Extracting Value Coding Features from Individual Serotonin Neurons

Emerson F. Harkin[1–3], **Jean-Claude Béïque**[1–3], **and Richard Naud**[1–3,4,✉]

[1]Department of Cellular and Molecular Medicine, University of Ottawa, K1H 8M5, Ottawa, Canada
[2]Centre for Neural Dynamics, University of Ottawa, K1H 8M5, Ottawa, Canada
[3]Brain and Mind Institute, University of Ottawa, K1H 8M5, Ottawa, Canada
[4]Department of Physics, University of Ottawa, K1H 8M5, Ottawa, Canada

Abstract

Adaptive behaviour requires animals to continually re-evaluate the appetitive or aversive quality of their surroundings. Dorsal raphe serotonin neurons, the main source of serotonergic input to the forebrain, have been implicated in both signaling the quality of an animal's environment and regulating reward-seeking and punishment-avoiding behaviour, but the precise quantity signaled by these neurons has remained unclear, as well as how these neurons relate with behaviour. Using open-access recordings of serotonergic neurons of the dorsal raphe nucleus while animals perform a dynamic Pavlovian task, we compare firing rate and behavioral data with a model that considers reward history accumulated over a tunable timescale. Our Bayesian parameter estimation supports that serotonergic neurons are consistent with reward history being estimated over about a hundred trials on average, with a heterogeneity across individual neurons spanning 30 to 300 trials. Anticipatory licking also correlated with reward history at multiple timescale, but could not be dissociated from that of a time/thirst nuisance variable and otherwise mostly on a timescale faster than seen in serotonergic cells. These results provide a more precise picture of the dynamics of serotonergic cells under a dynamic Pavlovian task.

✉ rnaud@uottawa.ca

Keywords: serotonin, reinforcement learning, neural coding

1 INTRODUCTION

Adaptive behaviour requires animals to continually re-evaluate the quality of their environment. The serotonin neurons of the dorsal raphe nucleus (DRN) have been implicated in regulating (1–8) and signaling (9–13) behaviours that relate to environment quality, including fleeing, feeding and mating. Conceptually similar, the concept of state value, central to many reinforcement learning algorithms, is a precisely defined proxy of environment quality. State value has been a central quantity for models of Pavlovian conditioning for decades (14). This quantity was recently shown to predict features of DRN neuron activity, supporting the idea that bulk serotonergic activity in the brain signals state value (12). The precise relationship between single serotonergic neuron activity and the ever changing state value, however, remains unclear as it would require a quantitative model fitting approach. This is challenging because, in reinforcement learning, state value is dynamic, continually updated based on prior experience due to both a reward time horizon (value discounting timescale) and a value estimation timescale. The problem therefore requires that the timescales as well as the relevance of the code be assessed simultaneously. Going further, given that serotonergic and dopaminergic activity follow reward association with different learning timescales (15), the question arises of whether features of the behaviour correlates with value code at the learning timescale found in the majority, only some, or none of the serotonergic neurons. This question can be addressed by fitting anticipatory licking with a model of value evolving over multiple timescales and comparing the timescale of behaviour adaptation with that of the adaptation of responses in serotonergic cells.

Approaches to model fitting and model selection are separated in two broad categories. Given multiple hypotheses, the frequentist approach assigns probability to data and will select the model for which the data is most likely. The Bayesian approach assigns instead a probability to the hypotheses. This is done by building a sufficiently general model and estimating the confidence interval on the parameters that regulate the switch between different hypotheses. Here we used the Bayesian approach to relate a state value model with serotonin neuron activity recorded in mice performing a task with fluctuating reward delivery. We first developed a mathematical framework able to fit a value model on single neuron responses. The result show that different neurons have heterogeneous coding in 1) the degree to which their firing is consistent with state value and 2) the time scale with which the value signal is learned. While different neurons show different learning timescales, the average learning timescale we found was around 100 trials (10-20 min), consistent with a prior report based on population responses (15). Then, we leveraged simultaneous recording of anticipatory licking in the same experiments to ask if this aspect of behaviour was related to value coding at the timescale recorded in the serotonergic neurons. Adapting our Bayesian approach to behaviour, we found that only the value estimated on timescales faster than those associated on average with the value code in the DRN were correlated with anticipatory licking. While some serotonin neurons faster learning timescales than others,

they still did not track the evolution of behaviour. Hence our work provides a quantitative report of potency of the value code and its learning timescale in individual serotonin neurons, an observed timescale that generally does not match the timescales correlated with behaviour.

2 RESULTS

We used an openly accessible data set where mice performed a dynamic Pavlovian task (Fig. 1; Ref. (16); see Methods). In this task, rewards were associated with a cue based on a probability of reward that changed in piecewise constant blocks. The activity of genetically identified serotonergic cells were recorded using extracellular electrodes. Anticipatory licking was also recorded. While the activity during the randomly initiated trials has previously been shown to relate with state value (12), we focused on the activity in the inter-trial period because the greater number of action potentials makes it more amenable to track learning-related changes and it allows us to ignore the effect of the discounting timescale in a value model since this timescale is shorter than the trial duration (10, 12). This approach is supported by both theory and previous observations, whereby the value of the inter-trial period scales with the value of the trace period (17) (see Discussion).

2.1 Reward, behaviour, and neural activity evolve slowly over time

As a first step, we sought to determine the timescales over which behaviour and neural activity evolve in this experiment. Frequency analysis of the autocorrelation patterns of reward collection, anticipatory licking, and serotonin neuron firing collectively revealed peaks at two frequencies: one near $1/70\,\mathrm{cycles\,trial}^{-1}$, corresponding to the upper limit of the block length, and another near $1/350\,\mathrm{cycles\,trial}^{-1}$, corresponding to the mean duration of a recording session (Fig. 2). Whereas previous analysis of this data focused on short-term fluctuations over the course of fewer than ten trials (12, 16), these results suggest that both behaviour and neural activity mainly evolve over significantly longer timescales.

Interestingly, fluctuations corresponding to block length and session length were not represented equally in the dynamics of reward collection, anticipatory licking, and serotonin neuron activity. The power spectrum of reward collection showed a pronounced peak at $1/63\,\mathrm{cycles\,trial}^{-1}$ and much less power at lower frequencies (Fig. 2A), suggesting that the dynamics of reward collection are dominated by the block structure of the experiment rather than session-long trends. In contrast, the power spectrum of anticipatory licking showed peaks at both $1/63\,\mathrm{cycles\,trial}^{-1}$ and $1/377\,\mathrm{cycles\,trial}^{-1}$ (Fig. 2B). This pattern is not consistent with the idea that anticipatory licking simply reflects the number of rewards collected in the recent past, since frequency analysis of an exponentially-weighted moving

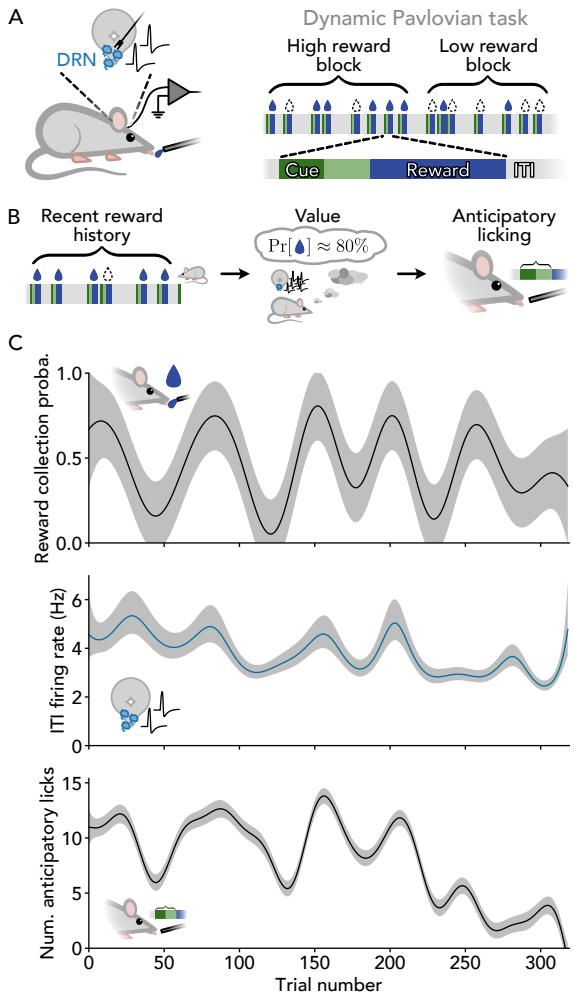

Figure 1. Overview of experimental setup and conceptual model.
A Experimental setup in the dynamic Pavlovian conditioning task of Grossman et al. (16). *In vivo* electrophysiological recordings of the activity of individual genetically-identified dorsal raphe nucleus (DRN) serotonin neurons were collected from head-fixed mice receiving stochastic water rewards in a trace conditioning task. Trials consisted of a $1\,\mathrm{s}$ odour cue followed by a $1\,\mathrm{s}$ trace period and $3\,\mathrm{s}$ reward collection window. Inter-trial interval (ITI) durations followed an exponential distribution with mean $3.3\,\mathrm{s}$. Rewards were delivered with probability $20\,\%$, $50\,\%$, or $80\,\%$ in a block structure with random uncued transitions. See Methods and ref. 16 for details.
B Conceptual model. Recent reward history is used to calculate value, defined as an estimate of the probability of receiving a reward in the upcoming trial. We hypothesize that high value is reflected in both an increase in the ITI firing rates of serotonin neurons and an increase in anticipatory licking.
C Time-course of reward collection probability (top), serotonin neuron firing rate (middle), and number of anticipatory licks per trial (bottom) for an example recording session. For illustration purposes, all data are smoothed using Gaussian process regression (squared exponential kernel with length scale $20\,\mathrm{trials}$, observation standard deviation set to $\sqrt{45(100-45)}$ percent, $62\,\mathrm{ms}$, and $\sqrt{2}$ licks for top, middle, and bottom, respectively). Bands represent an approximate $95\,\%$ CI. Note that firing rates were obtained by analyzing the mean inter-spike interval (ISI) in each ITI, which more closely follows a normal distribution, and inverting the results.

average of collected rewards with a learning timescale of $\tau = 4.5\,\mathrm{trials}$ reveals only a single peak corresponding to the block length (Fig. 2C). Finally, analysis of serotonin neuron activity reveals a single peak at $1/339\,\mathrm{cycles\,trial}^{-1}$ (Fig. 2D). This is also not consistent with a short-term history of collected rewards (Fig. 2C), but could arise from a medium-term reward history with a timescale similar to the average block length of $45\,\mathrm{trials}$ (Fig. 2E). Overall, we find that the dynamics of reward collection are dominated by fluctuations on a timescale that corresponds to the block length and serotonin neuron activity is dominated by session-long trends, while anticipatory licking displays a mixture of both timescales.

2.2 Serotonin neuron activity is consistent with a slowly-evolving estimate of value

To directly examine the connection between serotonin neuron activity and reward history, we constructed a generative model of inter-trial interval firing statistics as a function of value v_t

$$\rho_{t,[i]} = \beta_{0,[i]} + \beta_{1,[i]}\, v_{t,[i]} + \epsilon_{t,[i]} \tag{1}$$

where $\rho_{t,[i]}$ is the firing rate of neuron i during inter-trial interval t; $\beta_{0,[i]}$ is the firing rate intercept, which represents the baseline firing rate when value is zero; $\beta_{1,[i]}$ is the firing rate gain, which controls how quickly firing increases or decreases with increasing value; and $\epsilon_{t,[i]}$ is a noise term (see Section 5.2.3). We operationally define value as an estimate of the reward probability in the upcoming trial $v_{t,[i]} \approx \Pr\left[R_{t+1,[i]} = 1\right]$ learned according to the Rescorla–Wagner rule

$$v_{t,[i]} = v_{t-1,[i]} + \underbrace{\left(1 - e^{-1/\tau_{[i]}}\right)}_{\text{learning rate}} \underbrace{\left(r_{t,[i]} - v_{t-1,[i]}\right)}_{\text{reward prediction error}}, \tag{2}$$

where $r_{t,[i]}$ is the reward received in trial t, $v_{t,[i]}$ is the value during the inter-trial interval immediately following $r_{t,[i]}$, and $\tau_{[i]}$ is the learning timescale. Importantly, value learned in this way is equivalent to an exponentially-weighted moving average of past rewards with learning timescale τ and initial condition v_0

$$v_{t,[i]} = e^{\frac{-t}{\tau_{[i]}}} v_{0,[i]} + \left(1 - e^{\frac{-t}{\tau_{[i]}}}\right) \sum_{j=0}^{t-1} e^{\frac{-t}{\tau_{[i]}}} r_{t-j,[i]}, \tag{3}$$

as used in the previous section. Overall, our model assumes that serotonin neuron activity reflects a moving average of recent rewards, where the meaning of "recent" is governed by the learning timescale τ. Thus Eq. 2 defines our value model and Eq. 1 defines how it is fit to the observed firing rate (Fig. 3A-B).

To assess whether this value model provides a meaningful account of the activity patterns of serotonin neurons in this experiment, we compared its ability to predict firing in held-out trials against two control models: one which assumes a constant firing rate and the other which assumes a session-wide linear trend. The value model performed significantly better than the two control models in terms of the average error across all neurons in this dataset (mean squared error on the mean inter-spike interval in held-out trials of $0.0209\,\mathrm{s}^2$ for the value model, $0.0219\,\mathrm{s}^2$ for the constant firing rate model, and $0.0249\,\mathrm{s}^2$ for the trend model; SEM $<0.0001\,\mathrm{s}^2$ in each case; mean five-fold cross-validated accuracy, 10 repeats, $N = 37\,\mathrm{neurons}$ and $12\,387\,\mathrm{trials}$), indicating that the population-level activity patterns of serotonin neurons are best understood as a code for value, consistent with our previous work (12).

The aggregate performance statistics presented above conceal significant heterogeneity in the activity patterns of serotonin neurons. Some neurons exhibited trial-to-trial fluctuations in inter-trial firing that are clearly consistent with value (Fig. 3C), while other individual neurons had a seemingly constant firing rate, and still others exhibited a slow decrease in firing over the course of the recording session that visually resembles a linear trend (Fig. 3E). Under our value coding model, this heterogeneity is explained in terms of variability in 1) the strength of the effect of value on firing (termed normalized firing rate gain $\beta_1' = \beta_1/\beta_0$) and 2) the learning timescale τ. The heterogeneity in learning timescale spanned a large range, from 5 min to almost one hour, with the majority of neurons displaying a learning timescale between 17 and 33 min (100-200 trials). The non-overlapping credibility intervals in Fig. 3E shows that for those neurons, the learning timescale was different.

Across the different cells, the distribution of maximum a posteriori estimates (MAP) of the correlation with value coding (β_1) shifted to higher gain from the distribution used as a prior (Fig. 3F). This indicates a general tendency of cells to correlate with value estimated at different timescales. More specifically, to calculate the fraction of cells that were positively correlated with value, we computed the MAP distribution for the fraction of cells having a gain β_0 larger than zero. We found a distribution concentrated around 80% of the cells (Fig. 3G). These estimates are show more consistent value coding than could be extracted from a previous study using classical trace conditioning (9), as shown in gray in Fig. 3F-G.

Turning to the statistics of the learning timescale across cells, Figure 3H shows samples from the prior and the posterior for the learning timescale parameter. As illustrated in Fig. 3D, the posterior estimates concentrated close to a timescale corresponding to 100 trials. This is made more precise by computing the posterior distribution for the mean learning timescale of the full population of serotonin neurons. Our posterior peaked sharply at a value of 125 trials, such that the average estimation timescale across the serotonergic population matches quite closely the estimate of 130 trials of similar duration from Ref. (15).

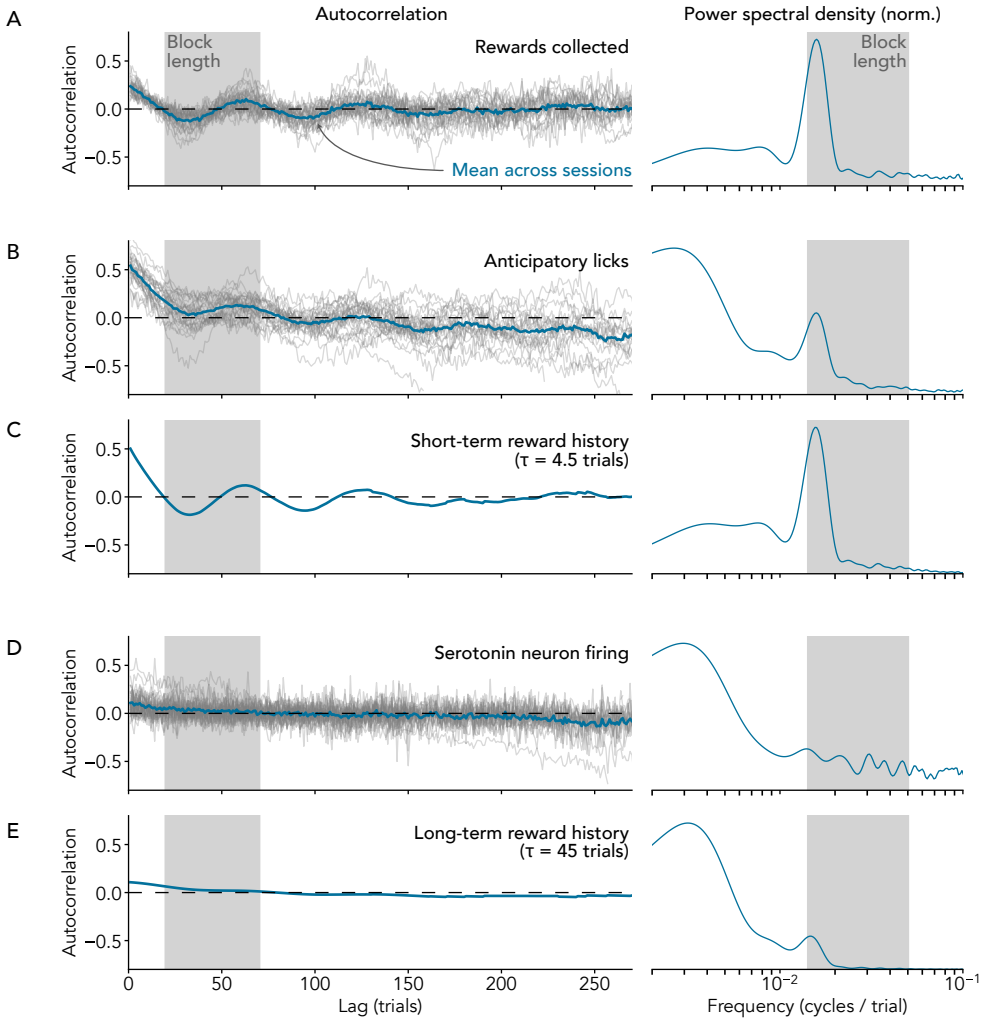

Figure 2. Reward collection, anticipatory licking behaviour, and serotonin neuron firing rates evolve slowly over time in a dynamic Pavlovian task.
Left column shows sample autocorrelation (individual sessions or neurons in gray, mean across sessions or neurons in blue), right column shows the estimated power spectrum (computed from the corresponding mean autocorrelation, see Section 5.3.3).
A Autocorrelation and power spectrum of reward collection. The power spectrum has a peak at $1/63\,\mathrm{cycles\,trial}^{-1}$. $N = 28$ sessions.
B Autocorrelation and power spectrum of anticipatory licking. The power spectrum has peaks at $1/377\,\mathrm{cycles\,trial}^{-1}$ and $1/63\,\mathrm{cycles\,trial}^{-1}$. $N = 28$ sessions.
C Autocorrelation and power spectrum of an exponentially-weighted moving average of reward collection using a time constant of $\tau = 4.5\,\mathrm{trials}$. Only the mean across $N = 28$ sessions is shown. The power spectrum has a peak at $1/64\,\mathrm{cycles\,trial}^{-1}$. Compare with B.
D Autocorrelation and power spectrum of serotonin neuron firing activity, quantified as the mean inter-spike interval during each inter-trial interval. The power spectrum has a peak at $1/339\,\mathrm{cycles\,trial}^{-1}$. $N = 37$ neurons.
E Autocorrelation and power spectrum of an exponentially-weighted moving average of reward collection using a time constant of $\tau = 45\,\mathrm{trials}$. Only the mean across $N = 28$ sessions is shown. The power spectrum has a peaks at $1/319\,\mathrm{cycles\,trial}^{-1}$ and $1/69\,\mathrm{cycles\,trial}^{-1}$. Compare with D.
Autocorrelations in C and E are normalized such that the autocorrelations at lag one match the means shown in B and D, respectively. For clarity, the autocorrorelation at zero lag is not shown. The 20–$70\,\mathrm{trial}$ block length used in the experiment and the corresponding frequency range of $1/70$–$1/20\,\mathrm{cycles\,trial}^{-1}$ are indicated in gray.

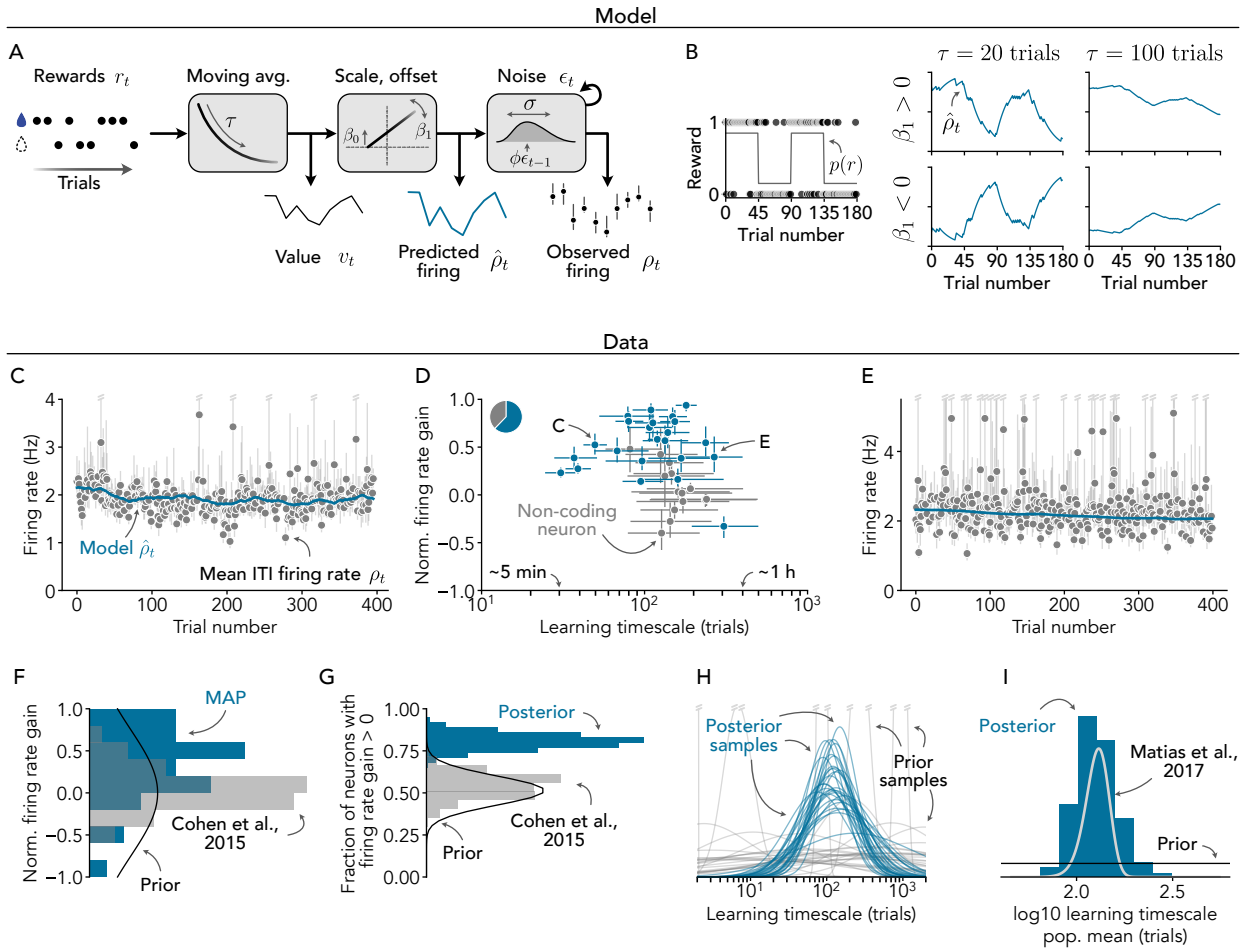

Figure 3. Value coding features of serotonin neurons.

A Generative model of serotonin neuron activity. Rewards are passed into an exponentially-weighted moving average with timescale τ to estimate value v_t (equivalent to learning value according to the Rescorla–Wagner rule) which is then scaled (by gain factor β_1) and offset (by intercept β_0) to produce an estimated firing rate $\hat{\rho}_t$. The observed firing rate ρ_t is a noisy reflection of $\hat{\rho}_t$. Note that the noise is autocorrelated, causing the observed firing rate to drift. See Methods for details.

B Effects of varying firing rate gain β_1 and learning timescale τ on the predicted firing rate $\hat{\rho}_t$. Simulated rewards from alternating $85\,\%$ and $15\,\%$ reward probability blocks are shown at left, and model outputs are shown at right. Note that for illustration purposes, here we have used more extreme reward probabilities and more regular block structure than in the actual experiment.

C Example of a value coding neuron with a short estimated learning timescale. Points and error bars indicate mean and SEM firing rate in each inter-trial interval. Blue line indicates the MAP value model prediction. For clarity, one outlier trial with very high estimated firing rate and the corresponding model prediction are not shown (note small gap in blue line near $375\,\mathrm{trials}$).

D Estimated normalized firing rate gain and learning timescale for each neuron. Points and error bars represent the median and IQR of the posterior, respectively. Neurons for which the $95\,\%$ credibility interval for the normalized firing rate gain includes zero are designated putatively non-value-coding and are indicated in gray; the fraction of neurons that neurons that fall in this category ($37.8\,\%$) is indicated in the pie chart. Exemplar neurons from C and E are labeled. For reference, timescales that correspond to approximately $5\,\mathrm{min}$ and $1\,\mathrm{h}$ are highlighted based on the average trial plus ITI duration ($9.3\,\mathrm{s}$).

E Example of a value coding neuron with a long estimated learning timescale. Data and model presented as in A. For clarity, two outlier trials with very high estimated firing rate and corresponding model predictions are not shown.

F Distribution of normalized firing rate gain across neurons. Distribution from the MAP model prediction is shown in blue, and the Bayesian prior used for model fitting is shown in black. For comparison, the distribution of normalized firing rate gain across neurons in ref. 9 is shown in gray.

G Fraction of $N = 37$ neurons with positive firing rate gain. Positive firing rate gain indicates that the neuron is activated when value is high. Full posterior distribution is shown in blue, and the Bayesian prior used for model fitting is shown in black. For comparison, the estimated fraction of neurons in the data of ref. 9 are activated when value is high is shown in gray (bootstrap distribution).

H Distribution of learning timescales in the full population of serotonin neurons. To illustrate uncertainty in this estimate, 30 draws from the posterior are shown. Samples from the prior are shown in gray.

I Mean learning timescale in the full population of serotonin neurons. Full posterior distribution is shown in blue, and the flat prior is shown in black. For comparison, the estimated mean from ref. 15 is shown in gray. Note close agreement with our results.

3 THE ROLE OF THIRST AS A CONFOUNDING FACTOR

It is possible, however, that confounding factors are integrated in the inference of the timescale and associated value code potency. We thus consider a control variable that relates to elapsed time or thirst, which we refer to as 'thirst' for short. It is defined as an inverse tally of the number of water rewards consumed (Section 5.2.2). This type of of variable can influence motivation, behaviour, and perception of reward in this task. Although the rate of reward collection, and therefore the rate of thirst reduction, fluctuates over the course of the experiment, the overall time-course of thirst generally resembles a linear trend (Fig. 4B). As expected, this means that if the learning timescale τ is large, value and thirst closely resemble each other (Fig. 4C), leading to a high degree of correlation/confounding between the two variables when $\tau \geq 45\,\mathrm{trials}$ (Fig. 4D). Thus, trend-like activity in a subset of serotonin neurons identified by our value coding model may actually be due to the effects of thirst. The timecourse of modeled thirst and value estimated on a long timescale are so similar, we sought to estimate what proportion of the neuron encode value in a manner that cannot be account for by thirst. To this end, we developed a second Bayesian model that can assess the degree with which a given neuron correlates with value or if its activity can be equally explained by value or thirst (see Methods; Mixture definition).

Fitting this mixture model to our data, we estimated that approximately half of neurons encode value (posterior mean $48.1\,\%$, $95\,\%$ CrI 27.3–$69.9\,\%$), while a large minority encode thirst (posterior mean $41.7\,\%$, $95\,\%$ CrI 21.0–$63.1\,\%$) and a small fraction are null coding (posterior mean $10.2\,\%$, $95\,\%$ CrI 0.5–$26.5\,\%$; see Fig. S1B). The high level of uncertainty on each of these estimates is mainly due to the moderate number of neurons in this dataset rather than statistical confounding between value, thirst, and null coding (for comparison, a frequentist $95\,\%$ CI on a proportion of $50\,\%$ would be 33.7–$66.3\,\%$ for $N = 37$ even in the absence of confounding). Null coding is less common than value and thirst coding (posterior probability $97.4\,\%$, prior probability $33.3\,\%$), but it is unclear whether value coding is more common than thirst coding or vice versa (posterior probability that value coding is more common $61.2\,\%$, prior probability $50.0\,\%$). We conclude that a large fraction of serotonin neurons encode value, but that a larger scale experiment is needed to determine whether this is the most common tuning feature.

In the subset of serotonin neurons that unambiguously encode value, we estimated that rewards are typically integrated over several dozen trials in order to estimate value (posterior mean learning timescale $72.9\,\mathrm{trials}$, $95\,\%$ CrI 45.1–$132.5\,\mathrm{trials}$; Fig. S1C). We still observed a very high level of variability in this feature across neurons, with the timescale over which serotonin neuron activity integrates rewards spanning roughly an order of magnitude (posterior mean standard deviation of the learning timescale 0.289 log10 units, $95\,\%$ CrI 0.136–0.558; Fig. S1D). Specifically, we estimate that the fastest $10\,\%$ of value coding serotonin neurons integrate rewards over fewer than $31.1\,\mathrm{trials}$ (posterior mean, $95\,\%$ CrI 14.2–$52.3\,\mathrm{trials}$; Fig. S1E) while the slowest $10\,\%$ integrate rewards over more than $170.8\,\mathrm{trials}$ (posterior mean, $95\,\%$ CrI 85.0–$574.2\,\mathrm{trials}$; Fig. S1F), with $80\,\%$ of neurons falling in the middle. Consistent with our model-free analysis (Fig. 2D and E), we estimate that $76.2\,\%$ (posterior mean, $95\,\%$ CrI 50.1–$94.7\,\%$) of value coding neurons integrate rewards over a period longer than the average block length in this experiment ($45\,\mathrm{trials}$). In sum, as one would expect by the act of removing those cells which are associated with a slow value estimation, the presence of thirst as a confounding factor reduces the proportion of neurons that unambiguously code for value and their estimation timescale is shorter.

3.1 Anticipatory licking depends on thirst and short-term reward history

We next sought to identify possible behavioural correlates of serotonin neuron activity. To determine whether long-term reward history or thirst influence the rate of anticipatory licking, we fitted a model in which licking is driven by a latent variable that reflects a mixture of value estimated over three different timescales, a fast timescale (1 trial), a medium timescale (10 trials) and a slow timescale of which matches the distribution of timescales observed in serotonin neurons (Fig. S1), and thirst (illustrated in Fig. 5A). Importantly, neither the fast nor the medium timescale match that observed in serotonergic cells (shortest timescale in Fig. 3D is 30 trials). Since this experiment contains both normal rewarded trials and a small number of cued unrewarded catch trials, we further modeled the lick rate as being dependent on the animal's perception of trial type (see Methods for details). Animals correctly perceived the trial type $93.6\,\%$ of the time (posterior mean, $95\,\%$ CrI 88.3–$96.8\,\%$, population parameter estimated using a hierarchical model fitted to $9693\,\mathrm{trials}$ across 28 sessions), licking in response to the cue for rewarded trials but not in response to the cue for catch trials. The fact that animals correctly perceived catch trials and adjusted the rate of anticipatory licking accordingly confirms that this behaviour is at least partially related to value.

The fitted behavioural model successfully extracted the main features of anticipatory licking visible in the raw data (Fig. 5B). In particular, fitted model predictions exhibited large modulations that reflect the block structure of the experiment, session-wide decreasing trends that may reflect satiety/thirst, and trial-to-trial fluctuations that reflect very short-term reward history. The first two of these features are consistent with our model-free analysis that pointed to anticipatory licking evolving over two timescales, one that reflects block structure and another that reflects a long-

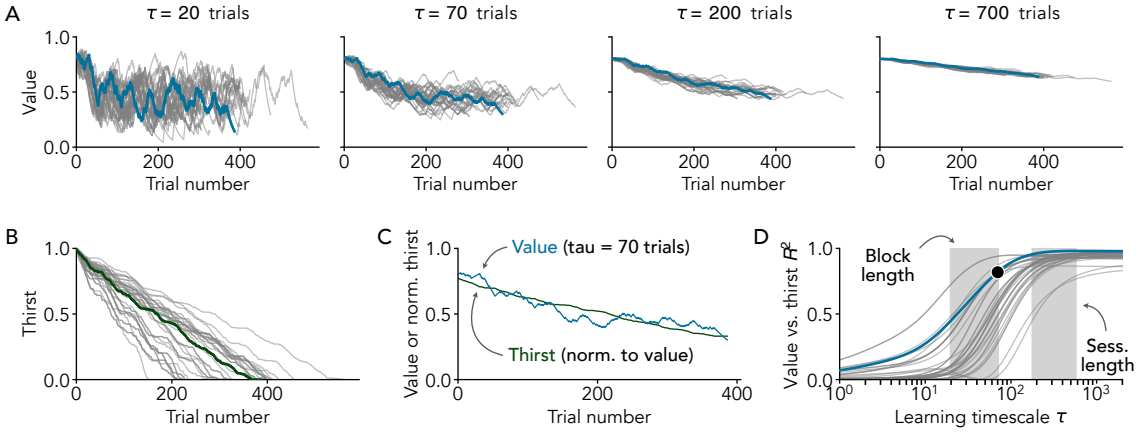

Figure 4. Value is confounded with thirst when learning is slow.
A Dynamics of value as a function of the learning timescale τ. Each line represents one experimental session. An exemplar session is highlighted in blue.
B Dynamics of thirst, defined as a quantity that decreases each time the animal consumes a water reward, starting at one and decreasing to zero over the course of a session. Each line represents one experimental session. The exemplar session from A is highlighted in green.
C Comparison of value and thirst for the exemplar session highlighted in A and B. Value (blue) is shown using a learning timescale of $\tau = 70\,\mathrm{trials}$. Thirst (green) is normalized so that its mean and standard deviation match those of value.
D Confounding between value and thirst as a function of the learning timescale τ. Confounding is measured using the squared coefficient of determination, representing the fraction of variance in value that can be explained by thirst (and vice-versa). Each line represents one session. The exemplar session from A and B is highlighted in blue, and the black dot corresponds to the curves shown in C.

term trend (Fig. 2). Overall, our model explained the majority of the observed variation in the rate of anticipatory licking ($67.2 \pm 14.3\,\%$ variance explained by MAP parameter estimates, mean $\pm$ SD, $N = 28$ sessions), further confirming that this behaviour depends on some mixture of value and thirst, while leaving open the possibility that other factors may play an important role.

Using the parameters estimated by our behavioural model, we next sought to determine the relative contributions of the value and thirst factors to anticipatory licking. We found that factors consistent with serotonin neuron activity, namely slow value and thirst, were collectively given a weight of $40.8\,\%$ ($95\,\%$ CrI 34.0–$47.7\,\%$) even though these factors both evolve over timescales considerably longer than the block length in this experiment. Within the non-serotonergic factors (collectively weighted at $59.2\,\%$, $95\,\%$ CrI 52.3–$66.0\,\%$; see Fig. 5C), fast value was given slightly more than half of the weight ($60.4\,\%$, $95\,\%$ CrI 51.1–$69.4\,\%$; see Fig. 5D), indicating that anticipatory licking depends meaningfully on value estimated over short and medium timescales. Within the serotonergic factors, the overwhelming majority of the weight was given to thirst ($93.6\,\%$, $95\,\%$ CrI 87.2–$97.7\,\%$; see Fig. 5E), indicating that value estimated over a timescale significantly longer than the average block length correlates less with anticipatory licking than our simple model of thirst. Taken together, anticipatory licking is explained partially by value integration occurring at a faster time scale than seen in serotonergic cells and partially by value integration happening at a timescale that matches the serotonergic cells, but the part of the behaviour matching the timescale seen in sertonergic cells is better explained by a model of thirst than a model of value.

4 DISCUSSION

We have reported a Bayesian model fitting approach for how individual serotonin neurons encode the evolving state value over time. Our model accounts for additional factors such as thirst or elapsed time and reveals that neurons vary in how strongly their firing reflects state value and in the timescales over which they update this value, with an average learning timescale of between 10 and 20 minutes. By also analyzing simultaneous behavioral data, specifically anticipatory licking, we find that behaviour is in most part linked to value signals learned on faster timescales than those encoded by serotonin neurons, while the part that is consistent in the serotonergic estimation timescale can be entirely explained by thirst as a confounding factor. These results offer a detailed quantification of how individual serotonin neurons represent value and its learning dynamics, uncovering a mismatch between the value estimation timescales in serotonergic cells and those influencing behaviour.

Relationship between neural activity and behaviour

There is a significant amount of variation in both neural activity and behaviour that is not explained by the models considered here. We have not attempted to determine whether the unexplained variation is correlated between neural activity and behaviour in the corresponding sessions, leaving open the possibility that some signal other than thirst or value may be responsible for driving both neural activity and behaviour. In principle, we could figure this out by looking for correlations in the residuals of the neural activity and behavioural models, but we determined that this was not feasible due to technical challenges. In particular, the

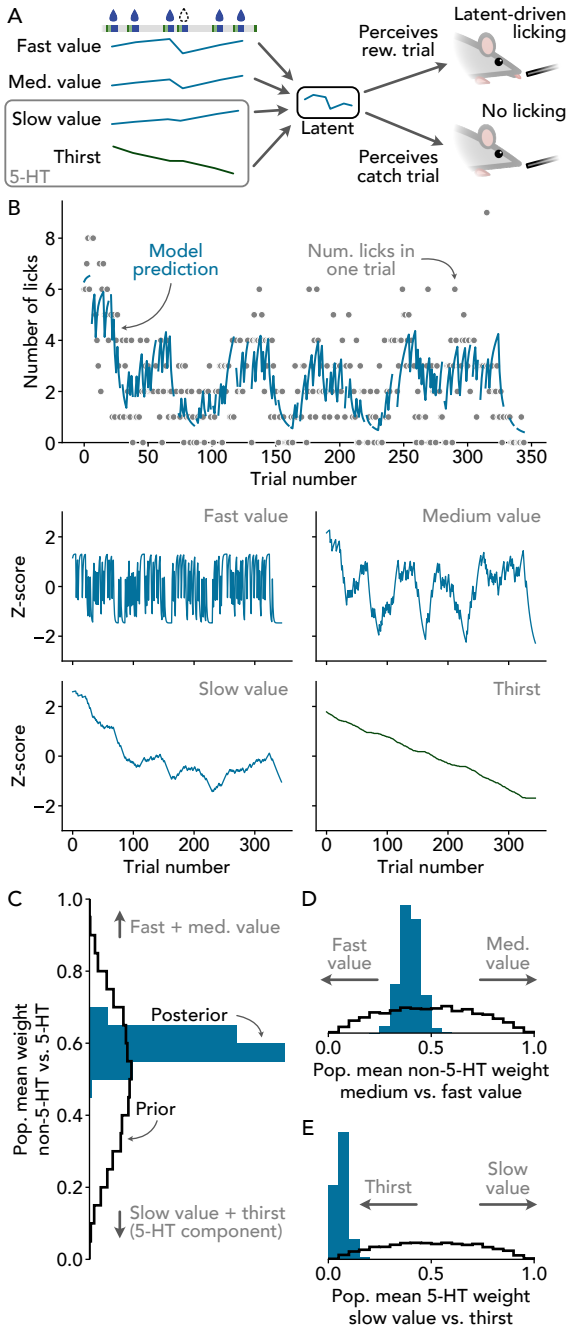

Figure 5. Analysis of anticipatory licking.
A Overview of behavioural model. Value computed based on a short reward history (fast value, $\tau = 1\,\mathrm{trial}$), medium reward history (medium value, $\tau = 10\,\mathrm{trials}$), long reward history (slow value, τ chosen to match distribution found in serotonin neurons, mean of approximately $72\,\mathrm{trials}$), and thirst are combined to form a latent that drives anticipatory licking if the animal believes it is in a regular, rewarded trial rather than a catch trial. See Methods for details.
B Data, MAP model prediction, and corresponding latent components for an exemplar session. For clarity, perceived catch trials are not shown (note gaps in blue line).
C Mean weight assigned to 5-HT-related (slow value and thirst) vs. non-5-HT-related (fast and medium value) latent components across $N = 28$ sessions.
D Mean weight assigned to medium value vs. fast value across $N = 28$ sessions.
E Mean weight assigned to slow value vs. thirst across $N = 28$ sessions.

fact that the residuals in the neural model depend on the coding scheme used by each neuron, which is treated as uncertain in our approach, and the fact that the experiment contains catch trials which give the behavioral residuals a structure that is very difficult to model.

Value of the trace period period We have compared trial-to-trial dynamics of anticipatory licking, which mostly takes place in the trace period, with the trial-to-trial fluctuations of serotonergic neuron activity, which were recorded during the inter-trial period. This approach is supported by the prospective value theory, where the value in the inter-trial period scales with the value in the trace period. This is statements, however, arises from the dynamics of value after learning. The dynamics of value during learning depends heavily on the type of learning algorithm that is being used. For example, the use eligibility traces in the learning algorithm will alter the speed of value learning. Importantly, some learning algorithms can, in principle, make the learning timescale of periods far from the reward different from that close to the reward. Experimental observations of Ref. (12) do not support this possibility as, on a 5 trial horizon, the speed at which the trace period activity of serotonin neurons changed matched the speed measured in the inter-trial period.

Estimation timescale of dopaminergic activity Matias et al. (15) found that the reward associations reflected in bulk serotonergic activity changed over the course of approximately 130 trials. This was an order of magnitude slower that the reward associations reflected in bulk dopaminergic activity, which changed over the course of approximately 20 trials (see also Ref. (10)). Given that their trials were about the same duration as the trials in the experiment analyzed here (9-14 s in Ref. (15) vs. an average of 9 s in Ref. (16)), we can expect that dopaminergic activity would change over the time course of 20 trials in the mice recorded here. It is interesting to note that this learning timescale, by beeing close to the medium timescale used in Fig. 5, showed some correlation with anticipatory (Fig. 5D), suggesting that the value estimation timescale reflected in dopaminergic cells is more closely associated with behaviour than the value estimation timescale seen in serotonergic cells (15).

Capturing activity fluctuations in vivo Mathematical models have been developed for predicting activity fluctuations of single neurons in vitro (18–24) and in vivo (25–28). While it is difficult to compare model performance across different recording modalities and different metrics (29), we note that models of cortical neurons tend to achieve high prediction accuracy in slice (20, 21, 30, 31) but very low prediction accuracy in vivo (25). This discrepancy may be attributed to many missing latent factors (27, 32), population multiplexing of information (33–35) as well as the difficulty of capturing the relationship with cortical cell's inputs

(36–42). Serotonergic neurons on the other hand are not particularly well captured by integrate-and-fire models in vitro (24). In this context, the variance explained achieved by our model for predicting activity fluctuations in vivo is surprisingly good. In contrast to the cortex, serotonergic neurons of the dorsal raphe are not numerous and may the population may not be able to support population multiplexing (33). These neurons are also much less bursty and cortical neurons (43), indicating that they may also not be able to support burst multiplexing (35).

5 METHODS

5.1 *In vivo* experiment

The Pavlovian conditioning experiment with dynamic reward probabilities has been reported previously by Grossman et al. (16). A brief summary follows.

Odour-cued water rewards were delivered to five water-restricted C57/BL6 mice (four male, one female) over the course of 28 recording sessions, each of which lasted $\sim 1\,\mathrm{h}$. Rewards were delivered stochastically with a probability of 0.2, 0.5, or 0.8 which changed dynamically according to a block structure. Blocks lasted 20–$70\,\mathrm{trials}$ (uniform distribution), and the first block of each session always had a reward probability of 0.8. Catch trials consisting of a different odour cue followed by no reward were randomly interleaved with regular trials throughout the experiment at an average rate of $1/20\,\mathrm{trials}$. Catch trials and trials in which a reward was available but not collected were deemed unrewarded for the purposes of our analysis. Mice collected 0.45 ± 0.04 rewards per trial and completed 356 ± 88 trials per session (mean $\pm$ SD across sessions). Inter-trial intervals followed an exponential distribution with mean $3.\overline{3}\,\mathrm{s}$. Spike times of optogenetically-tagged serotonin transporter-expressing neurons were collected using tetrode recordings at a rate of 1–$3\,\mathrm{neurons}$ per session (mean $1.32\,\mathrm{neurons}$).

Experimental and surgical procedures were performed in compliance with the *National Institutes of Health Guide for the Care and Use of Laboratory Animals* and were approved by the Animal Care and Use Committee at Johns Hopkins University.

5.2 Models

5.2.1 Value

We define value as the expected reward in the upcoming trial, estimated using the Rescorla–Wagner learning rule

$$v_t = v_{t-1} + \underbrace{(1 - e^{-1/\tau})}_{\text{learning rate}}(r_t - v_{t-1}), \tag{4}$$

where v_t is the value during the inter-trial interval immediately following the reward r_t on trial t, and τ is the learning timescale.

The model is written in terms of the learning timescale τ instead of the more conventional learning rate α because the learning timescale can be interpreted as the number of trials over which past rewards are integrated in order to estimate the expected reward in the upcoming trial. However, it is important to note that the two are equivalent: $\tau = -1/\ln(1-\alpha)$.

5.2.2 Thirst

We define thirst as a quantity h_t that decreases by a fixed amount each time the animal consumes a water reward. We arbitrarily set $h_0 = 1$ and $h_T = 0$, where T is the number of trials, such that

$$h_t = 1 - \frac{\sum_{i=1}^{t} r_i}{\sum_{j=1}^{T} r_j}, \tag{5}$$

where $r_x \in \{0, 1\}$ is the binary water reward received on trial x. Reward collection had the tendency to follow the time of since the start of the experiment, but showed individual differences across mice Fig. S3.

5.2.3 Latent variable model of neural activity

We model neural activity $\bar{\iota}_t$ in terms of a latent firing rate ρ_t driven by a latent input x_t. We assume a linear effect of the input on the firing rate

$$\rho_t = \beta_0 + \beta_1 x_t, \tag{6}$$

where β_0 is the baseline firing rate when the input is zero and β_1 is the gain. In turn, we model the mean inter-spike interval $\bar{\iota}_t = 1/M_t \sum_{j=1}^{M_1} \iota_{t,j}$, where $\iota_{t,1}, \iota_{t,2}, \ldots, \iota_{t,M_t}$ are individual inter-spike intervals in trial t, in terms of the inverse firing rate, or frequency $1/\rho_t$

$$\bar{\iota}_t = \frac{1}{\rho_t} + \epsilon_t, \tag{7}$$

where ϵ_t is the residual at time t due to a combination of spiking noise, measurement noise, and drift. This quantity could only be estimated when there were at least two spikes in the data (Fig. S4). We found that this method was more stable in such sparse firing than using spike trains and their associated likelihood functions (as in Refs. (20, 21, 24, 44)).

Residual models To simplify comparisons with null models based on constant firing rates or linear time trends (see Section 2.2), for the purposes of cross-validation we model the residuals ϵ_t as Gaussian white noise with time-varying variance

$$\epsilon_t \sim \mathcal{N}\left(0, \sigma_\iota^2 / M_t\right), \tag{8}$$

where σ_ι is the standard deviation of a single inter-spike interval and M_t is the number of inter-spike intervals in trial t (such that $\mathrm{Var}[\bar{\iota}_t] = \mathrm{Var}[\iota_t]/M_t$).

During inference, we model the residuals according to a third-order autoregressive process in order to properly account for drift (Fig. S5). The process is defined as

$$\epsilon_t = \xi_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \phi_3 \epsilon_{t-3}, \tag{9}$$

where $\xi_t \sim \mathcal{N}(0, \sigma_\iota^2/M_t)$ is Gaussian white noise driving the residuals and ϕ_j are weights constrained to $[-1, 1]$.

In both cases, we use Gaussian noise as the basis of our residual model. We have chosen Gaussian noise for two reasons: 1) inter-spike intervals recorded *in vitro* in identified serotonin neurons are approximately Gaussian (data not shown), and 2) the Lyapunov central limit theorem guarantees that the *mean* inter-spike interval $\bar{\iota}$ that is the focus of our analysis converges to a Gaussian distribution (with variance proportional to $1/M_t$) under mild conditions.

5.3 Data analysis

5.3.1 Quantification of anticipatory licking

We measure anticipatory licking as the number of licks during the $2\,\mathrm{s}$ period between cue and reward onset.

5.3.2 Quantification of neural activity

Neural activity is measured as the mean inter-spike interval $\bar{\iota}_t$ during each inter-trial interval. For inter-trial intervals with fewer than two spikes, the mean inter-spike interval is considered missing data.

5.3.3 Estimation of power spectra

To estimate the power spectra of reward collection, anticipatory licking behaviour, and serotonin neuron firing activity, we consider each of these as stochastic processes and extract the power spectra as the smoothed Fourier cosine transforms of the corresponding mean sample autocorrelation functions (45), ignoring missing data.

For example, taking $x_{1,[i]}, x_{2,[i]}, \ldots, x_{T_{[i]},[i]}$ to be the number of anticipatory licks on trials 1 to $T_{[i]}$ for mouse i, we compute the autocorrelation at lag k up to 400 as

$$c_{k,[i]} = \frac{1}{s^2_{[i]}\tilde{T}_{[i]}} \sum_{j=1}^{T_{[i]}-k} \left(x_{j,[i]} - \bar{x}_{[i]}\right)\left(x_{j+k,[i]} - \bar{x}_{[i]}\right), \quad (10)$$

where $s^2_{[i]}$ is the sample variance of $x_{[i]}$, $\bar{x}_{[i]}$ is the sample mean, and $\tilde{T}_{[i]}$ is the number of $(x_{j,[i]}, x_{j+k,[i]})$ pairs for which neither $x_{j,[i]}$ nor $x_{j+k,[i]}$ are missing values. Note that we take the term inside the sum to be zero if either $x_{j,[i]}$ or $x_{j+k,[i]}$ is missing, and we consider $c_{k,[i]}$ itself to be missing if $\tilde{T} < 50$. In the case of anticipatory licking, we treat the number of licks during catch trials as missing data. For firing activity, we treat the mean inter-spike interval during inter-trial intervals with fewer than two spikes as missing data. Reward data is never missing.

Given the sample autocorrelations of anticipatory licking $c_{k,[1]}, c_{k,[2]}, \ldots, c_{k,[N]}$ for $N = 28$ recording sessions, we estimate the underlying autocorrelation function of the licking process as the mean of the sample autocorrelations

$$\bar{c}_k = \frac{1}{\tilde{N}_k} \sum_{i=1}^{N} c_{k,[i]}, \quad (11)$$

where $\tilde{N}_k$ is the number of $c_{k,[i]}$ that are not missing, and where the term inside the sum is taken to be zero if $c_{k,[i]}$ is missing, as above. For firing activity, $N = 37$ neurons rather than 28 sessions.

Finally, the power spectrum is estimated from the autocorrelation function $\bar{c}_k$ using the smoothed Fourier cosine transform

$$\hat{p}(f) = 2\left[1 + 2\sum_{k=1}^{K} \lambda_k \bar{c}_k \cos(2\pi f k)\right], \quad (12)$$

where $\hat{p}(f)$ denotes the estimated power at frequency f rather than a probability (following the notation of ref. 45), $K = 399$ is the maximum lag, and the λ_k are smoothing coefficients. The smoothing coefficients are set according to $\lambda_k = \exp[-(k^4/(K/2.5)^4)]$, which is close to one for lags less than 100 and close to zero for lags greater than 200.

5.3.4 Hierarchical Bayesian model of value coding

Notation This section describes a model of data from multiple neurons. Sub-scripts in square brackets $_{[\cdot]}$ are used to indicate quantities specific to a particular neuron. For example, $\tau_{[i]}$ is the learning timescale for neuron i.

Posterior definition The log posterior density ℓ_p of our model of serotonin neuron activity is defined as

$$\ell_p = C + \ln[p(\theta)] + \sum_{i}^{N} \sum_{t=1}^{T_{[i]}} \ln\left[p\left(\bar{\iota}_{t,[i]} \mid \theta\right)\right], \quad (13)$$

where C is a constant term that represents the log probability density of the data (the denominator in Bayes' rule), $p(\theta)$ is the prior, $N = 37$ neurons, $T_{[i]}$ is the number of trials for neuron i, and $p\left(\bar{\iota}_{t,[i]} \mid \theta\right)$ is the likelihood of the mean inter-spike interval for neuron i on trial t given the parameters θ. Whenever $\bar{\iota}_{t,[i]}$ is missing, we take the corresponding likelihood to be one.

Parameterization and priors The choice of model parameterization has a dramatic effect on whether/how quickly the Hamiltonian Monte-Carlo algorithm used for Bayesian inference will converge (46), but does not affect the inferences that can be drawn about quantities of interest at convergence. For technical reasons, we have parameterized the model in terms of the following quantities:

- The learning timescale on a log10 scale $\log \tau_{[i]}$.
- Normalized initial value for each neuron $v'_{0,[i]} = \frac{v_{0,[i]} - 0.555}{1 - 0.555}$.
- Firing rate intercept population log shape and scale parameters $\ln a$, $\ln 1/b$.
- Firing rate intercept for each neuron $\beta_{0,[i]}$.
- Normalized firing rate slope for each neuron $\beta'_{1,[i]} = \beta_{1,[i]}/\beta_{0,[i]}$.
- Estimated ISI standard deviation for each neuron $\sigma_{\iota,[i]}$.
- Drifting noise process coefficients for each neuron $\phi_{1,[i]}, \phi_{2,[i]}, \phi_{3,[i]}$ (only for inference).

Priors for all parameters are listed in Table 1.

Likelihood The likelihood $p\left(\bar{\iota}_{t,[i]} \mid \theta\right)$ is derived by substituting the value v_t from Eq. (4) into the inter-spike interval model given by Eq. (7) evaluating the probability density of the residuals ϵ_t under the appropriate noise model. For cross-validation, we use the Gaussian white noise given in Eq. (8). For Bayesian inference, we use the drifting noise process given in Eq. (9).

5.3.5 Hierarchical Bayesian mixture model of neural activity

Notation This section describes a model of data from multiple neurons in terms of multiple possible coding schemes. In addition to sub-scripts in square brackets used to indicate quantities specific to a particular neuron, superscripts in round brackets $^{(\cdot)}$ are used to indicate quantities or functions specific to a particular coding scheme. For example, $p^{(\text{val})}\left(x_{[2]}\right)$ denotes the probability density of x from neuron 2 under the value coding scheme.

Mixture definition To measure the value and thirst coding features of serotonin neurons, we fitted a mixture model in which the time-dependent activity of each neuron $\bar{\iota}_{t,[i]}$ reflects either value $v_{t,[i]}$, thirst $h_{t,[i]}$, or nothing (i.e., the underlying firing rate is constant), with corresponding likelihoods $p^{(\text{val})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$, $p^{(\text{thr})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$, and $p^{(\text{null})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$. Given the prior $p(\theta)$, the overall log posterior density of the mixture model is therefore given by

$$\begin{aligned}\ell_p = C + \ln\left[p(\theta)\right] + \sum_{i=1}^{N} \ln\Bigg[\lambda^{(\text{val})} \prod_{t=1}^{T_{[i]}} p^{(\text{val})}\left(\bar{\iota}_{t,[i]} \mid \theta\right) \\ + \lambda^{(\text{thr})} \prod_{t=1}^{T_{[i]}} p^{(\text{thr})}\left(\bar{\iota}_{t,[i]} \mid \theta\right) \\ + \lambda^{(\text{null})} \prod_{t=1}^{T_{[i]}} p^{(\text{null})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)\Bigg], \quad (14)\end{aligned}$$

where C is a constant that represents the log probability density of the data (the denominator in Bayes' rule), $N = 37$ neurons, $T_{[i]}$ is the number of trials for neuron i, and the λ are mixing coefficients on the unit simplex $\sum_z \lambda^{(z)} = 1$. When $\bar{\iota}_{t,[i]}$ is missing, we take the corresponding likelihoods to be 1. Details on the prior and likelihoods are given below.

Parameterization The mixture model is parameterized in terms of the following quantities:

- Mixing coefficients $\lambda^{(\text{val})}$, $\lambda^{(\text{thr})}$, $\lambda^{(\text{null})}$.
- Learning timescale population parameters $\mu_{\log\tau}$, $\sigma_{\log\tau}$.
- Standardized learning timescale for each neuron $\log\tau'_{[i]} = \frac{\log\tau_{[i]} - \mu_{\log\tau}}{\sigma_{\log\tau}}$. This parameterization is chosen to avoid a Neal's funnel effect when $\sigma_{\log\tau}$ is small (46).
- Normalized initial value for each cell $v'_{0,[i]} = \frac{v_{0,[i]} - 0.555}{1 - 0.555}$.
- Firing rate intercept population shape and scale parameters $\ln a$, $\ln 1/b$.
- Estimated mean firing rate for each neuron $B_{[i]}$.
- Normalized firing rate slopes for each neuron under the value and thirst models $\beta^{(\text{val})\prime}_{1,[i]}$, $\beta^{(\text{thr})\prime}_{1,[i]}$, where $\beta'_1 = \beta_1/\beta_0$.
- Estimated ISI standard deviation for each neuron $\sigma_{\iota,[i]}$.
- Drifting noise process coefficients for each neuron $\phi_{1,[i]}, \phi_{2,[i]}, \phi_{3,[i]}$ (only for inference).

Firing rate intercepts The firing rate intercepts $\beta^{(z)}_{0,[i]}$ are an important source of prior information in our model (see Table 1) even though they do not appear directly in its parameterization (see above). Under the value model, the firing rate intercept is derived from the estimated mean firing rate $B_{[i]}$ and normalized firing rate slope $\beta^{(\text{val})\prime}_{1,[i]}$ according to

$$\beta^{(\text{val})}_{0,[i]} = f\left(B_{[i]}\right) = \frac{B_{[i]}}{1 + \beta^{(\text{val})\prime}_{1,[i]}\left\langle v_{t,[i]}\right\rangle}, \quad (15)$$

where $\langle v_{t,[i]}\rangle$ is the mean value across trials with at least one ISI for neuron i. Note that the absolute derivative of this transformation is

$$\left|\frac{d}{dB_{[i]}} f\left(B_{[i]}\right)\right| = \frac{1}{\left|1 + \beta^{(\text{val})\prime}_{1,[i]}\left\langle v_{t,[i]}\right\rangle\right|}. \quad (16)$$

Under the thirst and null models, the firing rate intercept is equal to the baseline firing rate

$$\beta^{(\text{thr})}_{0,[i]} = \beta^{(\text{null})}_{0,[i]} = B_{[i]}. \quad (17)$$

To incorporate prior information about the firing rate intercepts $\beta^{(\text{val})}_{0,[i]}, \beta^{(\text{thr})}_{0,[i]}, \beta^{(\text{null})}_{0,[i]}$ into our model, we need to decompose the overall prior $p(\theta)$ into a set of terms that depend on the β_0. First, we factorize $p(\theta)$ to obtain the prior probability of the mean firing rates $B_{[i]}$ conditional on the remaining parameters θ^{c}

$$p(\theta) = p\left(\theta^{\text{c}}\right) \prod_{i=1}^{N} p\left(B_{[i]} \mid \theta^{\text{c}}\right). \quad (18)$$

Next, we marginalize the conditional probability of the mean firing rate over the coding schemes

$$\begin{aligned}p\left(B_{[i]} \mid \theta^{\text{c}}\right) = \frac{1}{3} p^{(\text{val})}\left(B_{[i]} \mid \theta^{\text{c}}\right) \\ + \frac{1}{3} p^{(\text{thr})}\left(B_{[i]} \mid \theta^{\text{c}}\right) + \frac{1}{3} p^{(\text{null})}\left(B_{[i]} \mid \theta^{\text{c}}\right),\end{aligned}$$

where we have assigned each of the three coding schemes an equal prior probability of $1/3$. Using Eqs. (15) to (17), the above can be rewritten in terms of the β_0 under each model

$$p\left(B_{[i]} \mid \theta^{\mathsf{c}}\right) = \frac{1}{3} p\left(\beta_{0,[i]}^{(\mathsf{val})} \,\middle|\, \theta^{\mathsf{c}}\right) \frac{1}{\left|1 + \beta_{1,[i]}^{(\mathsf{val})\prime} \langle v_{t,[i]} \rangle\right|} + \frac{1}{3} p\left(\beta_{0,[i]}^{(\mathsf{thr})}\right) + \frac{1}{3} p\left(\beta_{0,[i]}^{(\mathsf{null})}\right), \quad (19)$$

where the first term includes an adjustment of $\left|\frac{d}{dB_{[i]}} f\left(B_{[i]}\right)\right|$ to account for the change of variables from $B_{[i]}$ to $\beta_{0,[i]}^{(\mathsf{val})} = f\left(B_{[i]}\right)$. The second and third terms do not need an adjustment for the change of variables, nor do the slopes depend on other parameters θ^{c}.

Priors Using Eqs. (18) and (19), the overall prior for the mixture model can be written as

$$p(\theta) = p\left(\lambda^{(\mathsf{val})}, \lambda^{(\mathsf{thr})}, \lambda^{(\mathsf{null})}, \mu_{\log \tau}, \sigma_{\log \tau}, \ln a, \ln b\right) \prod_{i=1}^{N} p\left(\log \tau'_{[i]}, v'_{0,[i]}, \beta_{1,[i]}^{(\mathsf{val})\prime}, \beta_{1,[i]}^{(\mathsf{thr})\prime}, \sigma_{\iota,[i]}\right) p\left(B_{[i]} \,\middle|\, v_{0,[i]}, \log \tau'_{[i]}, \mu_{\log \tau}, \sigma_{\log \tau}, \beta_{1,[i]}^{(\mathsf{val})\prime}, \ln a, \ln b\right),$$

where all of the parameters are conditionally independent except for $\lambda^{(\mathsf{val})}, \lambda^{(\mathsf{thr})}, \lambda^{(\mathsf{null})}$ due to the simplex constraint. The exact form of the prior on each parameter is listed in Table 1.

Table 1. Summary of quantities of interest in the hierarchical Bayesian mixture model of serotonin neuron activity.

Name	Description	Symbol(s)	Prior	Justification for prior
Mixing coefficients	Proportion of neurons that encode value, thirst, or nothing.	$\lambda^{(\text{val})}, \lambda^{(\text{thr})}, \lambda^{(\text{null})}$	Uniform on the unit 2-simplex	Non-informative default prior.
Learning timescale	Period over which rewards are integrated to estimate value for each neuron.	$\tau_{[i]}$	$\log \tau_{[i]} \sim \mathcal{N}(\mu_{\log \tau}, \sigma^2_{\log \tau})\ \forall i$	Single neuron fits suggested log normal distribution.
Learning timescale pop. mean	Mean learning timescale in the population of value-coding serotonin neurons on a log10 scale.	$\mu_{\log \tau}$	$\mu_{\log \tau} \sim \text{Uniform}[0, \log 2000]$	Restricted to $>1\,\text{trial}$ and $<2000\,\text{trials}$, but otherwise non-informative.
Learning timescale pop. SD	Standard deviation of learning timescales on a log10 scale across the population of value-coding serotonin neurons.	$\sigma_{\log \tau}$	$\sigma_{\log \tau} \sim \text{Uniform}[0, 1]$	Negative SD is impossible, and SD above one order of magnitude is nonsensical for a 200–$500\,\text{trial}$ experiment. Non-informative otherwise.
Initial value	Estimate of value before the start of the first trial for each neuron.	$v_{0,[i]}$	$\frac{v_{0,[i]}-0.555}{1-0.555} \sim \text{Beta}[2.45, 2]\ \forall i$	Weakly-informative prior with mean 0.8 reflects ground truth $v_{0,[i]} = 0.8\ \forall i$. We further restrict $v_{0,[i]} > 0.555$ because the model is not identified if $v_{0,[i]}$ can be below the mean of $v_{t,[i]}$ (0.37–0.54), and $v_{0,[i]} < 1$ by definition.
Firing rate intercept	Firing rate when value is zero, when thirst is equal to its average level, or the average firing rate of non-coding neurons.	$\beta^{(\text{val})}_{0,[i]}, \beta^{(\text{thr})}_{0,[i]}, \beta^{(\text{null})}_{0,[i]}$	$\beta^{(z)}_{0,[i]} \sim \text{Gamma}[a, b]\ \forall z$	Strongly informative. Directly constrained to prior data from ref. 9.
Firing rate intercept pop. parameters	Shape and rate parameters for Gamma-distributed firing rate intercepts.	a, b	$\begin{bmatrix} \ln a \\ \ln 1/b \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0.793 \\ 0.430 \end{bmatrix}, \begin{bmatrix} 0.0555 & -0.0492 \\ -0.0492 & 0.0500 \end{bmatrix} \right)$	Strongly informative. Directly constrained to prior data from ref. 9. Covariance obtained via bootstrap. See Supplementary Note B.
Firing rate slope	Change in firing rate (in Hz) for a unit change in value or normalized thirst.	$\beta^{(\text{val})}_{1,[i]}, \beta^{(\text{thr})}_{1,[i]}$	$\frac{\beta^{(z)}_{1,[i]}}{\beta^{(z)}_{0,[i]}} \sim \mathcal{N}(0, 0.75^2)\ \forall i, z$ $\lvert\beta^{(\text{val})}_{1,[i]}\rvert < 1, \quad \lvert\beta^{(\text{thr})}_{1,[i]}\rvert < 4$	Symmetry around zero reflects default assumption. Restrictions ensure that firing rate is non-negative given maximum possible value and thirst. SD is weakly informative based on data from ref. 9.
Mean firing rate	Estimated mean firing rate of neuron i across all trials.	$B_{[i]}$	Determined by priors on $\beta^{(\text{val})}_{0,[i]}, \beta^{(\text{val})}_{1,[i]}, \beta^{(\text{thr})}_{0,[i]}, \beta^{(\text{null})}_{0,[i]}$ using $\lambda^{(z)} = 1/3\ \forall z$, subject to $B_{[i]} > 0\ \forall i$	
ISI SD	Standard deviation of a single ISI in neuron i.	$\sigma_{\iota,[i]}$	$\sigma_{\iota,[i]} \sim \text{Uniform}[0, \infty]\ \forall i$	Non-informative default prior.

Value likelihood The likelihood under the value model $p^{(\text{val})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$ is derived by substituting $v_{t,[i]}$ from Eq. (4) (using $v_{0,[i]}$ for the initial condition) in for x_t in Eq. (6), yielding

$$\rho^{(\text{val})}_{t,[i]} = \beta^{(\text{val})}_{0,[i]} + \beta^{(\text{val})}_{1,[i]} v_{t,[i]}.$$

The firing rate $\rho^{(\text{val})}_{t,[i]}$ is then substituted into the inter-spike interval model given in Eq. (7), yielding

$$\bar{\iota}^{(\text{val})}_{t,[i]} \sim \mathcal{N}\left(1/\rho^{(\text{val})}_{t,[i]}, \sigma^2_{\iota,[i]}/M_{t,[i]}\right),$$

which defines the Gaussian likelihood of $\bar{\iota}^{(\text{val})}_{t,[i]}$.

Thirst likelihood The likelihood under the thirst model $p^{(\text{thr})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$ is derived analogously to the value likelihood described above, using $x_t = \left(h_{t,[i]} - \langle h_{t,[i]} \rangle\right) \times 0.381$ as the input to the firing rate model from Eq. (6). The thirst signal is centered so that $\beta^{(\text{thr})}_{0,[i]} = B_{[i]}$ and scaled by a factor of 0.381 to approximately match the average dynamic range of value $v_{t,[i]}$ using $\tau_{[i]} = 70\,\text{trials}$ and $v_{0,[i]} = 0.8\ \forall\, i$.

Non-coding likelihood The likelihood under the non-coding model $p^{(\text{null})}\left(\bar{\iota}_{t,[i]} \mid \theta\right)$ is derived analogously to the value and thirst likelihoods using $x_t = 0\ \forall\, t$ in Eq. (6).

5.3.6 Hierarchical Bayesian model of anticipatory licking

Notation Similar to previous sections, we use subscripts in square brackets $\cdot_{[i]}$ to denote quantities that are specific to session i.

Since this experiment contains a mixture of regular trials with stochastic rewards and catch trials in which reward was delivered, we model the number of anticipatory licks in trial t of session i, $l_{t,[i]}$, as being dependent on whether the animal perceives that it is in a regular or catch trial. To indicate this dependence, we use the indicator $z_{t,[i]} = 1$ to denote the event that the animal is in a regular trial and $z^{\text{c}}_{t,[i]} = 0$ to denote the event that the animal is in a catch trial. Similarly, we use $\hat{z}_{t,[i]} = 1$ to indicate that the animal *perceives* that it is in a regular trial, and $\hat{z}^{\text{c}}_{t,[i]} = 0$ to indicate that the animal *perceives* that it is in a catch trial.

Perceptual model of licking We model the number of anticipatory licks $l_{t,[i]}$ a nonlinear function of a standardized latent variable $f\left(\tilde{\xi}_{t,[i]}\right)$ (defined below) if the animal perceives that it is in a regular trial, and as being near zero if the animal perceives that is in a catch trial

$$l_{t,[i]} = \begin{cases} f\left(\tilde{\xi}_{t,[i]}\right) + \epsilon_{t,[i]} & \text{if } \hat{z}_{t,[i]} \\ \epsilon_{t,[i]} & \text{if } \hat{z}^{\text{c}}_{t,[i]} \end{cases} \tag{20}$$

where $\epsilon_{t,[i]} \sim \mathcal{N}\left(0, \sigma^2_{l,[i]}\right)$ is noise. The log posterior probability density of the licking model is

$$\begin{aligned} \ell_p = C + \ln[p(\theta)] + \sum_{i=1}^{N} \sum_{t=1}^{T_{[i]}} \ln \Bigg[& \left[\pi_{[i]} p(l_{t,[i]} \mid \hat{z}_{t,[i]}) + (1-\pi_{[i]}) p(l_{t,[i]} \mid \hat{z}^{\text{c}}_{t,[i]})\right] z_{t,[i]} \\ & + \left[(1-\pi_{[i]}) p(l_{t,[i]} \mid \hat{z}_{t,[i]}) + \pi_{[i]} p(l_{t,[i]} \mid \hat{z}^{\text{c}}_{t,[i]})\right] z^{\text{c}}_{t,[i]} \Bigg], \end{aligned} \tag{21}$$

where C is a constant that represents the log probability of the data; $p(\theta)$ is the prior; N is the number of recording sessions; $T_{[i]}$ is the number of trials in session i; $\pi_{[i]}$ is the probability that the animal perceives the trial type correctly in session i; and the likelihoods come from the normal densities implied by the noise terms in Eq. (20). We model the variability in $\pi_{[i]}$ across sessions using a beta distribution

$$\pi_{[i]} \sim \text{Beta}[a_\pi, b_\pi] \tag{22}$$

from which we can estimate the population mean correct perception probability reported in the main text.

Licking in perceived regular trials For perceived regular trials, we model the number of anticipatory licks as a non-linear function of a standardized latent variable $\tilde{\xi}_{t,[i]}$ (defined below)

$$l_{t,[i]} = \beta^{(\text{lick})}_{2,[i]} \left(\text{logit}^{-1}\left[\beta^{(\text{lick})}_{1,[i]} \tilde{\xi}_{t,[i]} - \beta^{(\text{lick})}_{0,[i]}\right]\right)^{k_{[i]}} + \epsilon_{t,[i]}, \tag{23}$$

where $\text{logit}^{-1}[\cdot]$ is the logistic function; $\beta^{(\text{lick})}_{0,[i]}, \beta^{(\text{lick})}_{1,[i]}, \beta^{(\text{lick})}_{2,[i]}$ are the intercept, input gain, and output gain, respectively; and $k_{[i]}$ is a parameter to control the shape of the nonlinear function ($k_{[i]} > 0$). The subscript in square brackets $_{[i]}$ denotes a quantity that is specific to recording session i, analogous to the subscripts that denote neuron-specific quantities in the sections above.

For ease of fitting, we reparameterize the intercept $\beta^{(\text{lick})}_{0,[i]}$ and input gain $\beta^{(\text{lick})}_{1,[i]}$ in the above model in terms of two anchor points $A_{1,[i]}, A_{2,[i]}$ that we constrain to be somewhere in the non-saturated range of the logistic function

$$\begin{aligned} A_{1,[i]} &= \text{logit}^{-1}\left[-\beta^{(\text{lick})}_{1,[i]} - \beta^{(\text{lick})}_{0,[i]}\right]^{k_{[i]}} \\ A_{2,[i]} &= \text{logit}^{-1}\left[\beta^{(\text{lick})}_{1,[i]} - \beta^{(\text{lick})}_{0,[i]}\right]^{k_{[i]}} \end{aligned} \tag{24}$$

and use these to back-calculate the input gain $\beta^{(\text{lick})}_{1,[i]}$ and intercept $\beta^{(\text{lick})}_{0,[i]}$. This procedure ensures that when the standardized latent $\tilde{\xi}_{t,[i]}$ varies between -1 and 1, the output of the logistic function will vary between $A_{1,[i]}$ and $A_{2,[i]}$, which we constrain to $0 < A_{1,[i]} < 1/2$ and $1/2 < A_{2,[i]} < 1$.

To obtain the input gain and intercept from the anchor points, we first solve Eq. (24) for $\beta_{1,[i]}^{(\text{lick})}$, yielding

$$\begin{aligned}\beta_{1,[i]}^{(\text{lick})} &= \ln\left[A_{1,[i]}^{-1/k_{[i]}} - 1\right] - \beta_{0,[i]}^{(\text{lick})} \\ \beta_{1,[i]}^{(\text{lick})} &= \beta_{0,[i]}^{(\text{lick})} - \ln\left[A_{2,[i]}^{-1/k_{[i]}} - 1\right].\end{aligned} \tag{25}$$

Subtracting these two equations and rearranging yields an equation for $\beta_{0,[i]}^{(\text{lick})}$

$$\beta_{0,[i]}^{(\text{lick})} = \frac{\ln\left[A_{1,[i]}^{-1/k_{[i]}} - 1\right] + \ln\left[A_{2,[i]}^{-1/k_{[i]}} - 1\right]}{2}, \tag{26}$$

which can be substituted back into Eq. (25) to obtain $\beta_{1,[i]}^{(\text{lick})}$.

Latent The unstandardized latent variable $\xi_{t,[i]}$ is defined as a mixture of thirst and value computed with three different learning timescales

$$\begin{aligned}\xi_{t,[i]} = \lambda_{[i]}^{(\text{val,fast})}\tilde{v}_{t,[i]}^{(\text{fast})} + \lambda_{[i]}^{(\text{val,med})}\tilde{v}_{t,[i]}^{(\text{med})} \\ + \lambda_{[i]}^{(\text{val,slow})}\tilde{v}_{t,[i]}^{(\text{slow})} + \lambda_{[i]}^{(\text{thr})}\tilde{h}_{t,[i]}\end{aligned} \tag{27}$$

where $\lambda_{[i]}^{(\text{val,fast})}, \lambda_{[i]}^{(\text{val,med})}, \lambda_{[i]}^{(\text{val,slow})}, \lambda_{[i]}^{(\text{thr})}$ are mixing coefficients constrained to the unit simplex (*i.e.*, sum to one). We use a tilde $\tilde{\cdot}$ to indicate a standardized variable

$$\tilde{x}_t = \frac{x_t - \langle x \rangle}{\sqrt{\langle (x_t - \langle x \rangle)^2 \rangle}},$$

where $\langle \cdot \rangle$ denotes the mean across trials, as above. Note that the variable $\tilde{\xi}_{t,[i]}$ used to drive licking in Eq. (23) is a standardized version of Eq. (27). We model variability in the mixing coefficients across sessions using a Dirichlet distribution

$$\begin{bmatrix}\lambda_{[i]}^{(\text{val,fast})} \\ \lambda_{[i]}^{(\text{val,med})} \\ \lambda_{[i]}^{(\text{val,slow})} \\ \lambda_{[i]}^{(\text{thr})}\end{bmatrix} \sim \text{Dirichlet}\left(\begin{bmatrix}a_{\text{val, fast}} \\ a_{\text{val, med}} \\ a_{\text{val, slow}} \\ a_{\text{thr}}\end{bmatrix}\right) \tag{28}$$

from which we can estimate the population mean of each mixing coefficient reported in the main text.

The unstandardized values $v^{(x)}$ used to drive the latent are parameterized by the initial value $v_0^{(x)}$ and learning timescale $\tau^{(x)}$ (recall Eq. (4)). The initial value is taken from a beta distribution and scaled to lie between 0.555 and 1

$$\frac{v_{0,[i](x)} - 0.555}{0.555} \sim \text{Beta}[2.45, 2]\ \forall\, x, i, \tag{29}$$

as in the neural activity models (Table 1). The fast values $v^{(\text{fast})}$ use $\tau = 1\,\text{trial}$ and the medium values $v^{(\text{med})}$ use $\tau = 10\,\text{trials}$. The slow values $v^{(\text{slow})}$, which are intended to reflect the potential influence of value-coding serotonin neurons on behaviour, use learning timescales taken from a log-normal distribution that closely approximates the posterior population distribution of learning timescales from the neural data

$$\log \tau_{[i]}^{(\text{slow})} \sim \mathcal{N}\left(\mu_{\log\tau,\text{lick}}, \sigma_{\log\tau,\text{lick}}^2\right), \tag{30}$$

subject to $\tau_{[i]}^{(\text{slow})} > 10\,\text{trials}$, where

$$\begin{aligned}\mu_{\log\tau,\text{lick}} &\sim \mathcal{N}\left(1.8625, 0.1182^2\right) \\ \sigma_{\log\tau,\text{lick}} &\sim \text{Beta}\left[5.1867, 12.6473\right].\end{aligned} \tag{31}$$

The parameters of the $\mu_{\log\tau,\text{lick}}$ and $\sigma_{\log\tau,\text{lick}}$ distributions were obtained via maximum likelihood fits to the corresponding posterior distributions from the neural activity mixture model. This constrains the $\tau_{[i]}^{(\text{slow})}$ in the behavioural model to match the neural activity model, while accounting for uncertainty inherent in the neural model.

Parameterization and priors For the purposes of fitting, the model is parameterized in terms of the following quantities:

- The correct perception probabilities $\pi_{[i]}$ and their population distribution parameters a_π, b_π.
- Normalized initial value for each session $v'_{0,[i]}$, as in the neural activity models.
- Standardized slow learning timescale for each session $\log \tau_{[i]}^{(\text{slow})\prime}$ and corresponding population distribution parameters $\mu_{\log\tau,\text{lick}}, \sigma_{\log\tau,\text{lick}}$, as in the hierarchical neural activity model.
- Mixing coefficients for each session $\lambda_{[i]}^{(\text{val,fast})}, \lambda_{[i]}^{(\text{val,med})}, \lambda_{[i]}^{(\text{val,slow})}, \lambda_{[i]}^{(\text{thr})}$ and corresponding population distribution parameters $a_{\text{val, fast}}, a_{\text{val, med}}, a_{\text{val, slow}}, a_{\text{thr}}$.
- Nonlinear function anchor points for each session $A_{1,[i]}, A_{2,[i]}$.
- Nonlinear function shape parameter on a log10 scale $\log k_{[i]}$ for each session.
- Nonlinear function output gain $\beta_{2,[i]}^{(\text{lick})}$ for each session.
- Lick count standard deviation for each session $\sigma_{l,[i]}$.

REFERENCES

[1] Melissa R Warden, Aslihan Selimbeyoglu, Julie J Mirzabekov, Maisie Lo, Kimberly R Thompson, Sung-Yon Kim, Avishek Adhikari, Kay M Tye, Loren M Frank, and Karl Deisseroth. A prefrontal cortex–brainstem neuronal projection that controls response to behavioural challenge. *Nature*, 492(7429):428–432, 2012.

[2] Changwoo Seo, Akash Guru, Michelle Jin, Brendan Ito, Brianna J Sleezer, Yi-Yun Ho, Elias Wang, Christina Boada, Nicholas A Krupa, Durgaprasad S Kullakanda, et al. Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science*, 363(6426):538–542, 2019.

[3] Aaron S Andalman, Vanessa M Burns, Matthew Lovett-Barron, Michael Broxton, Ben Poole, Samuel J Yang, Logan Grosenick, Talia N Lerner, Ritchie Chen, Tyler Benster, et al. Neuronal dynamics regulating brain and behavioral state transitions. *Cell*, 177(4): 970–985, 2019.

[4] João C Marques, Meng Li, Diane Schaak, Drew N Robson, and Jennifer M Li. Internal state dynamics shape brainwide activity and foraging behaviour. *Nature*, 577(7789):239–243, 2020.

[5] Zhixiang Liu, Jingfeng Zhou, Yi Li, Fei Hu, Yao Lu, Ming Ma, Qiru Feng, Ju-en Zhang, Daqing Wang, Jiawei Zeng, et al. Dorsal raphe neurons signal reward through 5-ht and glutamate. *Neuron*, 81(6):1360–1374, 2014.

[6] Patricia A Correia, Eran Lottem, Dhruba Banerjee, Ana S Machado, Megan R Carey, and Zachary F Mainen. Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *Elife*, 6:e20975, 2017.

[7] Eran Lottem, Dhruba Banerjee, Pietro Vertechi, Dario Sarra, Matthijs Oude Lohuis, and Zachary F Mainen. Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature communications*, 9(1):1000, 2018.

[8] Katsuhiko Miyazaki, Kayoko W Miyazaki, Akihiro Yamanaka, Tomoki Tokuda, Kenji F Tanaka, and Kenji Doya. Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. *Nature communications*, 9(1):2048, 2018.

[9] Jeremiah Y Cohen, Mackenzie W Amoroso, and Naoshige Uchida. Serotonergic neurons signal reward and punishment on multiple timescales. *Elife*, 4: e06346, 2015.

[10] Weixin Zhong, Yi Li, Qiru Feng, and Minmin Luo. Learning and stress shape the reward response patterns of serotonin neurons. *Journal of Neuroscience*, 37(37):8863–8875, 2017.

[11] Grace E Paquelet, Kassandra Carrion, Clay O Lacefield, Pengcheng Zhou, René Hen, and Bradley R Miller. Single-cell activity and network properties of dorsal raphe nucleus serotonin neurons during emotionally salient behaviors. *Neuron*, 110(16):2664–2679, 2022.

[12] Emerson F Harkin, Cooper D Grossman, Jeremiah Y Cohen, Jean-Claude Béïque, and Richard Naud. A prospective code for value in the serotonin system. *Nature*, pages 1–8, 2025.

[13] Michael B Lynn, Sean D Geddes, Mohamad Chahrour, Sébastien Maillé, Léa Caya-Bissonnette, Emerson Harkin, Érik Harvey-Girard, Samir Haj-Dahmane, Richard Naud, and Jean-Claude Béïque. Nonlinear recurrent inhibition through facilitating serotonin release in the raphe. *Nature Neuroscience*, pages 1–14, 2025.

[14] R. Sutton. Learning to predict by the method of temporal differences. *Machine learning*, 3:9–44, 1988.

[15] Sara Matias, Eran Lottem, Guillaume P Dugué, and Zachary F Mainen. Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife*, 6:e20552, 2017.

[16] Cooper D Grossman, Bilal A Bari, and Jeremiah Y Cohen. Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, 32(3):586–599, 2022.

[17] Emerson F Harkin, Cooper D Grossman, Jeremiah Y Cohen, Jean-Claude Béïque, and Richard Naud. Serotonin predictively encodes value. *biorxiv*, 2023.

[18] L Paninski, B Lau, and A Reyes. Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing*, 52(54):877–883, 2003.

[19] J.W. Pillow, L. Paninski, and E.P. Simoncelli. Maximum likelihood estimation of a stochastic integrate-and-fire model. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16, pages 1311–1318, 2004.

[20] S. Mensi, R. Naud, M. Avermann, C. C. H. Petersen, and W. Gerstner. Parameter extraction and classification of three neuron types reveals two different adaptation mechanisms. *Journal of Neurophysiology*, 107: 1756–1775, 2012.

[21] Christian Pozzorini, Richard Naud, Skander Mensi, and Wulfram Gerstner. Temporal whitening by power-law adaptation in neocortical neurons. *Nat. Neurosci.*, 16(7):942–948, 2013.

[22] Christian Pozzorini, Skander Mensi, Olivier Hagens, Richard Naud, Christof Koch, and Wulfram Gerstner.

Table 2. Summary of quantities of interest in hierarchical Bayesian model of anticipatory licking.

Name	Description	Symbol(s)	Prior	Justification for prior

Automated high-throughput characterization of single neurons by means of simplified spiking models. *PLoS Comp. Biol.*, 11(6):e1004275, 2015.

[23] Corinne Teeter, Ramakrishnan Iyer, Vilas Menon, Nathan Gouwens, David Feng, Jim Berg, Aaron Szafer, Nicholas Cain, Hongkui Zeng, Michael Hawrylycz, et al. Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature communications*, 9(1):709, 2018.

[24] Emerson F Harkin, Michael B Lynn, Alexandre Payeur, Jean-François Boucher, Léa Caya-Bissonnette, Dominic Cyr, Chloe Stewart, André Longtin, Richard Naud, and Jean-Claude Béïque. Temporal derivative computation in the dorsal raphe network revealed by an experimentally-driven augmented integrate-and-fire modeling framework. *eLife*, 12:e72951, 2023.

[25] Wilson Truccolo, Leigh R Hochberg, and John P Donoghue. Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nature Neuroscience*, 13(1):105–111, Jun 2010.

[26] Rick L Jenison, Antonio Rangel, Hiroyuki Oya, Hiroto Kawasaki, and Matthew A Howard. Value encoding in single neurons in the human amygdala during decision making. *Journal of Neuroscience*, 31(1):331–338, 2011.

[27] Vernon Lawhern, Wei Wu, Nicholas Hatsopoulos, and Liam Paninski. Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of neuroscience methods*, 189(2):267–280, 2010.

[28] Antoine D Madar, Anqi Jiang, Can Dong, and Mark EJ Sheffield. Synaptic plasticity rules driving representational shifting in the hippocampus. *Nature neuroscience*, pages 1–13, 2025.

[29] R Naud and W Gerstner. *Computational Systems Neurobiology*, chapter The Performance (and limits) of Simple Neuron Models: Generalizations of the Leaky Integrate-and-Fire Model. Springer, 2012.

[30] Renaud Jolivet, R Kobayashi, A Rauch, R Naud, S Shinomoto, and Wulfram Gerstner. A benchmark test for a quantitative assessment of simple neuron models. *Journal of Neuroscience Methods*, 169:417–424, 2008.

[31] Renaud Jolivet, F Schürmann, TK Berger, R Naud, W Gerstner, and A Roth. The quantitative single-neuron modeling competition. *Biological Cybernetics*, 99(4):417–426, 2008.

[32] Valerio Francioni, Vincent D Tang, Enrique HS Toloza, Norma J Brown, and Mark T Harnett. Vectorized instructive signals in cortical dendrites during a brain-computer interface task. *bioRxiv*, pages 2023–11, 2025.

[33] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.

[34] João D Semedo, Amin Zandvakili, Christian K Machens, M Yu Byron, and Adam Kohn. Cortical areas interact through a communication subspace. *Neuron*, 102(1):249–259, 2019.

[35] Richard Naud and Henning Sprekeler. Sparse bursts optimize information transmission in a multiplexed neural code. *Proc. Nat. Acad. Sci. (U.S.A.)*, 2018. ISSN 0027-8424. doi: $10.1073/\mathrm{pnas}.1720995115$.

[36] Daniel LK Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111 (23):8619–8624, 2014.

[37] Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. *Science*, 364(6439):eaav9436, 2019.

[38] Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, page 407007, 2018.

[39] Grace W Lindsay. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10):2017–2031, 2021.

[40] Yara Otor, Shay Achvat, Nathan Cermak, Hadas Benisty, Maisan Abboud, Omri Barak, Yitzhak Schiller, Alon Poleg-Polsky, and Jackie Schiller. Dynamic compartmental computations in tuft dendrites of layer 5 neurons during motor behavior. *Science*, 376(6590): 267–275, 2022.

[41] Alexandre Payeur, Jean-Claude Béïque, and Richard Naud. Classes of dendritic information processing. *Current opinion in neurobiology*, 58:78–85, 2019.

[42] Yuan Zhang, Yanhe Liu, Yachuang Hu, Zhenkun Zhang, Jie Du, Lele Cui, Qi Liu, Lin Zhong, Yu Xin, Ruiming Chai, et al. Dendritic computation for rule-based flexible categorization. *bioRxiv*, pages 2025–06, 2025.

[43] Zachary Friedenberger, Emerson Harkin, Katalin Tóth, and Richard Naud. Silences, spikes and bursts: Three-part knot of the neural code. *The Journal of Physiology*, 601(23):5165–5193, 2023.

[44] Wulfram Gerstner, Werner Kistler, Richard Naud, and Liam Paninski. *Neuronal Dynamics*. Cambridge University Press, Cambridge, 2014.

[45] George E.P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*, chapter 2: Autocorrelation function and spectrum of stationary processes. Wiley, 2008.

[46] M.J. Betancourt and M. Girolami. Hamiltonian monte carlo for hierarchical models. *arxiv*, 2013.

[47] Larry V Hedges. Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9(1):61–85, 1984.

[48] Anton Kühberger, Astrid Fritz, Eva Lermer, and Thomas Scherndl. The significance fallacy in inferential statistics. *BMC Research Notes*, 8(1):84, March 2015. ISSN 1756-0500. doi: 10.1186/s13104-015-1020-4.

[49] Ben van Calster, Ewout W. Steyerberg, Gary S. Collins, and Tim Smits. Consequences of relying on statistical significance: Some illustrations. *European Journal of Clinical Investigation*, 48(5):e12912, 2018. ISSN 1365-2362. doi: 10.1111/eci.12912.

[50] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*, chapter 22: Finite mixture models, page 523. 3rd edition, 2021.

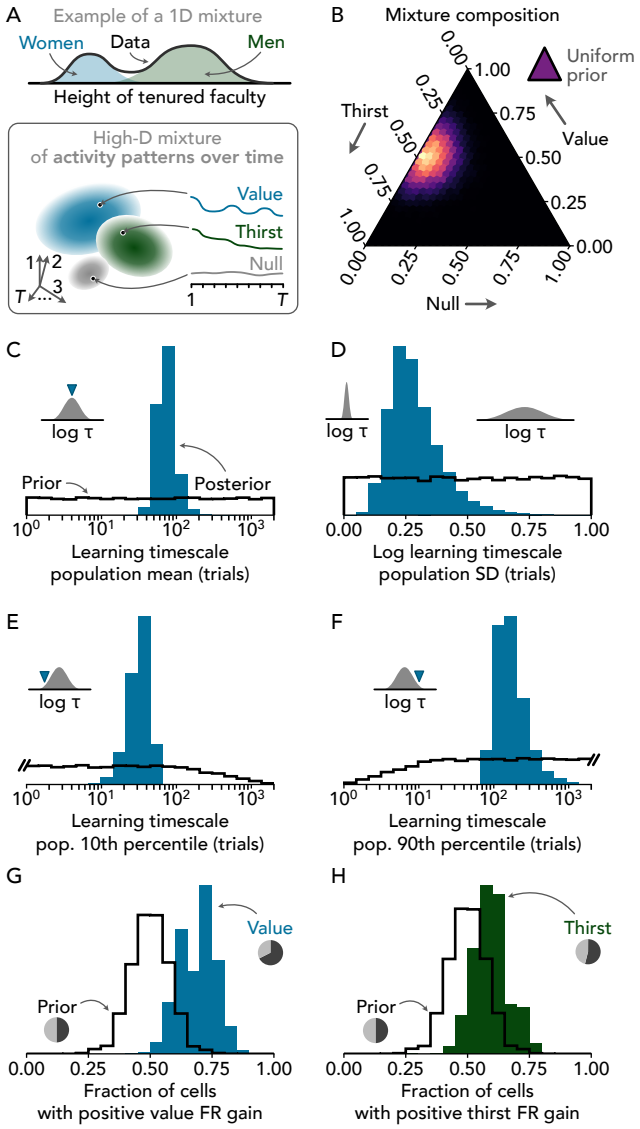

Figure S1. Features of value and thirst coding in serotonin neurons.

A Conceptual introduction to the value, thirst, and null coding mixture model. A hypothetical one-dimensional dataset of hairline measurements in tenured faculty is illustrated at top. A mixture model can infer the average hairlines of male and female faculty as well as the proportion of faculty that are male or female, even when gender information is missing from the data. Similarly, the mixture model of serotonin neuron coding features can infer the properties of value, thirst, and null coding neurons as well as the proportion of neurons of each type (bottom).

B Inferred proportion of neurons that are value, thirst, or null coding. Brighter colours in the main plot indicate higher posterior density. The uniform prior is illustrated schematically at top right.

C Mean learning timescale τ across the population of value-coding serotonin neurons.

D Standard deviation of the learning timescale τ across the population of value-coding serotonin neurons on a $\log_{10}$ scale.

E 10th percentile of the distribution of the learning timescale τ across the population of value-coding serotonin neurons.

F 90th percentile of the distribution of the learning timescale τ across the population of value-coding serotonin neurons.

G Fraction of cells in this dataset with positive firing rate gain for value $\beta_1^{(\text{val})}$. Pie plot shows the MAP estimate of the proportion of neurons with positive (negative) firing rate gain in black (gray). Note that the prior (black line) is based on a default assumption that positive and negative gains are equally common in the population.

H Fraction of cells in this dataset with positive firing rate gain for thirst $\beta_1^{(\text{thr})}$. Presented as in G.

All parameter estimates are based on a model fitted to $12\,387$ trials from 37 neurons. Black lines and coloured histograms in C–H indicate the prior and posterior, respectively, for each parameter or derived feature. B–D show parameters that are directly included in the model, while E–H are derived from fitted parameters. Partial-pooling across neurons was used for the quantities shown in C–F, full pooling was used for the parameters in B, and no pooling was used for G or H. See Section 5.3.5 for details.

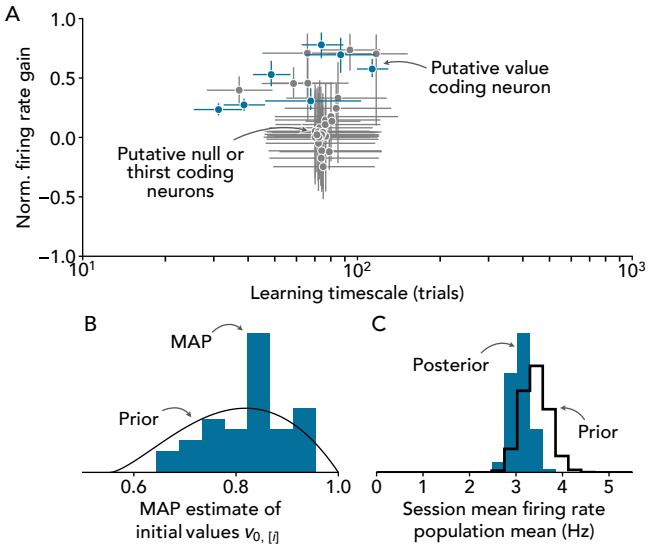

Figure S2. Additional features of value, thirst, and null coding mixture model. Related to Fig. S1.

A Estimated normalized firing rate gain and learning timescale for each neuron. Points and error bars represent the median and IQR of the posterior. Neurons for which the $95\,\%$ credibility interval for the normalized firing rate gain includes zero are indicated in gray.

B Distribution of estimated initial values $v_{0,[i]}$ from the *maximum a posteriori* (MAP) sample (blue histogram) compared with the prior (black line).

C Posterior distribution of mean firing rate in the population (blue) compared with the prior (black).

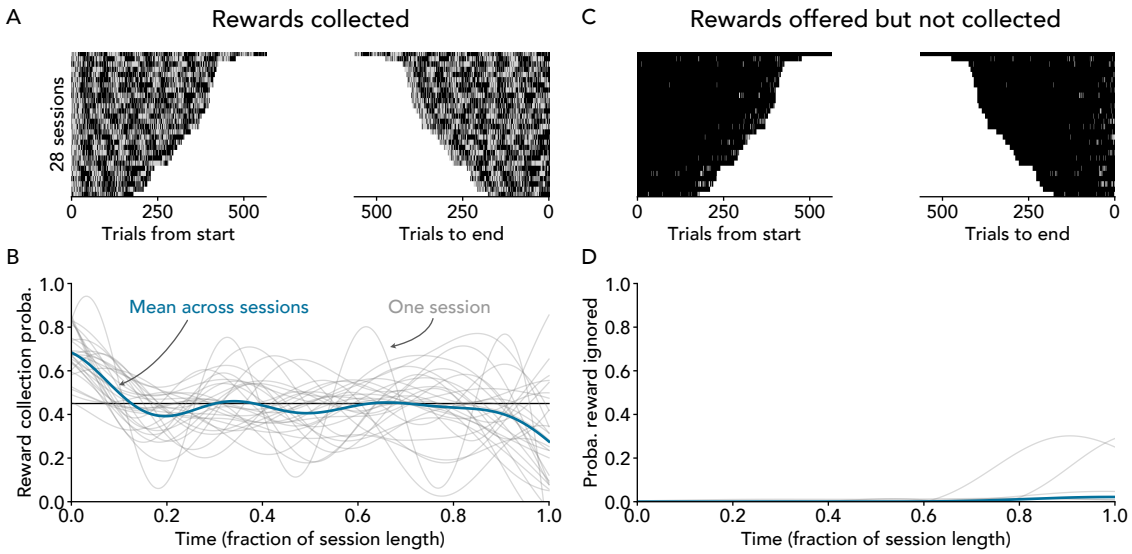

Figure S3. Time-course of reward collection across recording sessions.
A Heatmaps showing whether or not a reward was collected on each trial, aligned to either the start (left) or end (right) of the experiment. Trials in which a reward was collected are shown in gray, no-reward trials are shown in black.
B Time-course of reward collection aligned to fraction of session length. Each gray line represents data from one row of A smoothed using Gaussian process regression (observation variance set to $0.45 \times (1 - 0.45)$, squared exponential kernel with length scale $1/6$). Black line represents the expected mean reward probability of $45\,\%$. Note that the reward collection probability is high at the start of the experiment because the first block of trials always has a reward probability of $80\,\%$.
C Heatmaps showing trials where a reward was offered but not collected (ignored rewards), aligned as in A. Trials in which a reward was offered but not collected are shown in gray, trials in which a reward was either not offered or offered and collected are shown in black.
D Time-course of the rate of ignored rewards, presented as in B.

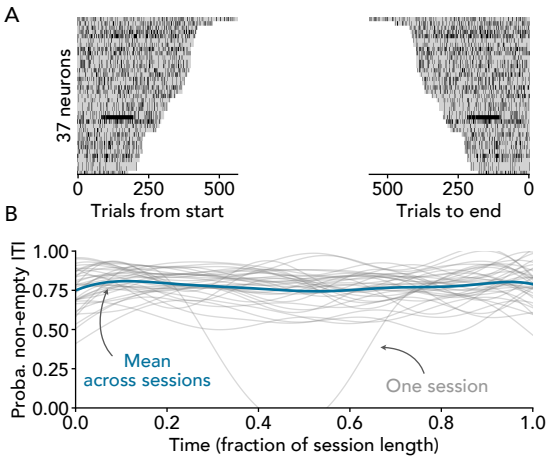

Figure S4. Time-course of inter-trial intervals (ITIs) with missing inter-spike interval (ISI) data across neurons. Missing ISI data is the result of ITIs with fewer than two spikes. This can occur if the ITI is shorter than the mean ISI (the minimum ITI duration is $0\,\mathrm{s}$ while serotonin neuron ISIs are typically approx. 0.2–$1\,\mathrm{s}$), or if the neuron is lost due to spike sorting or other recording issues.
A Heatmaps showing whether or not each ITI has at least one ISI, aligned to either the start (left) or end (right) of the experiment. ITIs with zero ISIs are shown in black, ITIs with at least one ISI are shown in gray.
B Time-course of missing ISI data aligned to fraction of session length. Each gray line represents data from one row of A smoothed using Gaussian process regression (observation variance set to $p \times (1 - p)$, where p is the fraction of non-empty ITIs in the corresponding session; squared exponential kernel with length scale $1/6$).

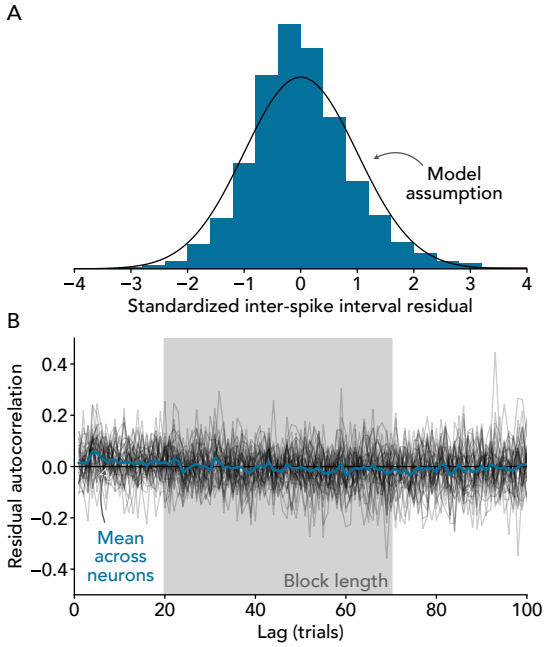

Figure S5. Value coding model residuals compared with model assumptions.
A Histogram of standardized inter-spike interval residuals (blue) compared with the normal distribution assumed by the model (black line).
B Autocorrelation of standardized residuals for each neuron (gray lines) compared with the model assumption of zero autocorrelation (black line). For clarity, the autocorrelation at zero lag is not shown.
Residuals are net of drift captured using a third-order autoregressive model.

A JUSTIFICATION FOR BAYESIAN APPROACH: BASELINE FIRING RATE EXAMPLE

A limitation of the value coding model presented in Eqs. (1) and (2) is that the firing rate gain and intercept terms β_1 and β_0 are degenerate when the learning timescale τ is very long. This can make it difficult to determine whether a particular neuron with a long estimated learning timescale is activated by value ($\beta_1 > 0$) and has a low baseline firing rate, is inhibited by value ($\beta_1 < 0$) and has a high baseline firing rate, or even whether it is insensitive to value ($\beta_1 \approx 0$) and has a moderate baseline firing rate purely based on the data likelihood. To see why, note from Eq. (3) that $\lim_{\tau\to\infty} v_t = v_0$, which implies that

$$\lim_{\tau\to\infty} \rho_t = \beta_0 + \beta_1 v_0 + \epsilon_t. \tag{32}$$

Since there are infinitely many combinations of β_0 and β_1 that satisfy the above equation, the likelihood of the firing rate given the slope and intercept $p(\rho_t \mid \beta_0, \beta_1)$ is not sufficient to identify the values of these parameters. This means that we need additional information beyond what is available in the data in order to estimate β_0 and β_1. Fortunately, previous work has already established the normal range of baseline firing rates in serotonin neurons. Incorporating this knowledge into our model in the form of a prior distribution for β_0 makes it possible to estimate the values of the remaining parameters, even if the data likelihood is insufficient.

B PRIOR FOR FIRING RATE INTERCEPTS

Background To establish a prior for the firing rate intercepts β_0 in our analysis, we took advantage of previous experiments reported in ref. 9. These experiments are similar to the dynamic Pavlovian experiment analyzed in the main text in important respects:

- Both sets of data consist of *in vivo* tetrode recordings of identified dorsal raphe serotonin neurons in C57BL/6J mice.
- Both sets of data are from head-fixed trace conditioning experiments.

The main differences between the two sets of experiments are as follows:

- The experiments from ref. 9 include reward and punishment conditions, while the dynamic Pavlovian experiment includes only rewards.
- Reward and punishment block transitions were cued in ref. 9, while changes in reward probability in the dynamic Pavlovian experiment were not cued.
- Unconditioned stimuli were deterministic in ref. 9 and stochastic in the dynamic Pavlovian experiment.

In view of these similarities and differences, we took the average baseline firing rate across reward $\bar{\rho}^{(\text{rew})}$ and punishment $\bar{\rho}^{(\text{pun})}$ conditions from ref. 9 to be analogous to the baseline firing rate in our model β_0

$$\beta_{0,[i]} \approx \frac{\bar{\rho}^{(\text{rew})}_{[i]} + \bar{\rho}^{(\text{pun})}_{[i]}}{2}, \tag{33}$$

allowing us to estimate the distribution of the firing rate intercept β_0 across neurons from the $(\bar{\rho}^{(\text{rew})}, \bar{\rho}^{(\text{pun})})$ pairs reported in ref. Figs. 3C and 6D.

Data extraction and pooling Using Web Plot Digitizer (`https://apps.autometris.io/wpd/`), we extracted measurements of the mean pre-trial firing rates of $N = 29\,\text{neurons}$ during water reward and air puff punishment blocks of the trace conditioning experiment reported in ref. Fig. 3C, and $N = 21\,\text{neurons}$ during water or chocolate milk reward and quinine solution punishment blocks of the trace conditioning experiment reported in ref. Fig. 6D. Data were pooled across the two experiments because we did not observe significant differences in the distributions of firing rates across the two experiments (differences in $\bar{\rho}^{(\text{rew})}$ are symmetric about zero, Mann–Whitney U-test $p = 0.74$, differences in $\bar{\rho}^{(\text{pun})}$ are symmetric about zero, Mann–Whitney U-test $p = 0.68$).

Results The distribution of firing rate intercepts implied by ref. 9 is approximately gamma with mean $3.41\,\text{Hz}$ and SD $2.34\,\text{Hz}$ ($\beta_0 \sim \text{Gamma}[2.12, 1.61]$ using the shape α scale θ parameterization; maximum likelihood estimates obtained with `scipy.stats.gamma.fit`; Fig. S6A). Since our goal is to arrive at a prior distribution for the firing rate intercepts in a new sample of neurons, we quantified the sampling-related uncertainty in our estimated intercept distribution using bootstrapping (Fig. S6B). We found that the bootstrap distribution of the natural log shape and scale parameters is a close match to a multivariate normal

$$\begin{bmatrix} \ln\alpha \\ \ln\theta \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0.793 \\ 0.430 \end{bmatrix}, \begin{bmatrix} 0.0555 & -0.0492 \\ -0.0492 & 0.0500 \end{bmatrix} \right),$$

as shown in Fig. S6C, which we therefore used as a prior for the distribution of firing rate intercepts in our analysis.

Verification The analogy between the cross-condition average firing rate in ref. 9 and the firing rate intercept β_0 given in Eq. (33) depends on the assumption that the effects of rewards and punishments on the firing rates of serotonin neurons are of roughly equal magnitude and opposite sign, such that they cancel out when averaged. To guard against the possibility that this assumption is wrong, we reserved the baseline firing rates reported in ref. 9 Fig. 8S2A (measured under unspecified conditions) as a test dataset to check the baseline firing rate distribution implied by ref. Figs. 3C and 6D. The distribution of firing rate intercepts implied by Eq. (33) shown in Fig. S6 closely agrees

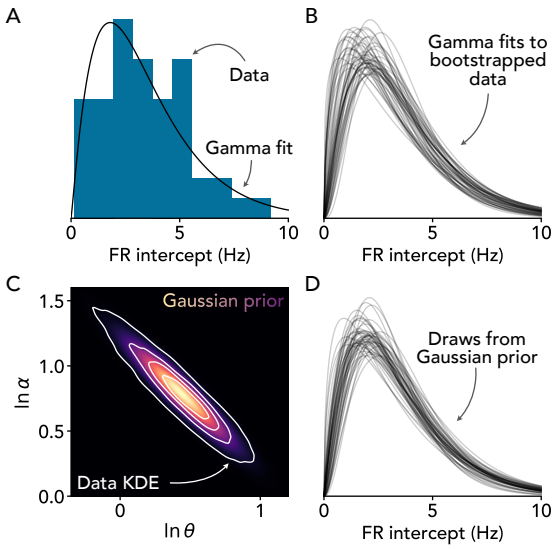

Figure S6. Firing rate intercept prior.
A Histogram of estimated firing rate intercepts β_0 from ref. 9 with fitted gamma distribution. $N = 50\,\text{neurons}$ from ref. Figs. 3C and 6D.
B Gamma distributions fitted to bootstrapped firing rate intercept data from A. Note variability due to sampling error.
C Estimated distribution of gamma distribution parameters across bootstrap samples. Contours show a kernel density estimate of the joint distribution of $\ln\alpha$ (natural log shape) and $\ln\theta$ (natural log scale) parameters across bootstrap samples. Heatmap shows fitted multivariate Gaussian distribution to be used as a prior.
D Draws from Gaussian prior distribution over $\ln\alpha$ and $\ln\theta$. Note close resemblance to the bootstrap distribution in B.

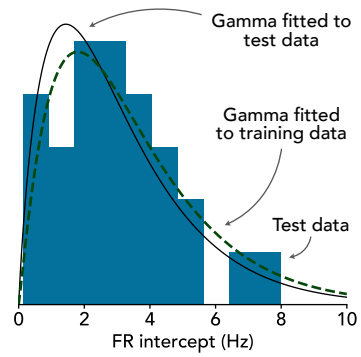

Figure S7. Verification of firing rate intercept prior distribution. Histogram of firing rate intercepts from ref. 9 Fig. 8S2A (test data; $N = 28\,\text{neurons}$) and fitted gamma distribution (black line) compared with the distribution fitted to data from ref. Figs. 3C and 6D (dashed green line; $N = 50\,\text{neurons}$).

with the distribution implied by the test data (Fig. S7), validating our approach.

C PRIOR FOR FIRING RATE GAIN

Using the same data as in Supplementary Note B, we established a prior for the normalized firing rate gain $\beta_1' = \beta_1/\beta_0$ using

$$\beta'_{1,[i]} \approx \frac{\bar{\rho}^{(\text{rew})}_{[i]} - \beta_{0,[i]}}{\beta_{0,[i]}}, \tag{34}$$

where $\beta_{0,[i]}$ is estimated according to Eq. (33) and where we use $\bar{\rho}^{(\text{rew})}_{[i]} - \beta_{0,[i]}$ as an estimate of β_1. This estimate is sensitive to mis-estimation of β_1, and, unlike with β_0 (see Supplementary Note B), we are unaware of data that could be used to check our estimate. We therefore use Eq. (34) to set only a weak prior of $\beta_1' \sim \mathcal{N}(0, 0.75^2)$ truncated to $[-1, 1]$, as shown in Fig. S8.

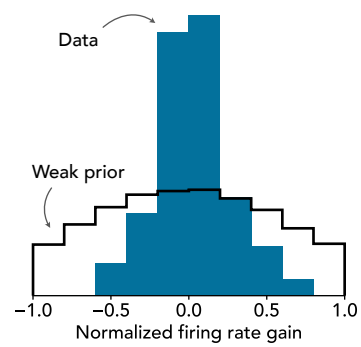

Figure S8. Prior for normalized firing rate gain. Histogram of normalized firing rate gains β_1' calculated according to Eq. (34) using data from ref. 9 Figs. 3C and 6D compared with truncated $\mathcal{N}(0, 0.75^2)$ prior.

D ALTERNATIVES TO MIXTURE MODELLING APPROACH

Here we have chosen to infer the properties of value, thirst, and null coding serotonin neurons using a mixture model due to important limitations of alternative approaches. In this section, we briefly outline the alternatives considered and why they were not chosen.

Problem definition The goals of this work are 1) to measure the proportion of serotonin neurons that encode value and 2) to infer the properties of value coding neurons. To accomplish these goals, we have analyzed a dataset of measurements of the activity of a number of serotonin neurons. The challenges we faced are that 1) the only information we can use to determine whether an individual serotonin neuron encodes value is its activity over time, 2) the time course of activity in a serotonin neuron that encodes value is not necessarily very different than the time course of activity in a serotonin neuron that encodes thirst or where the activity drifts randomly (null coding), and 3) value, thirst, and null coding are particularly difficult to tell apart if the firing rate gain is small and/or if the learning timescale associated with value coding is long. To address these challenges, we needed to select an analysis approach that can cope with unlabeled data (challenge 1), high noise (challenge 2), and confounding that is uneven throughout the parameter space (challenge 3).

Statistical null hypothesis testing The most familiar way to address the first challenge listed above would be to use statistical null hypothesis testing to rule out thirst and null coding in a subset of neurons. For example, to rule out null coding, we might have adopted the circular trial permutation test from our previous work (12) and selected only the subset of neurons with activity patterns that are very unlikely under null coding (e.g. at the $p < 0.05$ level) as being non-null-coding. A similar procedure could have been repeated for thirst coding to obtain a set of non-null-, non-thirst-coding neurons. By process of elimination, we would then consider these neurons to be value coding and analyze them further. The basic pattern of this approach will be familiar to most readers: apply a statistical test, label the neurons with $p < 0.05$ as having the coding feature of interest, and then analyze only that subset.

While this approach adequately addresses the first challenge posed above, it does not address challenges two or three. Specifically, because it is generally difficult to tell apart serotonin neurons with value, thirst, or null coding features, and because typical statistical tests focus on limiting the rate of false positives rather than false negatives (including the circular trial permutation test from our previous work, ref. 17), it is likely that a meaningful fraction of value coding serotonin neurons would be discarded. Thus, a bias towards false negatives due to low statistical power could cause the fraction of value coding neurons to be meaningfully underestimated. A further problem with this approach is that the subset of serotonin neurons that *are* selected as value coding will be the neurons with value coding features that are the most different from the activity patterns observed in thirst and/or null coding neurons. In other words, there will be a bias towards selecting serotonin neurons with unusually high firing rate gains and short learning timescales. This phenomenon, whereby selecting data points according to statistical significance causes model parameters to be biased towards extremes, is well documented in the statistical literature (47–49). Overall, taking this approach would likely have led us to conclude that value coding serotonin neurons are less abundant and encode value more strongly and over a shorter timescale than is actually the case.

Model comparison A more sophisticated way to address the first challenge listed above would be to fit a value, thirst, and null coding model to each cell, compare their predictive accuracy (using cross-validation, AIC, BIC, etc.), and then label each cell as following the coding scheme corresponding to the best fitting model. Compared with the null hypothesis testing approach outlined above, this approach does not have a strong bias towards thirst or null coding, eliminating the tendency to conclude that value coding neurons are rarer and respond more strongly to recent rewards than is actually the case.

While the model comparison approach represents an improvement over null hypothesis testing in terms of our two main goals, it does not fully address challenges 2 and 3. Specifically, because this approach involves labeling individual neurons as being value, thirst, or null coding in order to conduct further analysis, information about the uncertainty of these labels is lost. This could pose a problem if the sample of serotonin neurons in the dataset analyzed here happens to be enriched in neurons with activity patterns that are only slightly more consistent with value coding than thirst coding, for example, simply due to chance. Since the model comparison approach does not take the strength of evidence for one coding scheme over another into consideration, we would be at risk of confidently concluding that a large fraction of neurons encode value with a low firing rate gain and long learning timescale. The general problem of label-based methods ignoring uncertainty in data labels is briefly discussed in ref. 50.