Supplementary Notes

Note 1: Setting of eligibility trace time constant. It is intuitively clear that the eligibility trace time constant T needs to be set to match the timescales operating in the environment. This is because if the eligibility trace decays too quickly, there will be no memory of past events, and if it decays too slowly, it will take a long time to correctly learn event rates in the environment. Further, the asymptotic value of the baseline memory trace of event x, $M_{\leftarrow x-}$ for an event train at a constant rate λ_x with average period t_x is $T/t_x = T\lambda_x$. This means that the neural representation of $M_{\leftarrow x-}$ will need to be very high if T is very high and very low if T is very low. Since every known neural encoding scheme is non-linear at its limits with a floor and ceiling effect (e.g., firing rates can't be below zero or be infinitely high), the limited neural resource in the linear regime should be used appropriately for efficient coding. A linear regime of operation for $M_{\leftarrow x-}$ is especially important in ANCCR since the estimation of the successor representation by Bayes' rule depends on the ratio of $M_{\leftarrow x-}$ for different event types. Such a ratio will be highly biased if the neural representation of $M_{\leftarrow x-}$ is in its non-linear range. Assuming without loss of generality that the optimal value of $M_{\leftarrow x-}$ is M_{opt} for efficient linear coding, we can define a simple optimality criterion for the eligibility trace time constant T. Specifically, we postulate that the net sum of squared deviations of $M_{\leftarrow x-}$ from M_{opt} for all event types should be minimized at the optimal T. The net sum of squared deviations, denoted by SS, can be written as

$$SS = \sum_x (M_{\leftarrow x-} - M_{opt})^2 = \sum_x (T\lambda_x - M_{opt})^2 \quad \textbf{(5)}$$

Where the second equality assumes asymptotic values of $M_{\leftarrow x-}$. The minimum of SS with respect to T will occur when $\frac{\partial SS}{\partial T} = 0$. It is easy to show that this means that the optimal T is:

$$T_{opt} = M_{opt} \frac{\sum_x \lambda_x}{\sum_x \lambda_x^2} \quad \textbf{(6)}$$

For typical cue-reward experiments with each cue predicting reward at 100% probability, $\lambda_{cue} = \lambda_{reward} = \frac{1}{IRI}$. Substituting into the above equation, we get:

$$T_{opt} = M_{opt}.IRI \quad \textbf{(7)}$$

Thus, in typical experiments with 100% reward probability, the eligibility trace time constant should be proportional to the IRI or the total trial duration, which is determined by the ITI—the experimental proxy that we manipulate. Please do note, however, that the above relationship is not strictly controlled by the ITI, but by the frequency of repeating events in the environment (i.e., environmental timescale).

Note 2: Higher cue-offset induced anticipatory licking with short ITI. We observed empirically that the animals showed higher anticipatory licking following cue offset (i.e., 8 seconds after cue onset) during the short ITI condition compared to the long ITI condition (Fig 1f) (though this is only weakly significant). We believe that this simply reflects the fact that the cue onset is relatively much farther to reward delivery compared to the inter-cue interval during the short ITI condition compared to the long ITI condition (ratio of 9s to 9+8s in short ITI vs 9s to 9+55s in the long ITI). Therefore, in the short ITI condition, the cue offset provides a stronger signal indicating relative proximity to reward.

Note 3: The implications of assuming that internal states may serve the role of external cues in ANCCR. Some readers will note that we have previously argued against the assumption of internal states serving the role of externally signaled events in learning theories(46). It may therefore seem that our speculation that internal states can serve this role during timing tasks is problematic. However, there is a critical difference between our earlier position and the current speculation. Our earlier position was that assuming fixed internal states that pre-exist and provide a scaffold for learning, such as in temporal difference learning, is problematic. This is because these pre-existing states would need to already incorporate information that can only be acquired during the course of learning. Unlike this position, here we are merely speculating that after learning, an internal progression of states can serve the function of externally signaled events. Similarly, we have previously postulated that such an internal state exists during omission of a predicted reward, but only after learning of the cue-reward association.

Note 4: Some discrepancies between ANCCR simulations and experiments. We performed the ANCCR simulations not to explicitly fit the experiments, but to motivate them. Accordingly, there are many details of the experimental conditions that we did not include in the simulations. First, animals were trained initially using a long (Pavlovian) or medium ITI (VR), thereby establishing that the cue onset is a meaningful event before switching to the short ITI. Second, animals are unlikely to discriminate each change in tone frequency in the dynamic tone (80 Hz every 200 ms). Thus, we simplified the simulation and used a 1 s interval between sensory cues under the assumption that 400 Hz would be discriminable. The potential sensory noise in detection of frequency changes was not modeled in the simulation. Third, we did not explicitly model potential trial-by-trial changes in eligibility trace time constant, sensory noise, or internal threshold. Fourth, we did not simulate any biophysical mechanisms controlling dopamine release, or sensor dynamics. Thus, we did not expect to capture all experimental observations in the motivating simulation. One particular discrepancy is worth noting: the cue onset response in the short ITI condition is small but positive in the experiment but negative in ANCCR. This may potentially reflect the fact that the cue onset was already learned to be meaningful prior to the short ITI experiment.

Supplementary Table 1: Statistical Details.

Figure	Description	Test	Statistic	p value	Sample size
1g	Anticipatory lick rate across ITI (long, short) and tone (fixed, dynamic)	Two-way ANOVA	ITI: F(1) = 9.30 Tone: F(1) = 0.30 ITI x Tone: F(1) = 0.029	ITI: **p = 0.00457 Tone: p = 0.586 ITI x Tone: p = 0.865	n = 9 mice
1j	Cue onset peak dLight between conditions (LD, SD)	Paired t-test	t(8) = 6.31	***$p = 2.31 \times 10^{-4}$	n = 9 mice
1l	Slope between days (LD condition last day, SD condition first day)	One-sided (LD < SD) paired t-test	t(8) = -2.07	*p = 0.0363	n = 9 mice
1l	Slope between days (SD condition last day, SF condition first day)	One-sided (SD > SF) paired t-test	t(8) = 2.35	*p = 0.0233	n = 9 mice
1m	Slope across conditions (LF, LD, SD, SF)	One-way ANOVA	F(3) = 8.89	***$p = 1.98 \times 10^{-4}$	n = 9 mice
1m	Slope across conditions (LF, LD, SD, SF)	Tukey HSD test for multiple comparison of means	q = 3.83	LD vs LF: p = 0.762 LD vs SD: **p = 0.00266 LD vs SF: p = 0.952 LF vs SD: ***$p = 1.70 \times 10^{-4}$ LF vs SF: p = 0.445 SD vs SF: *p = 0.0107	n = 9 mice
2b	Trial slope regression β given previous ITI (SD condition only)	One-sided (< 0) one sample t-test	t(8) = -2.17	*p = 0.0308	n = 9 mice
2c	Trial slope given previous ITI (SD condition only)	Linear regression	t(2671) = -4.13 R^2 = 0.00634	***$p = 3.77 \times 10^{-5}$	n = 2672 trials
3e	Change in velocity at trial onset (long ITI condition only)	One-sided (> 0) one sample t-test	t(8) = 6.40	***$p = 1.05 \times 10^{-4}$	n = 9 mice
3e	Change in velocity at trial onset (short ITI condition only)	One-sided (> 0) one sample t-test	t(8) = 7.93	***$p = 2.33 \times 10^{-5}$	n = 9 mice
3e	Change in velocity at trial onset between conditions (long, short)	Paired t-test	t(8) = 4.25	**p = 0.00281	n = 9 mice
3g	Pre-reward velocity between conditions (long, short)	Paired t-test	t(8) = 0.71	p = 0.497	n = 9 mice
3i	Cue onset peak dLight between conditions (long, short)	Paired t-test	t(8) = 7.59	***$p = 6.34 \times 10^{-5}$	n = 9 mice
3l	Session slope given session IRI (both long and short ITI conditions)	Linear regression	t(53) = -2.61 R^2 = 0.116	*p = 0.0118	n = 54 sessions
3m	Slope between conditions (long, short)	One-sided (long < short) paired t-test	t(8) = -2.09	*p = 0.0349	n = 9 mice

Figure	Description	Test	Statistic	p value	Sample size
Ext 2d	Lick rate during ramp window across conditions (LF, LD, SD, SF)	One-way ANOVA	F(3) = 0.81	p = 0.498	n = 9 mice
Ext 3b	Trial slope regression β given previous ITI (LD condition only)	One-sided (< 0) one sample t-test	t(8) = 0.36	p = 0.637	n = 9 mice
Ext 3c	Trial slope given previous ITI (LD condition only)	Linear regression	t(1050) = 0.64 $R^2 = 3.85 \times 10^{-4}$	p = 0.525	n = 1051 trials
Ext 4c	Trial duration between conditions (long, short)	Paired t-test	t(8) = 1.02	p = 0.336	n = 9 mice
Ext 4e	Anticipatory lick rate (long ITI condition only)	One-sided (> 0) one sample t-test	t(8) = 2.89	*p = 0.0101	n = 9 mice
Ext 4e	Anticipatory lick rate (short ITI condition only)	One-sided (> 0) one sample t-test	t(8) = 4.38	**p = 0.00118	n = 9 mice
Ext 4e	Anticipatory lick rate between conditions (long, short)	Paired t-test	t(8) = 1.08	p = 0.311	n = 9 mice
Ext 5b	Trial slope regression β given previous IRI (short ITI condition only)	One-sided (< 0) one sample t-test	t(8) = -0.48	p = 0.321	n = 9 mice
Ext 5c	Trial slope given previous IRI (short ITI condition only)	Linear regression	t(1301) = -2.11 $R^2 = 0.00339$	*p = 0.0355	n = 1302 trials

Table 1. Statistical Details.