# Mesolimbic dopamine ramps reflect environmental timescales

**Joseph R Floeder[1], Huijeong Jeong[2], Ali Mohebi[2], and Vijay Mohan K Namboodiri[1,2,3,4]**

[1]Neuroscience Graduate Program, University of California, San Francisco, CA, USA
[2]Department of Neurology, University of California, San Francisco, CA, USA
[3]Weill Institute for Neurosciences, Kavli Institute for Fundamental Neuroscience, Center for Integrative Neuroscience, University of California, San Francisco, CA, USA
[4]Correspondence to VijayMohan.KNamboodiri@ucsf.edu

**Mesolimbic dopamine activity occasionally exhibits ramping dynamics, reigniting debate on theories of dopamine signaling. This debate is ongoing partly because the experimental conditions under which dopamine ramps emerge remain poorly understood. Here, we show that during Pavlovian and instrumental conditioning, mesolimbic dopamine ramps are only observed when the inter-trial interval is short relative to the trial period. These results constrain theories of dopamine signaling and identify a critical variable determining the emergence of dopamine ramps.**

## Introduction

Mesolimbic dopamine activity was classically thought to operate in either a "phasic" or a "tonic" mode(1–3). Yet, recent evidence points to a "quasi-phasic" mode in which mesolimbic dopamine activity exhibits ramping dynamics(4–16). This discovery reignited debate on theories of dopamine function because it appeared inconsistent with the dominant theory that dopamine signaling conveys temporal difference reward prediction error (RPE)(17), since ramping dopamine would paradoxically be a "predictable prediction error"(18). Recent work has hypothesized that dopamine ramps reflect the value of ongoing states, serving as a motivational signal(1, 4–6). Others have argued that ramping dopamine indeed reflects RPE under some assumptions, namely correction of uncertainty via sensory feedback(11, 15, 19), representational error(20), or memory lapses(21). Still others have proposed that dopamine ramps reflect a causal influence of actions on rewards in instrumental tasks(12). This debate has been exacerbated in part because there is no clear understanding of why dopamine ramps appear only under some experimental conditions. Accordingly, uncovering a unifying principle of the conditions under which dopamine ramps appear will provide important constraints on theories of dopamine function(1, 12, 18–42).

To investigate the necessary conditions for dopamine ramps, we turned to our recent work proposing that dopamine acts as a teaching signal for causal learning by representing the Adjusted Net Contingency for Causal Relations (ANCCR)(32, 42, 43). The crux of the ANCCR model is that learning to predict a future meaningful event (e.g., reward) can occur by looking back in time for potential causes of that event. Critically, the ability to learn by looking backwards depends on how long one holds on to past events in memory. If the past is quickly forgotten, there is little ability to identify causes that occurred long before a reward. On the other hand, maintaining memory for too long is computationally inefficient and would allow illusory associations across long delays. Thus, for optimal learning, the timescale for memory maintenance should flexibly depend on "environmental timescales" set by the overall rates of events. In ANCCR, this is achieved by controlling the duration of a memory trace of past events with the "eligibility trace" time constant (illustrated in **Extended Data Fig 1**). When there is a high average rate of events, the eligibility trace time constant is small. Accordingly, we successfully simulated dopamine ramping dynamics assuming two conditions: a dynamic progression of cues that signal temporal proximity to reward, and a small eligibility trace time constant relative to the trial period(32). However, whether these conditions are sufficient to experimentally produce mesolimbic dopamine ramps *in vivo* remains untested.

In this study, we designed experiments to address the influence of environmental timescales on dopamine ramps. Specifically, we sought to test the key prediction that dopamine ramps would be observed for a small, but not large, eligibility trace time constant, which is hypothesized to emerge for high overall event rate. To do so, we manipulated the inter-trial interval (ITI) duration in both an auditory Pavlovian conditioning paradigm and a virtual reality navigation task. The results confirmed our key prediction from ANCCR, providing a clear constraint on theoretical explanations for the controversial phenomenon of dopamine ramps.

## Results

We first measured mesolimbic dopamine release in the nucleus accumbens core using a dopamine sensor (dLight1.3b)(44) in an auditory cue-reward task. We varied both the presence or absence of a progression of cues indicating reward proximity ("dynamic" vs "fixed" tone) and the inter-trial interval (ITI) duration (short vs long ITI). Varying the ITI was critical because our theory predicts that the ITI is a variable controlling the eligibility trace time constant, such that a short ITI would produce a small time constant relative to the cue-reward interval (**Supplementary Note 1**, **Fig 1a-e**). In all four experimental conditions, head-fixed mice learned to anticipate the sucrose reward, as reflected by anticipatory licking (**Fig 1f-g**). In line with our earlier work, we showed that simulations of ANCCR exhibit a larger cue onset response when the ITI is long and exhibit ramps only when the ITI is short (**Fig 1h**). Consistent with these simulated predictions, experimentally measured mesolimbic dopamine release had a much higher cue

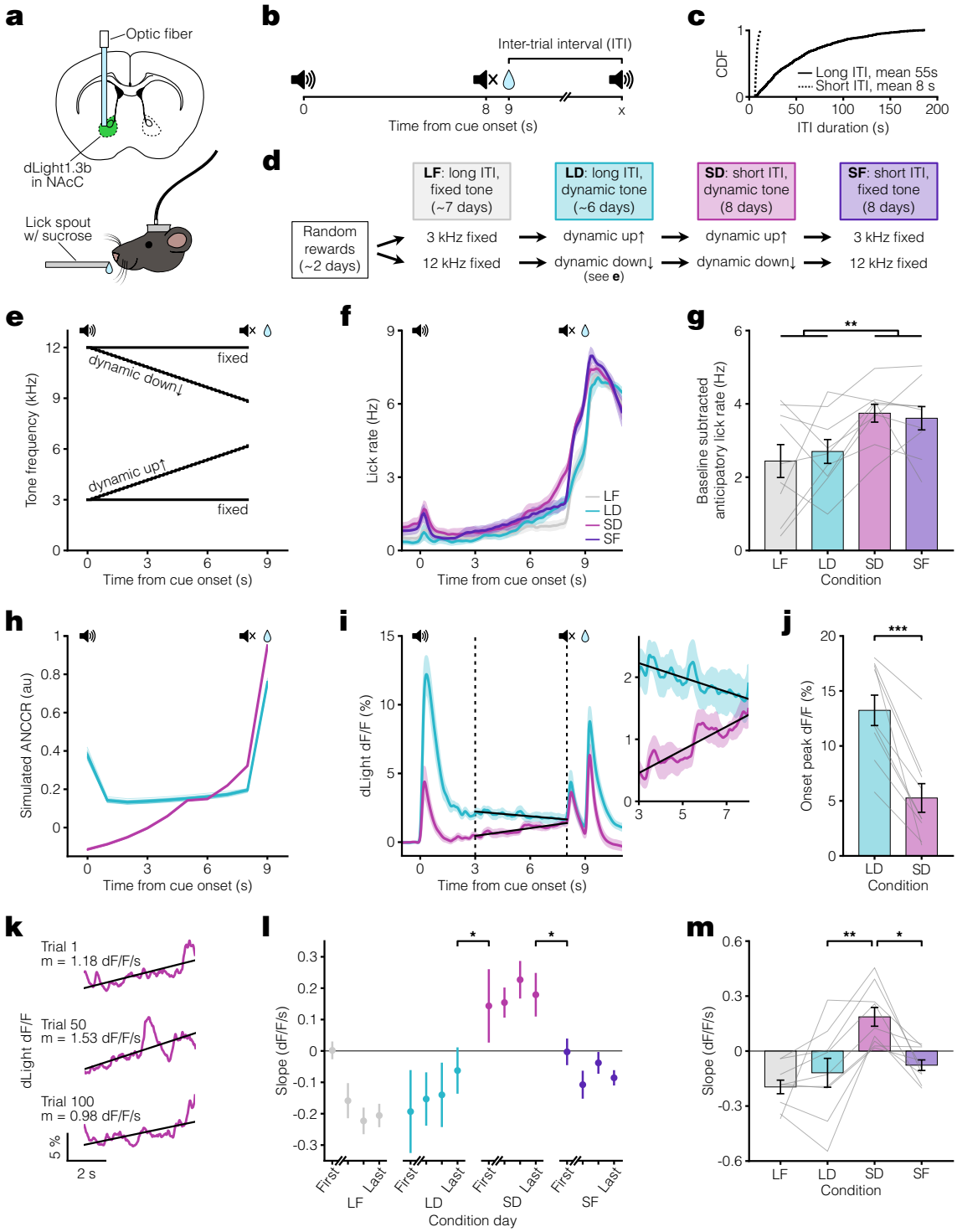



**Fig. 1. Pavlovian conditioning dopamine ramps depend on ITI. a**. Top, fiber photometry approach schematic for nucleus accumbens core (NAcC) dLight recordings. Bottom, head-fixed mouse. **b**. Pavlovian conditioning experimental setup. Trials consisted of an 8 s auditory cue followed by sucrose reward delivery 1 s later. **c**. Cumulative Distribution Function (CDF) of ITI duration for long (solid line, mean 55 s) and short ITI (dashed line, mean 8 s) conditions. **d**. Experimental timeline. Mice were divided into groups receiving either a 3 kHz fixed and dynamic up↑ tone or a 12 kHz fixed and dynamic down↓ tone. **e**. Tone frequency over time. **f**. Peri-stimulus time histogram (PSTH) showing average licking behaviors for the last 3 days of each condition (n = 9 mice). **g**. Average anticipatory lick rate (baseline subtracted) for 1 s preceding reward delivery (two-way ANOVA: long ITI vs short ITI $F(1) = 9.3$, $^{**}p = 0.0045$). **h**. ANCCR simulation results from an 8 s dynamic cue followed by reward 1 s later for long ITI (teal) and short ITI (pink) conditions. Bold lines show the average of 20 iterations. **i**. Left, average dLight dopamine signals. Vertical dashed lines represent the ramp window from 3 to 8 s after cue onset, thereby excluding the influence of the cue onset and offset responses. Solid black lines show linear regression fit during window. Right, closeup of dopamine signal during window. **j**. Average peak dLight response to cue onset for LD and SD conditions (paired t-test: $t(8) = 6.3$, $^{***}p = 2.3\times10^{-4}$). **k**. dLight dopamine signal with linear regression fit during ramp window for example SD trials. Reported m is slope. **l**. Session average per-trial slope during ramp window for the first day and last 3 days of each condition (one-sided [last day LD < first day SD] paired t-test: $t(8) = -2.1$, $^{*}p = 0.036$; one-sided [last day SD > first day SF] paired t-test: $t(8) = 2.4$, $^{*}p = 0.023$). **m**. Average per-trial slope for last 3 days of each condition (Tukey HSD test: $q = 3.8$, LD vs SD $^{**}p = 0.0027$, SD vs SF $^{*}p = 0.011$). All data presented as mean ± SEM. See **Supplementary Table 1** for full statistical details.

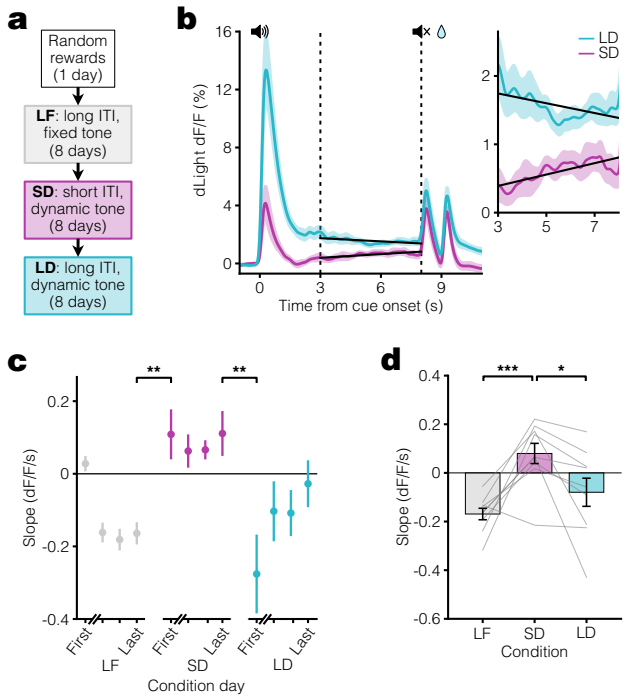



**Fig. 2. Pavlovian conditioning dopamine ramps do not depend on training order. a**. Experimental timeline in which the SD condition occurs before the LD condition. **b**. Left, average dLight dopamine signals for SD and LD conditions. Vertical dashed lines represent the ramp window from 3 to 8 s after cue onset. Solid black lines show linear regression fit during window. Right, closeup of dopamine signal during window (n = 9 mice). **c**.Session average per-trial slope during ramp window for the first day and last 3 days of each condition (last day LF to first day SD: **p = 0.0046, last day SD to first day LD: **p = 0.0067). **d**. Average per-trial slope for last 3 days of each condition (LF vs SD: ***p = 9.1 $\times 10^{-4}$, SD vs LD: *p = 0.010).

onset response for long ITI (**Fig 1i-j**). Furthermore, dopamine ramps were observed only when the ITI was short and the tone was dynamic (**Fig 1i**, **k-m**, **Extended Data Fig 2**). Indeed, dopamine ramps—quantified by a positive slope of dopamine response vs time within trial over the last five seconds of the cue—appeared on the first day after transition from a long ITI/dynamic tone condition to a short ITI/dynamic tone condition and disappeared on the first day after transition from a short ITI/dynamic tone condition to a short ITI/fixed tone condition (**Fig 1l**). These results confirm the key prediction of our theory in Pavlovian conditioning.

While these results are consistent with the idea that dopamine ramps are shaped by the ITI, an alternative explanation could be differences in behavioral learning across experimental conditions. To test this possibility, we repeated the same Pavlovian conditioning paradigm with a counterbalanced training order in a second cohort of mice (**Fig 2a**). Despite the shuffled training order, this cohort behaved similarly and showed robust anticipatory licking across conditions (**Extended Data Fig 3a-b**). As with the previous cohort, dopamine ramps were only observed in the short ITI/dynamic tone condition, rapidly appearing on the first day of this condition and disappearing on the first day of the subsequent long ITI/dynamic tone condition (**Fig 2b-d**, **Extended Data Fig 4**, **Extended Data Fig 5**). Critically, the presence of dopamine ramps during the last five seconds of the cue could not be explained by variations in behavior; during this period, anticipatory licking was similar across all conditions, and there was no difference in the slope of the lick rate between the dynamic tone conditions (**Extended Data Fig 3c-d**). Taken together, these results rule out any effects of differential learning across conditions on dopamine ramps.

Although the difference in dopamine ramp slope seems to be well explained by the ITI condition, it might instead reflect differences in post-reward dopamine dynamics, which drop below baseline. As dopamine levels recover from this drop over several seconds, it could appear as a dopamine ramp on the subsequent trial given a sufficiently short ITI. The lack of dopamine ramps in the short ITI/fixed tone condition serves as a control for this, however (**Fig 1l-m**). Furthermore, there is no significant difference in the pre-cue dopamine slope between conditions nor is there a correlation between the pre-cue dopamine slopes and the dopamine ramp slopes during the cue in the short ITI/dynamic condition (**Extended Data Fig 6**). As such, our results cannot be captured by a natural ramp in the dopamine signal following reward.

Given the speed with which dopamine ramps appeared and disappeared, we next tested whether the slope of dopamine ramps in the short ITI/dynamic tone condition depended on the previous ITI duration on a trial-by-trial basis. We found that there was indeed a statistically significant trial-by-trial correlation between the previous ITI duration and the current trial's dopamine response slope in the short ITI/dynamic condition with ramps, but not in the long ITI/dynamic condition without ramps (**Fig 3a-c**). The dependence of a trial's dopamine response slope with previous ITI was significantly negative, meaning that a longer ITI correlates with a weaker ramp on the next trial. This finding held when analyzing either animal-by-animal (**Fig 3a-b**) or the pooled trials across animals while accounting for mean animal-by-animal variability (**Fig 3c**). This relationship was only significant for a single previous trial, however, and did not hold for a broader estimate of average previous ITIs (**Extended Data Fig 7**). In addition, we quantified how the relative change in ITI duration between consecutive trials correlates with changes in dopamine ramp slope (**Fig 3d**). We found a significantly negative relationship between the change in dopamine slope and change in ITI (**Fig 3e-f**). Furthermore, the change in slope was significantly greater for relative decreases in ITI compared to relative increases in ITI, indicating that a relatively shorter ITI tends to have a stronger ramp (**Fig 3g**). These results suggest that the eligibility trace time constant adapts rapidly to changing ITI in Pavlovian conditioning.

Due to the robust relationship between ITI and dopamine ramp slope on a per-trial basis, we next sought to explore the potential relationships between other important dopaminergic and behavioral variables. Though the dopamine cue onset response is significantly greater in the long compared to short ITI/dynamic tone condition, there is no apparent relationship between the cue onset response and dopamine ramp slope in either condition (**Extended Data Fig 8a-b**). Furthermore, neither the cue onset response nor the dopamine ramp slope correlates with the per-trial behavior quantified as lick slope (**Extended Data Fig 8c-f**). Finally, unlike the ramping

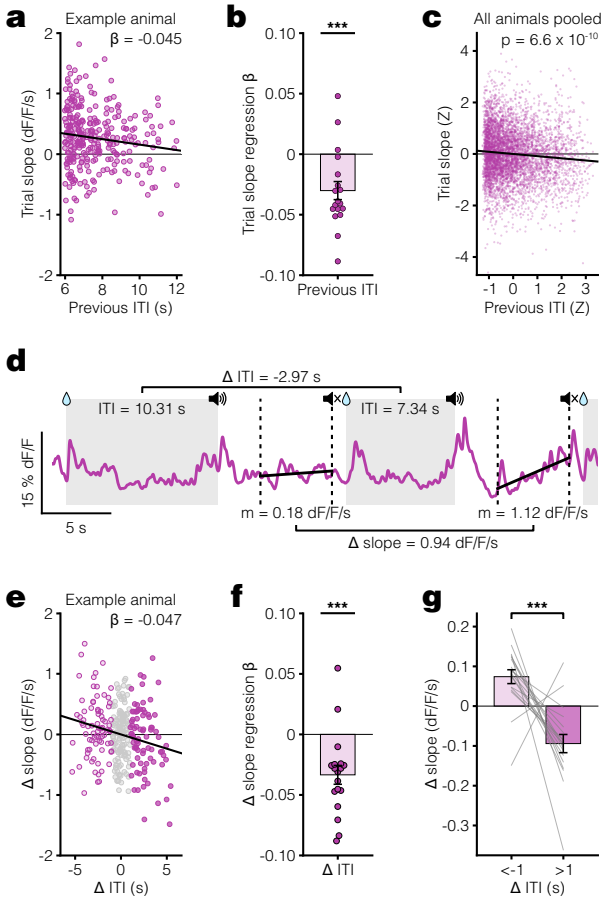



**Fig. 3. Per-trial dopamine ramps correlate with previous ITI. a**. Scatter plot for an example animal showing the relationship between dopamine response slope within a trial and previous ITI for all trials in the last 3 days of SD condition. Plotted with linear regression fit (black line) used to find this animal's β coefficient of -0.045. **b**. Linear regression β coefficients for previous ITI vs. trial slope calculated per animal (***p = 5.6 $\times 10^{-4}$). **c**. Scatter plot of Z-scored trial slope vs. previous ITI pooled across mice for all trials in the last 3 days of SD condition (***p = 6.6 $\times 10^{-10}$). The Z-scoring per animal removes the effect of variable means across animals on the slope of the pooled data. **d**.dLight dopamine signal for two consecutive example SD trials showing the change in ITI and change in slope. The grey shaded regions indicate ITIs, and the vertical dashed lines mark the ramp window period. Reported m is slope. **e**. Scatter plot for the same example animal in **a** showing the relationship between the change in dopamine slopes and the chance in ITI across all trials in the last 3 days of SD condition. Plotted with linear regression fit (black line). Dot colors indicate magnitude of Δ ITI: light pink for Δ ITI below -1 s; grey for Δ ITI between -1 s and 1 s; dark pink for Δ ITI above 1 s. **f**. Linear regression β coefficients for Δ ITI vs. Δ slope calculated per animal (***p = 3.2 $\times 10^{-4}$). **g**. Comparison of the average Δ slope for Δ ITI below -1 s vs above 1 s (*** p = 2.3 $\times 10^{-4}$).

dopamine slope, this ramping lick slope did not correlate with ITI duration (**Extended Data Fig 8g-h**). The fact that this exploration of additional variables yielded no significant relationships highlights the unique, specific influence of ITI on dopamine ramp slope.

We next tested whether the results from Pavlovian conditioning could be reproduced in an instrumental task. In keeping with prior demonstrations of dopamine ramps in head-fixed mice, we used a virtual reality (VR) navigational task in which head-fixed mice had to run towards a destination in a virtual hallway to obtain sucrose rewards(11, 15, 19, 45) (**Fig 4a-b**, **Extended Data Fig 9**). At reward delivery, the screen turned blank during the ITI and remained so until the next trial onset. After training animals in this task using a medium ITI, we changed the ITI duration to short or long for eight days before switching to the other (**Fig 4c**). We found evidence that mice learned the behavioral requirement during the trial period, as they significantly increased their running speed during trial onset (**Fig 4d-e**) and reached a similarly high speed prior to reward in both ITI conditions (**Fig 4f-g**). Consistent with the results from Pavlovian conditioning, the dopamine response to the onset of the hallway presentation was larger during the long ITI compared to the short ITI condition (**Fig 4h-i**), and dopamine ramps were observed only in the short ITI condition (**Fig 4j-m**). Unlike the Pavlovian conditioning, the change in the ITI resulted in a more gradual appearance or disappearance of ramps (**Fig 4l**), but there was still a weak overall correlation between dopamine response slope on a trial and the previous inter-reward interval (**Extended Data Fig 10**). These results are consistent with a more gradual change in the eligibility trace time constant in this instrumental task. Collectively, the core finding from Pavlovian conditioning that mesolimbic dopamine ramps are present only during short ITI conditions was reproduced in the instrumental VR task.

## Discussion

Our results provide a general framework for understanding past results on dopamine ramps. According to ANCCR, the fundamental variable controlling the presence of ramps is the eligibility trace time constant. Based on first principles, this time constant depends on the ITI in common task designs (**Supplementary Note 1**). Thus, the ITI is a simple proxy to manipulate the eligibility trace time constant, thereby modifying dopamine ramps. In previous navigational tasks with dopamine ramps, there was no explicitly programmed ITI(4, 10, 16). As such, the controlling of the pace of trials by these highly motivated animals likely resulted in short effective ITI compared to trial duration. An instrumental lever pressing task with dopamine ramps similarly had no explicitly programmed ITI(7), and other tasks with observed ramps had short ITIs(8, 11, 13–15). One reported result that does not fit with a simple control of ramps by ITI is that navigational tasks produce weaker ramps with repeated training(10). These results are generally inconsistent with the stable ramps that we observed in Pavlovian conditioning across eight days (**Fig 1l**). A speculative explanation might be that when the timescales of events vary considerably (e.g., during early experience in instrumental tasks due to variability in action timing), animals use a short eligibility trace time constant to account for the potential non-stationarity of the environment. With repeated exposure, the experienced stationarity of the environment might increase the eligibility trace time constant, thereby complicating its relationship with the ITI. Alternatively, as suggested previously(7, 10), repeated navigation may result in automated behavior that ignores the progress towards reward, thereby minimizing the calculation of associations of spatial locations with reward.

While the focus of this study was environmental timescales set by the ITI, we also assumed that a dynamic sequence of external cues signaling temporal or spatial

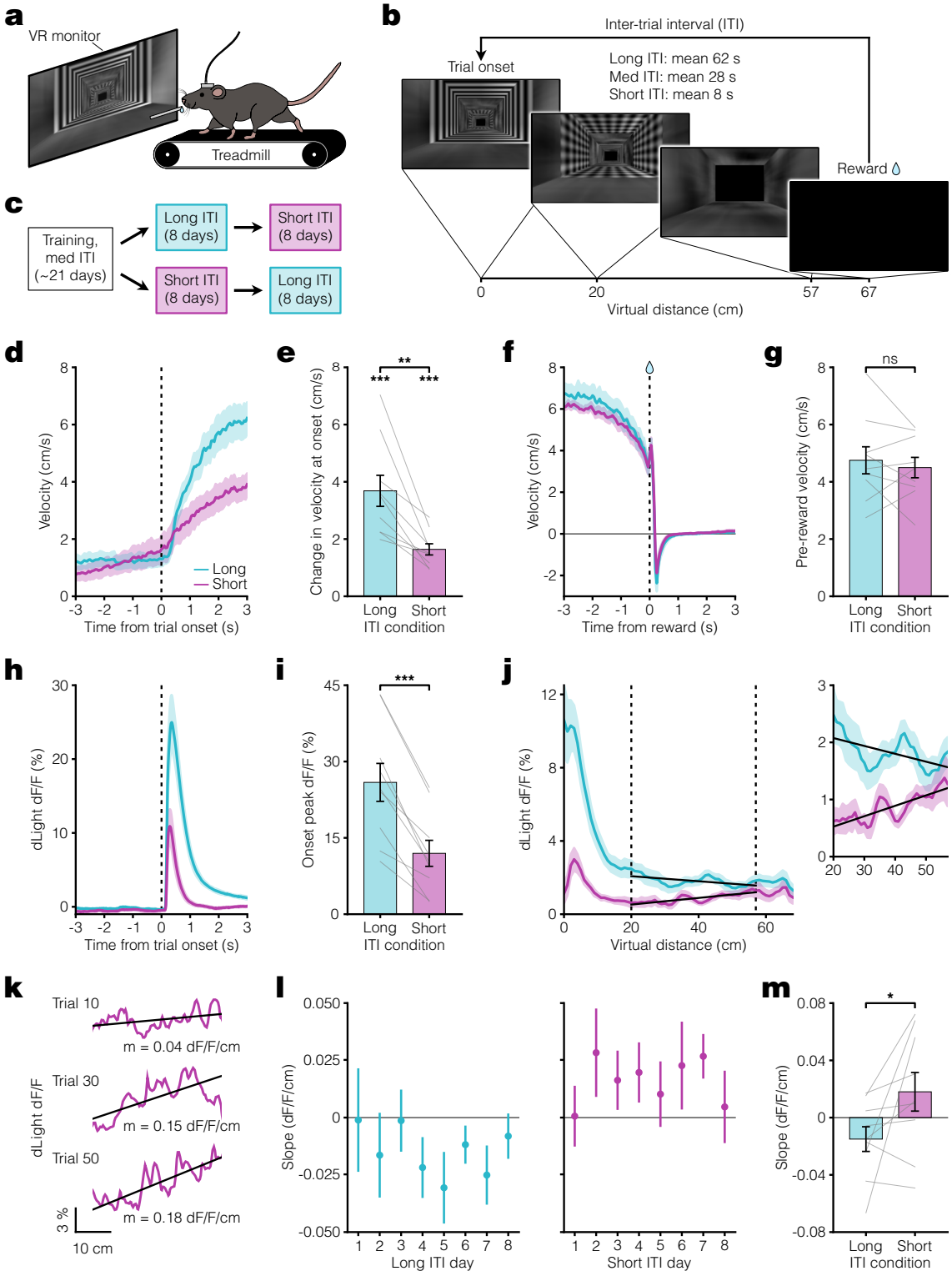



**Fig. 4. VR navigation dopamine ramps depend on ITI. a**. Head-fixed VR approach schematic. **b**. VR navigation task experimental setup. Trials consisted of running down a patterned virtual hallway to receive sucrose reward. VR monitor remained black during the ITI. **c**. Experimental timeline. Following training, mice were assigned to either long or short ITI conditions for 8 days before switching. **d**. Velocity PSTH aligned to trial onset for long (teal) and short (pink) ITI conditions (n = 9 mice). **e** Average change in velocity at trial onset. Bottom asterisks indicate both conditions significantly differ from zero (long: ***p = $1.0 \times 10^{-4}$; short: ***p = $2.3 \times 10^{-5}$). Top asterisks indicate significant difference between conditions (**p = 0.0028). **f**. Velocity PSTH aligned to reward delivery. **g**. Average velocity during 1 s preceding reward (p = 0.50). **h**. PSTH showing average dLight dopamine signal aligned to trial onset. **i**. Comparison of peak dLight onset response (***p = $6.3 \times 10^{-5}$). **j**. Left, average dLight dopamine signal across distances spanning the entire virtual corridor. Vertical dashed lines represent the ramp window from 20 to 57 cm (10 cm before end of track). Solid black lines show linear regression fit during window. Right, closeup of dopamine signal during window. **k**. dLight dopamine signal with linear regression fit during ramp window for example short ITI trials. Reported m is slope. **l**. Session average per-trial slope during ramp window for all days of each condition. **m**. Comparison of average per-trial slope during ramp window for last 3 days of both conditions (*p = 0.035).

proximity to reward would be required for dopamine ramps to occur. Indeed, our results that dopamine ramps occur in the short ITI/dynamic, but not fixed, tone condition corroborate this assumption, as well as results from other Pavlovian conditioning experiments utilizing dynamic cues in a head-fixed setup(11, 19). In contrast, experiments involving freely moving animals do not require explicitly dynamic cues because the sensory feedback from navigating though the environment presumably functions in the same way to indicate proximity to reward. Future experiments can investigate this further by characterizing the role of sensory feedback indicating reward proximity in mediating dopamine ramps. One set of observations superficially inconsistent with our assumption of the necessity of a sequence of external cues is that ramping dopamine dynamics can be observed even when only internal states signal reward proximity (e.g., timing a delayed action)(10, 13, 14). In these cases, however, animals were required to actively keep track of the passage of time, which therefore strengthens an internal progression of neural states signaling temporal proximity to reward. We speculate that once learned, these internal states could serve the role of external cues in the ANCCR framework. Previously, we argued against this kind of assumption in learning theories(46). Our earlier position was that it is problematic to assume fixed internal states that pre-exist and provide a scaffold for learning, such as in temporal difference learning. This is because these pre-existing states would need to already incorporate information that can only be acquired during the course of learning(46). 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(32).

Though the experiments in this study were motivated by the ANCCR framework, they were not conducted to discriminate between theories. As such, it is also possible to rationalize these results in the context of other models of dopamine ramps. In the value model, dopamine is thought to represent the discounted sum of future rewards(4–6, 18). The shape of this value function, and thus the predicted dopamine dynamics, is determined by the discount factor, $\gamma$. If there is a low $\gamma$ (i.e., greater discounting), then the corresponding value function produces a steeper ramp. Consequently, it is possible to use the value model to explain our results if one assumes that a shorter ITI causes greater temporal discounting. The basis for such an assumption is unclear, though it has been suggested that the overall temporal discounting in an environment depends on reward rates(47, 48). Furthermore, we do not find evidence for dopamine ramps acting as a value signal to directly increase motivation. This is because we find similar trial-related behaviors in conditions with and without dopamine ramps. Thus, ITI dependent emergence of dopamine ramps for the same trial parameters provides strong constraints for the motivational role of dopamine ramps.

Substantial efforts have been made to account for the phenomenon of dopamine ramps as a temporal difference RPE(11, 18–20, 36). As with the value model, simulated dopamine responses in temporal difference RPE models are also modulated by the discount factor, $\gamma$. It has been proposed that temporal discounting in the dopamine system depends on the cue-reward delay(35). In our experiment, however, the cue-reward delay is not the key variable determining the presence of ramps; instead, it is the ITI. Another work has also proposed that a spectrum of discount factors can explain diverse activity profiles of single dopamine neurons(36). Specifically, monotonic upward ramps were simulated using a high $\gamma$ (i.e., weaker discounting). Therefore, in this model, one would need to assume that shorter ITIs cause weaker temporal discounting to produce steeper ramps. Notably, this is in the opposite direction as the value model. Overall, it is unclear whether any fundamental principle predicts an ITI dependent change in temporal discounting in the dopamine system to allow RPE to explain our results. Similarly, whether other models of dopamine ramps(12, 18, 33) can capture an ITI dependent emergence of dopamine ramps remains to be explored.

While it is thus possible to rationalize our results using alternative theories of dopamine, ANCCR provides a principled and parsimonious explanation. Given that the foundation of ANCCR is looking back in time for causes of rewards, it is clear that differences in memory maintenance via eligibility traces will have profound implications on predicted dopamine signaling. For example, when the ITI is short and rewards are being frequently delivered, it intuitively makes sense that the eligibility trace time constant needs to be small; this is because the time window over which one would want to search for potentially causal cues is going to be shorter in this situation. We formalize this intuition by postulating that the eligibility trace time constant adapts to the overall event rates for efficient coding (**Supplementary Note 1**). In the case of a dynamic progression of cues, the cues closer in time to the reward will have higher causal power, and thus higher ANCCR, resulting in a dopamine ramp.

Our ANCCR simulations motivated the experiments, but we did not explicitly intend to fit the data. Accordingly, there are several details of the experiments that we did not include in the simulations. First, animals were trained initially using a long ITI (Pavlovian) or medium ITI (VR). This may explain a discrepancy between the simulations and experimental results: the cue onset response in the short ITI condition is small but positive in the experiment but negative in ANCCR. This discrepancy may be because the cue onset was already learned to be meaningful prior to the short ITI condition, thereby resulting in a stronger cue onset response in the experimental data. Further, we did not explicitly model potential trial-by-trial changes in eligibility trace time constant, sensory noise, internal threshold, local mechanisms controlling dopamine release, or sensor dynamics. Thus, we did not expect to capture all experimental observations in the motivating simulations. Regardless of such considerations, the current results provide a clear constraint for dopamine theories and demonstrate that an underappreciated experimental variable determines the emergence of mesolimbic dopamine ramps.

**ACKNOWLEDGEMENTS**

We thank J. Berke and members of the Namboodiri laboratory for helpful discussions. This project was supported by the NIH (grants R00MH118422 and R01MH129582 to V.M.K.N.), the NSF (graduate research fellowship to J.R.F.), the UCSF Discovery Fellowship (J.R.F.), and the Scott Alan Myers Endowed Professorship (V.M.K.N.). The authors have no competing interests.

**AUTHOR CONTRIBUTIONS**

J.R.F. and V.M.K.N. conceived the project. J.R.F. performed experiments and analyses. H.J. performed simulations. A.M. helped with the design and instrumentation of the VR experiment. V.M.K.N. oversaw all aspects of the study. J.R.F. and V.M.K.N. wrote the manuscript with help from all authors.

# Bibliography

1. Joshua D Berke. What does dopamine mean? *Nature Neuroscience*, 21(6):787–793, 2018.
2. Anthony A Grace. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, 41(1): 1–24, 1991.
3. Yael Niv, Nathaniel D Daw, Daphna Joel, and Peter Dayan. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, 191:507–520, 2007.
4. Mark W Howe, Patrick L Tierney, Stefan G Sandberg, Paul EM Phillips, and Ann M Graybiel. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, 500(7464):575–579, 2013.
5. Arif A Hamid, Jeffrey R Pettibone, Omar S Mabrouk, Vaughn L Hetrick, Robert Schmidt, Caitlin M Vander Weele, Robert T Kennedy, Brandon J Aragona, and Joshua D Berke. Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, 19(1):117–126, 2016.
6. Ali Mohebi, Jeffrey R Pettibone, Arif A Hamid, Jenny-Marie T Wong, Leah T Vinson, Tommaso Patriarchi, Lin Tian, Robert T Kennedy, and Joshua D Berke. Dissociable dopamine dynamics for learning and motivation. *Nature*, 570(7759):65–70, 2019.
7. Anne L Collins, Venuz Y Greenfield, Jeffrey K Bye, Kay E Linker, Alice S Wang, and Kate M Wassum. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Scientific Reports*, 6(1):20231, 2016.
8. Emilie CJ Syed, Laura L Grima, Peter J Magill, Rafal Bogacz, Peter Brown, and Mark E Walton. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nature Neuroscience*, 19(1):34–36, 2016.
9. Ben Engelhard, Joel Finkelstein, Julia Cox, Weston Fleming, Hee Jae Jang, Sharon Ornelas, Sue Ann Koay, Stephan Y Thiberge, Nathaniel D Daw, David W Tank, et al. Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. *Nature*, 570 (7762):509–513, 2019.
10. Akash Guru, Changwoo Seo, Ryan J Post, Durga S Kullakanda, Julia A Schaffer, and Melissa R Warden. Ramping activity in midbrain dopamine neurons signifies the use of a cognitive map. *bioRxiv*, 2020.
11. HyungGoo R Kim, Athar N Malik, John G Mikhael, Pol Bech, Iku Tsutsui-Kimura, Fangmiao Sun, Yajun Zhang, Yulong Li, Mitsuko Watabe-Uchida, Samuel J Gershman, et al. A unified framework for dopamine signals across timescales. *Cell*, 183(6):1600–1616, 2020.
12. Arif A Hamid, Michael J Frank, and Christopher I Moore. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell*, 184(10):2733–2749, 2021.
13. Allison E Hamilos, Giulia Spedicato, Ye Hong, Fangmiao Sun, Yulong Li, and John A Assad. Slowly evolving dopaminergic activity modulates the moment-to-moment probability of reward-related self-timed movements. *eLife*, 10:e62583, 2021.
14. Zilong Gao, Hanqing Wang, Chen Lu, Tiezhan Lu, Sean Froudist-Walsh, Ming Chen, Xiao-Jing Wang, Ji Hu, and Wenzhi Sun. The neural basis of delayed gratification. *Science Advances*, 7(49):eabg6611, 2021.
15. Karolina Farrell, Armin Lak, and Aman B Saleem. Midbrain dopamine neurons signal phasic and ramping reward prediction error during goal-directed navigation. *Cell Reports*, 41(2), 2022.
16. Timothy A Krausz, Alison E Comrie, Ari E Kahn, Loren M Frank, Nathaniel D Daw, and Joshua D Berke. Dual credit assignment processes underlie dopamine signals in a complex spatial environment. *Neuron*, 111(21):3465–3478, 2023.
17. Yael Niv. Dopamine ramps up. *Nature*, 500(7464):533–535, 2013.
18. Kevin Lloyd and Peter Dayan. Tamping ramping: Algorithmic, implementational, and computational explanations of phasic dopamine signals in the accumbens. *PLOS Computational Biology*, 11(12):e1004622, 2015.
19. John G Mikhael, HyungGoo R Kim, Naoshige Uchida, and Samuel J Gershman. The role of state uncertainty in the dynamics of dopamine. *Current Biology*, 32(5):1077–1087, 2022.
20. Samuel J. Gershman. Dopamine ramps are a consequence of reward prediction errors. *Neural Computation*, 26(3):467–471, 2014.
21. Kenji Morita and Ayaka Kato. Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. *Frontiers in Neural Circuits*, 8, 2014.
22. Peter Redgrave and Kevin Gurney. The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, 7(12):967–975, 2006.
23. Kent C Berridge. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, 191:391–431, 2007.
24. Ethan S Bromberg-Martin, Masayuki Matsumoto, and Okihide Hikosaka. Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, 68(5):815–834, 2010.
25. Scott Waddell. Reinforcement signalling in drosophila; dopamine does it all after all. *Current Opinion in Neurobiology*, 23(3):324–329, 2013.
26. Matthew PH Gardner, Geoffrey Schoenbaum, and Samuel J Gershman. Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B: Biological Sciences*, 285(1891):20181645, 2018.
27. Benjamin T Saunders, Jocelyn M Richard, Elyssa B Margolis, and Patricia H Janak. Dopamine neurons create pavlovian conditioned stimuli with circuit-defined motivational properties. *Nature Neuroscience*, 21(8):1072–1083, 2018.
28. Melissa J Sharpe, Hannah M Batchelor, Lauren E Mueller, Chun Yun Chang, Etienne JP Maes, Yael Niv, and Geoffrey Schoenbaum. Dopamine transients do not act as model-free prediction errors during associative learning. *Nature Communications*, 11(1):106, 2020.
29. Ryan N Hughes, Konstantin I Bakhurin, Elijah A Petter, Glenn DR Watson, Namsoo Kim, Alexander D Friedman, and Henry H Yin. Ventral tegmental dopamine neurons control the impulse vector during motivated behavior. *Current Biology*, 30(14):2681–2694, 2020.
30. Etienne J P Maes, Melissa J Sharpe, Alexandra A Usypchuk, Megan Lozzi, Chun Yun Chang, Matthew PH Gardner, Geoffrey Schoenbaum, and Mihaela D Iordanova. Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nature Neuroscience*, 23(2):176–178, 2020.
31. Munir Gunes Kutlu, Jennifer E Zachry, Patrick R Melugin, Stephanie A Cajigas, Maxime F Chevee, Shannon J Kelly, Banu Kutlu, Lin Tian, Cody A Siciliano, and Erin S Calipari. Dopamine release in the nucleus accumbens core signals perceived saliency. *Current Biology*, 31(21):4748–4761, 2021.
32. Huijeong Jeong, Annie Taylor, Joseph R Floeder, Martin Lohmann, Stefan Mihalas, Brenda Wu, Mingkang Zhou, Dennis A Burke, and Vijay Mohan K Namboodiri. Mesolimbic dopamine release conveys causal associations. *Science*, 378(6626):eabq6740, 2022.
33. Anthony Jakob, John G Mikhael, Allison E Hamilos, John A Assad, and Samuel J Gershman. Dopamine mediates the bidirectional update of interval timing. *Behavioral Neuroscience*, 136(5):445, 2022.
34. Luke T Coddington, Sarah E Lindo, and Joshua T Dudman. Mesolimbic dopamine adapts the rate of learning from action. *Nature*, 614(7947):294–302, 2023.
35. Margarida Sousa, Pawel Bujalski, Bruno F Cruz, Kenway Louie, Daniel McNamee, and Joseph J Paton. Dopamine neurons encode a multidimensional probabilistic map of future reward. *bioRxiv*, 2023.
36. Paul Masset, Pablo Tano, HyungGoo R Kim, Athar N Malik, Alexandre Pouget, and Naoshige Uchida. Multi-timescale reinforcement learning in the brain. *bioRxiv*, 2023.
37. Yuji K Takahashi, Thomas A Stalnaker, Lauren E Mueller, Sevan K Harootonian, Angela J Langdon, and Geoffrey Schoenbaum. Dopaminergic prediction errors in the ventral tegmental area reflect a multithreaded predictive model. *Nature Neuroscience*, 26(5):830–839, 2023.
38. Jeffrey E Markowitz, Winthrop F Gillis, Maya Jay, Jeffrey Wood, Ryley W Harris, Robert Cieszkowski, Rebecca Scott, David Brann, Dorothy Koveal, Tomasz Kula, et al. Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature*, 614(7946):108–117, 2023.
39. Francis Carter, Marie-Pierre Cossette, Ivan Trujillo-Pisanty, Vasilios Pallikaras, Yannick-André Breton, Kent Conover, Jill Caplan, Pavel Solis, Jacques Voisard, Alexandra Yaksich, et al. Does phasic dopamine release cause policy updates? *European Journal of Neuroscience*, 2023.
40. Jonathan CY Tang, Vitor Paixao, Filipe Carvalho, Artur Silva, Andreas Klaus, Joaquim Alves da Silva, and Rui M Costa. Dynamic behaviour restructuring mediates dopamine-dependent credit assignment. *Nature*, 626(7999):583–592, 2024.
41. Ana C Sias, Yousif Jafar, Caitlin M Goodpaster, Kathia Ramírez-Armenta, Tyler M Wrenn, Nicholas K Griffin, Keshav Patel, Alexander C Lamparelli, Melissa J Sharpe, and Kate M Wassum. Dopamine projections to the basolateral amygdala drive the encoding of identity-specific reward memories. *Nature Neuroscience*, pages 1–9, 2024.
42. Eric Garr, Yifeng Cheng, Huijeong Jeong, Sara Brooke, Laia Castell, Aneesh Bal, Robin Magnard, Vijay Mohan K Namboodiri, and Patricia H Janak. Mesostriatal dopamine is sensitive to changes in specific cue-reward contingencies. *Science Advances*, 10(22): eadn4203, 2024.
43. Dennis A Burke, Huijeong Jeong, Brenda Wu, Seul Ah Lee, Joseph R Floeder, and Vijay Mohan K Namboodiri. Few-shot learning: temporal scaling in behavioral and dopaminergic learning. *bioRxiv*, 2023.
44. Tommaso Patriarchi, Jounhong Ryan Cho, Katharina Merten, Mark W Howe, Aaron Marley, Wei-Hong Xiong, Robert W Folk, Gerard Joey Broussard, Ruqiang Liang, Min Jee Jang, et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science*, 360(6396):eaat4422, 2018.
45. Gonçalo Lopes, Karolina Farrell, Edward AB Horrocks, Chi-Yu Lee, Mai M Morimoto, Tomaso Muzzu, Amalia Papanikolaou, Fabio R Rodrigues, Thomas Wheatcroft, Stefano Zucca, et al. Creating and controlling visual environments using bonvision. *eLife*, 10:e65541, 2021.
46. Vijay Mohan K Namboodiri. How do real animals account for the passage of time during associative learning? *Behavioral Neuroscience*, 136(5):383–391, 2022.
47. Vijay Mohan K Namboodiri, Stefan Mihalas, and Marshall GH Shuler. Rationalizing decision-making: understanding the cost and perception of time. *Timing & Time Perception Reviews*, 1, 2014.
48. Douglas A Williams, Travis P Todd, Chrissy M Chubala, and Elliot A Ludvig. Intertrial unconditioned stimuli differentially impact trace conditioning. *Learning & Behavior*, 45:49–61, 2017.
49. Mingkang Zhou, Brenda Wu, Huijeong Jeong, Dennis A Burke, and Vijay Mohan K Namboodiri. An open-source behavior controller for associative learning and memory (b-calm). *Behavior Research Methods*, pages 1–16, 2023.

## Methods

**Animals.** All experimental procedures were approved by the Institutional Animal Care and Use Committee at UCSF and followed guidelines provided by the NIH Guide for the Care and Use of Laboratory Animals. A total of 27 adult wild-type C57BL/6J mice (#000664, Jackson Laboratory) were divided between experiments: nine mice (4 females, 5 males) were used for the first cohort of Pavlovian conditioning, nine mice (4 females, 5 males) were used for the second cohort of Pavlovian conditioning, and nine mice (6 females, 3 males) were used for the VR task. Following surgery, mice were single housed in a reverse 12-hour light/dark cycle. Mice received environmental enrichment and had *ad libitum* access to standard chow. To increase motivation, mice underwent water deprivation. During deprivation, mice were weighed daily and given enough fluids to maintain ~85% of their baseline weight.

**Surgeries.** Surgical procedures were always done under aseptic conditions. Induction of anesthesia was achieved with 3% isoflurane, which was maintained at 1-2% throughout the duration of the surgery. Mice received subcutaneous injections of carprofen (5 mg/kg) for analgesia and lidocaine (1 mg/kg) for local anesthesia of the scalp prior to incision. A unilateral injection (Nanoject III, Drummond) of 500 nL of dLight1.3b(44) (AAVDJ-CAG-dLight1.3b, $2.4 \times 10^{13}$ GC/mL diluted 1:10 in sterile saline) was targeted to the NAcC using the following coordinates from bregma: AP 1.3, ML ±1.4, DV -4.55. The glass injection pipette was held in place for 10 minutes prior to removal to prevent the backflow of virus. After viral injection, an optic fiber (NA 0.66, 400 µm, Doric Lenses) was implanted 100 µm above the site of injection. Subsequently, a custom head ring for head-fixation was secured to the skull using screws and dental cement. Mice recovered and were given at least three weeks before starting behavioral experiments. After completion of experiments, mice underwent transcardial perfusion and subsequent brain fixation in 4% paraformaldehyde. Fiber placement was verified using 50 µm brain sections under a Keyence microscope for subsequent visualization (**Extended Data Fig 2a**, **Extended Data Fig 4a**, **Extended Data Fig 9a**).

**Behavior.** All behavioral experiments took place during the dark cycle in dark, soundproof boxes with white noise playing to minimize any external noise. Prior to starting Pavlovian conditioning, water-deprived mice underwent 1-2 days of random rewards training to get acclimated to our head-fixed behavior setup(49). In a training session, mice received 100 sucrose rewards (~3 µL, 15% in water) at random time intervals taken from an exponential distribution averaging 12 s. Mice consumed sucrose rewards from a lick spout positioned directly in front of their mouths. This same spout was used for lick detection. After completing random rewards, mice were trained on Pavlovian conditioning. An identical trial structure was used across all conditions, consisting of an auditory tone lasting 8 s followed by a delay of 1 s before sucrose reward delivery. Two variables of interest were manipulated—the length of the ITI (long or short) and the type of auditory tone (fixed or dynamic)—resulting in four conditions: long ITI/fixed tone (LF), long ITI/dynamic tone (LD), short ITI/dynamic tone (SD), and short ITI/fixed tone (SF). In the first cohort (**Fig 1**), mice began with the LF condition (mean 7.4 days, range 7-8) before progressing to the LD condition (mean 6.1 days, range 5-11), the SD condition (8 days), and finally the SF condition (8 days). In the second cohort (**Fig 2**), the experimental order was switched such that mice began with the LF condition before moving on to the SD condition and ending with the LD condition (8 days for each condition). The ITI was defined as the period between reward delivery and the subsequent trial's cue onset. In the long ITI conditions, the ITI was drawn from a truncated exponential distribution with a mean of 55 s, maximum of 186 s, and minimum of 6 s. The short ITIs were similarly drawn from a truncated exponential distribution, averaging 8 s with a maximum of 12 s and minimum of 6 s. While mice had 100 trials per day in the short ITI conditions, long ITI sessions were capped at 40 trials due to limitations on the amount of time animals could spend in the head-fixed setup. For the fixed tone conditions, mice were randomly divided into groups presented with either a 3 kHz or 12 kHz tone. While the 12 kHz tone played continuously throughout the entire 8 s, the 3 kHz tone was pulsed (200 ms on, 200 ms off) to make this lower frequency tone more obvious to the mice. For the dynamic tone conditions, the tone frequency either increased (dynamic up↑ starting at 3 kHz) or decreased (dynamic down↓ starting at 12 kHz) by 80 Hz every 200ms, for a total change of 3.2 kHz across 8 s. Mice with the 3 kHz fixed tone had the dynamic up↑ tone, whereas mice with the 12 kHz fixed tone had the dynamic down↓ tone. This dynamic change in frequency across the 8 s was intentionally designed to indicate to the mice the temporal proximity to reward, which is thought to be necessary for ramps to appear in a Pavlovian setting.

For the VR task, water-deprived mice were head fixed above a low-friction belt treadmill. A magnetic rotary encoder attached to the treadmill was used to measure the running velocity of the mice. In front of the head-fixed treadmill setup, a virtual environment was displayed on a high-resolution monitor (20" screen, 16:9 aspect ratio) to look like a dead-end hallway with a patterned floor, walls, and ceiling. The different texture patterns in the virtual environment were yoked to running velocity such that it appeared as though the animal was travelling down the hallway. Upon reaching the end of the hallway, the screen would turn fully black and mice would receive sucrose reward delivery from a lick spout positioned within reach in front of them. The screen remained black for the full duration of the ITI until the reappearance of the starting frame of the virtual hallway signaled the next trial onset. To train mice to engage in this VR task, they began with a 10 cm long virtual hallway. This minimal distance requirement was chosen to make it relatively easy for the mice to build associations between their movement on the treadmill, the corresponding visual pattern movement displayed on the VR monitor, and reward deliveries. Based on their performance throughout training, the distance requirement progressively increased by increments of 5-20 cm across days until reaching a

maximum distance of 67 cm. Training lasted an average of 21.4 days (range 11-38 days), ending once mice could consistently run down the full 67 cm virtual hallway for three consecutive days. The ITIs during training ("med ITI") were randomly drawn from a truncated exponential distribution with a mean of 28 s, maximum of 90 s, and minimum of 6 s. Following training, mice were randomly divided into two groups with identical trials but different ITIs (long or short). Again, both ITIs were randomly drawn from truncated exponential distributions: long ITI (mean 62 s, max 186 s, min 6 s) and short ITI (mean 8 s, max 12 s, min 6 s). After 8 days of the first ITI condition, mice switched to the other condition for an additional 8 days. There were 50 trials per day in both the long and short ITI conditions.

**Fiber Photometry.** Beginning three weeks after viral injection, dLight photometry recordings were performed with either an open-source (PyPhotometry) or commercial (Doric Lenses) fiber photometry system. Excitation LED light for wavelengths of 470 nm (dopamine dependent dLight signal) and 405 nm (dopamine independent isosbestic signal) were sinusoidally modulated via an LED driver and integrated into a fluorescence minicube (Doric Lenses). The same minicube was used to detect incoming fluorescent signals at a 12 kHz sampling frequency before demodulation and downsampling to 120 Hz. Excitation and emission light passed through the same low autofluorescence patchcord (400 µm, 0.57 NA, Doric Lenses). Light intensity at the tip of this patchcord was consistently ~40 µW across days. For Pavlovian conditioning, the photometry software received a TTL signal for the start and stop of the session to align the behavioral and photometry data. For alignment in the VR task, the photometry software received a TTL signal at each reward delivery.

**Data Analysis.** Behavior: Licking was the behavioral readout of learning used in Pavlovian conditioning. The lick rate was calculated by binning the number of licks every 100 ms. A smoothed version produced by Gaussian filtering is used to visualize lick rate in PSTHs (**Fig 1f**, **Extended Data Fig 5a**, **Extended Data Fig 9d**). Anticipatory lick rate for the last three days combined per condition was calculated by subtracting the average baseline lick rate during the 1 s before cue onset from the average lick rate during the trace period 1 s before reward delivery (**Fig 1g**, **Extended Data Fig 5b**). The same baseline subtraction method was used to calculate the average lick rate during the 3 to 8 s post cue onset period (**Ext Data Fig 5c**).

Running velocity, rather than licking, was the primary behavioral readout of learning for the VR task. Velocity was calculated as the change in distance per time. Distance measurements were sampled every 50 ms throughout both the trial and ITI periods. Average PSTHs from the last three days per condition were used to visualize velocity aligned to trial onset (**Fig 4d**) and reward delivery (**Fig 4f**). The change in velocity at trial onset was calculated by subtracting the average baseline velocity (baseline being 1 s before trial onset) from the average velocity between 1-2 s after trial onset (**Fig 4e**). Pre-reward velocity was the mean velocity during the 1 s period before reward delivery (**Fig 4g**).

The inter-trial interval (ITI) used throughout is defined as the time period between the previous trial reward delivery and the current trial onset (**Fig 1c**, **Ext Data Fig 9b**). The inter-reward interval (IRI) is defined as the time period between the previous trial reward delivery and the current trial reward delivery (**Ext Data Fig 9b**). For the previous IRI vs trial slope analysis (**Ext Data Fig 10**), IRI outliers were removed from analysis if they were more than three standard deviations away from the mean of the original IRI distribution. Finally, trial durations in the VR task were defined as the time it took for mice to run 67 virtual cm from the start to the end of the virtual hallway (**Ext Data Fig 9b-c**).

Dopamine: To analyze dLight fiber photometry data, first a least-square fit was used to scale the 405 nm signal to the 470 nm signal. Then, a percentage dF/F was calculated as follows: dF/F = (470 – fitted 405) / (fitted 405) * 100. This session-wide dF/F was then used for subsequent analysis. The onset peak dF/F (**Fig 1j**, **Fig 4i**) was calculated by finding the maximum dF/F value within 1 s after onset and then subtracting the average dF/F value during the 1 s interval preceding onset (last three days per condition combined). For each trial in Pavlovian conditioning, the time aligned dLight dF/F signal during the "ramp window" of 3 to 8 s after cue onset was fit with linear regression to obtain a per-trial slope. These per-trial slopes were then averaged for each day separately (**Fig 1l**) or for the last three days in each condition (**Fig 1m**) for subsequent statistical analysis. A smoothing Gaussian filter was applied to the group average (**Fig 1i**, **Fig 2b**) and example trial (**Fig 1k**) dLight traces for visualization purposes.

Distance, rather than time, was used to align the dLight dF/F signal in the VR task. Virtual distances were sampled every 30 ms, while dF/F values were sampled every 10 ms. To sync these signals, the average of every three dF/F values was assigned to the corresponding distance value. Any distance value that did not differ from the previous distance value was dropped from subsequent analysis (as was its mean dF/F value). This was done to avoid issues with averaging if the animal was stationary. For each trial in the VR task, the distance aligned dLight dF/F signal during the "ramp window" of 20 to 57 cm from the start of the virtual hallway was fit with linear regression to obtain a per-trial slope. These per-trial slopes were then averaged for each day separately (**Fig 4l**) or for the last three days in each condition (**Fig 4m**) for subsequent statistical analysis. To visualize the group averaged distance aligned dLight trace (**Fig 4j**) and example trial traces (**Fig 4k**), the mean dF/F was calculated for every 1 cm after rounding all distance values to the nearest integer.

**Simulations.** We previously proposed a learning model called Adjusted Net Contingency of Causal Relation (ANCCR)(32), which postulates that animals retrospectively search for causes (e.g., cues) when they receive a meaningful event (e.g., reward). ANCCR measures this retrospective association, which we call predecessor representation contingency (PRC), by comparing

the strength of memory traces for a cue at rewards ($M_{\leftarrow cr}$; Equation 1) to the baseline level of memory traces for the same cue updated continuously ($M_{\leftarrow c-}$; Equation 2).

$$M_{\leftarrow cr} \equiv M_{\leftarrow cr} + \alpha[E_{\leftarrow cr} - M_{\leftarrow cr}] \tag{1}$$

$$M_{\leftarrow c-} \equiv M_{\leftarrow c-} + \alpha_0[E_{\leftarrow c-} - M_{\leftarrow c-}] \tag{2}$$

$$PRC_{\leftarrow cr} \equiv M_{\leftarrow cr} - M_{\leftarrow c-} \tag{3}$$

$\alpha$ and $\alpha_0$ are learning rates and the baseline samples are updated every $dt$ seconds. $E_{\leftarrow ci}$ represents eligibility trace of cue ($c$) at the time of event $i$ and $E_{\leftarrow c-}$ represents eligibility trace of cue ($c$) at baseline samples updated continuously every $dt$ seconds. The eligibility trace ($E$) decays exponentially over time depending on decay parameter $T$ (Equation 4).

$$E_{\leftarrow i}(t) = \sum_{t_i \leq t} e^{-\frac{t-t_i}{T}} \tag{4}$$

where $t_i \leq t$ denotes the moments of past occurrences of event i. In **Supplementary Note 1**, we derived a simple rule for the setting of $T$ based on event rates. For the tasks considered here, this rule translated to a constant multiplied by IRI. We have shown in a revised version of a previous study(43) that $\alpha = 1-(1-\alpha_0)^{\frac{IRI}{dt}}$ during initial learning. To mimic the dynamic tone condition, we simulated the occurrence of 8 different cues in a sequence with a 1 s interval between each cue. We used 1 s intervals between cues because real animals are unlikely to detect the small change in frequency occurring every 200ms in the dynamic tone, and we assumed that a frequency change of 400 Hz in 1 s was noticeable to the animals. We included the offset of the last cue as an additional cue. This is based on observation of animal behavior, which showed a sharp rise in anticipatory licking following the offset of the last cue (Fig 1f-g). Inter-trial interval was matched to the actual experimental conditions, averaging 2 s for the short dynamic condition and 49 s for long dynamic condition, with an additional 6 s fixed consummatory period. This resulted in 17 s IRI for short dynamic condition and 64 s IRI for long dynamic condition on average. 1000 trials were simulated for each condition, and the last 100 trials were used for analysis. Following parameters were used for simulation: $w$=0.5, $b_{cues}$=0, $b_{reward}$=0.5, $threshold$=0.2, $T$=0.2 IRI, $\alpha_0 = 5\times10^{-3}$, $\alpha_R = 1$, $dt$=0.2 s.

**Statistics.** All statistical tests were run on Python 3.11 using the scipy (version 1.10) package. Full details related to statistical tests are included in **Supplementary Table 1**. Data presented in figures with error bars represent mean ± SEM. Significance was determined using 0.05 for α. *p < 0.05, **p < 0.01, ***p < 0.001, ns p > 0.05.

## Supplementary 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 $\lambda_x$ 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

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

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:

$$T_{opt} = M_{opt} \frac{\sum_x \lambda_x}{\sum_x \lambda_x^2} \tag{6}$$

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:

$$T_{opt} = M_{opt}.IRI \tag{7}$$

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).

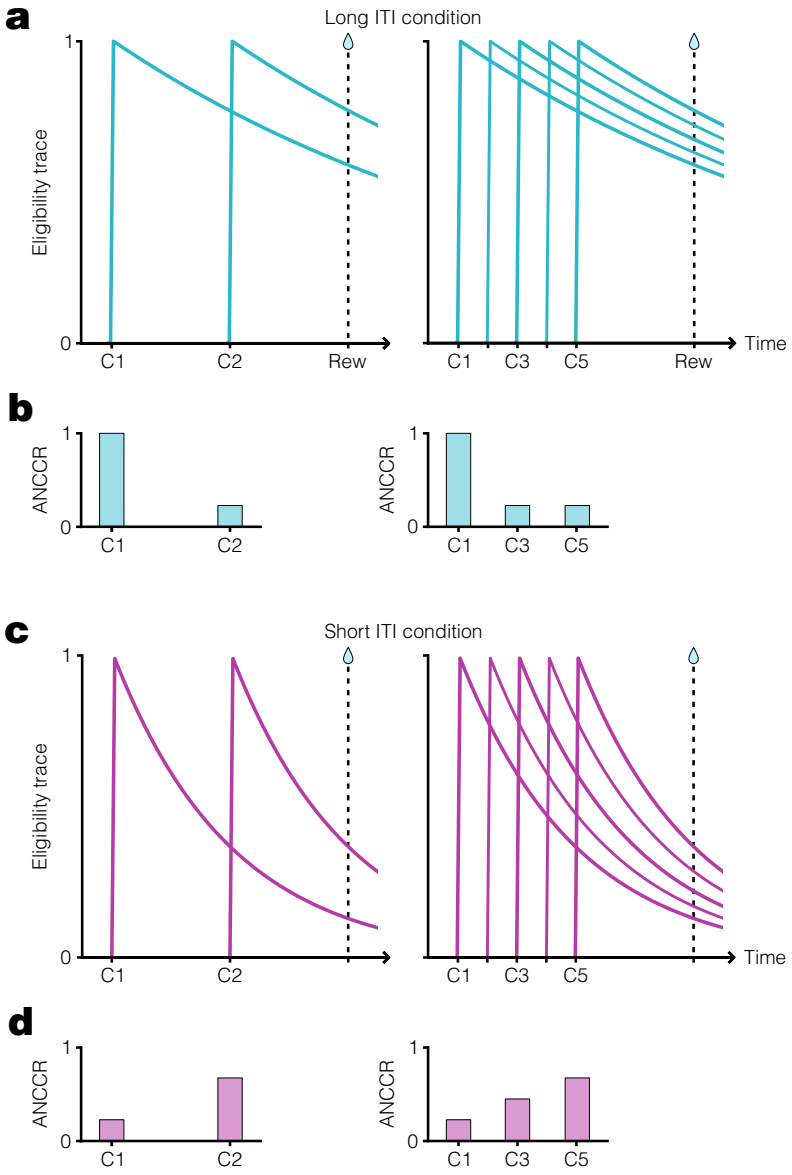
## Supplementary Table 1: Statistical Details.

| 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<br>Tone: F(1) = 0.30<br>ITI x Tone: F(1) = 0.029 | ITI: **p = 0.00457<br>Tone: p = 0.586<br>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<br>LD vs SD: **p = 0.00266<br>LD vs SF: p = 0.952<br>LF vs SD: ***p = $1.70 \times 10^{-4}$<br>LF vs SF: p = 0.445<br>SD vs SF: *p = 0.0107 | n = 9 mice |
| 2c | Slope between days (LF condition last day, SD condition first day) | One sided (LF < SD) paired t-test | t(8) = -3.42 | **p = 0.00455 | n = 9 mice |
| 2c | Slope between days (SD condition last day, LD condition first day) | One sided (SD > LD) paired t-test | t(8) = 3.16 | **p = 0.00666 | n = 9 mice |
| 2d | Slope between conditions (LF, SD) | One sided (LF < SD) paired t-test | t(8) = -4.57 | ***p = $9.07 \times 10^{-4}$ | n = 9 mice |
| 2d | Slope between conditions (SD, LD) | One sided (SD > LD) paired t-test | t(8) = 2.88 | *p = 0.0103 | n = 9 mice |
| 3b | Trial slope regression $\beta$ given previous ITI (SD condition only) | One-sided (< 0), one sample t-test | t(17) = -3.91 | ***p = $5.58 \times 10^{-4}$ | n = 18 mice |
| 3c | Trial slope given previous ITI (SD condition only) | Linear regression | t(5343) = -6.19<br>$R^2$ = 0.00711 | ***p = $6.57 \times 10^{-10}$ | n = 5345 trials |
| 3f | $\Delta$ slope regression $\beta$ given $\Delta$ ITI (SD condition only) | One-sided (< 0), one sample t-test | t(17) = -4.17 | ***p = $3.23 \times 10^{-4}$ | n = 18 mice |
| 3g | $\Delta$ slope between $\Delta$ ITI <-1 and $\Delta$ ITI >1 (SD condition only) | One-sided (($\Delta$ ITI <-1) > ($\Delta$ ITI >1)) paired t-test | t(17) = 4.32 | ***p = $2.31 \times 10^{-4}$ | n = 18 mice |
| 4e | 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 |
| 4e | 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 |

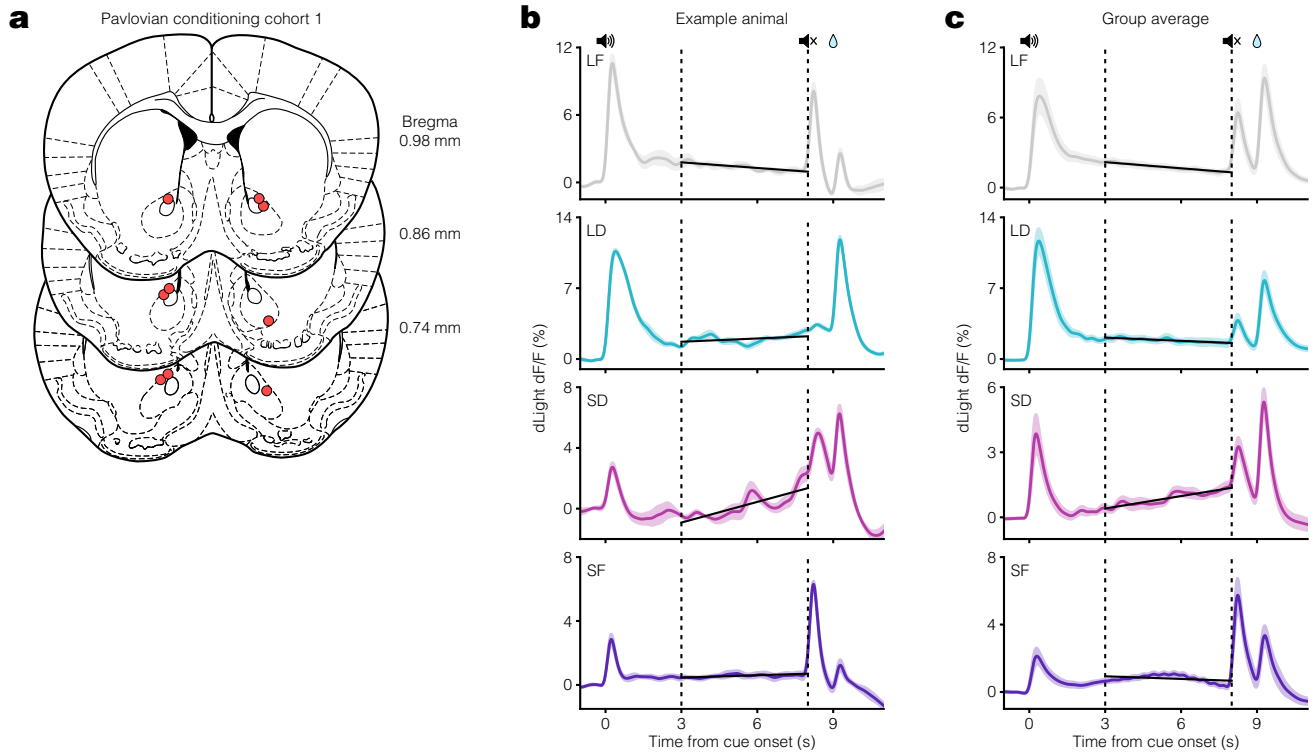
| Figure | Description | Test | Statistic | p value | Sample size |
|---|---|---|---|---|---|
| 4e | Change in velocity at trial onset between conditions (long, short) | Paired t-test | t(8) = 4.25 | **p = 0.00281 | n = 9 mice |
| 4g | Pre-reward velocity between conditions (long, short) | Paired t-test | t(8) = 0.71 | p = 0.497 | n = 9 mice |
| 4i | Cue onset peak dLight between conditions (long, short) | Paired t-test | t(8) = 7.59 | ***p = $6.34 \times 10^{-5}$ | n = 9 mice |
| 4m | Slope between conditions (long, short) | One-sided (long < short) paired t-test | t(8) = -2.09 | *p = 0.0349 | n = 9 mice |
| Ext 3b | Anticipatory lick rate across conditions (LF, SD, LD) | One-way ANOVA | F(2) = 0.26 | p = 0.770 | n = 9 mice |
| Ext 3c | Lick rate during ramp window across conditions (LF, LD, SD, SF) | One-way ANOVA | F(3) = 2.24 | p = 0.0929 | n = 18 mice |
| Ext 3d | Lick slope during ramp window across ITI (long, short) and tone (fixed, dynamic) | Two-way ANOVA | ITI: F(1) = 1.39<br>Tone: F(1) = 12.03<br>ITI x Tone: F(1) = 0.044 | ITI: p = 0.244<br>Tone: ***p = $9.83 \times 10^{-4}$<br>ITI x Tone: p = 0.865 | n = 18 mice |
| Ext 6b | Pre-cue dLight slope across conditions (LF, LD, SD, SF) | One-way ANOVA | F(3) = 1.06 | p = 0.375 | n = 18 mice |
| Ext 6c | Ramp dLight slope regression β given pre-cue dLight slope (SD condition only) | One-sided (< 0), one sample t-test | t(17) = -1.46 | p = 0.0816 | n = 18 mice |
| Ext 6d | Ramp dLight slope given pre-cue dLight slope (SD condition only) | Linear regression | t(5397) = -1.38<br>$R^2$ = 3.53x10-4 | p = 0.168 | n = 5399 trials |
| Ext 7a | Trial dLight slope regression β given average ITI for previous X trials (SD condition only) | One-sided (< 0 for previous trial, > 0 for previous 2-10 trials) one sample t-test<br><br>With Benjamini-Hochberg Procedure | Prev 1: t(17) = -3.91<br>Prev 2: t(17) = 0.80<br>Prev 3: t(17) = 1.31<br>Prev 4: t(17) = 1.48<br>Prev 5: t(17) = 1.23<br>Prev 6: t(17) = 1.09<br>Prev 7: t(17) = 1.32<br>Prev 8: t(17) = 1.05<br>Prev 9: t(17) = 1.02<br>Prev 10: t(17) = 0.76 | Prev 1: **p = 0.00558<br>Prev 2: p = 0.229<br>Prev 3: p = 0.202<br>Prev 4: p = 0.202<br>Prev 5: p = 0.202<br>Prev 6: p = 0.202<br>Prev 7: p = 0.202<br>Prev 8: p = 0.202<br>Prev 9: p = 0.202<br>Prev 10: p = 0.229 | n = 18 mice |
| Ext 7b | Trial dLight slope regression β given average ITI for previous X trials (LD condition only) | One-sided (< 0) one sample t-test<br><br>With Benjamini-Hochberg Procedure | Prev 1: t(17) = -0.42<br>Prev 2: t(17) = -3.34<br>Prev 3: t(17) = -1.94<br>Prev 4: t(17) = -1.62<br>Prev 5: t(17) = -1.94<br>Prev 6: t(17) = -1.61<br>Prev 7: t(17) = -1.65<br>Prev 8: t(17) = -1.81<br>Prev 9: t(17) = -1.31<br>Prev 10: t(17) = -1.06 | Prev 1: p = 0.339<br>Prev 2: *p = 0.0194<br>Prev 3: p = 0.0901<br>Prev 4: p = 0.0901<br>Prev 5: p = 0.0901<br>Prev 6: p = 0.0901<br>Prev 7: p = 0.0901<br>Prev 8: p = 0.0901<br>Prev 9: p = 0.129<br>Prev 10: p = 0.170 | n = 18 mice |

| Figure | Description | Test | Statistic | p value | Sample size |
|---|---|---|---|---|---|
| Ext 8a | dLight slope regression β given dLight onset peak (SD condition only) | One-sided (< 0), one sample t-test<br><br># With Benamini-Hochberg (BH) Procedure for all t-tests in Ext Fig 8 | t(17) = -0.55 | p = 0.538 | n = 18 mice |
| Ext 8a | dLight slope given dLight onset peak (SD condition only) | Linear regression<br><br>## With BH Procedure for all linear regression in Ext Fig 8 | t(5397) = -0.53<br>$R^2 = 5.28 \times 10^{-5}$ | p = 0.839 | n = 5399 trials |
| Ext 8b | dLight slope regression β given dLight onset peak (LD condition only) | One-sided (< 0), one sample t-test<br># | t(17) = -0.19 | p = 0.985 | n = 18 mice |
| Ext 8b | dLight slope given dLight onset peak (LD condition only) | Linear regression<br>## | t(2157) = -0.55<br>$R^2 = 1.41 \times 10^{-4}$ | p = 0.839 | n = 2159 trials |
| Ext 8c | dLight slope regression β given lick slope (SD condition only) | One-sided (< 0), one sample t-test<br># | t(17) = -0.019 | p = 0.519 | n = 18 mice |
| Ext 8c | dLight slope given lick slope (SD condition only) | Linear regression<br>## | t(5397) = -0.34<br>$R^2 = 2.13 \times 10^{-5}$ | p = 0.153 | n = 5399 trials |
| Ext 8d | dLight slope regression β given lick slope (LD condition only) | One-sided (< 0), one sample t-test<br># | t(17) = -0.88 | p = 0.519 | n = 18 mice |
| Ext 8d | dLight slope given lick slope (LD condition only) | Linear regression<br>## | t(2157) = -2.53<br>$R^2 = 2.96 \times 10^{-3}$ | p = 0.440 | n = 2159 trials |
| Ext 8e | dLight onset peak regression β given lick slope (SD condition only) | One-sided (> 0), one sample t-test<br># | t(17) = 1.30 | p = 0.539 | n = 18 mice |
| Ext 8e | dLight onset peak given lick slope (SD condition only) | Linear regression<br>## | t(5397) = 2.07<br>$R^2 = 7.96 \times 10^{-4}$ | p = 0.839 | n = 5399 trials |
| Ext 8f | dLight onset peak regression β given lick slope (LD condition only) | One-sided (< 0), one sample t-test<br># | t(17) = 0.071 | p = 0.519 | n = 18 mice |
| Ext 8f | dLight onset peak given lick slope (LD condition only) | Linear regression<br>## | t(2157) = 0.18<br>$R^2 = 1.43 \times 10^{-5}$ | p = 0.0914 | n = 2159 trials |
| Ext 8g | Lick slope regression β given previous ITI (SD condition only) | One-sided (< 0), one sample t-test<br># | t(17) = -1.01 | p = 0.539 | n = 18 mice |
| Ext 8g | Lick slope given previous ITI (SD condition only) | Linear regression<br>## | t(5343) = -1.39<br>$R^2 = 3.61 \times 10^{-4}$ | p = 0.861 | n = 5345 trials |

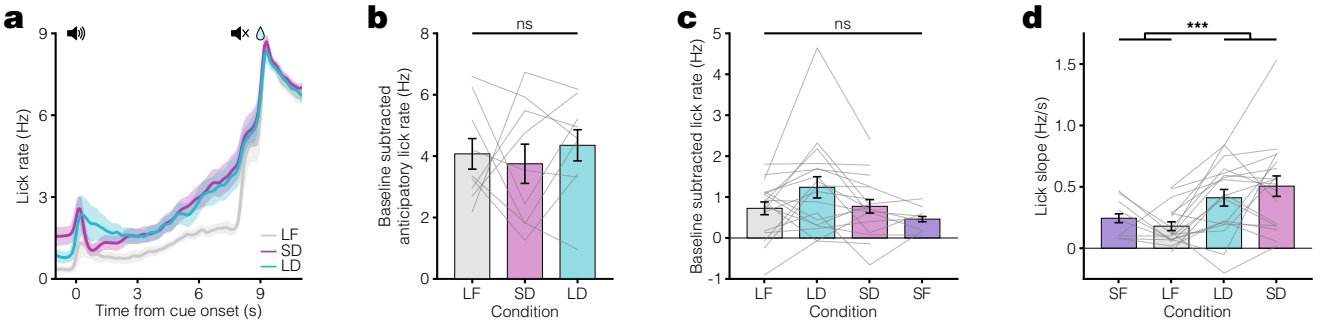
| Figure | Description | Test | Statistic | p value | Sample size |
|---|---|---|---|---|---|
| Ext 8h | Lick slope regression β given previous ITI (LD condition only) | One-sided (< 0), one sample t-test # | t(17) = -0.43 | p = 0.538 | n = 18 mice |
| Ext 8h | Lick slope given previous ITI (LD condition only) | Linear regression ## | t(2103) = -0.40<br>$R^2 = 7.46 \times 10^{-5}$ | p = 0.839 | n = 2105 trials |
| Ext 9c | Trial duration between conditions (long, short) | Paired t-test | t(8) = 1.02 | p = 0.336 | n = 9 mice |
| Ext 9e | Session slope given session IRI (both long & short ITI conditions) | Linear regression | t(52) = -2.61<br>$R^2 = 0.116$ | *p = 0.0118 | n = 54 sessions |
| Ext 10b | 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 10c | Trial slope given previous IRI (short ITI condition only) | Linear regression | t(1301) = -2.11<br>$R^2 = 0.00339$ | *p = 0.0355 | n = 1302 trials |



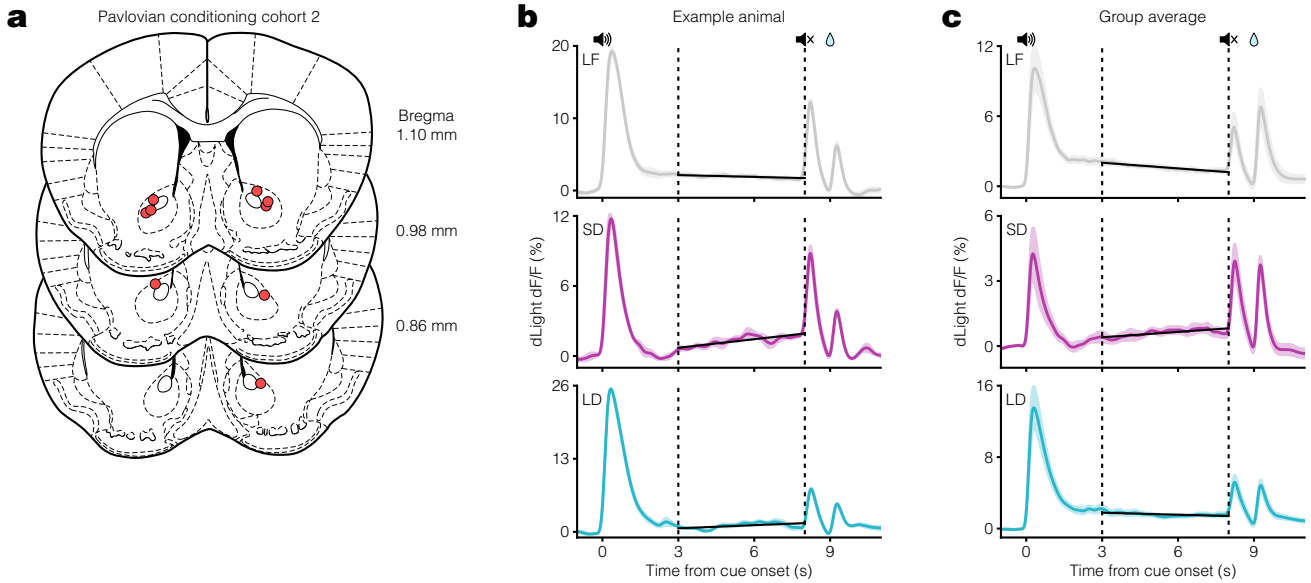

**Extended Data Figure 1. Dependence of ANCCR on eligibility trace time constant. a**.Schematic showing exponential decay of cue eligibility traces for two-cue sequential conditioning (left) and multi-cue conditioning (right) with a long inter-trial interval (ITI). In this case, a long ITI results in a proportionally large eligibility trace time constant, *T*, producing slow eligibility trace decay (**Supplementary Note 1**). Reward delivery time indicated by vertical dashed line. **b**. Schematized ANCCR magnitudes (arbitrary units) for cues in the two-cue (left) and multi-cue (right) conditioning paradigms with a long ITI. Since the eligibility trace for the first cue is still high at reward time, there is a large ANCCR at this cue. The remaining cues are preceded consistently by earlier cues associated with the reward, thereby reducing their ANCCR. **c**. Same conditioning trial structure as in **a**, but with a short ITI and smaller *T*, producing rapid eligibility trace decay. **d**. Schematized ANCCR magnitudes for cues in both conditioning paradigms with a short ITI. Since the eligibility trace for the first cue is low at reward time, there is a small ANCCR at this cue. Though the remaining cues are preceded consistently by earlier cues associated with the reward, the eligibility traces of these earlier cues decay quickly, thereby resulting in a higher ANCCR for the later cues.



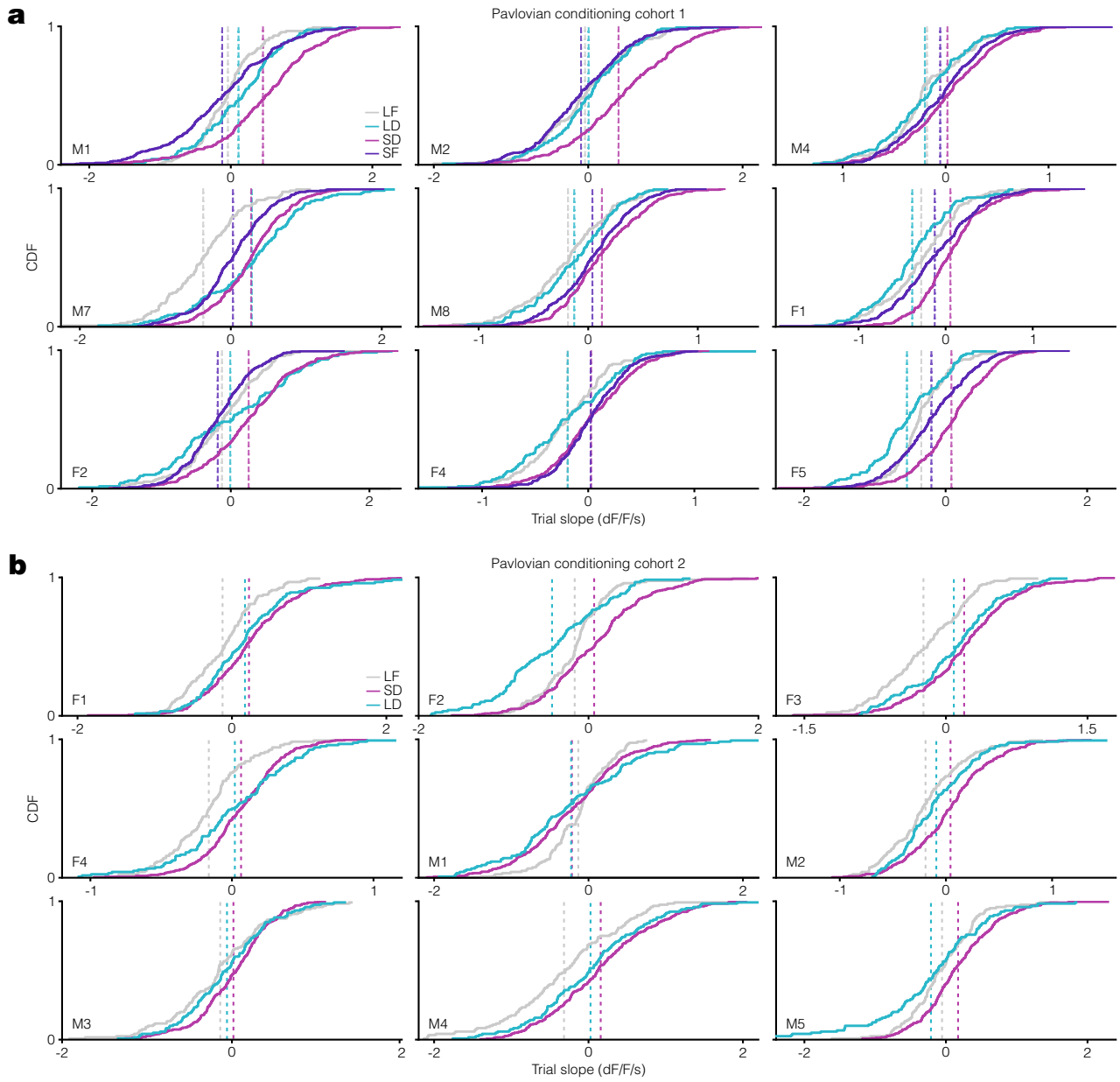

**Extended Data Figure 2. Pavlovian conditioning cohort 1 histology and dopamine responses. a**. Mouse coronal brain sections showing reconstructed locations of optic fiber tips (red circles) in NAcC for Pavlovian conditioning cohort 1. **b**. Example average dLight traces for the last three days of all conditions. Vertical dashed lines at 3 and 8 s represent the ramp window period. Black lines display the linear regression fit during this period. **c**. Same as in b but for the average dLight traces across all animals (n = 9 mice).



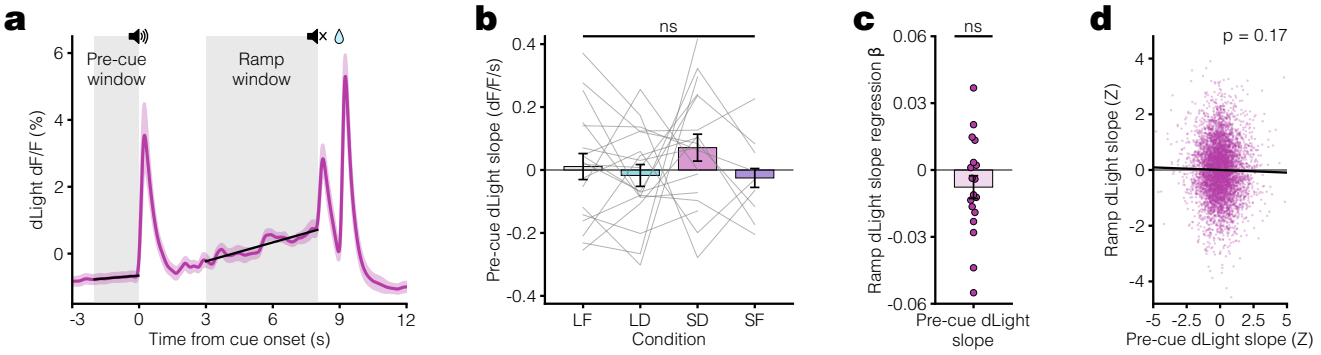

**Extended Data Figure 3. Pavlovian conditioning licking behavior data. a**. PSTH showing average licking behavior for the last 3 days of each condition for Pavlovian conditioning cohort 2 (n = 9 mice). **b**. Average anticipatory lick rate (baseline subtracted) for 1 s trace preceding reward delivery for cohort 2 (p = 0.77). **c**. Comparison of average baseline subtracted lick rate during the ramp window (3 to 8 s after cue onset) across all conditions for both cohorts (p = 0.093, n = 18 mice). **d**. Comparison of average lick slope during the ramp window across all conditions (fixed tone vs dynamic tone: ***p = $9.8 \times 10^{-4}$).



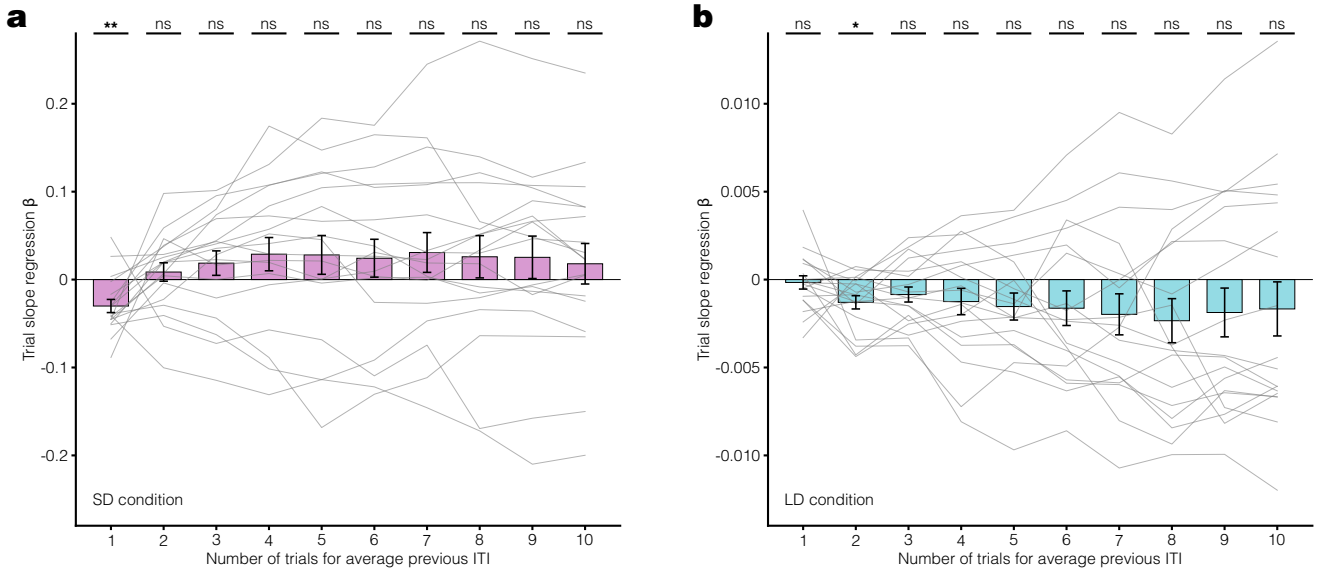

**Extended Data Figure 4. Pavlovian conditioning cohort 2 histology and dopamine responses. a**. Mouse coronal brain sections showing reconstructed locations of optic fiber tips (red circles) in NAcC for Pavlovian conditioning cohort 2. **b**. Example average dLight traces for the last three days of all conditions. Vertical dashed lines at 3 and 8 s represent the ramp window period. Black lines display the linear regression fit during this period. **c**. Same as in b but for the average dLight traces across all animals (n = 9 mice).



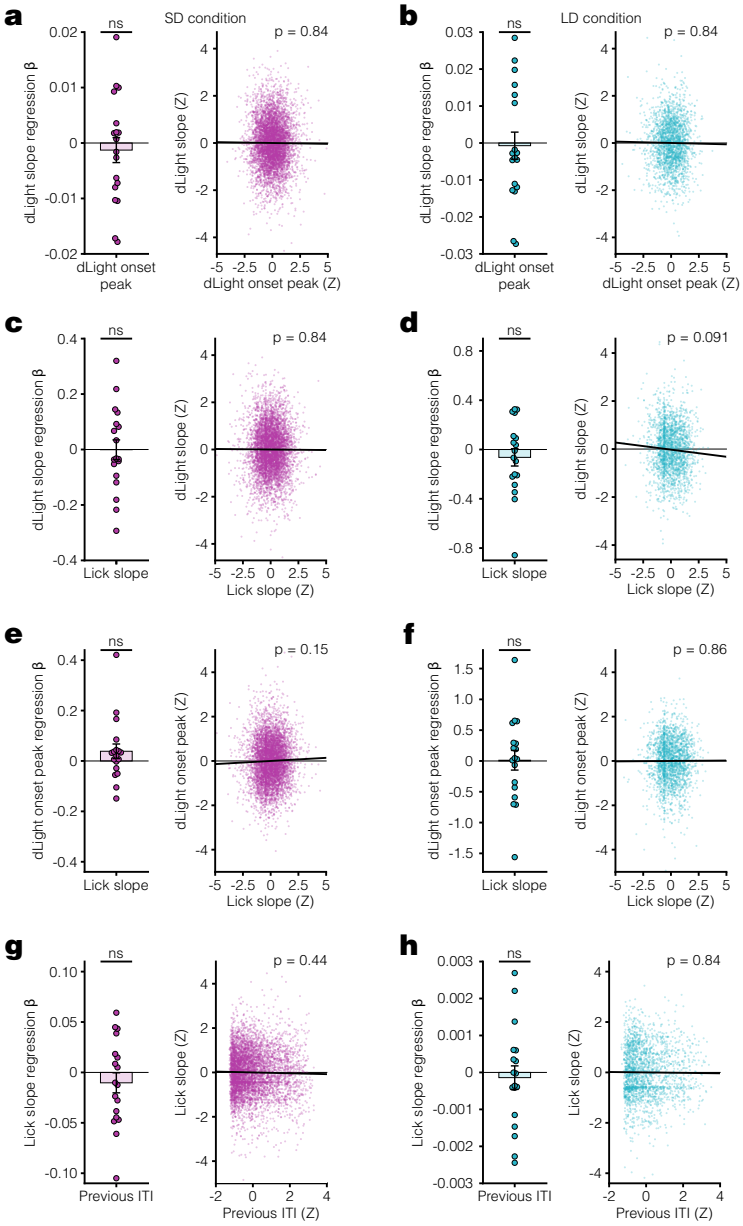

**Extended Data Figure 5. Pavlovian conditioning cumulative dopamine response data. a**. Individual plots for each mouse from Pavlovian conditioning cohort 1 displaying the cumulative distribution of per-trial slopes for the last three days in all conditions. Vertical dashed lines indicate the average trial slope for LF (grey), LD (teal), SD (pink), and SF (purple) conditions. **b**. Same as in **a** but for Pavlovian conditioning cohort 2 mice.



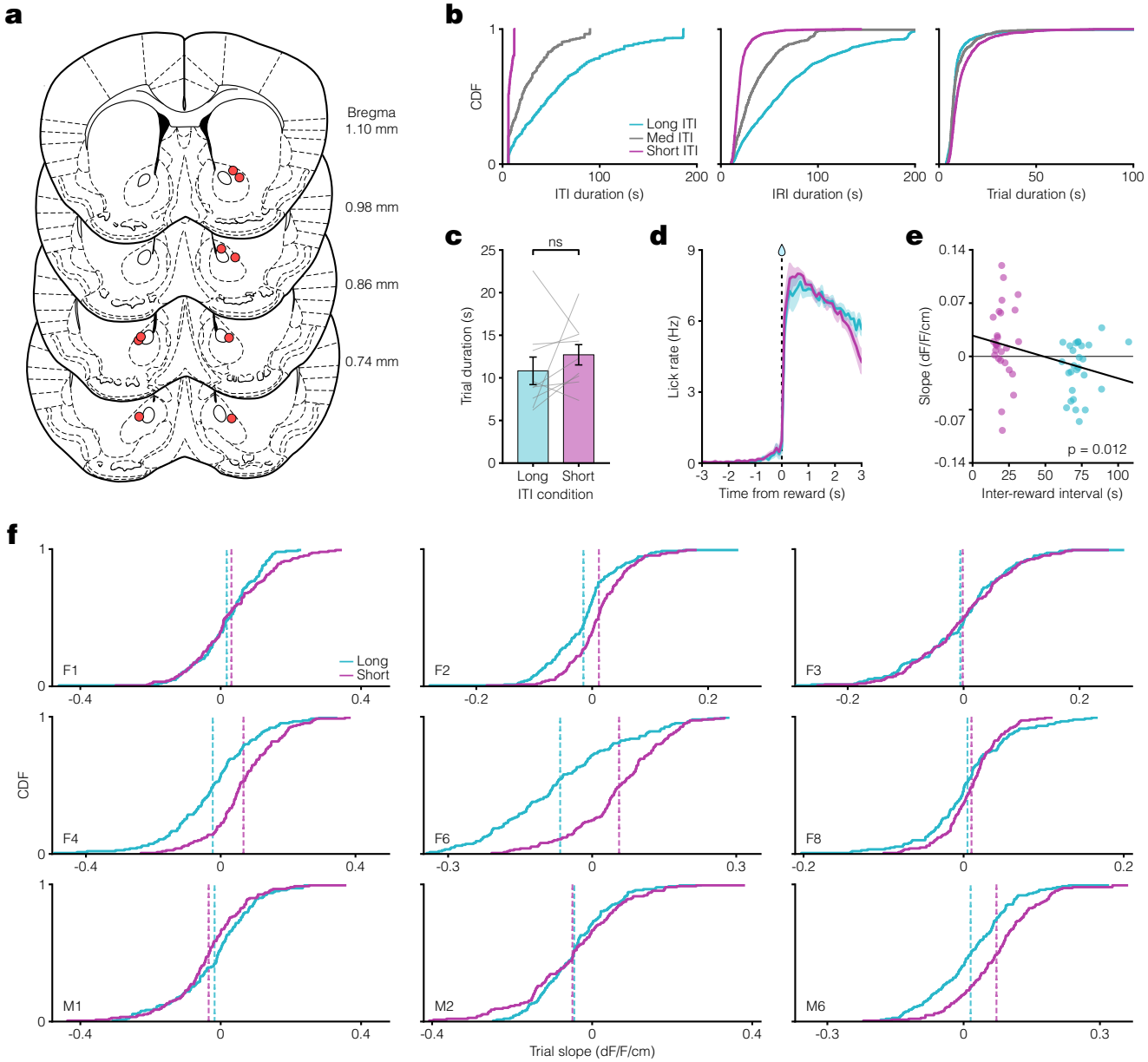

**Extended Data Figure 6. Pavlovian conditioning dopamine ramps do not correlate with pre-cue dopamine activity. a**. Average dLight dopamine signal for Pavlovian conditioning cohort 1. Black lines represent linear regression fit during the pre-cue window and ramp window, each marked with grey shaded regions. **b**. Average per-trial dLight slope during the pre-cue window for the last 3 days of each condition (p = 0.37). **c**. Linear regression β coefficients for per-trial ramp dLight slope vs. pre-cue dLight slope in SD condition calculated per animal (p = 0.082). **d**. Scatter plot with linear regression fit (black line) of Z-scored ramp slope vs pre-cue slope pooled across mice for all trials in the last 3 days of SD condition (p = 0.17).



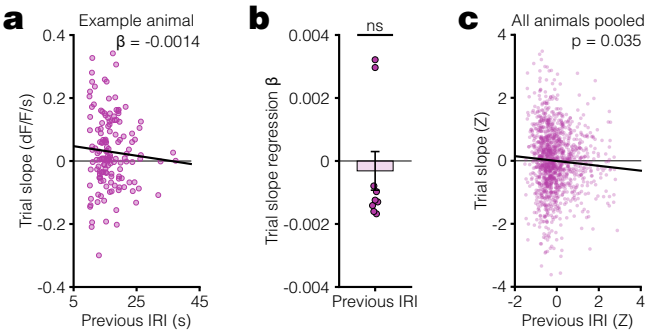

**Extended Data Figure 7. Pavlovian conditioning dopamine responses do not correlate with broader estimates of ITI. a**. Linear regression β coefficients for trial dLight slope vs. average previous ITI for the past 1 through 10 ITIs calculated per animal for all trials in the last 3 days of the SD condition (**p = 0.0056, ns p > 0.05; using Benjamini-Hochberg Procedure). **b**.Same as in **a** but for the LD condition (*p = 0.019, ns p > 0.05).




**Extended Data Figure 8. No significant correlations exist between additional dopamine and behavior measurements. a**.Left, linear regression β coefficients for dLight slope vs. dLight onset peak calculated per animal in SD condition (p = 0.54). Right, scatter plot with linear regression fit (black line) of Z-scored dLight slope vs dLight onset peak pooled across mice for all trials in the last 3 days of SD condition (p = 0.84). **b**. Same as in **a** but for LD condition (left p = 0.99, right p = 0.84). **c**. Left, linear regression β coefficients for dLight slope vs. lick slope calculated per animal in SD condition (p = 0.52). Right, scatter plot with linear regression fit (black line) of Z-scored dLight slope vs lick slope pooled across mice for all trials in the last 3 days of SD condition (p = 0.15). **d**. Same as in **c** but for LD condition (left p = 0.52, right p = 0.44). **e**. Left, linear regression β coefficients for dLight slope vs. dLight onset peak calculated per animal in SD condition (p = 0.54). Right, scatter plot with linear regression fit (black line) of Z-scored dLight slope vs dLight onset peak pooled across mice for all trials in the last 3 days of SD condition (p = 0.84). **f**. Same as in **e** but for LD condition (left p = 0.52, right p = 0.091). **g**. Left, linear regression β coefficients for dLight slope vs. dLight onset peak calculated per animal in SD condition (p = 0.54). Right, scatter plot with linear regression fit (black line) of Z-scored dLight slope vs dLight onset peak pooled across mice for all trials in the last 3 days of SD condition (p = 0.86). **h**. Same as in **g** but for LD condition (left p = 0.54, right p = 0.84). Benjamini-Hochberg Procedure used for p values from all t-tests and p values from all linear regression separately for all comparisons in this figure.




**Extended Data Figure 9. VR navigation task histology and responses. a**. Mouse coronal brain sections showing reconstructed locations of optic fiber tips (red circles) in NAcC for VR navigation task. **b**. Left, CDF of ITI duration for long (teal), medium (grey), and short (pink) ITI conditions. Middle, CDF plot of inter-reward interval (IRI) durations for each condition. Right, CDF plot of trial durations for each condition. **c**.Comparison of average trial duration for long and short ITI conditions (p = 0.34). **d**. Lick rate PSTH aligned to reward delivery indicates minimal anticipatory licking behavior. **e**. Scatter plot showing relationship between average per-session slope and inter-reward interval for the last 3 days in long (teal) and short (pink) ITI conditions. Black line indicates linear regression fit (*p = 0.012). **f**. CDF plots for each mouse separately showing the distribution of per-trial slopes for the last three days in both conditions. Vertical dashed lines indicate the average trial slope for long (teal) and short (pink) ITI conditions.




**Extended Data Figure 10. Trial-by-trial correlation of dopamine response slope vs previous inter-reward interval (IRI) in the VR task. a**. Scatter plot for an example animal showing the relationship between dopamine response slope within a trial and previous inter-reward interval (IRI) for all trials in the last 3 days of the short ITI condition. Plotted with linear regression fit (black line) used to find this animal's β coefficient of -0.0014. Here, we are measuring the environmental timescale using IRI instead of ITI because the trial duration in this task (see **Supplementary Note 1**) depends on the running speed of the animals, which varies trial to trial. Thus, IRI measures the net time interval between successive trial onsets. In the Pavlovian conditioning experiment, IRI and ITI differ by a constant since the trial duration is fixed. **b**.Linear regression β coefficients for previous IRI vs trial slope calculated per animal ($p = 0.32$). **c**. Scatter plot of Z-scored trial slope vs. previous IRI pooled across mice for all trials in the last 3 days of the Short ITI condition (*$p = 0.035$).