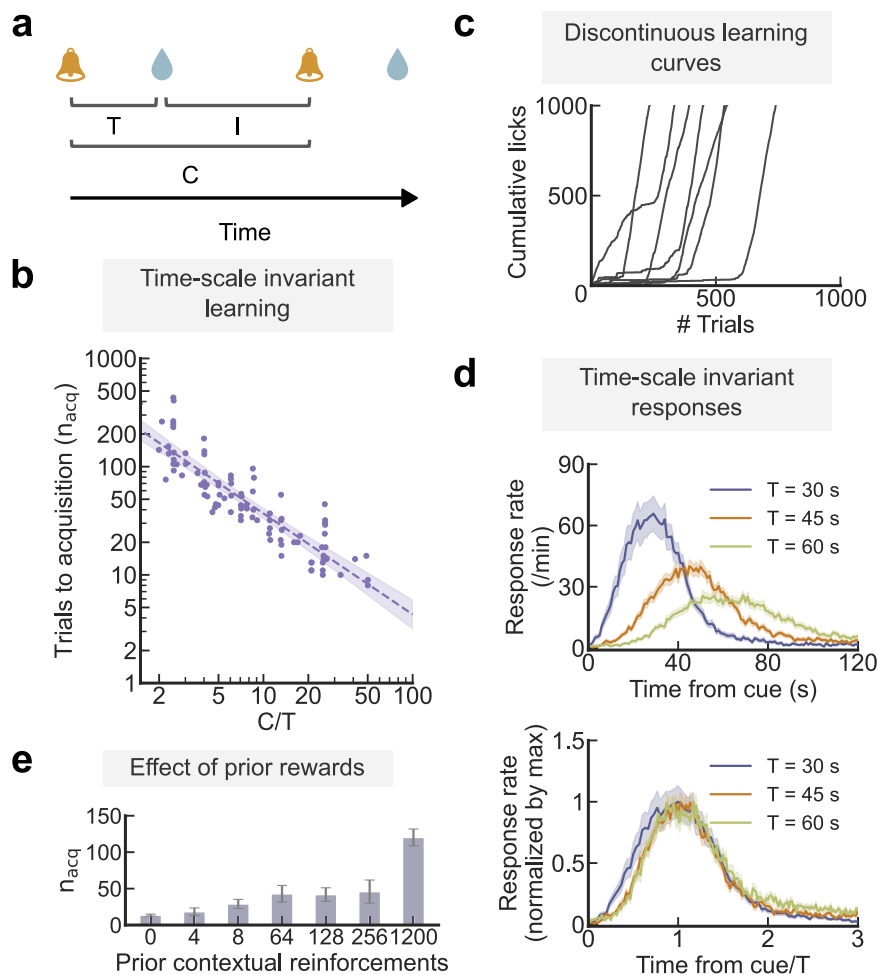


Fig. 1 | Learning phenomena in a classical conditioning paradigm. **a** Schematic representation of the delay-conditioning protocol. A reward (such as water) is presented a fixed interval after the cue (such as a bell). The interval between the cue and the reward is T , between the reward and the next cue is I , while $C = I + T$ is the interval between consecutive cues/rewards. **b** Timescale invariant learning. The number of trials required for acquisition n_{acq} is plotted against the C/T ratio on a logarithmic scale. Data are compiled from studies across different laboratories, as listed in Gallistel & Gibbon¹⁶. A linear fit is shown, highlighting the power-law relationship between n_{acq} and C/T . The shaded area corresponds to the 99% confidence interval for the linear regression. The dashed line in the center of the shaded area shows the estimated regression fit. **c** Discontinuous learning curves. Animals trained in a Pavlovian conditioning framework learn to anticipate rewards by licking in response to a cue. Cumulative lick counts are displayed, with traces shown up to the 1000th lick per animal. Data adapted from Jeong et al.²¹.

d Timescale invariance in response profiles. In the top panel, response rates are plotted for various cue-reward intervals T (left). In the bottom panel, both response rates and time are normalized by their respective maxima for each T . Data are adapted from Church et al. (1998)¹⁹ and sourced from the *Timing Database*⁶⁴. The shaded areas correspond to the standard error of the mean. **e** Effect of prior rewards. During pre-conditioning with rewards, animals receive prior contextual reinforcements before any cues are introduced. Increasing the number of contextual trials results in animals requiring more cues to form the association. The number of subjects was $n = 5$ for contextual trials 4, 8, 64, and 128; $n = 6$ for contextual trials 256; and $n = 8$ for contextual trials 0 and 1200. Average number of trials to acquisition and error bars (standard error of the mean) were obtained from Balsam et al.^{38,39}. Refer to the Methods section for a more detailed explanation of the data collection procedures.

reinvigorating efforts to develop a unified framework that reconciles RPE models with these phenomena.

Alternative models have been proposed to account for some of these phenomena^{14,16,21–28} (discussed further in Supplementary Note 1). Discontinuous learning curves can be explained based on the nature of TD signal propagation in structured environments¹⁴, or by assuming that animals implement approximate Bayesian inference by implementing a sampling algorithm²⁹. Building on rate estimation theory (RET)¹⁶, a line of work^{22,30} argues that the rate of acquisition is determined by the additional information (in bits) the cue provides about reward timing relative to the background context, which depends on C/T . A recent model grounds RET in learning theoretic terms and makes a link with RPE-like models, though key assumptions about temporal structure differ from ours²⁵. Another recent theoretical framework, called retrospective causal learning theory (RCT)²¹, proposes that animals learn causal associations using an eligibility trace

mechanism whose characteristic timescale is set by the inter-trial interval. RET and RCT help rationalize why n_{acq} depends primarily on the ratio C/T . Other models invoke cue competition²³ to highlight the influence of reward pre-conditioning, propose a noisy accumulator model to explain timescale invariance in response profiles²⁸, and predictive representations of stimulus-reward intervals^{12,26,31,32} to explain how an animal could form associations across multiple timescales.

Here, we show that prior models describing complementary aspects of associative learning can be synthesized within one common framework. The framework can be viewed as a version of model-based RL where learning temporal structure plays a central role. We present two main contributions. First, we formulate a general Bayesian framework for timing-based causal learning that describes how causal associations are learned and how these associations determine anticipatory responses. The inference process involves estimating two interdependent quantities: the distribution of intervals between

stimuli, and a probabilistic measure of the causal association between them. The key insight is that a single assumption of timescale invariance, formulated as a hierarchical generative model, quantitatively explains the phenomena described in Fig. 1. Second, we show that online algorithms for learning distributions of stimulus-reward intervals closely resemble prediction error models. We propose a learning rule for estimating causal associations between stimuli, which predicts neural signals with distinctive dynamics that encode causal associations. When applied to a common classical conditioning protocol (Fig. 1a), the model reproduces the core features of the Rescorla-Wagner model, including contingency degradation, blocking, extinction and a prediction error signal consistent with an RPE.

Results

A Bayesian framework for timing-based causal learning

We consider a scenario where the animal predicts when a rewarding stimulus (r) will appear based on the timing of past stimuli (Fig. 2a). The rewarding stimulus r appears at some (possibly stochastic) interval after a stimulus that causes it. The causal stimulus c may either be a previous occurrence of r itself or a previous occurrence of a stimulus c

amongst a set of possible non-rewarding stimuli (Fig. 2a). Our theory has two main features: (1) that the animal estimates the likelihood that one of the stimuli causes r based on statistical regularities in stimulus-reward intervals, and (2) that the animal displays an anticipatory response that maximizes long-term reward. The anticipatory response relies on a predictive map of when r will occur next given the historical record of when past stimuli have occurred.

The framework is cast as a hierarchical Bayesian model. Given the historical record until time t , Bayesian inference lets us compute the probability density that r will appear at time t . This probability is given by

$$p(\text{reward at } t) = \sum_c p(\text{reward at } t | \text{cause } c) p(\text{cause } c), \quad (1)$$

which we express as $p(t) = \sum_c \rho_c(t) \pi_c(t)$, where the sum over c includes r and all possible non-rewarding stimuli in the context. $\rho_c(t)$ encodes the information the animal has acquired about the distribution of intervals between c and r . $\pi_c(t)$ is the association, defined as the posterior probability that c causes r . Using Bayes' rule, the association $\pi_c(t)$ is

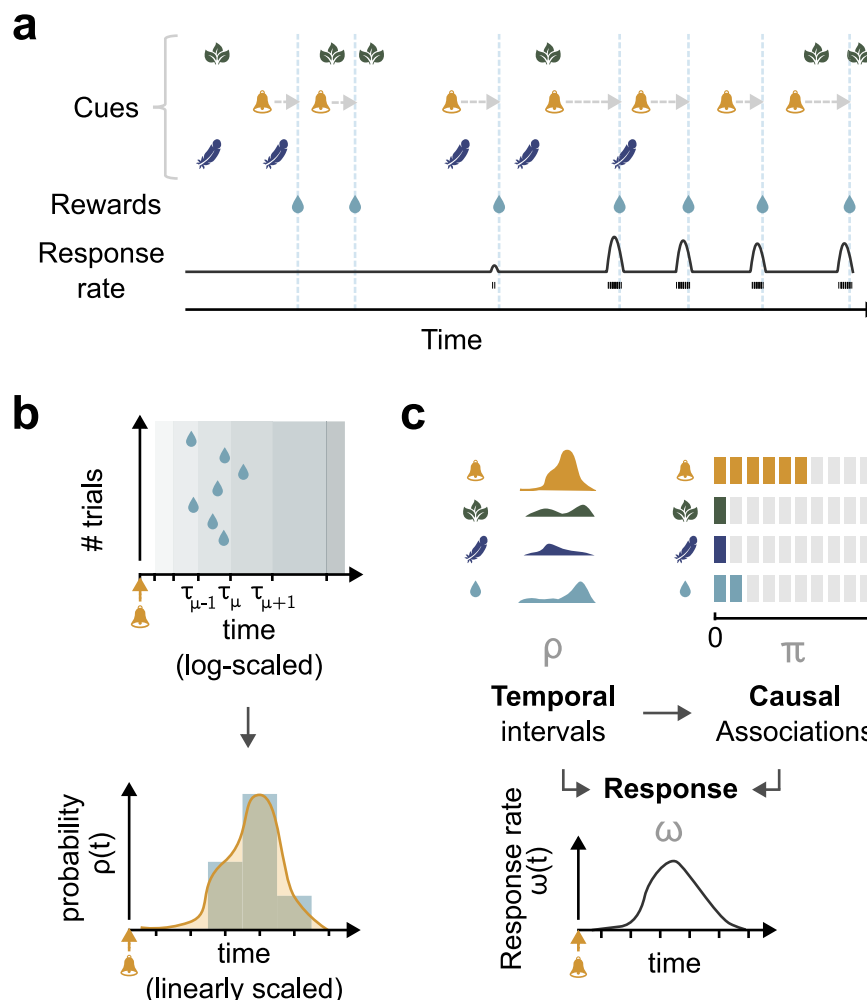


Fig. 2 | Overview of the model. **a** Schematic representation of the timing-based causal learning framework. Stimuli are point-events in time. Based on the timing between stimuli and rewards, the model agent learns to respond in anticipation of reward following stimuli that are predictive of rewards. In this example, the orange bell is the best predictor of when reward will occur. **b** (Top) Schematic representation of the learned histogram of the interval between the causal stimulus and reward. Rewards are placed in bins, where bin locations $\tau_1, \tau_2, \dots, \tau_K$ are uniformly spaced on a logarithmic time axis. (Bottom) The corresponding distribution of

intervals (yellow) between the causal stimulus and the reward is obtained by smoothing the histogram (blue). **c** Algorithmic steps of the proposed Bayesian learning theory. The agent learns to estimate distributions of intervals between each stimulus and reward (including the distribution of intervals from reward to reward). The agent uses these estimates of interval distributions to estimate the probability that each of the stimuli causes reward. The agent then combines the estimated interval distribution and causal associations to produce an anticipatory response to the reward.

proportional to the likelihood of observing the historical record until time t if c were the cause, weighted by the prior probability that c is the cause. The association can generally be expressed as

$$\pi_c(t) = \sigma(\varepsilon_c + \ell_c(t)), \quad (2)$$

where ε_c is the relative log prior, $\ell_c(t)$ is the relative log likelihood and $\sigma(x_c) = e^{x_c} / \sum_c e^{x_c}$ is the softmax function. ε_c and ℓ_c are respectively measured relative to the log prior and log likelihood that reward causes reward ($\varepsilon_c = \log \pi_c^0 - \log \pi_r^0$ and $\ell_c(t) = \log \rho_c(t) - \log \rho_r(t)$). The possibility that reward causes reward plays a similar role as a common assumption in prior models that a non-rewarding stimulus competes with a background contextual cue^{33,34}. The relative log prior ε_c captures a notion of “preparedness”³⁵, that is, the propensity for a particular stimulus to be associated with the reward based on the animal’s past experiences or innate biases. We now expand on the theory’s two key features, timescale invariance and reward maximization.

Timescale invariance. Timescale invariance is motivated by the viewpoint that an animal could form associations between contiguous events (separated by fractions of seconds) but also between distant events (separated by minutes or hours). One would expect a representation of time intervals that supports associations across timescales separated by orders of magnitude to be logarithmic. We will show that this assumption is also sufficient to explain experimental data. Specifically, we assume the interval distribution between the causal stimulus and reward is represented as a (smoothed) histogram, where the temporal locations of the K histogram bins are $\tau_1, \tau_2, \tau_3, \dots, \tau_K$ (Fig. 2b). The animal learns this smoothed histogram during conditioning. Importantly, the timescales are spaced uniformly on a logarithmic scale, $\tau_{\mu+1} - \tau_\mu = k\tau_\mu$ for $k \ll 1$.

This formulation of timescale invariance can be expressed mathematically as a generative model using a Dirichlet-multinomial distribution compounded with a scale-invariant emission function (Methods). The complete inferential framework is expressive enough to allow for multi-modal distributions of stimulus-reward intervals and partial reinforcement. Exact inference can be computationally hard in certain scenarios due to the many possible assignments between rewards and their causes when stimuli and rewards are interleaved. We discuss approximate algorithms for inference in Supplementary Note 2, noting however that the assignment problem is absent for the delay-conditioning protocol considered here.

In this model, the probability density that the reward will occur at t if the causal stimulus c appeared at $t - \delta$ is given by

$$\rho_c(t) = \sum_{\mu=1}^K w_\mu \phi(\delta/\tau_\mu) / \tau_\mu. \quad (3)$$

Intuitively, w_μ represents the probability that the reward will fall in the bin at timescale τ_μ and ϕ is a normalized basis function which smooths the estimated histogram (Fig. 2b). The weights w_μ thus encode information about the distribution of stimulus-reward intervals. A uniform prior over w_μ leads to a $1/\delta$ power-law prior distribution over stimulus-reward intervals. This scaling relation implies that shorter intervals are more likely, and that the relative likelihood of observing two intervals is equal to the ratio of those intervals.

Reward maximization. When an animal experiences a reward-predictive stimulus, it displays an anticipatory response (for example, by licking a water port) at a rate $\omega(t)$ that reflects its estimate of whether the reward will occur at time t ^{36,37}. We show that if each response has a rate-dependent cost and future rewards are discounted at a discount rate λ , the optimal anticipatory response rate $\omega^*(t)$

generally takes the form

$$\omega^*(t) = \lambda H\left(\frac{p(t)}{\lambda F(t)}\right), \quad (4)$$

where $F(t) = \int_t^\infty p(s)ds$ and H is a monotonic function (see Methods for the derivation). For example, if the cost of a response is independent of the rate ω , we find $H(x) = [\sqrt{\gamma x} - 1]_+$ up to a maximal response rate. γ is a constant that depends on the subjective value the animal receives from the reward relative to the cost of each response.

Equation (4) implies that any threshold criterion applied on the response rate to deem that the animal has acquired the association is equivalent to a criterion on the certainty with which reward is predicted to occur, that is, $p(t)/F(t) > \Theta$ for some threshold Θ . Importantly, since $p(t)$ depends on the product of $\rho_c(t)$ and $\pi_c(t)$, equation (4) further highlights that an association is acquired when the stimulus-reward interval is learned (ρ_c is sharply distributed) and when the stimulus is deemed causal ($\pi_c \approx 1$) (Fig. 2c).

Timescale invariance explains timing-related phenomena

We now examine the behavior of the model when applied to the commonly used delay-conditioning protocol shown in Fig. 1a. The experiment involves one unrewarding cue (CS) and a reward (US). The cue-reward and reward-reward intervals are fixed at T and C , respectively. The experiment begins with a pre-conditioning phase where the reward is delivered alone. The cue is introduced after n_p prior presentations of the reward. As in experiments, an association is deemed to be acquired when the rate of anticipatory response crosses a threshold.

Simulations successfully recapitulate discontinuous learning curves (Fig. 3a) and timescale invariance in response profiles (Fig. 3b). That is, response profiles across simulations with different T collapse when time since cue presentation is re-scaled by T and their amplitude is re-scaled by the maximum response value. Next, we examine how the number of trials for acquisition, n_{acq} , depends on n_p , C and T . We find that n_{acq} depends only on the ratio C/T . In particular, n_{acq} has an approximate power-law dependence on C/T , but tapers off for large values of C/T (Fig. 3c). The relative log prior ε_c has a weak influence on n_{acq} and C/T . As noted previously^{23,38,39}, the pre-conditioning phase has a strong influence on n_{acq} in experiments (Fig. 1e). This dependence is also captured by our model (Fig. 3d, Supplementary Fig. 1a).

A mathematical analysis of the model shows that timescale invariance in response profiles (Fig. 3b), the power-law scaling of n_{acq} with respect to C/T (Fig. 3c) and discontinuous learning curves (Fig. 3a) are generic consequences of timescale invariance of the stimulus-reward interval distribution. We summarize the main results derived from our analysis and refer to the Methods for mathematical details.

Acquisition can only occur once the relative log likelihood ℓ_c that the cue causes reward exceeds the relative log prior ε_c , $\ell_c > -\varepsilon_c$ (Eq. (2)). We show that ℓ_c has a non-monotonic dependence on the number of presented cues: starting from zero, it first declines and subsequently rises to a positive value (Supplementary Fig. 1). The initial drop in ℓ_c is due to the agent’s greater confidence in the reward-reward interval distribution acquired during the pre-conditioning phase. Longer pre-conditioning leads to a larger initial drop, which in turn leads to the significant dependence of n_{acq} on the number of prior rewards n_p .

Since shorter intervals are more likely, the shorter cue-reward interval ($T < C$) leads to a subsequent rapid rise in ℓ_c after a sufficient number of cue presentations. This rapid rise together with the sigmoidal dependence of the association π_c on ℓ_c (Eq. (2)) leads to an abrupt learning of the cue-reward association. If the threshold criterion on the response rate for acquisition is small, then acquisition immediately follows. Thus, the theory suggests that acquisition is primarily limited by the time taken for the animal to establish that the cue is causal rather than the time taken for the animal to fully learn the

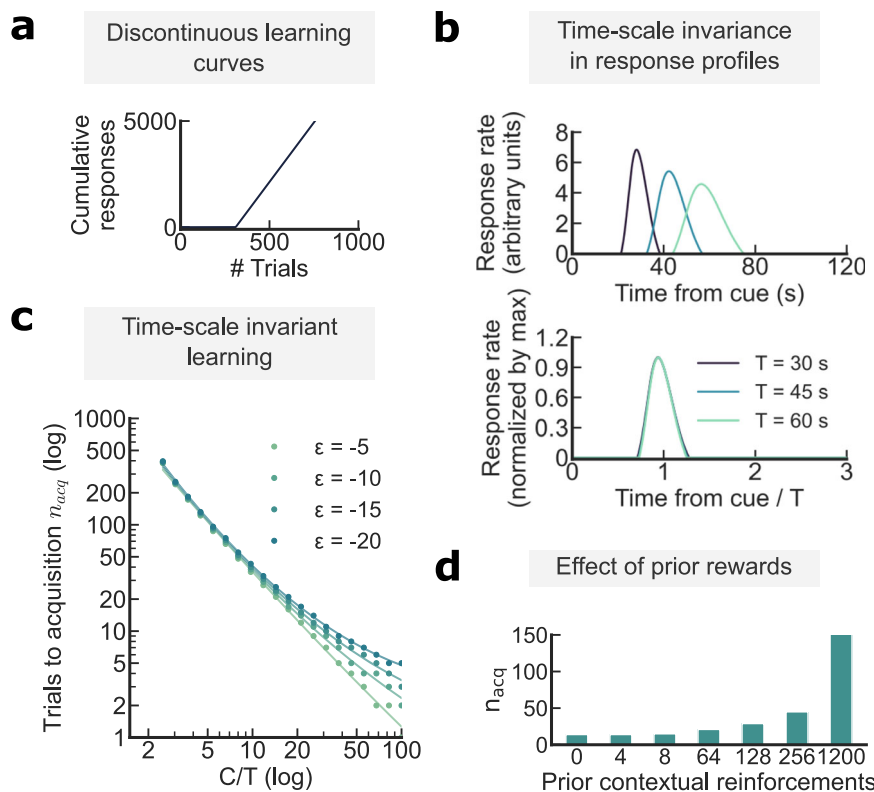


Fig. 3 | The Bayesian model reproduces timing-related phenomena.

a Discontinuous learning curves. In accordance with Fig. 1b, we plot the cumulative responses of a Bayesian agent during the delay-conditioning protocol described in Fig. 1a. **b** Timescale invariance of response profiles. Response rates of the Bayesian agent for different values of cue-reward interval T (top) and after normalizing the response rate by the maximum response rate and re-scaling time by T (bottom).

c Timescale invariant learning. The number of trials to acquisition, n_{acq} , for the Bayesian agent are plotted against C/T for different values of the relative log-prior (ε_c). Note the log-log scale. **d** Effect of prior rewards. In the Bayesian model, increasing the number of contextual trials results in the agent requiring more cues to form the association. See also Supplementary Fig. 1a.

cue-reward interval distribution (with some dependence on the acquisition criterion).

After the cue-reward association and interval are learned, the learned cue-reward interval distribution converges to $\rho_c(\delta) \approx \phi(\delta/T)/T$ (Eq. (3)). Since the response rate depends monotonically on ρ_c , re-scaling δ with T and re-scaling the amplitude of the profile with the maximal value leads to timescale invariance in the response profile. The shape of this invariant response profile is determined primarily by the basis function ϕ and the response function H . Timescale invariance in the response profile is not exact in our model unless H is linear; however, the approximation is excellent despite the nonlinear H used in simulations (Fig. 3c).

The analytical dependence of n_{acq} on C and T is in general non-trivial to obtain. We derive expressions for different parameter ranges (Methods). When $n_p, K \gg n_{\text{acq}} \gg 1$ in particular, we find

$$n_{\text{acq}} \approx \frac{-\varepsilon_c}{W\left(\frac{-\varepsilon_c C}{K e^{\frac{n_p}{T}} n_p + K}\right)}, \quad (5)$$

where W is the Lambert W function. $W(x) \approx x$ for $x \ll 1$ implies $n_{\text{acq}} \propto (C/T)^{-1}$ whenever the argument in the W function of equation (5) is small.

To explain this reciprocal dependence of n_{acq} on C/T , we first note that n_{acq} depends on the log likelihood ratio $\log(\rho_c(T)/\rho_r(C))$ (see Supplementary Note 3). Timescale invariance implies $\rho_c(T) \propto 1/T$ and $\rho_r(C) \propto 1/C$, and thus n_{acq} depends only on the ratio C/T . By itself, however, this argument would imply a linear scaling of the evidence ($\sim n \log C/T$) after n cue-reward presentations. This linear scaling

leads to an $n_{\text{acq}} \propto (\log C/T)^{-1}$ relation inconsistent with data. The $n_{\text{acq}} \propto (C/T)^{-1}$ relation comes about because the animal's estimate of the cue-reward interval also gets sharper with n . Specifically, $\rho_c(T)$ at the beginning of learning increases linearly with n : $\rho_c(T) \propto n/T$. The reward-reward interval is learned during the pre-conditioning phase, so that $\rho_r(C) = 1/C$. Acquisition follows soon after the likelihood that the cue is causal exceeds the likelihood that the reward is causal $\rho_c(T) \approx \rho_r(C)$, which leads to $n_{\text{acq}} \propto (C/T)^{-1}$.

Our assumption that animals learn the empirical histogram of stimulus-reward intervals is important to explain the relationship between n_{acq} and C/T . To emphasize this point, we repeat the above analysis supposing the animal learns the rate parameter (drawn from a scale invariant prior) of a Gamma distribution. We find that while n_{acq} depends only on the ratio C/T , the specific relation is inverse logarithmic $n_{\text{acq}} \propto (\log C/T)^{-1}$ (Methods).

A theory of temporal causal learning

Building on the Bayesian theory, we now derive a biologically plausible algorithm for inference, which we call temporal causal learning (TCL). TCL involves two mechanisms: an update rule for learning stimulus-reward interval distributions, and an update rule for learning causal associations.

Learning the interval distribution involves updating the stimulus-specific weights w_μ corresponding to each timescale τ_μ (Eq. (3)). For each stimulus, we consider the update rule

$$\frac{dw_\mu(t)}{dt} = \eta a_\mu(t) (f_r(t) - c_\mu w_\mu(t)), \quad (6)$$

where η is a learning rate and c_μ is a constant that ensures the weights are normalized (Methods). $f_r(t) = \sum_i \delta(t - t_i)$ represents the reward signal, where the t_i 's correspond to the times when the reward appeared in the past. $a_\mu(t)$ is a stimulus-specific gating signal that determines which w_μ is updated when the reward appears (Supplementary Fig. 2b). The predicted probability that the reward will appear at time t is given by a linear readout $\hat{p}_c(t) = \sum_\mu \phi_\mu(t) w_\mu(t)$, where ϕ_μ are normalized basis functions. The a_μ 's are set to zero after the reward appears until the next appearance of the stimulus.

With appropriate constraints on c_μ , the gating signal a_μ and the basis functions ϕ_μ , we show that Eq. (6) is a generic online kernel density estimation algorithm for learning distributions of intervals between two events (Methods). The kernel is specified by the choice of a_μ and ϕ_μ . The update rule is consistent with an interpretation of the weight w_μ as encoding the estimated probability that the reward appears within the interval $(\tau_\mu, \tau_{\mu+1})$. The sum $\sum_\mu w_\mu$ in turn encodes the probability that reward does indeed appear after the stimulus ($\sum_\mu w_\mu < 1$ in a partial reinforcement paradigm).

We now show how the general update rule (6) can be used to derive a timescale invariant density estimator implementable in biological networks. Specifically, the gating signals a_μ 's are derived from a set of eligibility traces ψ_μ 's associated with each stimulus (Supplementary Fig. 2a). The eligibility traces for each stimulus (say c) are updated as

$$\tau_\mu \frac{d\psi_\mu(t)}{dt} = -\psi_\mu(t) + f_c(t), \quad (7)$$

where $f_c(t)$ represents the stimulus train corresponding to stimulus c . A downstream network implements a soft winner-take-all operation, $a_\mu(t) = e^{\beta\psi_\mu(t)} / \sum_{\mu'=1}^K e^{\beta\psi_{\mu'}(t)}$. In the $\beta \rightarrow \infty$ limit, we find that $\Delta\tau_\mu = k\tau_\mu$ and $k \ll 1$ imply $a_\mu(t) \approx 1$ in the interval $(\tau_\mu, \tau_{\mu+1})$ after the stimulus and 0 otherwise. Thus, $a_\mu(t)$ represents the activity of “time cells” that are active in the interval τ_μ to $\tau_{\mu+1}$ after the stimulus is presented (Supplementary Fig. 2b). In simulations, we use normalized gamma functions with scale parameter τ_μ and a fixed shape parameter as basis functions ϕ_μ .

To derive an update rule for learning causal associations, we observe that the log likelihood L_n after n trials can generally be written as $L_n = \log P(\text{data at } n | \text{past data}) + L_{n-1}$. Based on this recursive equation, we propose an update rule for the relative log likelihood (ℓ_c) that the stimulus c is causal:

$$\frac{d\ell_c}{dt} = f_r(t) (\log \hat{p}_c(t) - \log \hat{p}_r(t) - \eta' \ell_c(t)), \quad (8)$$

where η' is a small constant that determines how many past events are averaged over when estimating ℓ_c . The pre-factor $f_r(t)$ in equation (8) indicates that the causal association is updated whenever the reward appears.

TCL reproduces both prediction-error and timing-related phenomena

The model when applied to the delay-conditioning protocol reproduces discontinuous learning curves and timescale invariance in response profiles (Supplementary Fig. 3). The model also recapitulates the approximate power-law scaling of n_{acq} with C/T and has an excellent match with the data (Fig. 4). All simulations of the model used the same set of parameters (see Methods). Notably, the approximate power-law behavior is preserved across a broad range of model parameters (Supplementary Fig. 4). Note that pre-conditioning trials are not strictly necessary to reproduce these effects (Supplementary Fig. 5).

Observing that the term in the parenthesis in Eqs. (6) and (8) resemble a prediction error, we hypothesized that TCL can reproduce phenomena attributed to RPE learning, such as extinction (Fig. 5a, b),

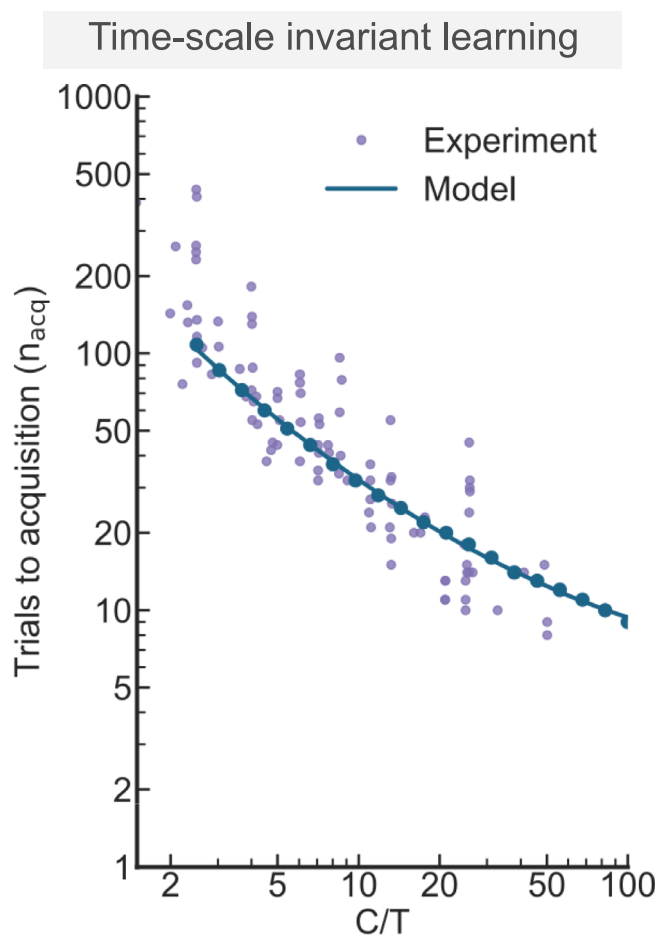


Fig. 4 | The online approximation version of our model (TCL) reproduces timescale invariant learning. Timescale invariant learning in the online model and in experiments¹⁷. The experimental points (purple dots) correspond to the data shown in Fig. 1c. The model was trained on the delay-conditioning protocol described in Fig. 1a with varying C/T ratios.

blocking (Fig. 5c) and contingency degradation (Fig. 5d). Indeed, in our model, we observe extinction of an acquired response when a previously expected reward is omitted (Fig. 5a). Consistent with experiments⁴⁰, the rate of extinction is independent of the C/T ratio (Fig. 5b). This effect arises due to the update (6), which decreases the weight associated with the cue-reward interval at a constant rate η . The TCL model also successfully captures blocking (Fig. 5c). Specifically, simulations show that the acquisition of a new stimulus-reward association is impaired when the reward has already been paired with a different stimulus, aligning with the classical blocking phenomenon. Blocking is a consequence of cue competition implicit in Eq. (8).

The TCL model reproduces contingency degradation, which is the reduction of an animal's anticipatory response when additional uncued rewards are introduced after the cue-reward association is learned^{41–43}. This effect arises in our model because the introduction of new rewards in the inter-trial period shortens the intervals between rewards, thus increasing the likelihood that rewards are caused by past rewards rather than past cues (Fig. 5d).

Finally, the model is also able to account for the scalar relationship between the response rate and the reinforcement schedule³⁰, and the observation that partial reinforcement does not affect the number of reinforcements to acquisition¹⁶ (Supplementary Fig. 6).

The dynamics of neural correlates during learning

The dynamics of the quantities related to learning intervals (Δw_μ , w_μ), associations ($\Delta\ell_c$, ℓ_c , π_c) and response (ω) are shown in Fig. 6 for the

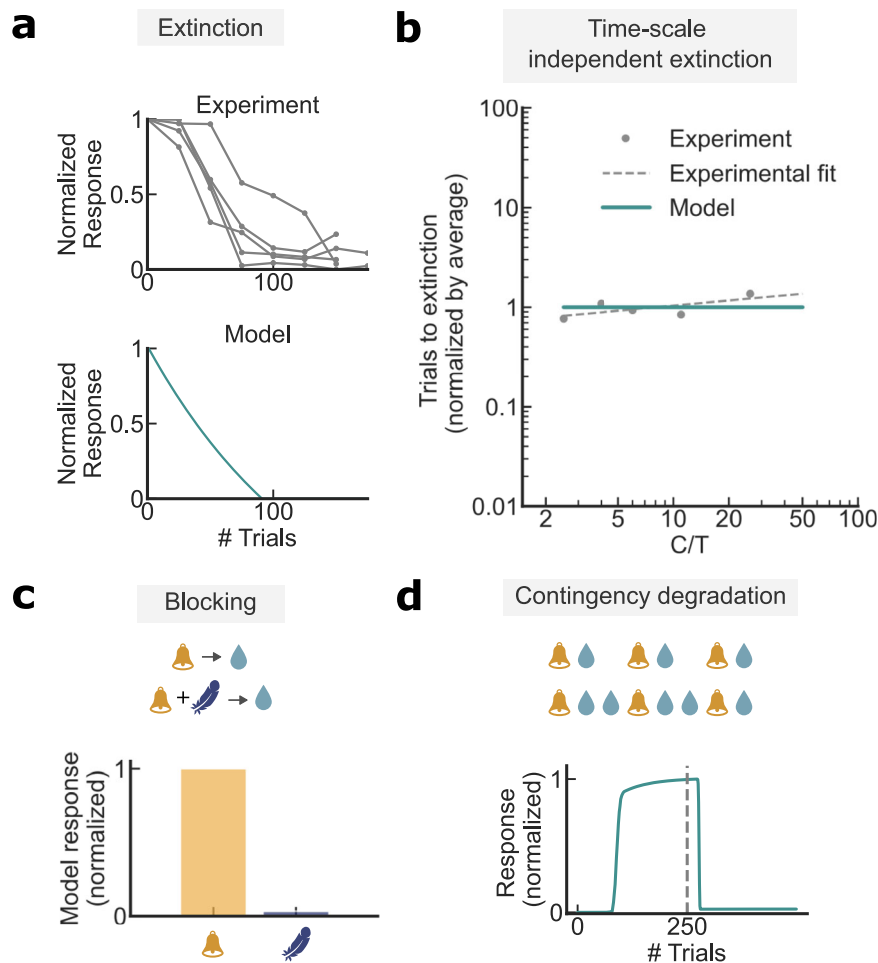


Fig. 5 | The TCL model is consistent with classical conditioning results.

a Extinction in experiments (top) and TCL simulations (bottom). Individual curves correspond to individual mice. The normalization is by the maximum response rate. Data obtained from²¹. **b** Extinction rates are not dependent on C/T in experiments and TCL simulations. Data obtained from⁶². **c** Schematic representation of the blocking paradigm and the blocking effect in the TCL model. The response was

normalized by the maximum response out of the two stimuli. **d** Schematic representation of the contingency degradation paradigm and of contingency degradation in the online model. The dashed gray line indicates when contingency degradation starts, which corresponds to the trial at which additional rewards are introduced in-between cue presentations. The response is normalized by the maximum response.

delay-conditioning protocol. The update rule for Δw_μ displays similar behavior as an RPE signal on the delay-conditioning protocol, though their behavior may differ when applied to other protocols. Specifically, the appearance of a reward at a certain interval triggers the positive update of weights corresponding to that interval, which eventually decays to zero while the interval distribution is learned (Fig. 6a). The absence of rewards after the cue-reward interval has been learned leads to a concomitant negative update.

In the Rescorla-Wagner model, the response is considered to be a direct reflection of the associative strength between the cue and reward. Our model aligns with this picture; the abrupt acquisition of the response coincides with the acquisition of the association (Fig. 6b). However, since the association increases together with the weights, the acquisition of the response will also be correlated with the weights that encode timing information (Fig. 6c). Thus, whether a neural signal encodes causal associations or timing could be challenging to disentangle in experiments.

The update rule for the relative log likelihood (Eq. (8)) predicts a non-monotonic reward-triggered signal, with the magnitude of the update peaking just before acquisition (Fig. 6d). The magnitude of the negative dip in ℓ_c increases with the number of prior rewards presented during the pre-conditioning phase. The response is acquired soon after ℓ_c becomes positive. We note however that Eq. (8) is not the

unique update rule that recapitulates this phenomenology. For example, it is possible that the log likelihoods for each stimulus (rather than the relative log likelihood) are represented independently, which are later mixed when determining the response.

Discussion

A longstanding puzzle is the absence of a fixed intrinsic timescale for how quickly animals acquire associations. Curiously, the inter-trial interval has a large influence on learning rate: scaling the inter-trial interval by a factor of ten reduces the number of trials required for acquisition (n_{acq}) by approximately the same factor^{16,30,44}. We propose a Bayesian causal learning framework to address this puzzle and other unresolved learning phenomena that are not easily explained by existing models. Our approach synthesizes features of prior models into one framework, and highlights the central role played by temporal structure learning for forming associations. The key insight is that a single assumption of timescale invariance, formulated in terms of how animals represent and learn distributions of stimulus-reward intervals when maximizing reward, can quantitatively account for abrupt learning curves, timescale invariance in response profiles and the quantitative relationship between n_{acq} and the ratio of the reward-reward and cue-reward intervals (C/T).

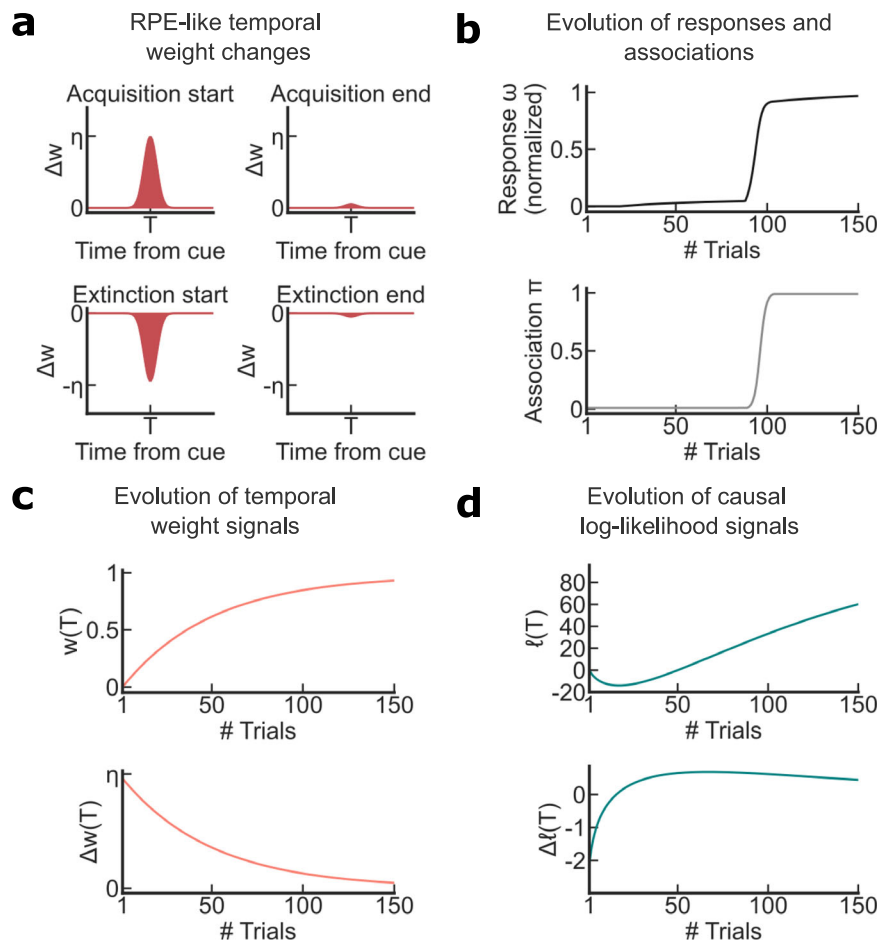


Fig. 6 | The TCL model predicts the existence of two learning signals: a classical RPE-like signal and a causal association signal. a RPE-like change of temporal weights during acquisition (top) and during extinction (bottom). **b** Evolution of the response (top) and of the association (bottom) across learning. **c** Evolution of the

weights (top) and of the change in weights (bottom) across learning. **d** Evolution of the relative log-likelihood (top) and of the change in relative log-likelihood (bottom) across learning.

Guided by the intuitive notion that animals form associations across many timescales, we propose that animals learn kernel density estimates of interval distributions while measuring time on a logarithmic scale. If all intervals on this scale are equally likely, then the probability of observing a stimulus-reward interval is inversely proportional to that interval. Using a reward maximization framework to link anticipatory response with the animal's temporal predictive map, we show that timescale invariance in interval distributions naturally leads to a Weber law scaling in response profiles. Further, since the likelihood of observing an interval is inversely proportional to the interval, the evidence (i.e., relative log likelihood) that the cue causes reward, and thus the number of trials to acquisition (n_{acq}), depends only on C/T .

The theory predicts a specific non-trivial relationship between n_{acq} and C/T that depends both on the animal's prior probability of forming the cue-reward association and on the animal's exposure to the reward prior to cue-reward pairing. For a broad parameter range, we show that this relation approximates the empirically observed power-law relation between n_{acq} and C/T , but we expect deviations from this law, particularly when C/T is large. The non-trivial relation between n_{acq} and C/T arises because the evidence that the cue is causal increases supralinearly ($\sim n \log(nC/T)$) with the number of cue-reward presentations (n). This supralinear relation combined with the non-linear relationship between evidence and response produces an abrupt learning effect akin to an "a-ha" moment when the evidence overcomes the prior. We predict that the relationship between n_{acq} and C/T

switches from the approximate $n_{\text{acq}} \propto (C/T)^{-1}$ scaling observed in experiments to an $n_{\text{acq}} \propto (\log C/T)^{-1}$ dependence (and a larger learning rate overall) if the animal is not significantly pre-conditioned to rewards.

Building on the Bayesian theory, we propose a biologically plausible model of inference, which we call temporal causal learning (TCL, Fig. 7). The TCL update rule for learning intervals (Eqs. (6), (7)) is closely related to a line of work highlighting the role of time cells and temporal context cells in associative learning^{24,45–47}. As in these models, a_μ reflects the activity of time cells and ψ_μ (the eligibility trace) reflects the activity of temporal context cells. However, the TCL update rule has certain key differences. First, we show that time cells can be derived from eligibility traces using a simple winner-take-all circuit rather than an approximate Laplace transform²⁴. The TCL update using eligibility traces (7) can be viewed as the Laplace representation of the derivative of the stimulus train, $df_c(t)/dt$. Next, our update rule has a precise interpretation as an online kernel density estimator for learning interval distributions, where the kernel is specified by a_μ and ϕ_μ . This connection to density estimation emphasizes that there are multiple update rules, corresponding to different choices of the kernel, that can approximate inference and explain data equally well. Thus, our proposed algorithm using eligibility traces and a downstream winner-take-all circuit is one of potentially many biologically plausible mechanisms for implementing temporal causal learning. Finally, the correspondence with the Bayesian framework highlights the need for another update rule (8) to learn causal associations and a

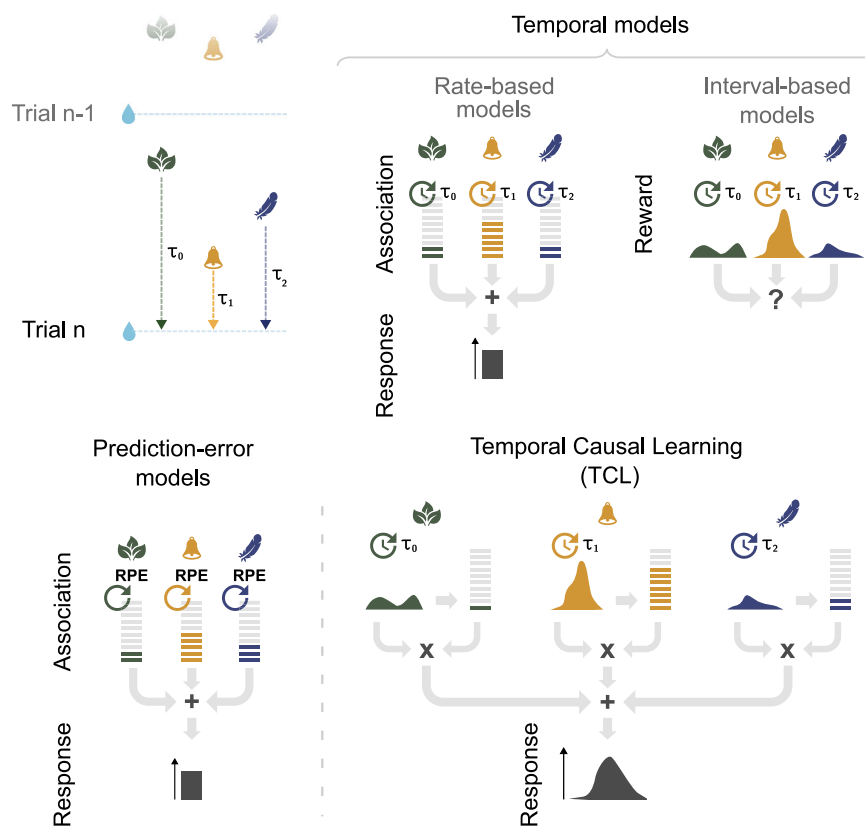


Fig. 7 | Associative learning theories. We distinguish two large classes of associative learning models: Prediction-error models (bottom left) and temporal models (right). Proposed temporal models can be classified into rate-based models and interval-based models. Rate-based models estimate the rates of cues and rewards to produce a response that depends on the ratios of these quantities. Interval-based models estimate the full distribution between cues and rewards. It is, however, unclear how, in a purely interval-based model, a response is computed. Classical prediction-error models (bottom left) do not estimate time, but update

the value of association on each trial based on the difference between the actual reward and the predicted reward. The response reflects the weighted sum of the different associative values. The model proposed in this study (bottom right) is a temporal model that estimates both the intervals between events and a causal association term. The response of the agent reflects the sum of the interval estimates weighted by the causal association. While this model explicitly constructs a representation of time, we show that this model can also be approximated by a prediction-error model.

reward maximization principle to connect interval estimation with response (4).

Our model is of similar complexity to previous models of associative learning^{2,16,28} with both the Bayesian and the TCL model relying on four and six free parameters, respectively. These are: (i) ϵ , the relative log prior; (ii) γ , a scaling factor for the response function; (iii) K , the number of timescales; (iv) k , which controls the spacing between adjacent timescales; (v, vi) η and η' the two learning rates for the TCL model. Among these, γ and k have minimal influence on the model's predictions, as shown analytically in (5). K and the two learning rates determine the overall rate at which an association is acquired, and ϵ influences the values of C/T beyond which n_{acq} deviates from the power-law scaling (see Supplementary Fig. 4 for example).

We predict a neural signal with distinctive learning dynamics that encodes the causal association between the cue and reward. Possible candidates include dopamine itself^{21,48} or other neurotransmitters such as acetylcholine which gates dopamine-dependent learning^{49–51}. In the delay-conditioning protocol, the dynamics of the associative signal are non-monotonic (Fig. 6d), with its trajectory over learning influenced by the animal's prior exposure to rewards. The prediction error term, $\Delta\ell_c$, associated with this update rule, rises sharply before saturating near zero. The association is acquired during the rising phase. Meanwhile, a parallel set of signals (Δw_μ) updates information about the distribution of cue-reward intervals, and exhibit behavior similar to reward prediction errors (Fig. 6a,c) in the commonly used delay-conditioning

protocol. Notably, the update associating with timing could be tied to serotonin, which regulates the temporal window during which stimulus-shock associations form in *Drosophila*⁵².

Consistent with the dynamics of dopamine observed in ref. ²¹, the evolution of both RPE signals ($\Delta\ell_c$ and Δw_μ) in the model is gradual and yet leads to the abrupt emergence of a response. The model predicts that the two prediction error signals display similar dynamics but with opposite signs. One potential approach to disentangle neural correlates of causal learning and temporal coding is to design an experiment in which the cue-reward interval follows an atypical (say bimodal) distribution. Our model predicts that distinct signals will code for different intervals, but a common signal will code for causal associations. Systematically varying the temporal structure (e.g., adjusting the likelihood of one mode in the distribution) and the cue-reward contingency (e.g., changing the proportion of rewards following the cue) would allow for examining whether specific neuromodulators convey timing information or convey causal associations.

The computations described in our model could thus involve, among other regions, the striatum, cortex, and the hippocampus. In contrast, the timing of learned responses on the scale of tens to hundreds of milliseconds has been shown to depend on the cerebellar cortex⁵³. These shorter timescales differ from those examined here, which span a few seconds to hundreds of seconds. Accordingly, our model's predictions are more closely associated with dopaminergic signaling and neural processes characterized by longer timescales.

Nonetheless, the presence of temporal basis functions in the cerebellum³⁴ suggests the possibility of a convergent timing mechanism that allows for forming long-range temporal associations.

A limitation of our model is that it does not account for second-order conditioning and temporal integration²⁰ (that is, the expression of temporal relationships between cues that were never paired together). As a result, it cannot explain how a predictive conditioned stimulus (CS) acquires “value” after learning, such that it elicits a dopaminergic response. One extension of the model to second-order conditioning would be analogous to the one proposed by²¹, where the learned CS will acquire the status of a “meaningful causal target” once the CS-US association is learned. In this extended framework, both prediction errors (i.e., dopaminergic reward responses) tend toward zero as the CS-US association is learned. At a certain point, the CS is granted the status of a meaningful causal target and will begin to elicit a positive dopamine response. Recent experiments suggest that this CS response of dopamine neurons encodes information about the timing and magnitude of future rewards^{55,56}, but these effects are beyond the scope of our current framework.

The theory does not explain how animals infer causes in realistic scenarios that involve many putative causal stimuli. Attentional mechanisms may play an important role in such scenarios^{57,58}. An attentional mask could be introduced as a scaling factor that modulates the rate at which the stimulus-reward interval is learned based on the probability that the stimulus is causal. Another extension to account for changing environments is to incorporate the influence of context, where each context is associated with a common temporal structure of causes and effects^{27,59,60}. While a full-fledged theory of temporal reinforcement learning⁶¹ incorporating attention, higher-order conditioning, actions and context remains to be fleshed out, this work establishes a link between reward prediction errors and the learning of temporal relationships, and thereby offers a foundational basis for such a theory.

Methods

Data curation

Experimental data used in the plots were obtained directly from openly accessible datasets whenever available. In cases where the datasets were not publicly accessible, data points were extracted from published figures using the online tool Automeris.io.

Below, we provide a detailed description of the experimental protocols corresponding to the datasets considered.

Discontinuous learning curves. The data for this section were obtained from one of the experiments reported by Jeong et al. (2022)²¹ (Fig. 4, Panel E in the original paper). In this experiment, mice were trained using a Pavlovian conditioning task. Each trial consisted of a 2-second auditory cue, followed by the delivery of a water reward 1 second after the cue ended. The cumulative number of anticipatory licks—licks occurring during the cue presentation—was recorded for seven individual mice.

Time-scale invariant learning. The data for this section were taken from Gallistel and Gibbon¹⁶, based on 12 experiments conducted on birds (primarily pigeons) across several laboratories. The subjects were trained using an autoshaping procedure, where a visual light cue predicted the delivery of a food pellet. Acquisition was defined as the first sequence of four successive trials during which a peck occurred on at least three trials. The fixed cue-reward interval is denoted by T , while I represents the mean of an approximately exponentially distributed set of intertrial intervals.

Time-scale invariance in response profiles. The data for this section were obtained from Church et al.¹⁹, in an experiment that tested four groups of rats (five rats per group) across 30 sessions using a peak

procedure paradigm. Each group experienced a different fixed interval (30 s, 45 s, or 60 s) between the onset of a white noise stimulus and food availability. Trials included both reinforced (food delivered after a lever press) and nonreinforced (no food) conditions, presented in random order. The response curves presented here are based exclusively on behavior during nonreinforced trials.

Effect of prior rewards. The data for this section were drawn from two studies: Balsam and Gibbon (1977) and Balsam and Schwartz^{38,40}, both conducted on birds (pigeons and doves). Before the introduction of auditory cues, the animals received unsignaled rewards delivered at regular intervals, referred to as prior context reinforcements. In both studies, the intertrial-to-trial duration ratio (I/T) was approximately 6, where I is the intertrial interval and T is the duration of the conditioned stimulus. The number of acquisition trials was measured as the number of trials required for animals to begin responding to the newly introduced cue. For Fig. 1e, individual animal data points could not be recovered; however, the average number of trials to acquisition and corresponding error bars (standard error of the mean) were extracted from the original figures using a graphical analysis tool (automeris.io).

Extinction. The data for this section were derived from the extinction experiment reported by Jeong et al.²¹ (Fig. 4, Panel K in the original paper). In this experiment, mice were first trained to associate a tone (cue) with a water reward. During the extinction phase, the cue was presented without the reward, breaking the learned association. The number of anticipatory licks—licks occurring during the cue presentation after the onset of extinction—was recorded for five individual mice.

Time-scale independence of extinction. The data for this section were obtained from Gibbon et al.⁶² (Fig. 4 of the original paper). In this experiment, pigeons underwent an autoshaping procedure in which a visual light cue predicted the delivery of a food pellet. Subjects were initially trained to associate the cue with the reward. During the extinction phase, the cue was presented without the reward, thereby breaking the learned association. Although the original study included groups exposed to partial reinforcement, we restricted our analysis to subjects that received continuous (full) reinforcement during acquisition. For these groups, the extinction trial was defined as the first trial in which the response rate dropped below 20% of the baseline rate measured prior to extinction.

Model simulations

All simulations of the biologically-inspired version of the model (TCL) presented in the main figures of the paper were conducted with the following parameters: learning rate $\eta = 1.5 \times 10^{-2}$, $\eta' = 10^{-2}$, and $\varepsilon = -20$. A notebook containing all the simulations and the code to plot the figures can be found on Github⁶³.

A Bayesian framework for estimating interval distributions

We present our generative model and discuss algorithms for inference. In this model, there are C possible stimuli generated by a point process. Our goal is to predict when the rewarding stimulus, indexed by r , will appear next, based on the times at which stimuli have appeared in the agent’s history. We assume that the rewarding stimulus r has a single cause, which we index as c . This causal stimulus c can be any one of the C stimuli, including r itself. The interval between r and its cause c is drawn probabilistically from a distribution described further below. We denote ε_i as the logarithm of the ratio between the prior probability that stimulus i causes r and the prior probability that stimulus r causes itself. Clearly, $\varepsilon_r = 0$.

We denote the posterior probability that stimulus i is the causal stimulus as π_i . Following our terminology in the main text, we call π_i the association of stimulus i to the reward r . After observing data, π_i is

given according to Bayes' rule as

$$\pi_i = \frac{e^{\varepsilon_i + \ell_i}}{\sum_{i=1}^C e^{\varepsilon_i + \ell_i}}, \quad (9)$$

where $\ell_i = \log\left(\frac{P(\text{data}|i)}{P(\text{data}|r)}\right)$ is the log-likelihood of observing the data given that i is the cause, relative to the log-likelihood of observing the data given that r is the cause.

The reward r appears after cause c with probability p , where p is drawn from a Beta prior $B(p; a, b)$ with hyperparameters a and b . If r does appear after c , the interval t is drawn from a distribution ρ_c . To sample the interval $t - \rho_c(t)$, we first sample an index μ from a Dirichlet-multinomial distribution. Specifically, μ (ranging from 1 to K) is drawn from a multinomial distribution with class probabilities $\mathbf{q} = (q_1, q_2, \dots, q_K)$. The probabilities $\mathbf{q}$ are in turn drawn from a Dirichlet prior $D(\mathbf{a})$, where $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_K)$. Given the sampled index μ , we then draw $t \sim \phi_\mu(t)$, where $\phi_\mu(t)$ is a normalized probability density function defined below.

We now specify the key features of timescale invariance required to recapitulate the behavioral phenomena of interest. Informally, we would like our model to capture the notion that the animal builds a histogram based on past stimulus-reward intervals. The histogram's bins are spaced uniformly on a logarithmic scale, and thus the width of a histogram bin is proportional to the bin's location. Formally,

1. Each class μ is associated with a timescale τ_μ . The timescales τ_μ are spaced uniformly on a logarithmic scale; that is, $\tau_{\mu+1} = (1+k)\tau_\mu$, where $k \ll 1$. The smallest and largest timescales are thus τ_1 and $\tau_K = (1+k)^{K-1}\tau_1$, respectively. We assume that the support of the distribution spans many orders of magnitude, i.e., $\tau_K/\tau_1 \gg 1$, which implies $K \gg 1$ if $k \ll 1$.
2. The emission probabilities $\phi_\mu(t)$ for all μ have the form $\phi_\mu(t) = \tau_\mu^{-1} \phi(t/\tau_\mu)$, where $\phi(x)$ is a density function that normalizes to one, $\int_0^\infty \phi(x) dx = 1$. Thus, the functions ϕ_μ tile the interval axis from τ_1 to τ_K , and the width of ϕ determines how much smoothing is applied when inferring the probability density from finite data. A reasonable choice is to require that ϕ_μ has width proportional to the difference in adjacent timescales, $\Delta\tau_\mu = k\tau_\mu$. Two possible choices for ϕ are: (a) a uniform distribution where $\phi_\mu(t) = 1/(k\tau_\mu)$ when $\tau_\mu \leq t \leq \tau_{\mu+1}$ and zero otherwise; and (b) a Gamma distribution $\phi_\mu(t) = \left(\frac{1}{\tau_\mu}\right) \left(\frac{k}{\Gamma(k)}\right) \left(\frac{t}{\tau_\mu}\right)^{k-1} e^{-k t/\tau_\mu}$ for $k \gg 1$.

Inference. Exact inference in this model is challenging due to the “assignment” problem. For instance, consider a scenario where the causal stimulus c (cue) appears thrice at times $t_1 < t_2 < t_3$, and the target stimulus r (reward) appears thrice afterward at times $s_1 < s_2 < s_3$, with $t_3 < s_1$. Exact inference would involve iterating over the six possible assignments of the three causal cues with the three rewards. Unless the separation between successive rewards is significantly larger than the cue-reward interval, the number of possible assignments increases exponentially with the number of presentations of c and r in the worst case scenario. We outline a method for performing approximate inference in Supplementary Note 2, although our analysis of the standard conditioning protocol later allows for exact inference due to its trial structure.

Optimal response rates

In this section, we derive the optimal anticipatory response rates (equation (4) in the main text) given the agent's estimate of the density $p(t)$ that the reward will appear at time t . We set $t = 0$ to be the moment when the most recent stimulus appeared. The agent's responses are samples from an inhomogeneous Poisson process with a time-dependent, controlled response rate $\omega(t)$. Whenever the agent responds, it incurs a rate-dependent cost $\kappa(\omega)$. The agent receives reward R once it

responds after the reward has appeared. We aim to find $\omega^*(t)$, the optimal response rate that maximizes long-term reward, provided that future rewards and costs are discounted at a rate λ . The discount rate motivates the agent to develop an anticipatory response, as the agent would prefer to obtain reward as quickly as possible after it appears.

We first compute the expected discounted reward minus the cost given that the reward appears at time t . The expected discounted cost incurred up to t is given by $\int_0^t \kappa(\omega(t')) e^{-\lambda t'} \omega(t') dt'$. Now, since the reward appears at time t , the expected discounted reward is determined by when the agent first responds after t . The expected discounted reward is then $R \int_0^\infty e^{-\lambda(t+t')} \omega(t) e^{-\omega(t)t'} dt' = R e^{-\lambda t} \omega(t) / (\omega(t) + \lambda)$. Note that we have ignored the variation in ω over the timescale of the first response. The expected discounted cost due to this response is $\kappa(\omega) e^{-\lambda t} \omega(t) / (\omega(t) + \lambda)$. Adding the expected discounted costs and rewards together and averaging over t , the net expected discounted reward (including rewards and costs) is given by

$$\mathcal{R} = \int_0^\infty \Omega(t) p(t) dt, \quad (10)$$

with

$$\Omega(t) = (r - \kappa(\omega(t))) e^{-\lambda t} \frac{\omega(t)}{\omega(t) + \lambda} - \int_0^t \kappa(\omega(t')) e^{-\lambda t'} \omega(t') dt'.$$

To optimize over ω , we take the functional derivative of $\mathcal{R}$ w.r.t $\omega(t)$ and set it to zero. This yields an expression for the optimal response rate, $\omega^*(t)$. Defining $F(t) = \int_t^\infty p(t') dt'$, a series of straightforward steps shows that ω^* satisfies

$$\frac{p(t)}{F(t)} \left(\frac{(R - \kappa)\lambda}{(\omega^* + \lambda)^2} - \frac{\kappa' \omega^*}{\omega^* + \lambda} \right) = \kappa' \omega^* + \kappa, \quad (11)$$

with the constraint $\omega^*(t) \geq 0$. Re-scaling ω^* with the discount rate λ , we get

$$\frac{p(t)}{\lambda F(t)} \left(\frac{(R - \kappa)}{(\omega^*/\lambda + 1)^2} - \frac{(\lambda \kappa')(\omega^*/\lambda)}{\omega^*/\lambda + 1} \right) = (\lambda \kappa')(\omega^*/\lambda) + \kappa. \quad (12)$$

Note that ω^* always appears as the ratio ω^*/λ and the time-dependence only appears as the ratio $p(t)/\lambda F(t)$. The optimal response rate thus generically takes the form

$$\omega^*(t) = \lambda H \left(\frac{p(t)}{\lambda F(t)} \right), \quad (13)$$

where H is a function obtained by solving (12) and depends on the specific form of $\kappa(\omega)$. If $\kappa(\omega) = C$ (a constant cost), equation (12) gives us

$$\frac{p(t)}{\lambda F(t)} \frac{(R - C)}{(\omega^*/\lambda + 1)^2} = C \quad (14)$$

Re-arranging, we get

$$\omega^*(s) = \lambda \left[\sqrt{\frac{(R - C)p(s)}{\lambda C F(s)}} - 1 \right]_+, \quad (15)$$

where $[\cdot]_+$ is the rectified linear function. Re-scaling the optimal response rate by λ , $\tilde{\omega} = \omega^*/\lambda$, and defining $\gamma \equiv (R - C)/C\lambda$ we get

$$\tilde{\omega}(t) = \left[\sqrt{\frac{\gamma p(t)}{F(t)}} - 1 \right]_+. \quad (16)$$

Behavior of the Bayesian model on the delay conditioning protocol

We now examine the behavior of the Bayesian model on the delay conditioning protocol with a single cue and reward. Specifically, the reward is presented n_p times during the pre-conditioning phase with reward-reward interval C . Subsequently, the cue and the reward are presented n times. The cue-reward interval is T and the reward-reward interval remains at C ($C > T$). Since $C > T$, the experiment has a trial structure, and n indexes the trial number.

Recall that π_c and $\pi_r = 1 - \pi_c$ are the associations of the cue and reward respectively. Our goal is to examine the behavior of π_c as the experiment progresses (increasing n) for different experimental protocols (different n_p , C , T). To derive analytical expressions, we assume $\phi_\mu(t)$ is a uniform distribution with width $\tau_{\mu+1} - \tau_\mu = k\tau_\mu$. We do not expect the results to change qualitatively for other scale-invariant choices of $\phi_\mu(t)$ as long as the density $\phi_\mu(t)$ is localized around τ_μ .

For convenience, we index π_c and ℓ_c with the trial number n rather than time t . From the definition of π_c , we have $\pi_c(n) = \sigma(\varepsilon_c + \ell_c(n))$, where ε_c is the relative log prior, $\ell_c(n)$ is the relative log likelihood after n trials and σ is the logistic function. The relative log prior ε_c is a constant. We now derive an approximate expression for $\ell_c(n)$.

To do this, we first derive an exact expression for the likelihood of observing a sequence of intervals $c_1, c_2, \dots, c_n$. Each of these intervals will fall into one of the K bins whose timescales are $\tau_1, \tau_2, \tau_3, \dots, \tau_K$. Recall that the timescales are logarithmically spaced and the width of the μ th bin is $\tau_{\mu+1} - \tau_\mu = k\tau_\mu$. Denote m_μ as the number of intervals amongst $c_1, c_2, \dots, c_n$ that fall in bin μ . Note $\sum_{\mu=1}^K m_\mu = n$.

The likelihood of observing $c_1, c_2, \dots, c_n$ with a uniform ϕ_μ is proportional to the likelihood of the counts $m_1, m_2, \dots, m_K$ for a Dirichlet-multinomial distribution. Suppose class μ has Dirichlet parameter α_μ . Using the expression for the likelihood of a Dirichlet-multinomial distribution, the likelihood of observing $c_1, c_2, \dots, c_n$ is given by

$$P(c_1, c_2, \dots, c_n) = \frac{\Gamma(\alpha_0)\Gamma(n+1)}{\Gamma(n+\alpha_0)} \prod_{\mu=1}^K \left(\frac{1}{k\tau_\mu} \right)^{m_\mu} \frac{\Gamma(m_\mu + \alpha_\mu)}{\Gamma(\alpha_\mu)\Gamma(m_\mu + 1)}, \quad (17)$$

where Γ is the Gamma function and $\alpha_0 \equiv \sum_{\mu=1}^K \alpha_\mu$.

We now apply the general expression (17) to compute the log likelihoods of observing the cue-reward intervals and the reward-reward intervals in the delay conditioning protocol. We assume $\alpha_\mu = 1$ for all μ . This choice corresponds to a uniform prior over the K timescales, though the calculation below can be generalized to arbitrary α_μ as long as they are not too large. Since all the cue-reward intervals fall into the same bin (say μ_c), we have $m_\mu = n$ if $\mu = \mu_c$ and 0 otherwise. Moreover, when $k \ll 1$, $\tau_{\mu_c} \approx T$. The likelihood (denoted $p_c(n)$) given n cue-reward intervals is thus

$$p_c(n) \approx \frac{\Gamma(K)\Gamma(n+1)}{\Gamma(n+K)} \left(\frac{1}{kT} \right)^n. \quad (18)$$

The likelihood of the reward-reward intervals is affected by the n_p prior contextual rewards. Suppose the index of the bin corresponding to the reward-reward interval C is μ_r . The effect of the prior contextual rewards is to update the Dirichlet prior for the reward-reward interval before the cue-reward pairing begins. Using properties of the Dirichlet distribution, updating the prior is equivalent to updating the Dirichlet parameter corresponding to the μ_r th bin from α_{μ_r} to $\alpha_{\mu_r} + n_p$. Assuming again that $\alpha_\mu = 1$ for all μ , the likelihood $p_r(n)$ of observing the reward-reward intervals is

$$p_r(n) \approx \frac{\Gamma(n_p + K)\Gamma(n + n_p + 1)}{\Gamma(n + n_p + K)\Gamma(n_p + 1)} \left(\frac{1}{kC} \right)^n. \quad (19)$$

The relative log likelihood $\ell_c(n) = \log p_c(n)/p_r(n)$ is then

$$\ell_c(n) \approx n \log \frac{C}{T} + \log \left(\frac{\Gamma(K)\Gamma(n+1)}{\Gamma(n+K)} \right) - \log \left(\frac{\Gamma(n_p + K)\Gamma(n + n_p + 1)}{\Gamma(n + n_p + K)\Gamma(n_p + 1)} \right).$$

We plot ℓ_c for different n , n_p , C/T values in Supplementary Fig. 1. Next, we find analytical expressions for ℓ_c in two relevant limits. Recall that $k \ll 1$ and $\tau_K/\tau_1 = (1+k)^{K-1}$ implies $K \gg 1$ as the minimal timescale τ_1 and the maximal timescale τ_K are separated by orders of magnitude.

Consider $n_p = 0$. The last two terms in (20) vanish when $n_p = 0$ and we have $\ell_c(n) \approx n \log \frac{C}{T}$. Denote n_{acq} the trial at which the likelihood of the data overcomes the prior, $\ell_c(n_{\text{acq}}) = -\varepsilon_c$. The non-trivial scenario is when $\varepsilon_c < 0$, i.e., when the prior probability that the cue is causal is small. When $\varepsilon_c > 0$, there is no learning required to form the association. We have

$$n_{\text{acq}} \approx \frac{-\varepsilon_c}{\log(C/T)}, \quad \text{when } n_p = 0. \quad (20)$$

We now consider the asymptotic limit $n_p, K \gg n \gg 1$. Intuitively, this corresponds to the scenario when the animal has learned the reward-reward interval (to a much better extent than the cue-reward interval) during the pre-conditioning phase and is going through the process of learning the cue-reward interval. Using Stirling's approximation, we get

$$\log \left(\frac{\Gamma(K)\Gamma(n+1)}{\Gamma(n+K)} \right) \approx \frac{1}{2} \log 2\pi n + n \log \left(\frac{n}{Ke} \right), \quad (21)$$

Similarly,

$$\log \left(\frac{\Gamma(n_p + K)\Gamma(n + n_p + 1)}{\Gamma(n + n_p + K)\Gamma(n_p + 1)} \right) \quad (22)$$

$$\begin{aligned} &= \log \left(\frac{\Gamma(n_p + K)\Gamma(n)}{\Gamma(n + n_p + K)} \right) + \log \left(\frac{\Gamma(n + n_p + 1)}{\Gamma(n)\Gamma(n_p + 1)} \right) \\ &\approx n \log \left(\frac{n}{(n_p + K)e} \right) - n \log \left(\frac{n}{n_p e} \right) \\ &= n \log \left(\frac{n_p}{n_p + K} \right) \end{aligned} \quad (23)$$

Combining (20), (21), (23), we have

$$\ell_c(n) \approx \frac{1}{2} \log 2\pi n + n \log \left(\left(\frac{n}{Ke} \right) \left(\frac{C}{T} \right) \left(\frac{n_p + K}{n_p} \right) \right). \quad (24)$$

Since the leading order term is $O(n \log n)$, we ignore the lowest order term $\frac{1}{2} \log(2\pi n)$ and solve for n_{acq} . After re-arranging terms, we get

$$n_{\text{acq}} \approx \frac{-\varepsilon_c}{W \left(\left(\frac{-\varepsilon_c}{Ke} \right) \left(\frac{C}{T} \right) \left(\frac{n_p}{n_p + K} \right) \right)}, \quad \text{when } n_p, K \gg n \gg 1, \quad (25)$$

where W is the Lambert W function ($x = W(y)$ if $xe^x = y$). From (25), we see that n_{acq} has weak dependence on n_p when $n_p \gg K$. Moreover, $W(x) \approx x$ for $x \ll 1$, which leads to

$$n_{\text{acq}} \approx Ke \left(\frac{C}{T} \right)^{-1} \left(\frac{n_p + K}{n_p} \right). \quad (26)$$

Thus, the $n_{\text{acq}} \propto \left(\frac{C}{T} \right)^{-1}$ relation indeed holds, but only when the term in the argument of the W function in (25) is small.

Other parameterizations of stimulus-reward interval distributions are inconsistent with data

Here, we show that other parameterizations of stimulus-reward interval distributions inadequately explain data. This analysis highlights the importance of our assumption that animals estimate Dirichlet-multinomial distributions (i.e., histograms) of stimulus-reward intervals. We focus our attention on Gamma distributions of intervals:

$$\Gamma(t; \nu, \lambda) = \lambda^\nu \frac{t^{\nu-1} e^{-\lambda t}}{\Gamma(\nu)}, \quad (27)$$

where ν is the shape parameter and λ is the rate parameter. We assume ν is fixed and λ is learned. We also consider a scale invariant prior of the rates $p_0(\lambda) \propto 1/\lambda$ (with the normalization constant determined by upper and lower cutoffs which are not important here). The assumption of fixed ν is reasonable as ν controls the maximal resolution (i.e., the mean to standard variation) of the distribution after the parameters converge. That is, we assume the animal cannot learn the interval to an arbitrarily high precision. The scale invariant prior p_0 captures time-scale invariance but also allows us to derive an exact expression for the data likelihood.

Following our analysis in the previous section, we first write down the likelihood of observing a sequence of intervals $c_1, c_2, \dots, c_n$ and then specialize to the delay conditioning protocol. The likelihood is given by

$$P(c_1, c_2, \dots, c_n) = \int d\lambda \frac{1}{\lambda} \prod_{i=1}^n \Gamma(c_i; \nu, \lambda), \quad (28)$$

$$= \int d\lambda \frac{\lambda^{n\nu-1}}{\Gamma(\nu)^n} \left(\prod_{i=1}^n c_i \right)^{\nu-1} e^{-n\lambda \bar{c}}, \quad (29)$$

$$= \frac{(\prod_{i=1}^n c_i)^{\nu-1}}{\Gamma(\nu)^n} \frac{\Gamma(n\nu)}{(n\bar{c})^{n\nu}}, \quad (30)$$

where we have defined $\bar{c} \equiv \frac{1}{n} \sum_{i=1}^n c_i$ and performed the integral in the second step recognizing that it can be written as a Gamma function.

We use this expression to compute the relative log likelihood $\ell_c(n)$ as in the previous section. After a few straightforward steps, we get

$$\begin{aligned} \ell_c(n) = & n \log \left(\frac{C}{T} \right) + \log \left(\frac{\Gamma(n\nu) \Gamma(n_p \nu)}{\Gamma((n+n_p)\nu)} \right) \\ & + \nu \log \left(\frac{(n+n_p)^{n+n_p}}{n^n n_p^{n_p}} \right). \end{aligned} \quad (31)$$

This is an exact expression. We now examine $\ell_c(n)$ when $n, n_p \gg 1$ (recall that in the previous section we assumed $n_p \gg n \gg 1$, so this is a weaker condition). Using Stirling's approximation, we have

$$\begin{aligned} \log \left(\frac{\Gamma(n\nu) \Gamma(n_p \nu)}{\Gamma((n+n_p)\nu)} \right) \approx & n\nu \log n + n_p \nu \log n_p \\ & - (n+n_p)\nu \log(n+n_p), \end{aligned} \quad (32)$$

which exactly cancels out with the third term in the exact expression for $\ell_c(n)$. The leading order term in $\ell_c(n)$ is thus $n \log(C/T)$. We therefore obtain $n_{\text{acq}} \approx -\varepsilon_c / \log(C/T)$ rather than a power-law scaling observed in data.

The biologically plausible model for timing estimation as an online kernel density estimator

In this section, we show that the biologically plausible model for learning stimulus-reward interval distributions can be interpreted as

an online kernel density estimation method. The estimated distribution of stimulus-reward intervals for a particular stimulus is encoded in stimulus-specific weights w_μ . Given w_μ , the estimated distribution $\hat{\rho}$ given that the stimulus appears at time $t = 0$ is

$$\hat{\rho}(t) = \sum_\mu w_\mu \phi_\mu(t), \quad (33)$$

where ϕ_μ is a basis function. We prescribe a update rule for updating weights w_μ

$$\frac{dw_\mu}{dt} = \eta a_\mu(t) (f_r(t) - c_\mu w_\mu(t)), \quad (34)$$

where η is a learning rate, a_μ represents the activity of “time cells” (discussed further below) and the constant c_μ is introduced to ensure the weights are normalized. $f_r(t) = \sum_i \delta(t - t_i)$ is the reward train, where the t_i s are the times when the reward appeared. We will show that the kernel estimator is defined by the specific choice of a_μ and ϕ_μ . Ensuring that the estimated density integrates to the true probability of reward given the causal cue imposes constraints on a_μ and ϕ_μ .

We assume a trial structure. If $\eta \ll 1$, integrating (34), the change in w_μ over one trial is

$$\Delta w_\mu = \eta \left(\int a_\mu(t) f_r(t) dt - w_\mu c_\mu \int a_\mu(t) dt \right), \quad (35)$$

where the integral is over the duration of a single trial and we have used $\eta \ll 1$ to ignore changes in w_μ during the trial. To remove the dependence on the second integral, we fix $c_\mu = (\int a_\mu(t) dt)^{-1}$. This leads to

$$\Delta w_\mu = \eta \left(\int a_\mu(t) f_r(t) dt - w_\mu \right), \quad (36)$$

After many trials, the “prediction error” term in the parenthesis converges to zero. After convergence, plugging in $w_\mu = \int dt' a_\mu(t') \langle f_r(t') \rangle = \int dt' a_\mu(t') \rho(t')$ into the expression for $\hat{\rho}$ gives

$$\hat{\rho}(t) = \int dt' \rho(t') \sum_\mu a_\mu(t') \phi_\mu(t) \quad (37)$$

$$= \int dt' \rho(t') K(t, t'), \quad (38)$$

where $K(t, t') \equiv \sum_\mu a_\mu(t') \phi_\mu(t)$ is the kernel. The estimated density is thus a smoothed version of the true density.

For any particular choice of a_μ and ϕ_μ , we would like to ensure that $\int dt \rho(t) = \int dt \hat{\rho}(t)$, i.e., the true probability of reward appearing within a trial matches the estimated probability. To enforce this constraint, we integrate both sides of (37) to get

$$\int dt \hat{\rho}(t) = \int dt' \rho(t') \sum_\mu a_\mu(t') \int \phi_\mu(t) dt \quad (39)$$

$$= \int dt' \rho(t') \sum_\mu \lambda_\mu a_\mu(t'), \quad (40)$$

where $\lambda_\mu \equiv \int \phi_\mu(t) dt$. The probability matching constraint (which should apply for arbitrary ρ) thus requires $\sum_\mu \lambda_\mu a_\mu(t) = 1$ for all t . Let's consider ϕ_μ that are normalized, i.e., $\lambda_\mu = 1$ for all μ , which implies $\sum_\mu a_\mu(t) = 1$ for all t . Recall that in the main text, we specify $a_\mu(t) = e^{\beta \psi_\mu(t)} / \sum_\mu e^{\beta \psi_\mu(t)}$ and $\phi_\mu(t) = \phi(t/\tau_\mu) / \tau_\mu$ where $\int_0^\infty dx \phi(x) = 1$. This choice does indeed satisfy $\lambda_\mu = 1$ for all μ and $\sum_\mu a_\mu(t) = 1$ for all t .

Requiring scale invariance would impose additional constraints on a_μ and ϕ_μ . In particular, we would like the kernel to be symmetric and translation invariant in logarithmic time coordinates, $x = \log t$. Suppose the true and estimated densities in x coordinates are $\zeta(x)$ and $\tilde{\zeta}(x)$, respectively. By the law of density transformations, we have

$$\tilde{\zeta}(x) = \int dx' \zeta(x') \sum_\mu \tilde{a}_\mu(x') \tilde{\phi}_\mu(x) = \int dx' \zeta(x') \tilde{K}(x', x), \quad (41)$$

where $\tilde{a}_\mu(x) \equiv a_\mu(e^x)$, $\tilde{\phi}_\mu(x) \equiv \phi_\mu(e^x)e^x$ and $\tilde{K}(x', x) \equiv \sum_\mu \tilde{a}_\mu(x') \tilde{\phi}_\mu(x) = \sum_\mu a_\mu(e^{x'}) \phi_\mu(e^x) e^x$. If ϕ_μ is normalized, then so is $\tilde{\phi}_\mu$. The kernel is symmetric if $\tilde{a}_\mu(x) = \xi_\mu \tilde{\phi}_\mu(x)$ for some constant positive coefficient ξ_μ . One can construct appropriate a_μ 's and ϕ_μ 's by first constructing a symmetric, translation invariant kernel $\tilde{K}$ in x (on a finite interval) and diagonalizing the kernel to obtain $\tilde{\phi}_\mu$'s and ξ_μ 's. Our particular choice of a_μ and ϕ_μ (which are derived from eligibility traces ψ_μ) is motivated by the simple linear update rule for updating ψ_μ .

Note that for the behavioral experiments considered in this paper, updating a single timescale at a time is sufficient, and a unimodal $a_\mu(t)$ (i.e., with a single peak) captures the essential dynamics. To model more complex scenarios, such as those where the cue-reward interval has a bimodal distribution, a more elaborate computation of $a_\mu(t)$ may be required.

Data availability

No experimental data was collected in this work. All datasets analyzed in this study are either publicly available from their original sources or were extracted from published figures using automeris.io. Both openly accessible and extracted datasets have been uploaded to the project's GitHub repository under the `data/` folder⁶³. The referenced datasets are listed below.

Code availability

A notebook containing all the simulations and the code to plot the figures can be found at⁶³.

References

- Pavlov, I. & Anrep, G. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. <https://books.google.co.uk/books?id=OfyUwuYTQE4C> (Oxford University Press: Humphrey Milford, 1927).
- Rescorla, R. A. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current theory and research/Appleton-Century-Crofts* (1972).
- Sutton, R. S. Learning to predict by the methods of temporal differences. *Mach. Learn.* **3**, 9–44 (1988).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (A Bradford Book, 2018).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Pearce, J. M. & Bouton, M. E. Theories of associative learning in animals. *Annu. Rev. Psychol.* **52**, 111–139 (2001).
- Niv, Y. Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154 (2009).
- Schultz, W. Neuronal reward and decision signals: From theories to data. *Physiol. Rev.* **95**, 853–951 (2015).
- Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* **40**, 373–394 (2017).
- Starkweather, C. K. & Uchida, N. Dopamine signals as temporal difference errors: recent advances. *Curr. Opin. Neurobiol.* **67**, 95–105 (2021).
- Gershman, S. J. & Uchida, N. Believing in dopamine. *Nat. Rev. Neurosci.* **20**, 703–714 (2019).
- Shankar, K. H. & Howard, M. W. A scale-invariant internal representation of time. *Neural Comput.* **24**, 134–193 (2012).
- Namboodiri, V. M. K. How do real animals account for the passage of time during associative learning? *Behav. Neurosci.* **136**, 383–391 (2022).
- Reddy, G. A reinforcement-based mechanism for discontinuous learning. *Proc. Natl. Acad. Sci. USA* **119**, e2215352119 (2022).
- Amo, R. et al. A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. *Nat. Neurosci.* **25**, 1082–1092 (2022).
- Gallistel, C. & Gibbon, J. Time, rate, and conditioning. *Psychol. Rev.* **107**, 289–344 (2000).
- Gibbon, J. & Balsam, P. Spreading association in time. *Autoshaping and Conditioning Theory* 219–253 (1981).
- Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: Implications of a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **101**, 13124–13131 (2004).
- Church, R. M., Lacourse, D. M. & Crystal, J. D. Temporal search as a function of the variability of interfood intervals. *J. Exp. Psychol. Anim. Behav. Process.* **24**, 291–315 (1998).
- Molet, M. & Miller, R. R. Timing: an attribute of associative learning. *Behav. Process.* **101**, 4–14 (2014).
- Jeong, H. et al. Mesolimbic dopamine release conveys causal associations. *Science* **378**, eabq6740 (2022).
- Gallistel, C. R., Craig, A. R. & Shahan, T. A. Contingency, contiguity, and causality in conditioning: applying information theory and weber's law to the assignment of credit problem. *Psychol. Rev.* **126**, 761–773 (2019).
- Kakade, S. & Dayan, P. Acquisition and extinction in autoshaping. *Psychol. Rev.* **109**, 533–544 (2002).
- Howard, M. W., Esfahani, Z. G., Le, B. & Sederberg, P. B. *Learning temporal relationships between symbols with laplace neural manifolds* <http://arxiv.org/abs/2302.10163> (2024).
- Gershman, S. J. Bridging computation and representation in associative learning. *Comput. Brain Behav.* <https://doi.org/10.1007/s42113-025-00242-y> (2025).
- Momennejad, I. & Howard, M. W. Predicting the future with multi-scale successor representations. *BioRxiv* 449470 (2018).
- Courville, A. C., Daw, N. D. & Touretzky, D. S. Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* **10**, 294–300 (2006).
- Luzardo, A., Alonso, E. & Mondragón, E. A rescorla-wagner drift-diffusion model of conditioning and timing. *PLoS Comput. Biol.* **13**, e1005796 (2017).
- Daw, N. & Courville, A. The pigeon as particle filter. *Adv. Neural Inf. Process. Syst.* **20**, 369–376 (2008).
- Harris, J. A. & Gallistel, C. R. Information, certainty, and learning. *eLife* **13**, RP102155 (2024).
- Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A. & Sederberg, P. B. The successor representation and temporal context. *Neural Comput.* **24**, 1553–1568 (2012).
- Tiganj, Z., Gershman, S. J., Sederberg, P. B. & Howard, M. W. Estimating scale-invariant future in continuous time. *Neural Comput.* **31**, 681–709 (2019).
- Goddard, M. J. & Jenkins, H. Blocking of a cs–us association by a us–us association. *J. Exp. Psychol. Anim. Behav. Process.* **14**, 177 (1988).
- Goddard, M. J. The role of us signal value in contingency, drug conditioning, and learned helplessness. *Psychonom. Bull. Rev.* **6**, 412–423 (1999).
- Seligman, M. E. On the generality of the laws of learning. *Psychol. Rev.* **77**, 406–418 (1970).
- Ghirlanda, S. & Enquist, M. How associations become behavior. *Neurobiol. Learn. Mem.* **205**, 107833 (2023).

37. Harris, J. A. Changes in the distribution of response rates across the cs-us interval: Evidence that responding switches between two distinct states. *J. Exp. Psychol. Anim. Learn. Cogn.* **41**, 217 (2015).
38. Balsam, P. D. & Schwartz, A. L. Rapid contextual conditioning in autoshaping. *J. Exp. Psychol. Anim. Behav. Process.* **7**, 382 (1981).
39. Balsam, P. & Gibbon, J. Formation of tone-US associations does not interfere with the formation of context-US associations in pigeons. *J. Exp. Psychol. Anim. Behav. Process.* **14**, 401–412 (1988).
40. Gibbon, J., Baldock, M. D., Locurto, C., Gold, L. & Terrace, H. S. Trial and intertrial durations in autoshaping. *J. Exp. Psychol. Anim. Behav. Process.* **3**, 264–284 (1977).
41. Rescorla, R. A. Pavlovian conditioning: It's not what you think it is. *Am. Psychol.* **43**, 151 (1988).
42. Escobar, M. & Miller, R. R. A review of the empirical laws of basic learning in pavlovian conditioning. *Int. J. Comparat. Psychol.* **17**, 2 (2004).
43. Qian, L. et al. The role of prospective contingency in the control of behavior and dopamine signals during associative learning <https://doi.org/10.1101/2024.02.05.578961> (2024).
44. Burke, D. A. et al. Few-shot learning: temporal scaling in behavioral and dopaminergic learning. *BioRxiv* <https://doi.org/10.1101/2023.03.31.535173> (2023).
45. Cao, R., Bladon, J. H., Chaczynski, S. J., Hasselmo, M. E. & Howard, M. W. Internally generated time in the rodent hippocampus is logarithmically compressed. *eLife* **11**, e75353 (2022).
46. Tsao, A. et al. Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**, 57–62 (2018).
47. Cao, R., Bright, I. M. & Howard, M. W. Ramping cells in the rodent medial prefrontal cortex encode time to past and future events via real laplace transform. *Proc. Natl Acad. Sci.* **121**, e2404169121 (2024).
48. Garr, E. et al. Mesostriatal dopamine is sensitive to specific cue-reward contingencies <https://doi.org/10.1101/2023.06.05.543690> (2023).
49. Ashby, F. G. & Crossley, M. J. A computational model of how cholinergic interneurons protect striatal-dependent learning. *J. Cogn. Neurosci.* **23**, 1549–1566 (2011).
50. Stalnaker, T. A., Berg, B., Aujla, N. & Schoenbaum, G. Cholinergic interneurons use orbitofrontal input to track beliefs about current state. *J. Neurosci.* **36**, 6242–6257 (2016).
51. Costa, K. M. et al. Dopamine and acetylcholine correlations in the nucleus accumbens depend on behavioral task states. *Curr. Biol.* **35**, 1400–1407.e3 (2025).
52. Zeng, J. et al. Local 5-HT signaling bi-directionally regulates the coincidence time window for associative learning. *Neuron* **111**, 1118–1135 (2023).
53. Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
54. Guo, C., Huson, V., Macosko, E. Z. & Regehr, W. G. Graded heterogeneity of metabotropic signaling underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. *Nat. Commun.* **12**, 5491 (2021).
55. Masset, P. et al. Multi-timescale reinforcement learning in the brain. *Nature* <https://doi.org/10.1038/s41586-025-08929-9> (2025).
56. Sousa, M. et al. A multidimensional distributional map of future reward in dopamine neurons. *Nature* <https://doi.org/10.1038/s41586-025-09089-6> (2025).
57. 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 (1980).
58. Esber, G. R. & Haselgrove, M. Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. *Proc. R. Soc. B.* **278**, 2553–2561 (2011).
59. Gershman, S. J., Blei, D. M. & Niv, Y. Context, learning, and extinction. *Psychol. Rev.* **117**, 197 (2010).
60. Gershman, S. J. & Daw, N. D. Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annu. Rev. Psychol.* **68**, 101–128 (2017).
61. Howard, M. W., Esfahani, Z. G., Le, B. & Sederberg, P. B. Foundations of a temporal rl. *ArXiv* (2023).
62. Gibbon, J., Farrell, L., Locurto, C., Duncan, H. & Terrace, H. Partial reinforcement in autoshaping with pigeons. *Anim. Learn. Behav.* **8**, 45–59 (1980).
63. Hamou, N., Gershman, S. & Reddy, G. Reconciling time and prediction error theories of associative learning. <https://doi.org/10.5281/zenodo.17193299> (2025).
64. Aydogan, T. et al. The timing database: an open-access, live repository for interval timing studies. *Behav. Res.* **56**, 290–300 (2023).

Acknowledgements

We wish to thank Randy Gallistel for sharing recent experimental data and for useful comments. G.R. was partially supported by a joint research agreement between Princeton University and NTT Research Inc. S.J.G. was partially supported by the Air Force Office of Scientific Research grant FA9550-20-1-041.

Author contributions

N.H., S.J.G. and G.R. designed research. N.H. and G.R. performed research. N.H., S.J.G. and G.R. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65137-9>.

Correspondence and requests for materials should be addressed to Gautam Reddy.

Peer review information *Nature Communications* thanks Marc Howard and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025