Title

Dynamic dopaminergic activity controls the timing of self-timed movement

Authors

Allison E. Hamilos[1†], Giulia Spedicato[1†], Ye Hong[1†], Fangmiao Sun[2‡], Yulong Li[2‡] & John A. Assad[1,3†]

Affiliations

[1]Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, USA. [2]State Key Laboratory of Membrane Biology, Peking University School of Life Sciences, Beijing, China. [3]Istituto Italiano di Tecnologia, Genova, Italy. Correspondence should be addressed to A.E.H. (ahamilos@mit.edu) or J.A.A. (jassad@hms.harvard.edu).

Current Addresses

[†] Harvard Medical School, 220 Longwood Avenue, Warren Alpert Building Room 222, Boston, Massachusetts, 02115, USA

[‡] PKU-IDG/McGovern Institute for Brain Research, Yi He Yuan Rd#5, Beijing, 100871, P.R. China.

Summary

Deciding when to move is a universal aspect of behavior. Pharmacological studies implicate the neurotransmitter dopamine as a regulator of self-timed movements, with increased dopamine availability generally leading to earlier movements, as if speeding an internal clock.

How dopamine affects self-timed movements is unclear; a recent study even suggested that increased activity in nigrostriatal dopamine neurons (DANs) is associated with *slower* internal timing[1]. Here we show the dynamics of DAN activity control the timing of self-timed movements in mice. Animals were trained to make a self-timed lick several seconds after a start-timing cue. Movement times were highly variable from trial-to-trial, typical for self-timed actions[2-6]. Higher pre-trial DAN signals predicted earlier movements, consistent with pharmacological studies. However, surprisingly, DAN signals ramped-up over seconds following the start-timing cue, with the steepness of ramping predicting the trial-by-trial movement time. Steeply ramping signals preceded early lick-times whereas shallow ramping preceded later lick-times, reminiscent of a ramp-to-threshold process. Optogenetic DAN activation during the timed interval caused systematic early-shifting of self-timed movements, whereas inhibition caused systematic late-shifting. These results reveal a novel, causal role for dynamic DAN activity unfolding over seconds-long timescales in controlling the moment-by-moment decision of when to move.

Main Text

Body movements can occur as short-latency reactions to external stimuli, but many movements are generated without obvious, abrupt prompting[7]. For example, s*elf-timed* movements come after a reference-timing cue, but their exact timing is highly variable from trial-to-trial relative to that cue[2-6]. Evidence from lesion studies and human disease implicate the nigrostriatal system in the generation of self-timed movements[4,5,8,9], and pharmacological manipulations of the dopamine neurotransmitter causally influence movement timing[5,8,9,10,11]. For example, decreased dopamine availability/efficacy (e.g., Parkinson's disease, neuroleptic drugs) produces late-shifted self-timed

movements[4,8], whereas high dopamine (e.g., amphetamines) produces early-shifted movements[10,11], suggesting the activity of nigrostriatal dopamine neurons (DANs) may affect the speed of the internal clock.

Here, we exploited the inherent variability in the timing of self-timed movements to examine how the moment-to-moment activity of nigrostriatal DANs relates to the timing of these movements. We trained head-fixed mice to make self-timed movements to receive juice rewards (Fig. 1a). Animals received an audio/visual start-timing cue and then had to decide when to lick in the absence of further cues. Animals only received juice if they waited a proscribed interval following the cue before making their first-lick (3.3s in most experiments). First-lick timing exhibited a broad distribution spanning several seconds, as expected from previous studies[3,4,5,6](Fig. 1b, Extended Data Figs. 1a,b). As mice executed the task, we employed fiber photometry to record the activity of genetically-defined DANs expressing the calcium-sensitive fluorophore GCaMP6f (12 mice, substantia nigra pars compacta (SNc); Fig. 1c-e, Extended Data Figs. 2,3,4a). We controlled for mechanical/optical artifacts by simultaneously recording fluorescence modulation of a co-expressed, calcium-insensitive fluorophore, tdTomato (tdt) (Fig. 1e), as well as body movements detected by neck EMG, high-speed video and a back-mounted accelerometer (Extended Data Fig. 5).

DAN signals correlate with self-timing. DAN GCaMP6f fluorescence typically exhibited brief transients following cue onset and immediately before movement onset, as observed in previous studies[12-16] (Fig. 1c). However, during the timed interval, we observed slow "ramping up" of fluorescence, with a minimum after the cue-locked transient and maximum just before the lick-

related transient. We asked whether this ramping differed between trials in which the animal moved relatively early or late. Strikingly, when we averaged signals by movement time, we observed systematic differences in the steepness of ramping that were highly predictive of movement timing (Fig. 1d,e). Trials with early first-licks exhibited steep ramping, whereas trials with later first-licks started from lower fluorescence levels and rose more slowly toward the time of movement. The fluorescence ramps terminated at nearly the same amplitude regardless of movement time. Similar ramping dynamics and baseline differences were found in the dorsal lateral striatal "lick area[17]" (DLS), in both the fluorescence of GCaMP6f in DAN axon terminals (Extended Data Fig. 4b), as well as $GRAB_{DA2m}$ (DA_{2m}) expressed in striatal cells (Fig. 1f, Extended Data Fig. 4c). DA_{2m} is a new, improved extracellular dopamine indicator derived from the dopamine-2-receptor[18]. Thus, ramping SNc GCaMP6f dynamics are played out at the axon terminal and are reflected in striatal dopamine accumulation, suggesting that DAN ramping dynamics may causally influence movement timing via the interaction of released dopamine with downstream striatal neurons. Similar ramping dynamics were also observed in GCaMP6f-expressing DAN cell bodies in the ventral tegmental area (VTA), reminiscent of ramping dynamics observed in VTA spiking and mesolimbic dopamine release during goal-oriented navigation tasks as animals approached a rewarded target[19,20](Extended Data Fig. 4d).

In addition to ramping dynamics, slowly-modulating DAN signals were correlated with first-lick timing even before cue-onset, with higher baseline fluorescence predicting earlier first-licks (Fig. 1e,f; Extended Data Figs. 4a-d, 6a-b). Because dF/F correction methods can potentially distort baseline measurements, we rigorously tested and validated three different dF/F methods, and we also repeated analyses with raw fluorescence values compared between pairs of sequential trials

with different movement times (Extended Data Fig. 3, Supplementary Methods A). All reported results, including the systematic baseline differences, were robust to dF/F correction. The systematic correlation of baseline signals with first-lick timing was not fully explained by prior trial outcome (reward/no reward) nor licking during the intertrial interval (ITI) (Extended Data Fig. 6c). In fact, although a reward on the prior trial tended to elevate signals during the ITI, there was an abrupt "resetting" of baseline signals during the random delay (after lamp-off but before the start-timing cue), such that baseline amplitude became abruptly and progressively better explained by the upcoming trial outcome (reward/no reward) compared to the prior trial (Extended Data Fig. 6b,c). Mice trained on a variant of the self-timing task without lamp-off/on events showed no systematic differences in their timing distributions, suggesting that although DAN resetting occurred at lamp-off, the mice still referenced their timing to the start-timing cue (Extended Data Fig. 1c).

Controlling for movement artifacts. The systematic ramping dynamics and baseline differences observed with GCaMP6f and DA_{2m} were not observed in the tdt optical control channel nor in any of the other movement-control channels (Fig. 1e; Extended Data Figs. 4e,5b), making it unlikely that these ramping dynamics could have arisen from optical artifacts. Nevertheless, because DANs show transient responses to salient cues and movements[12-16], it is possible GCaMP6f and DA_{2m} fluorescence could reflect the superposition of dopaminergic responses to multiple task events, including the cue, lick, ongoing spurious body movements, and hidden cognitive processes like timing. For example, accelerating spurious movements could, in principle, produce motor-related neural activity that "ramps up" during the timed interval, perhaps even at different rates on different trials. We thus derived a nested linear encoding model of single-trial GCaMP6f signals,

a data-driven, statistical approach designed to isolate and quantify the contributions of task events (timing-independent predictors) from processes predictive of movement timing (timing-dependent predictors)[21,22,23](Fig. 2a,b; Extended Data Fig. 7a-d). The model robustly detected task-event GCaMP6f kernels locked to cue, lick and EMG/accelerometer events (Fig. 2c; Extended Data Fig. 7e), but these timing-independent predictors alone were insufficient to capture the rich variability of GCaMP6f signals for trials with different self-timed movement times, especially the timing-dependent ramp-slope and baseline offset (68 sessions, Fig. 2c; Extended Data Fig. 7f,g). In contrast, two timing-dependent predictors robustly improved the model: 1) a baseline offset whose amplitude was linearly proportional to first-lick time; and 2) a "stretch" feature representing percentages of the interval following the cue, which predicted a ramp from cue-to-lick with slope inversely proportional to first-lick time (68 sessions, Fig. 2b,c; Extended Data Fig. 7e). Similar results were obtained for SNc DAN axon terminals in the DLS, DLS neurons expressing DA_{2m}, and VTA DAN cell bodies (Extended Data Fig. 7h).

In contrast to the GCaMP6f model, when the same procedure was applied to control photometry signals (tdt), the timing-independent predictors (which could potentially cause optical or mechanical artifacts—cue, first-lick, EMG/accelerometer) improved the model, but timing-dependent predictors did not improve the model (Fig. 2c; Extended Data Fig. 7f-h).

Principle component (PC) analysis revealed ramp-like and baseline-offset-like components that explained as much as 93% of the variance in GCaMP6f signals during the timing interval (mean: 66%, range: 16-93%), but similar PCs were not present in tdt signals (mean: 4%, range: 1.6-15%) (Extended Data Fig. 8a,b).

DAN signals predict single-trial timing. Given that ramping and baseline-offset signals were not explained by nuisance movements or optical artifacts, we asked whether DAN GCaMP6f fluorescence could predict first-lick timing on single trials. Using a simple threshold-crossing model[24], we found that the GCaMP6f signal was predictive of movement time even for low thresholds intersecting the "base" of the ramp, with the predictive value of the model progressively improving for higher thresholds (R^2 low: 0.34; mid: 0.64; high: 0.94, Fig. 3a). To more thoroughly determine the independent, additional predictive power of DAN baseline and ramping signals over other task variables (e.g., movement time on previous trial; presence/absence of reward on previous trial, etc.), we derived a nested decoding model for first-lick time (Fig. 3a; Extended Data Fig. 8c). All predictors contributed to the predictive power of the model. However, even when we accounted for the contributions of prior trial history, tdt artifacts and baseline GCaMP6f signals, GCaMP6f threshold-crossing time robustly dominated the model, alone explaining 10% of the variance in first-lick time on average (range: 1-27%) (Fig. 3b-d). Alternate versions of the decoding model showed similar results (Extended Data Fig. 8c).

SNc DANs causally influence self-timing. Because the DAN ramping signal robustly predicted first-lick timing and was apparently transmitted via dopamine release to downstream striatal neurons, ramping DAN activity may causally determine movement timing. If so, causally increasing the activity of DANs during timing should result in earlier self-timed movements, and vice-versa. We thus optogenetically activated or inhibited DANs (in separate experiments) on 30% of trials (Fig. 4a, Extended Data Fig. 9a,b). Activation significantly early-shifted the distribution of self-timed movements on stimulated trials compared to unstimulated trials (12 mice), whereas

inhibition produced significant late-shifting compared to unstimulated trials (4 mice). Stimulation of mice expressing no opsin produced no consistent effect on timing (5 mice, Figure 4b-d; Extended Data Fig. 9c-e). The direction of these effects was consistent across all animals tested in each category. Whereas bilateral stimulation of SNc DAN cell bodies caused early-shifting, the effects were generally larger and more consistent when activating SNc DAN terminals in DLS (2 mice, Extended Data Fig. 9c,d). Outside the context of the timing task, DAN activation did not elicit immediate licking, nor did inhibition prevent licking, suggesting optogenetic effects on timing did not result from direct triggering or suppression of movement[14], but rather were expressed through a "higher-level" cognitive process related to self-timing of the movement (Extended Data Fig. 10).

Discussion. Here, we found that both baseline and slowly ramping DAN signals predict the timing of self-initiated movements. Trial-by-trial differences in these signals were finely tuned to movement onset, whether these signals were recorded from SNc cell bodies, SNc terminals in the DLS, or VTA cell bodies. Moreover, slow DAN dynamics were reflected in dopamine release in DLS, demonstrating availability of this information to downstream striatal effectors positioned to influence when movement occurs. Consistent with the direction of these effects, optogenetic suppression and augmentation of DAN activity during the timing interval causally altered movement timing. Thus, DAN activity is poised to control the moment-to-moment decision of when to move.

A number of studies have reported short-latency (≤ ~500 ms) increases in DAN activity in response to sensory cues and immediately preceding self-initiated movements[12-16], similar to the sensory-

and motor-related transients we observed following the cue and preceding first-lick. However, the slow-timescale DAN signals we observed during self-timing were markedly different. First, the ramping signal unfolded over *seconds*, preceding the first-lick by as long as 10 s. Second, variations in baseline amplitude before the cue and the subsequent ramp-slope predicted the trial-by-trial timing of the first-lick. To effectively model DAN signals on single trials, we had to incorporate two time-dependent features: a baseline offset and a "stretch" parameter that scaled DAN signals along the time axis (Fig. 2). Moreover, these features predicted movement time independent of recent trial history (Fig. 3). Combined with the optogenetic results, these findings suggest that variations in slow DAN dynamics affect trial-by-trial movement timing.

These slow-timescale DAN signals could be unique to the timing requirement of our task. However, when we averaged DAN signals aligned to "spontaneous" licks during the ITI, we also observed slow ramping similar to that observed during the timing interval, with signal building over seconds from the offset of the previous lick up to the time of the next lick (Extended Data Fig. 6d). Thus, slowly evolving DAN signals may be integral to self-initiated movements more generally. It is possible that slow ramping dynamics predictive of movement timing would emerge in previous datasets if DAN signals were similarly averaged according to the interval between self-initiated movement bouts.

Previous studies have reported slow ramping signals in the mesolimbic system in certain behavioral contexts, including goal-directed navigation[19]; multi-step tasks culminating in reward[25,26]; and passive observation of dynamic visual cues indicating proximity to reward[20]. It has been proposed that slowly ramping mesolimbic DAN signals could encode increasing value as

animals approach reward[25,26] or alternatively could reflect moment-by-moment reward-prediction errors (RPE)[20,27]. The ramping signals we observed in the nigrostriatal system are consistent with either value or RPE interpretations. However, it has been unclear how the brain *employs* slowly ramping DAN signals in behavior. Our study moves beyond previous studies by finding that trial-by-trial variability in ramping dynamics explains the precise timing of a behavioral output—the self-timed lick—and that optogenetically manipulating SNc DAN activity causally alters the timing of that output. Thus, SNc ramping may not merely encode progress toward a goal, but could also play a *causal* role in the timing of movement initiation. This interpretation could be related to classic findings from Parkinson's disease, in which loss of nigrostriatal pathway DANs results in difficulty initiating movements[28,29].

Lesion and pharmacological studies have long suggested roles for the SNc and dopamine in timing[4,5]. Broadly speaking, conditions that increase dopamine availability, such as amphetamine administration, affect timing as if speeding an internal "pacemaker[10,11,30]," whereas conditions that decrease dopamine availability/efficacy generally have the opposite effect[4,8]. Our results—in both recordings and optogenetic manipulations of DANs—are consistent with this view. Moreover, the ramping signals and the anti-correlation of ramping slope with movement time bear striking resemblance to Pacemaker-Accumulator models of neural timing[5,9], a longstanding conceptual framework that captures canonical features of timing behavior.

Soares *et al*. recently reported findings complicating the standard view of dopamine in timing[1]. In mice performing a temporal bisection task, RPE-like transients in SNc DAN GCaMP6f signals were observed immediately after the stop-timing cue. These transients were smaller when animals

overestimated the timed interval, which was interpreted as evidence that *lower* DAN activity reflects a *faster* pacemaker, the *opposite* of our findings and most prior work[30]. This finding may be unique to the temporal bisection task, which has aspects of categorization as well as timing and thus is more complex than tasks that only rely on movement to produce timed intervals (as in our task).

However, a recently-proposed temporal difference learning framework for explaining dynamic DAN activity could provide a unified explanation for these findings[20,27,30](see Supplementary Discussion for details). The model assumes that DAN activity provides a continuous readout of RPE, which under conditions of state uncertainty (as in timing) is shown to reflect the moment-to-moment derivative of the value landscape[27]. In this framework, variation in interval timing arises from differences in the rate of traversing the internal model of the value landscape compared to veridical time, which can be modeled as stretching/compression of the subjective value function[30]. Critically, the amount of compression is taken to be controlled by a pacemaker whose speed is proportional to the tonic level of dopamine from trial-to-trial[30]. In both the self-timed movement and bisection tasks, we observed higher *baseline* DAN signals associated with relatively fast timekeeping, consistent with relatively high tonic DAN activity reflecting fast pacemaking in both tasks. Additionally, temporally-discounted value should increase as the time of reward approaches in both tasks, and the model predicts faster pacemaking on a given trial would *compress* this function. During the self-timed movement task, compression would cause faster increases in value and thus steeper ramping of the DAN signal, as we observed. In the bisection task, compression of the value function would produce higher estimated value just before the stop-cue, and thus a smaller change in value (RPE) following the stop-cue, resulting in a blunted DAN transient. Soares

et al. indeed observed smaller stop-cue-related transients when animals overestimated the timed interval—consistent with compression of the subjective value function during fast pacemaking.

Altogether, we argue that relatively high DAN activity reflects faster pacemaking across timing tasks, with the specific timing from trial-to-trial influenced by the dynamics of DAN signaling.

Main References

1. Soares, S., Atallah, B. & Paton, J. Midbrain DA neurons encode perception of time. *Science* **354,** 1273–1277 (2016).
2. Rakitin, B. C., Penney, T. B., Gibbon, J., Hinton, S. C., & Meek, W. H. Scalar Expectancy Theory and Peak-Interval Timing in Humans. *Journal of Experimental Psychology: Animal Behavior Processes* **24**, 15–33 (1998).
3. Gallistel, C. R. & Gibbon, J. Time, Rate and Conditioning. *Psychol. Rev*. **107,** 289–344 (2000).
4. Merchant, H., Harrington, D. L. & Meck, W. H. Neural Basis of the Perception and Estimation of Time. *Annu. Rev. Neurosci* 36, 313–36 (2013).
5. Meck, W. H. Neuroanatomical localization of an internal clock: A functional link between mesolimbic, nigrostriatal, and mesocortical dopaminergic systems. *Brain Res*. **1109,** 93–107 (2006).
6. Mello, G. B. M., Soares, S. & Paton, J. J. A scalable population code for time in the striatum. *Curr. Biol*. **25,** 1113–1122 (2015).
7. Lee, I. H. & Assad, J. A. Putaminal Activity for Simple Reactions or Self-Timed Movements. *J. Neurophysiol*. **89,** 2528–2537 (2003).

8. Meck, W. H. Affinity for the dopamine D2 receptor predicts neuroleptic potency in decreasing the speed of an internal clock. *Pharmacol. Biochem. Behav.* **25,** 1185–1189 (1986).

9. Lustig, C. & Meck, W. H. Chronic treatment with haloperidol induces deficits in working memory and feedback effects of interval timing. *Brain Cogn.* **58,** 9–16 (2005).

10. Dews, P. B. & Morse, W. H. Some observations on an operant in human subjects and its modification by dextro amphetamine. *J. Exp. Anal. Behav.* **1,** 359–364 (1958).

11. Schuster, C. & Zimmerman, J. Timing behavior during prolonged treatment with dl-amphetamine. *J. Exp. Anal. Behav.* **4,** 327–330 (1961).

12. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).

13. Dodson, P. D. et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc. Natl Acad. Sci. USA* **113**, E2180–E2188 (2016).

14. Coddington, L. T. & Dudman, J. T. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573 (2018).

15. da Silva, J. A., Tecuapetla, F., Paixão, V. & Costa, R. M. Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018).

16. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).

17. Sippy, T., Lapray, D., Crochet, S., & Petersen, C. C. H. Cell-Type-Specific Sensorimotor Processing in Striatal Projection Neurons during Goal-Directed Behavior. *Neuron* **88**, 298–305 (2015).

18. Sun, F. et al. New and improved GRAB fluorescent sensors for monitoring dopaminergic activity in vivo. Preprint at https://www.biorxiv.org/content/10.1101/2020.03.28.013722v2 (2020).

19. Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. M. & Graybiel, A. M. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).

20. Kim, H. R. et al. A unified framework for dopamine signals across timescales. Preprint at https://www.biorxiv.org/content/10.1101/803437v1 (2019).

21. Engelhard, B et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).

22. Park, I. M., Meister, M. L. R., Huk, A. C., & Pillow, J. W. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neuro.* **17**, 1395–1403 (2014).

23. Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. Distinct timescales of population coding across cortex. *Nature* **548**, 92–96 (2017).

24. Maimon, G. & Assad, J. A. A cognitive signal for the proactive timing of action in macaque LIP. *Nat. Neuro.* ***9***, 948–955 (2006).

25. Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).

26. Mohebi, A et al. Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).

27. Mikhael, J. G., Kim, H. R., Uchida, N., & Gershman, S. J. Ramping and State Uncertainty in the Dopamine Signal. Preprint at https://www.biorxiv.org/content/10.1101/805366v1 (2019).

28. Grillner, S. & Robertson, B. The Basal Ganglia Over 500 Million Years. *Curr. Biol.* R1088–R1100 (2016).

29. Freeze, B. S., Kravitz, A. V, Hammack, N., Berke, J. D. & Kreitzer, A. C. Control of Basal Ganglia Output by Direct and Indirect Pathway Projection Neurons. *J. Neurosci.* **33,** 18531–18539 (2013).

30. Mikhael, J. G. & Gershman, S. J. Adapting the flow of time with dopamine. *J. Neurophysiol.*, **121**, 1748–1760 (2019).

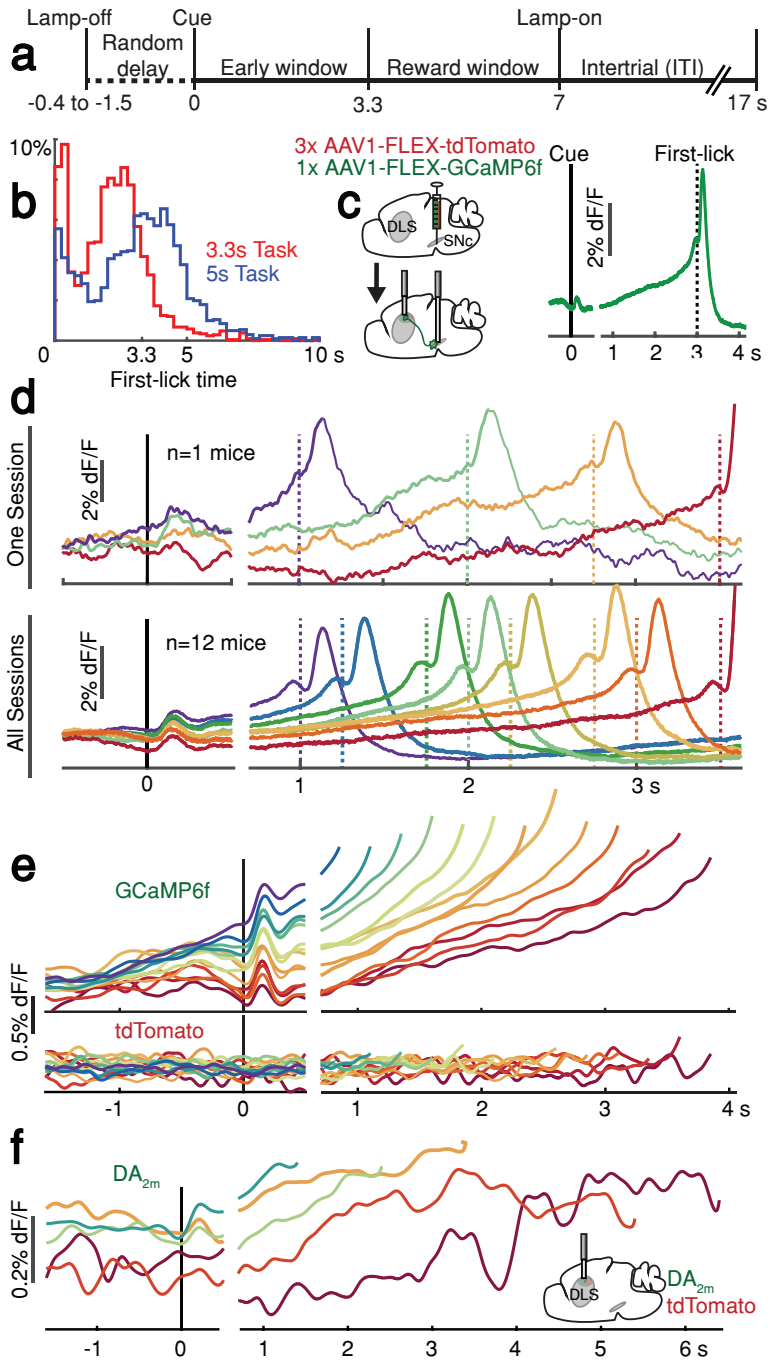


Figure 1 | SNc DAN signals preceding self-timed movement. a, Self-timed movement task. **b,** First-lick distribution for two task variants performed by the same mouse. Red: 3.3 s-reward boundary (4 sessions); Blue: 5 s-reward boundary (4 sessions). For all mice, see Extended Data Fig. 1. **c,** Left: surgical strategy for GCaMP6f/tdTomato fiber photometry. Right: average SNc DAN GCaMP6f response for first-licks between 3-3.25 s (12 mice). Left of plot: cue-aligned average; right of plot: first-lick-aligned average. Vertical dashed line: first-lick time. Break in axis indicates change in plot alignment. **d,** Average SNc DAN GCaMP6f responses for different first-lick times (12 mice). **e,** Comparison of average DAN GCaMP6f and tdTomato responses (12 mice). Traces plotted up to 150 ms before first-lick. **f,** Average DLS DA_{2m} signals (4 mice).

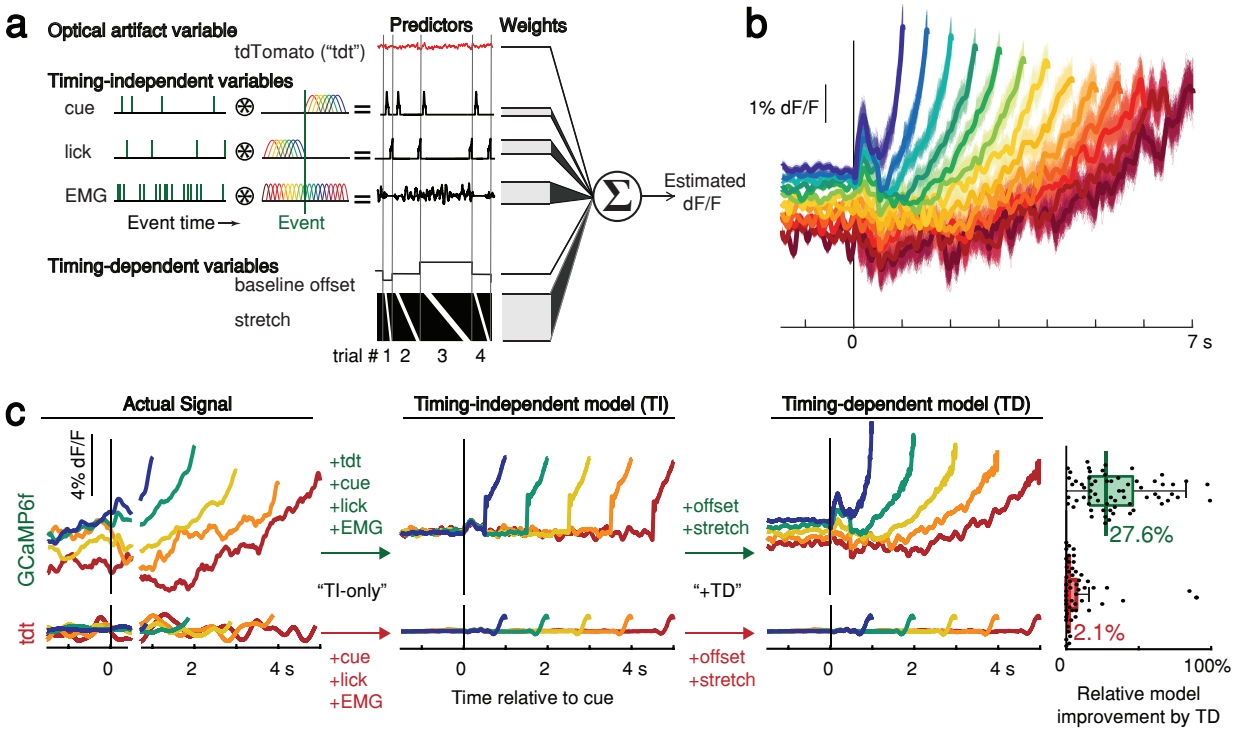


Figure 2 | Contribution of optical artifacts, task variables and ongoing movements to SNc GCaMP6f signals. a, Nested encoding model comparing the contribution of timing-independent predictors (TI) to the contribution of timing-dependent predictors (TD). **b,** Predicted dF/F signal for one session plotted up to time of first lick. Model error simulated 300x (shading). **c,** Nested encoding model for 1 session showing the recorded signal (left panel), the timing-independent model (middle panel), and the full, timing-dependent model with all predictors (right panel). Top: GCaMP6f; Bottom: tdTomato (tdt).

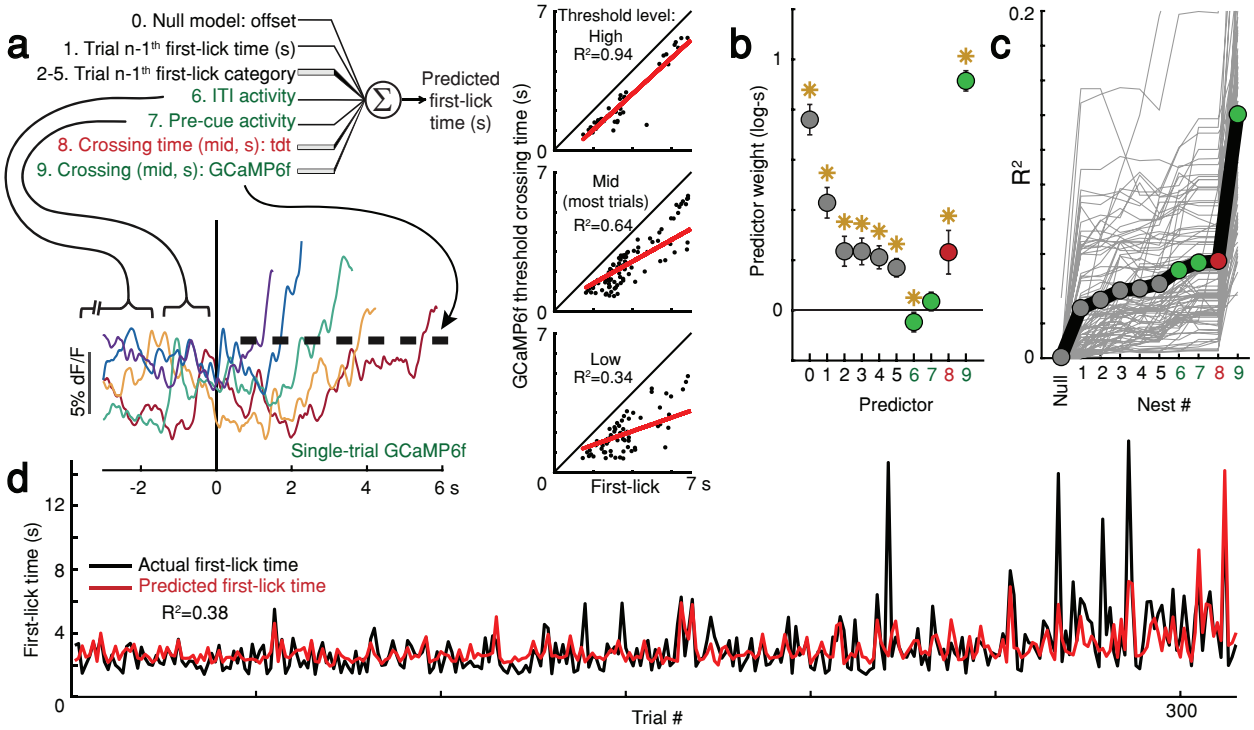


Figure 3 | Single-trial baseline and ramping DAN signals predict first-lick timing. a, Nested decoding model. Top-left: schematic. Bottom-left: single-trial cue-aligned SNc DAN GCaMP6f signals from one session (6 trials shown for clarity). Traces plotted up to first-lick. Right: threshold crossing model. Low/Mid/High: threshold amplitude. Grey dots: single trials. **b,** Model weights, 94 sessions. Error bars: 95% CI, *: p<0.05. Numbers indicate nesting-order. **c,** Variance explained by each model nest. Grey lines: single sessions; thick black line: average. For model selection, see **Extended Data Fig. 8c. d,** Predicted vs. actual first-lick time, same session as **a**.

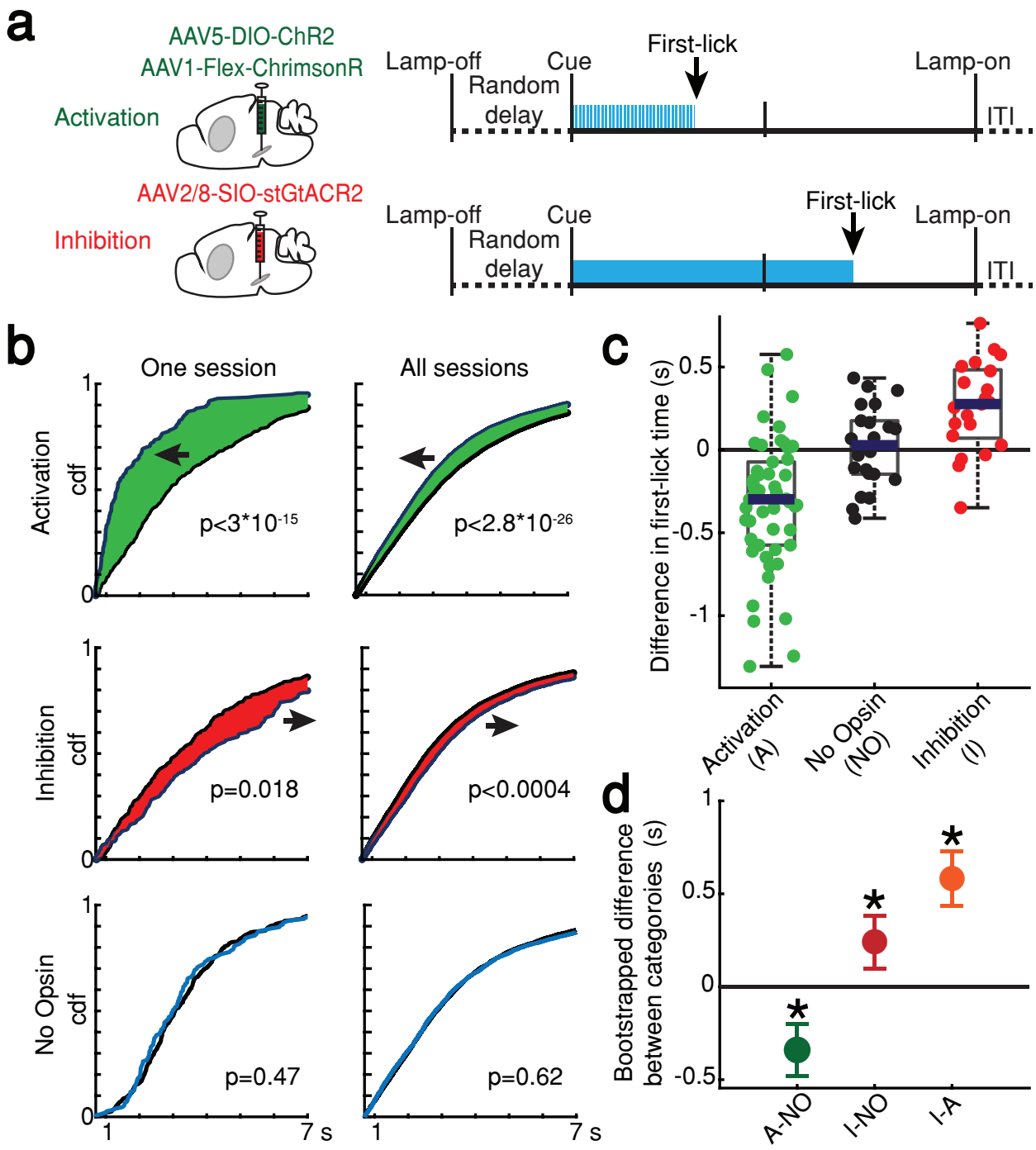


Figure 4 | Optogenetic DAN manipulation systematically and bidirectionally shifts the timing of self-timed movements. a, Strategy for optogenetic DAN activation or inhibition. Mice were stimulated from cue-onset until first-lick or 7 s. **b,** Empirical continuous probability distribution functions (cdf) of first-lick times for stimulated (blue line) versus unstimulated (grey line) trials**.** Arrow and shading show direction of effect. P-values calculated by Kolmogorov-Smirnov test (for other metrics, see Extended Data Fig. 9b-e). **c,** Mean bootstrapped difference in first-lick time, stimulated-minus-unstimulated trials. Dots: single sessions. **d,** Comparison of mean first-lick time difference across all sessions. Error bars: 95% confidence interval (*: $p < 0.05$).

Methods

Animals. Adult male and female hemizygous DAT-cre mice[31] (B6.SJL-Slc6a3$^{tm1.1(cre)Bkmm}$/J; The Jackson Laboratory) or *wt* C57/b6 mice were used in all experiments. Mice were housed on a reversed night/day cycle (12h dark/12h light) and behavioral sessions occurred during the dark cycle. All experiments and protocols were approved by the Harvard Institutional Animal Care and Use Committee and were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Surgery. Surgeries were conducted under aseptic conditions. Mice were anesthetized with isoflurane (0.5-2% at 0.8L/min). Analgesia was provided by *s.c.* 5mg/kg ketoprofen injection during surgery and once daily for 3d postoperatively. Virus was injected (50nL/min) and the pipet remained in place for 10min before removal. 200μm, 0.53NA blunt fiber optic cannulae (Doric) or tapered fiber optic cannulae (200μm, 0.60NA, 2mm tapered shank, OptogeniX) were positioned at SNc, VTA or DLS and secured to the skull with dental cement (Metabond). Neck EMG electrodes were constructed from two 32G pacemaker wires attached to a custom socket mounted in the dental cement. Sub-occipital neck muscles were exposed by blunt dissection and electrode tips embedded bilaterally.

Stereotaxic coordinates (from bregma and brain surface).

Virus:

SNc: 3.16mm posterior, +/- 1.4mm lateral, 4.2mm ventral

<u>VTA</u>: 3.1mm posterior, +/-0.6mm lateral, 4.2mm ventral

<u>DLS</u>: 0mm anterior, +/- 2.6mm lateral, 2.5mm ventral.

Fiber Optic Tips:

SNc/VTA: 4.0mm ventral (photometry) or 3.9mm ventral (optogenetics).

DLS: 2.311mm ventral (blunt fiber) or 4.0mm ventral (tapered fiber)

Virus.

Photometry:

tdTomato ("tdt"): AAV1-CAG-FLEX-tdT (UNC Vector Core), 100nL used alone or in mixture with other fluorophores (below), working concentration $5.3*10^{12}$gc/mL

gCaMP6f (at SNc or VTA): 100nL AAV1.Syn.Flex.GCaMP6f.WPRE.SV40 ($2.5*10^{13}$gc/mL, Penn Vector Core). Virus was mixed in a 1:3 ratio with tdt (200nL total).

DA_{2m} (at DLS): 200-300nL AAV9-hSyn-DA4.4(DA2m) (working concentration: *ca.* $3*10^{12}$gc/mL, Vigene) + 100nL tdt

Optogenetics (all bilateral at SNc):

ChR2: 1000nL AAV5-EF1a-DIO-hChR2(H134R)-EYFP-WPRE-pA ($3.2*10^{13}$gc/mL, UNC Vector Core)

ChrimsonR: 700nL AAV1-hSyn-FLEX-ChrimsonR-tdT ($4.1*10^{12}$gc/mL, UNC Vector Core)

stGtACR2: 300nL 1:10 AAV2/8-hSyn1-SIO-stGtACR2-FusionRed (working concentration $4.7*10^{11}$gc/mL, Addgene/Janelia Viral Core)

Water-deprivation and acclimation. Animals recovered 1wk postoperatively before water deprivation. Mice received daily water supplementation to maintain ≥80% initial body weight and fed *ad libitum*. Training commenced when mice reached the target weight (~8-9d post-surgery).

Histology. Mice were anesthetized with >400mg/kg pentobarbital (Somnasol) and perfused with 10mL 0.9% sodium chloride followed by 50mL ice-cold 4% paraformaldehyde in 0.1M phosphate buffer. Brains were fixed in 4% paraformaldehyde at 4°C for >24hr before being transferred to 30% sucrose in 0.1M phosphate buffer for >48hr. Brains were sliced in 50μm coronal sections by freezing microtome, and fluorophore expression was assessed by light microscopy. The sites of viral injections and fiber optic placement were mapped with an Allen Mouse Brain Atlas.

Behavioral rig, data acquisition and analysis. A custom rig provided cues, recorded events and delivered juice rewards under the control of a Teensy 3.2 microprocessor running a custom Arduino state-system behavioral program with MATLAB serial interface. Digital and analog signals were acquired with a CED Power 1400 data acquisition system/Spike2 software (Cambridge Electronic Devices). Photometry and behavioral events were acquired at 1000Hz; movement channels were acquired at 2000Hz. Video was acquired with FlyCap2 or Spinnaker at 30fps (FLIR). Data were analyzed with custom MATLAB statistics packages.

Self-timed movement task. Mice were head-fixed with a juice tube positioned in front of the tongue. During periods when rewards were not available, a houselamp was illuminated. At trial start, the houselamp turned off, and a random delay ensued (0.4-1.5s) before a cue (simultaneous LED flash and 3300Hz tone, 100ms) indicated start of the timing interval. The timing interval was

divided into two windows, early (0-3.333s in most experiments; 0-4.95s in others) and reward (3.333-7 s; 4.95-10s), followed by the intertrial interval (ITI, 7-17 s; 10-20s). The window in which the mouse first licked determined the trial outcome (early, reward, or no-lick). An early first-lick caused an error tone (440Hz, 200ms) and houselamp illumination, and the mouse had to wait until the full timing interval had elapsed before beginning the ITI. A first-lick during the reward window caused a reward tone (5050Hz, 200ms) and juice delivery, and the houselamp remained off until the end of the trial interval. If the timing interval elapsed with no lick, a time-out error tone played (131Hz, 2s), the houselamp turned on, and ITI commenced. During the ITI and pre-cue delay, there was no penalty for licking.

Mice learned the task in 3 stages (Extended Data Fig. 1a). On the first 1-4 days of training, mice learned a beginner-level task, which was modified in two ways: 1. To encourage participation, if mice did not lick before 5s post-cue, they received a juice reward at 5s, 2. Mice were not penalized for licking in reaction to the cue (within 500ms). When the mouse began self-triggering ≥50% of rewards (day 2-6 of training), the mouse advanced to the intermediate-level task, in which the training reward at 5s was omitted, and the mouse had to self-trigger all rewards. After completing >250 trials/day on the intermediate task, mice advanced to the mature task (no reaction licks permitted, day 4-7 of training). All animals learned the mature task and worked for ~400-1500 trials/session.

Online movement monitoring. Movements were recorded simultaneously during behavior with four movement-control measurements: neck EMG (band-pass filtered 50-2000Hz, 60Hz notch, amplified 100-1000x), back-mounted accelerometer (SparkFun), high-speed camera (30Hz, FLIR),

and tdt photometry. All control signals contained similar information, and thus only a subset of controls was used in some sessions.

Photometry. Fiber optics were illuminated with 475nm blue LED light (Plexon) (SNc/VTA: 50μW, DLS: 35μW) measured at patch cable tip with a light-power meter (Thor Labs). Green fluorescence was collected via a custom dichroic mirror (Doric) and detected with a Newport 1401 Photodiode. Fluorescence was allowed to recover ≥1d between recording sessions. To avoid crosstalk in animals with tdt expression, tdt was recorded at one of the 3 sites (SNc, VTA, or DLS, 550 lime LED, Plexon) while GCaMP6f or DA_{2m} was recorded simultaneously only at the other implanted sites.

dF/F. Raw fluorescence for each session was pre-processed by removing rare singularities (single points >15 STD from the mean) by interpolation to obtain F(t). To correct photometry signals for bleaching, dF/F was calculated as:

$$\frac{dF}{F}(t) = \frac{F(t) - F_0(t)}{F_0(t)}$$

where $F_0(t)$ is the 200 s moving average of F(t). We tested several other complementary methods for calculating dF/F and all reported results were robust to dF/F method (Supplementary Methods A). To ensure dF/F signal processing did not introduce artifactual scaling or baseline shifts, we also tested several complementary techniques to isolate undistorted F(t) signals where possible and quantified the amount of signal distortion when perfect isolation was not possible (Supplementary Methods A and Extended Data Fig. 3c.).

Baseline DAN signal encoding models. To determine whether baseline DAN signals were best explained by the prior trial outcome (n-1th), current trial outcome (nth), or some interaction between the two, we employed two "paired-trial" strategies, one using raw F(t) (to check for robustness to dF/F scaling artifacts, Supplementary Methods A) and the other using dF/F signals. Mean baseline activity was measured in sliding windows (100ms divisions, 2s windows) from the n-1th trial ITI-start to the nth trial cue-onset. We abbreviate trial outcomes as E="early" (first-lick between 0.7-3.333s, unrewarded) and R="rewarded" (first-lick at 3.334-7s, rewarded), with the position of the letters indicating the order of consecutive trials (e.g., "ER" or "RE").

Baseline differences between all paired, consecutive trials (EE, ER, RE, RR) were compared for dF/F signals, controlling for the prior trial outcome by 4-factor ANOVA (factor 1: previous trial outcome (E or R, 1 degree of freedom (df)); factor 2: subsequent trial outcome (E or R, 1 df); factor 3: presence of licks in window (df: 1); factor 4: session ID (df: 112), comparison: baseline dF/F activity within 2s sliding windows, Extended Data Fig. 6c, top). Relative influence on baseline dF/F by the n-1th and nth trial outcome was estimated as the relative size of the 4-way ANOVA F-statistic as a function of time (Extended Data Fig. 6c, bottom). Similar results were obtained on single sessions by 2-factor and 3-factor ANOVA. Similar results were also obtained comparing triplets of consecutive trials (EER, RRE, EEE, RRR). Both trial 2 and trial 3 in the triplet were preceded by a trial of the same outcome type, and thus we could compare baseline raw F(t) signals with minimal bleaching distortion.

DAN signal encoding model. To test the independent contribution of each task-related input to the photometry signal and select the best model, we employed a nested fitting approach, in which each dataset was fit multiple times (in "nests"), with models becoming progressively more complex in subsequent nests. The nests fit to the GCaMP6f photometry data employed the inputs $X^{(j)}$ at each jth nest:

Null Model: $X^{(0)} = x_0$

Nest 1: $X^{(1)} = X^{(0)} + \text{tdt}$

Nest 2: $X^{(2)} = X^{(1)} + \text{cue} + \text{first-lick}$

Nest 3: $X^{(3)} = X^{(2)} + \text{EMG/accelerometer}$

Nest 4: $X^{(4)} = X^{(3)} + \text{time-dependent baseline offset}$

Nest 5: $X^{(5)} = X^{(4)} + \text{stretch representing percentages of interval}$

Overfitting was penalized by ridge regression, and the optimal regularization parameter for each nest was obtained by 5-fold cross-validation to derive the final model fit for each session. Model improvement by each input was assessed by the percentage loss improvement at the nest where the input first appeared compared to the prior nest. The loss improvement of Nest 1 was compared to the Null Model (the average of the photometry timeseries). The nested model of tdt control photometry signals was the same, except Nest 1 was omitted.

The GLM for each nest takes the form:

$$Y = \Theta X^{(j)}$$

Where Y is the *1*x*n* vector of the photometry signal across an entire behavioral session (*n* is the total number of sampled timepoints); $X^{(j)}$ is the *d*x*n* design matrix for nest *j*, where the rows

correspond to the d_j predictors for nest j and the columns correspond to each of the n sampled timepoints of Y; and Θ is the dx1 vector of fit weights.

Y is the concatenated photometry timeseries taken from trial start (lights off) to the time of first lick. Because of day-to-day/mouse-to-mouse variation (ascribable to many possible sources, *e.g.*, different neural subpopulations, expression levels, behavioral states, *etc*.), each session was fit separately.

The d_j design matrix predictors were each scaled (maximum amplitude 1) and grouped by input to the model. The timing-independent inputs were: 1. Null offset (x_0, 1 predictor), 2. tdt (1 predictor), 3. cue (24 predictors), 4. first-lick (28 predictors), and 5. EMG/accelerometer (44 predictors). The timing-dependent inputs were: 6. timing-dependent baseline offset (1 predictor), 7. stretch (500 predictors).

To reduce the number of predictors, cue, first-lick and EMG/accelerometer predictors were composed from sets of basis kernels as described previously[22,23](Extended Data Fig. 7c). The cue basis kernels were spaced 0-500 ms post-cue and first-lick basis kernels were spaced -500ms-0ms relative to first-lick, the typically-observed windows of stereotypical sensory and motor-related neural responses. For nuisance movements (EMG/accelerometer), events were first discretized by thresholding (Extended Data Fig. 7b) and then convolved with basis kernels spanning -500 to 500 ms around the event. This window was consistent with the mean movement-aligned optical artifact observed in the tdt channel. The timing-dependent baseline offset was encoded as a constant offset spanning from lamp-off until first-lick, with amplitude taken as linearly proportional to the timed

interval on the current trial. The timing-dependent stretch input was composed of 500 predictors, with each predictor containing 1's tiling 0.05% of the cue-to-lick interval, and 0's otherwise (Extended Data Fig. 7d). Importantly, the stretch was not constrained in any way to form ramps.

Basis sets were optimized to minimize Training Loss, as calculated by mean squared error of the unregularized model:

$$\text{argmin}_{X^{(j)}}(\text{Training Loss}(\Theta) = 1/n\ \text{sum}(Y - \Theta X^{(j)})^2)$$

Superfluous basis set elements that did not improve Training Loss compared to the Null Model were not included in the final model. Goodness of the training fit was assessed by Akaiki Information Criterion (AIC), Bayesian Information Criterion (BIC), R^2, and Training Loss. The optimal, regularized model for each nest/session was selected by 5-fold cross-validation in which the regularization parameter, λ_j, was optimized for minimal average Test Loss:

$$\text{argmin}_{\lambda_j}\ (\text{Test Loss}(\Theta,\lambda_j) = 1/n\ \text{sum}(Y - \Theta X^{(j)})^2 + \lambda_j|\Theta|^2)$$

Test Loss for each optimal model was compared across nests to select the best model for each session. Models were refit with the optimal λ_j to obtain the final fit.

Model error was simulated 1000 times by redrawing Θ coefficients consistent with the data following the method described by Gelman and Hill[32], and standard errors were propagated across sessions. The absolute value of each predictor was summed and divided by the total number of predictors for that input to show the contribution of the input to the model (Extended Data Fig. 7g). To simulate the modeled session's photometry signal for each nest *j*, Yfit was calculated as

$\Theta X^{(j)}$ and binned by the time of first lick relative to the cue. The error in the simulation was shown by calculating $Yfit_{sim} = \Theta_{sim} X^{(j)}$ for 300 simulated sets of Θ_{sim}.

Principle component analysis (PCA)

Unsmoothed ramping intervals for GCaMP6f photometry timeseries were fit with PCA and reconstructed with the first three principle components (PCs). To derive a PCA fit matrix with ramping intervals of the same number of samples, the length of each trial was scaled up by interpolation to the maximum ramping interval duration:

7s–0.7s cue buffer–0.6s first-lick buffer=5.7s: 5700 sample ramping interval

Following PC-fitting, datasets were down-sampled to produce a fit of the correct time duration. Trials where the ramping interval was <0.1s were excluded to exclude noise from down-sampling.

First-lick time decoding model

A nested, generalized linear model was derived to predict the first-lick time on each trial in a session and quantify the contribution of previous reward history and photometry signals to the prediction. The model was of the form:

$$\log(y) = bx$$

where b is a vector of fit coefficients and x is a vector of predictors. The nested model was constructed such that predictors occurring further back in time (such as reward history) and confounding variables (such as tdt photometry signals) were added first to determine the additional variance explained by predictors occurring closer to the time of first-lick, which might otherwise obscure the impact of these other variables. The predictors, in order of nesting, were:

Nest 0: b0 (Null model, average log-first-lick time)

Nest 1:	b1 = b0 + first-lick time on previous trial
Nest 2-5:	b2 = b1 + previous trial outcome (1,0)*
Nest 6:	b3 = b2 + median photometry signal in 10s window before lamp-off
Nest 7:	b4 = b3 + median photometry signal from lamp-off to cue
Nest 9:	b5 = b4 + tdt threshold crossing time**
Nest 10:	b6 = b5 + GCaMP6f threshold crossing time**

where all predictors were normalized to be in the interval (0,1).

* Outcomes included (in order of nest): Reaction (first-lick before 0.5s), Early (0.5-3.333s), Reward (3.333-7s), ITI (7s-17s). No-lick was implied by all four outcomes encoded as zeros.
** Details on threshold-crossing time and alternative models included in Supplementary Methods B.

To exclude the sensory- and motor-related transients locked to the cue and the first-lick events in the threshold-crossing nests, the ramping interval was conservatively defined as 0.7s post-cue up until 0.6s before first-lick, and the minimum ramping interval for fitting was 0.1s. Thus, for a trial to be included in the model, the first lick occurred between 1.4s to 17s (end of trial).

Initial model goodness of fit was assessed by R^2, mean-squared loss and BIC. Models were 5-fold cross-validated with ridge regression at each nest to derive the final models, as described above. 95% confidence intervals on model coefficients were calculated by 2-sided t-test with standard errors propagated across sessions.

Optogenetics—naïve/expert control sessions. To determine whether optogenetic stimulation directly elicited or prevented licking, licking behavior was first tested outside the context of the self-timed movement task on separate sessions in the same head-fixed arena but with no cues or behavioral task. Opsin-expressing mice were tested before any exposure to the self-timed movement task ("Naïve") as well as after the last day of behavioral recording ("Expert"). In ChR2 control sessions, stimulation (5mW 425nm light, 3s duration, 10Hz, 20% duty cycle) was applied randomly at the same pace as in the self-timed movement task. stGtACR2 control sessions were conducted similarly (12mW 425mW light, 3s duration, constant illumination); but to examine if inhibition could block ongoing licking, we increased the baseline lick-rate by delivering juice rewards randomly (5% probability checked once every 5s).

Optogenetics—self-timed movement task. SNc DANs were optogenetically manipulated in the context of the 3.3s self-timed movement task. To avoid over-stimulation, light levels were adjusted to be subthreshold for eliciting overt movements[14], and mice were not stimulated on consecutive days.

<u>Activation</u>: SNc cell bodies were illuminated bilaterally (ChR2: 0.5-5mW 425nm blue LED light; ChrimsonR 550nm lime or 660nm crimson) on 30% of trials (10Hz, 20ms up-time starting at cue onset and terminating at first-lick). DAN terminals in DLS were stimulated bilaterally via tapered fiber optics on separate sessions.

<u>Inactivation</u>: SNc cell bodies were illuminated bilaterally (stGtACR2: 12 mW 425 nm blue light) on 30% of trials (constant illumination starting at cue onset and terminating at first lick).

Quantification of optogenetic effects. The difference in the distribution of trial outcomes between stimulated and unstimulated trials on *each session* was quantified in four ways.

1. 2-Sample Unsigned Kolmogorov-Smirnov Test.
2. Difference in empirical continuous probability distribution function (cdf). The difference in the integral of the stimulated and unstimulated cdf (dAUC) was calculated for each session from 0.7-7s. Effect size was quantified by permutation test, wherein the identity of each trial (stimulated or unstimulated) was shuffled, and the distribution of dAUCs for the permuted cdfs was calculated 10,000x. Results were reported for all sessions.
3. Difference in mean movement time. Movement times on stimulated and unstimulated trials were pooled and the distribution of movement time differences was determined by non-parametric bootstrap, in which a random stimulated and unstimulated trial were drawn from their respective pools 1,000,000x and the difference taken. The mean of each session's bootstrapped distribution was compared across sessions by the 1,000,000x bootstrapped difference of the mean between sessions of different categories.
4. Difference in median movement time. Same as above but with median.

Code availability

All custom behavioral software and analysis tools are available at https://github.com/harvardschoolofmouse.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Methods references

31. Backman, C. M *et al*. Characterization of a Mouse Strain Expressing Cre Recombinase From the 3' Untranslated Region of the Dopamine Transporter Locus. *Genesis* **44**, 383–390 (2006).

32. Gelman, A., & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, 43-45 (2006).

Acknowledgements

We thank D. Chicharro, J.G. Mikhael, S.J. Gershman, S. Panzeri and E.N. Brown for discussions on analytical methods; V. Berezovskii, J. LeBlanc, T. LaFratta, O. Mazor, P. Gorelik, J. Markowitz, A. Lutas, K.W. Huang, and N.E. Ochandarena for technical assistance; and B. Sabatini, C. Harvey, S.R. Datta, M. Andermann and W. Regehr for reagents. The work was supported by NIH grants UF-NS109177 and U19-NS113201, and NIH core grant EY-12196. A.E.H. was supported by a Harvard Lefler Predoctoral Fellowship and a Harvard Quan Predoctoral Fellowship.

Author Contributions

A.E.H. and J.A.A conceived the project. A.E.H. performed all experiments. G.S. assisted with experiments using tapered fiber optics; Y.H. assisted with optogenetics control experiments. F.S.

and Y.L. developed the dopamine sensor, DA_{2m}. A.E.H. and J.A.A. analysed the data and wrote the paper.

Competing Interests

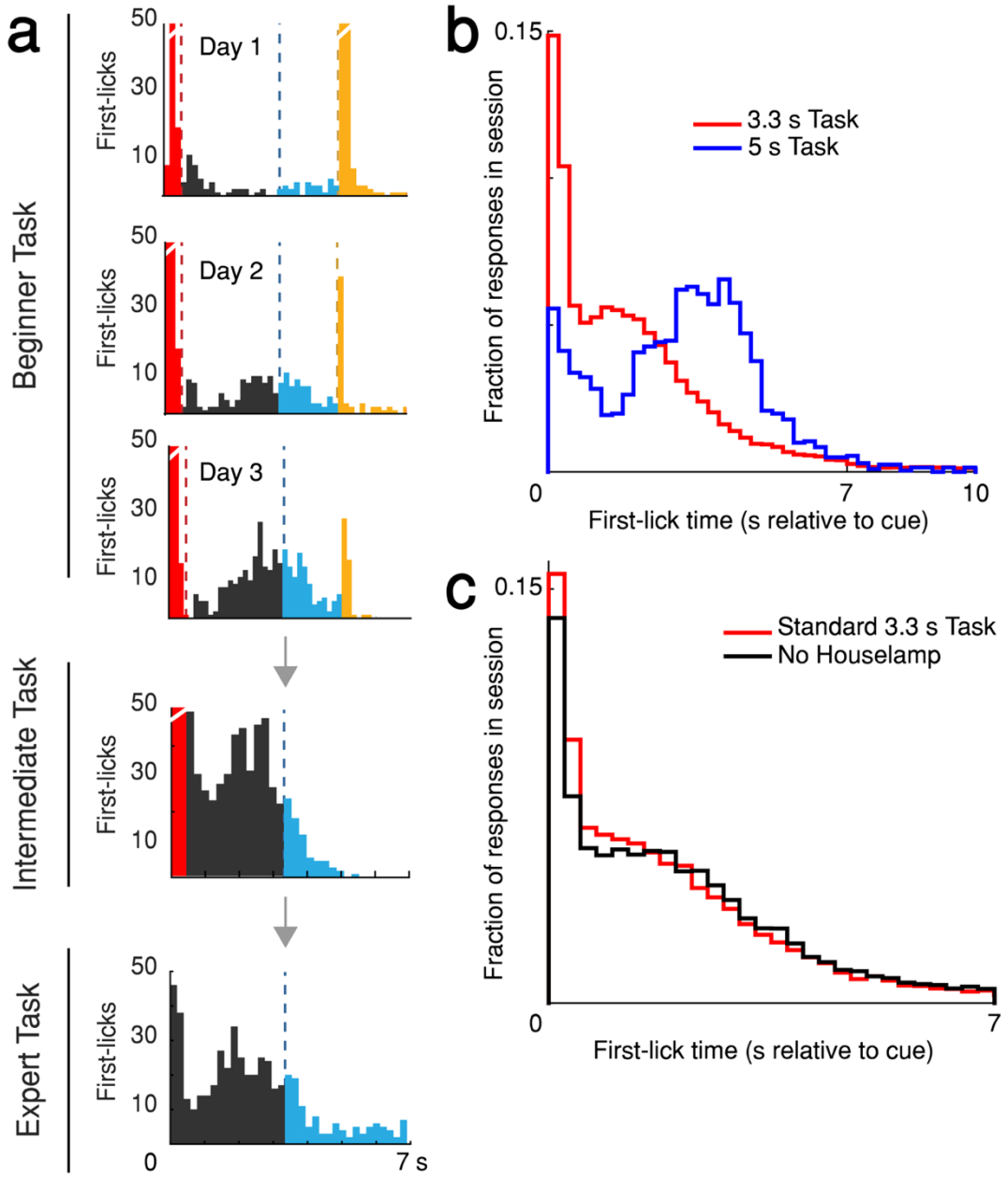
J.A.A. is a co-founder of OptogeniX, which produces the tapered optical fibers used in some experiments.

Extended Data is available for this paper.

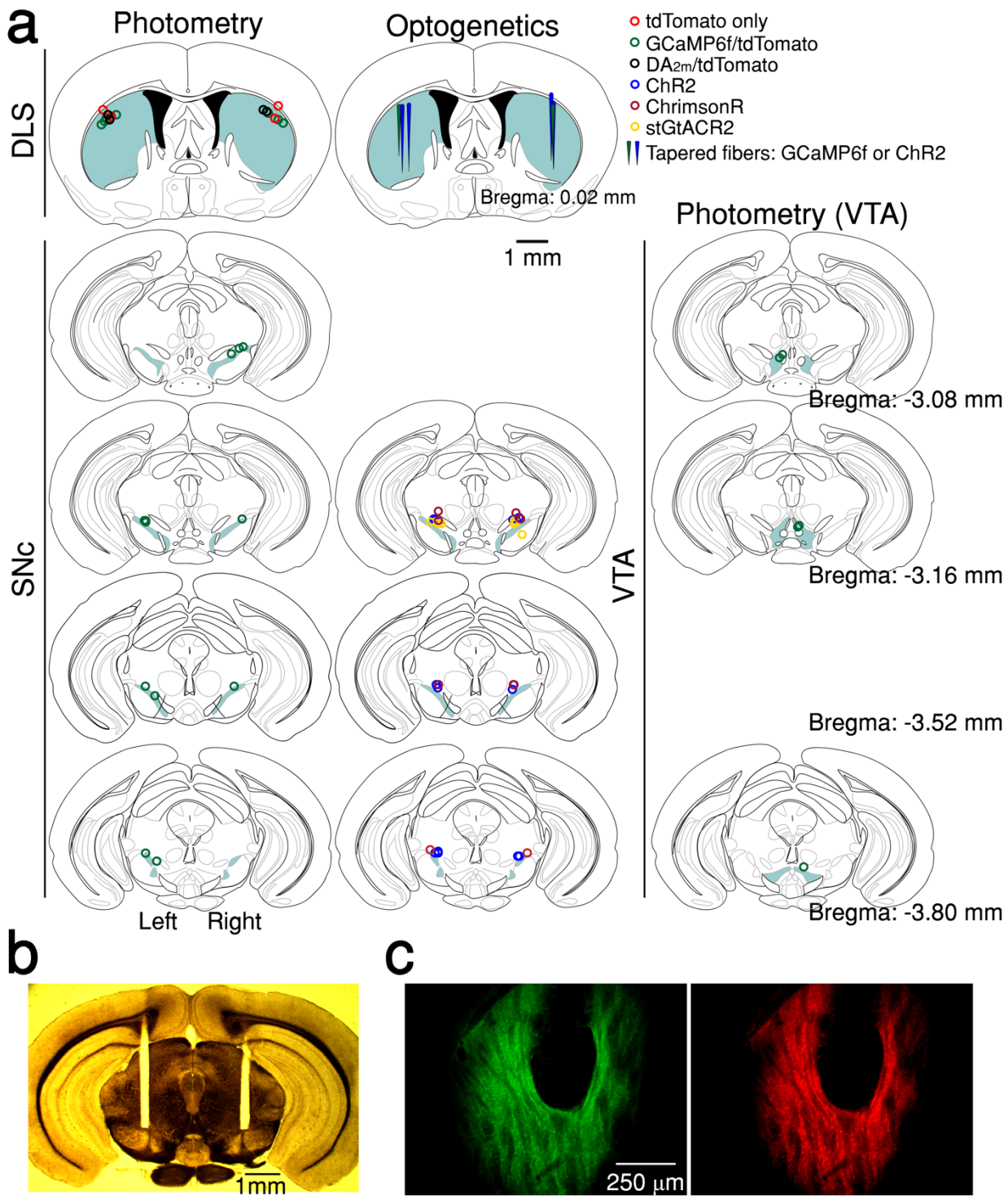
Supplementary Information including Supplementary Methods and Discussion are available for this paper.

Correspondence and requests for materials should be addressed to A.E.H. or J.A.A.

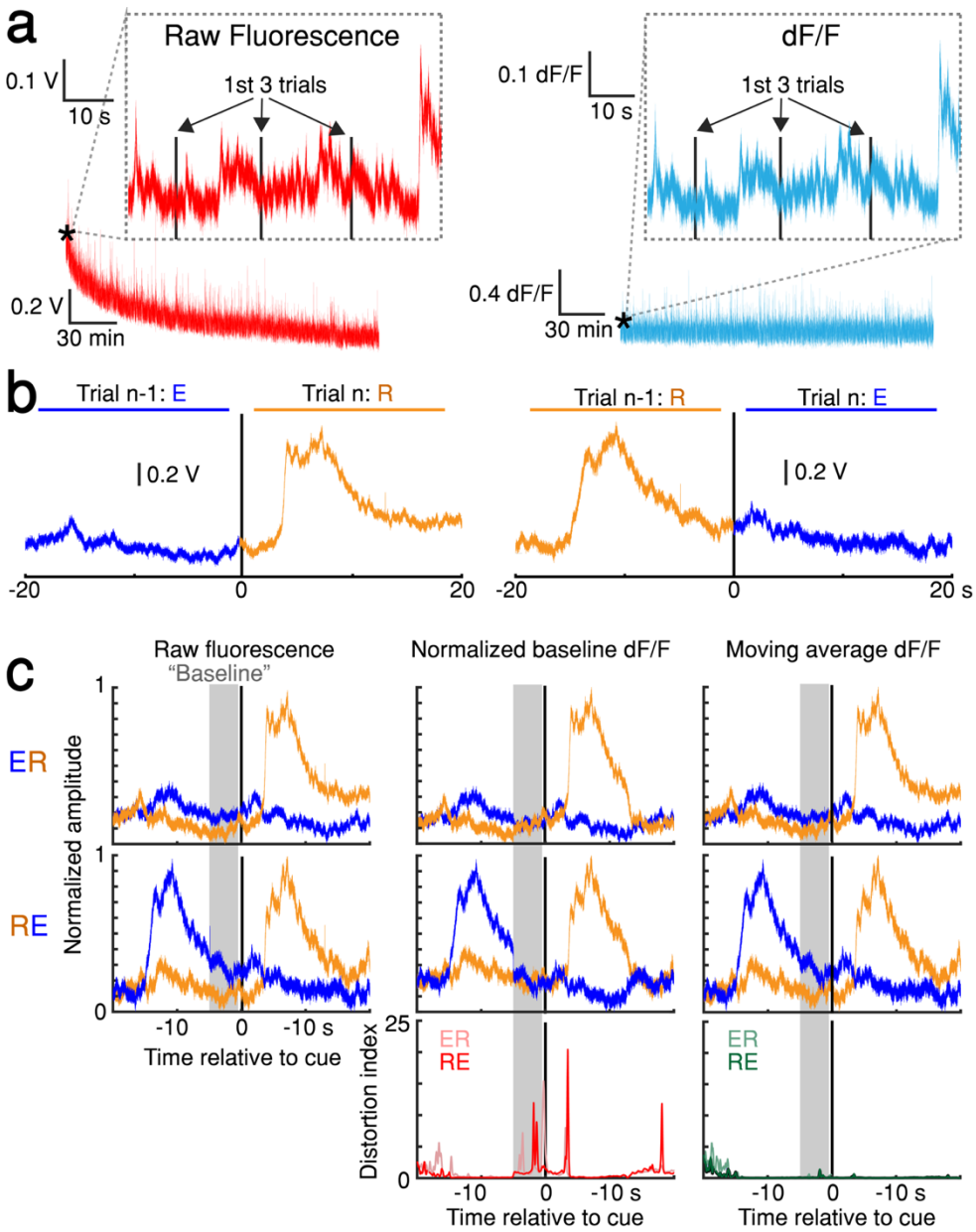
Extended Data



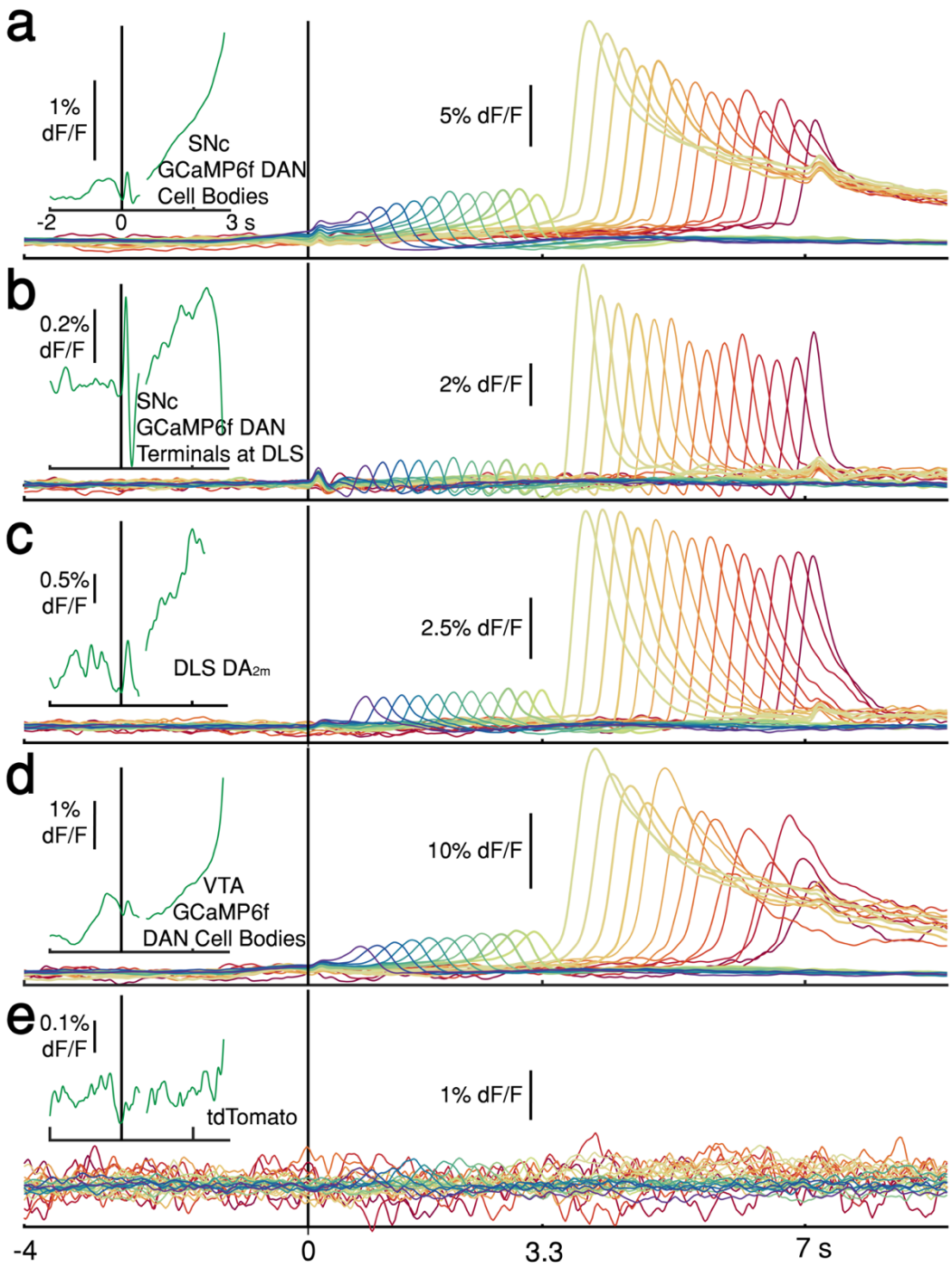
Extended Data Figure 1 | Self-timed movement task learning and variations. a, Task learning. Histogram of first-licks from single sessions at different stages of training. Bar color indicates first lick category (red: reaction, grey: early, blue: operant-rewarded, gold: Pavlovian-rewarded). Diagonal slash across bar top indicates bar height truncated for clarity. **b,** First-lick distributions from tasks with different target timing intervals (16 mice, 152 sessions). Red: 3.3 s reward-boundary. Blue: 5 s reward-boundary. Mice adjust behavior to the timing-contingencies of the task. **c,** First-lick distributions during behavior with and without houselamp cues. Red: standard 3.3 s task; Black: 3.3 s task omitting houselamp cues (4 mice, 4-5 sessions/mouse on each version of the task). Mice time their first-licks relative to the start cue, not the houselamp.



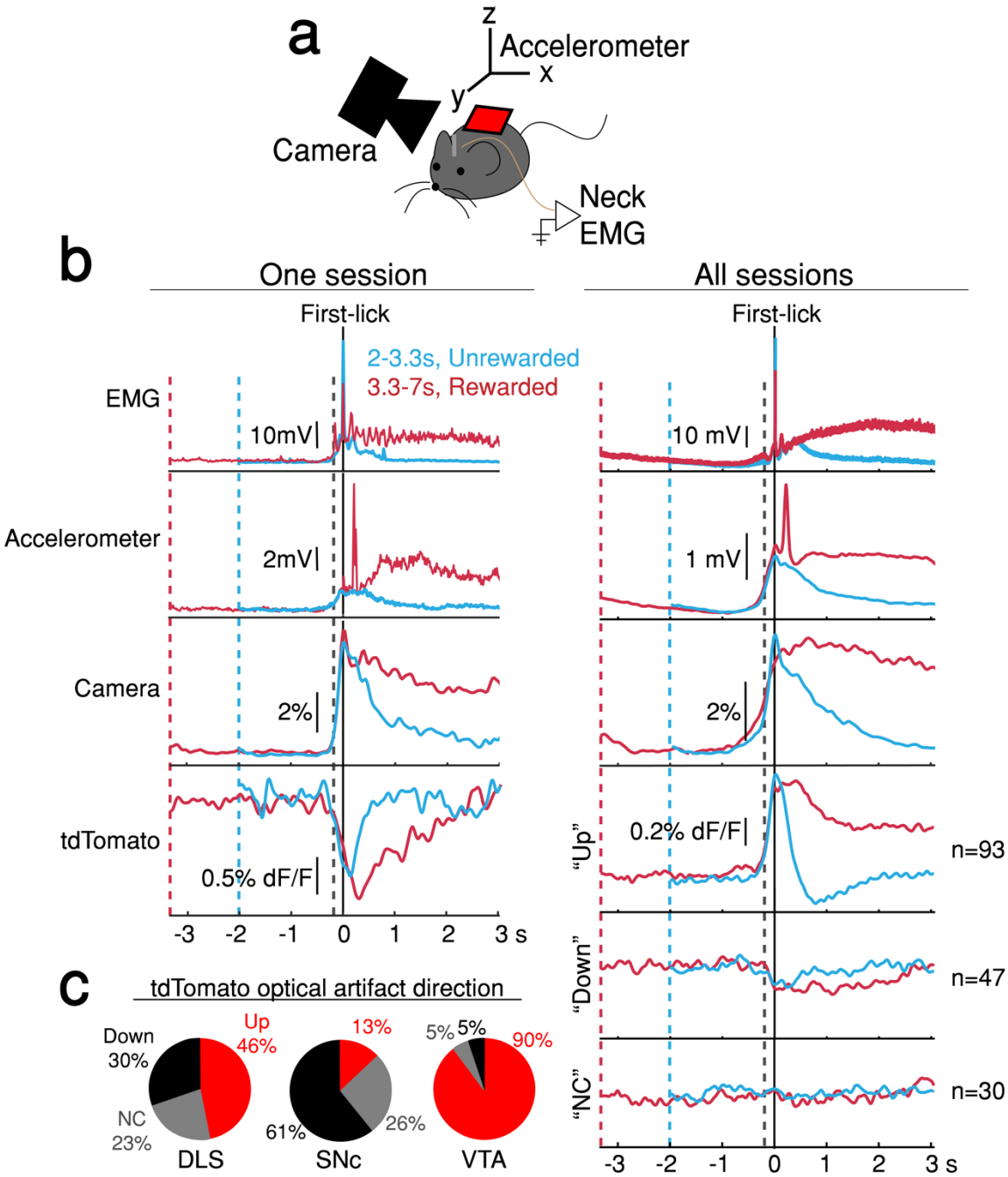
Extended Data Figure 2 | Fiber optic placement and histology. a, Approximate fiber positions, all mice. **b,** Brightfield microscopy with polarized filter on a freshly-cut brain slice showing bilateral fiber placement at SNc (stGtACR2). **c,** Example of co-expression of green and red fluorophores relative to fiber optic tip (Left: DA_{2m}, Right: tdTomato).



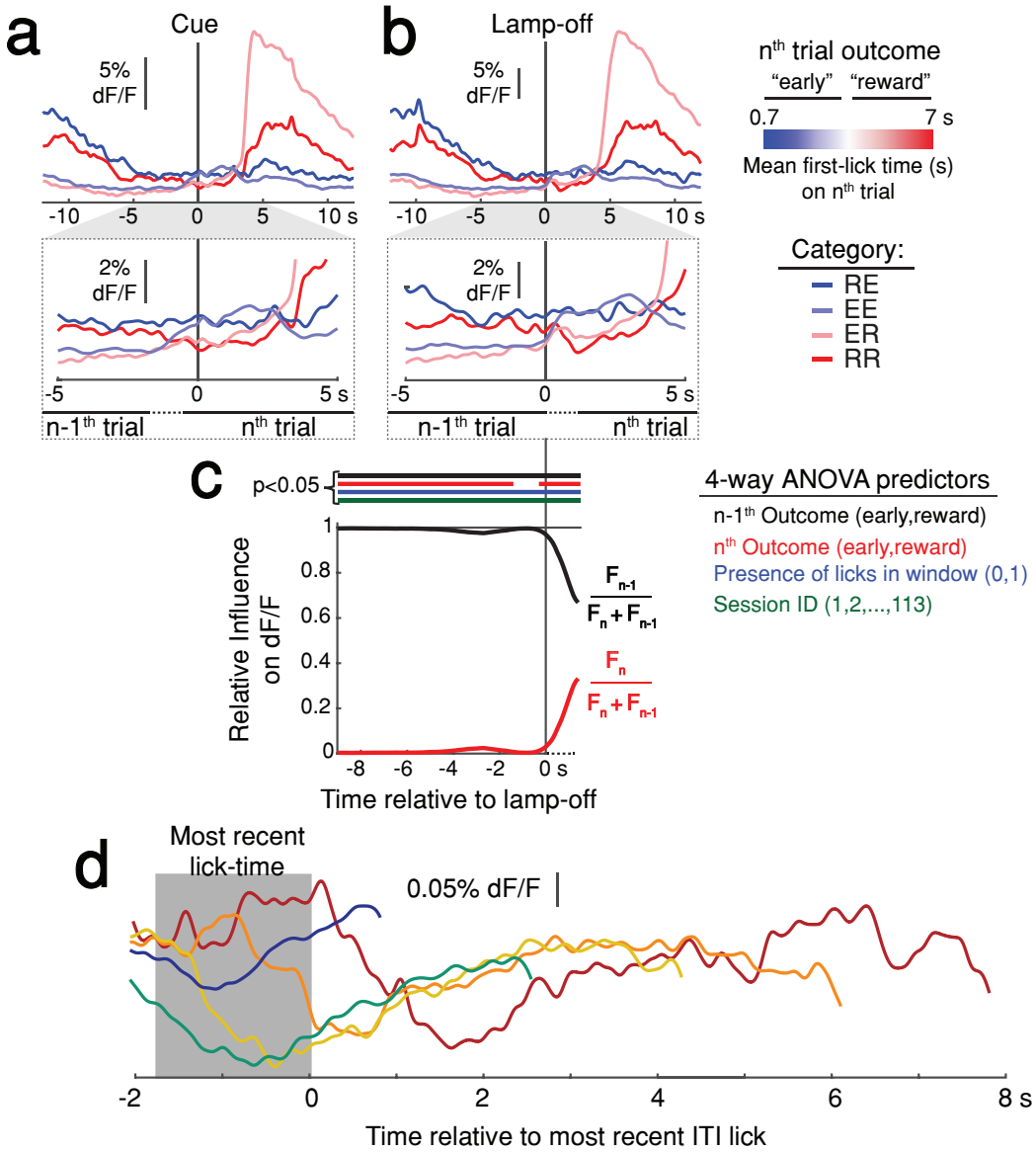
Extended Data Figure 3 | dF/F method validation. a, Minimal bleaching occurs over a 3-trial window. Left: slow, raw fluorescence bleaching across one session. Left inset: Comparison to bleaching across the first 3 trials (~1 min). Right: dF/F removes bleaching dynamics. Right inset: the same 3-trial window shown for dF/F-treated signal. **b,** Average raw fluorescence on paired, consecutive trials from one session aligned to cue on the nth trial. Left: n-1th trial was early, nth trial was rewarded ("ER" condition). Right: "RE" condition (See Supplementary Methods A). **c.** Comparison of baseline GCaMP6f signals on paired, consecutive trials aligned to cue. Columns: three different versions of the signal (Raw fluorescence, Normalized baseline dF/F method, Moving average dF/F method. Top row: ER condition; middle row: RE condition; bottom row: distortion index. Red distortion index plot shows only Normalized baseline method. Green distortion index plot shows overlay of Moving Average, Low-Pass Filter, and Multiple Baseline dF/F Methods because the difference in signal distortion between these methods was indistinguishable (See Supplementary Methods A).



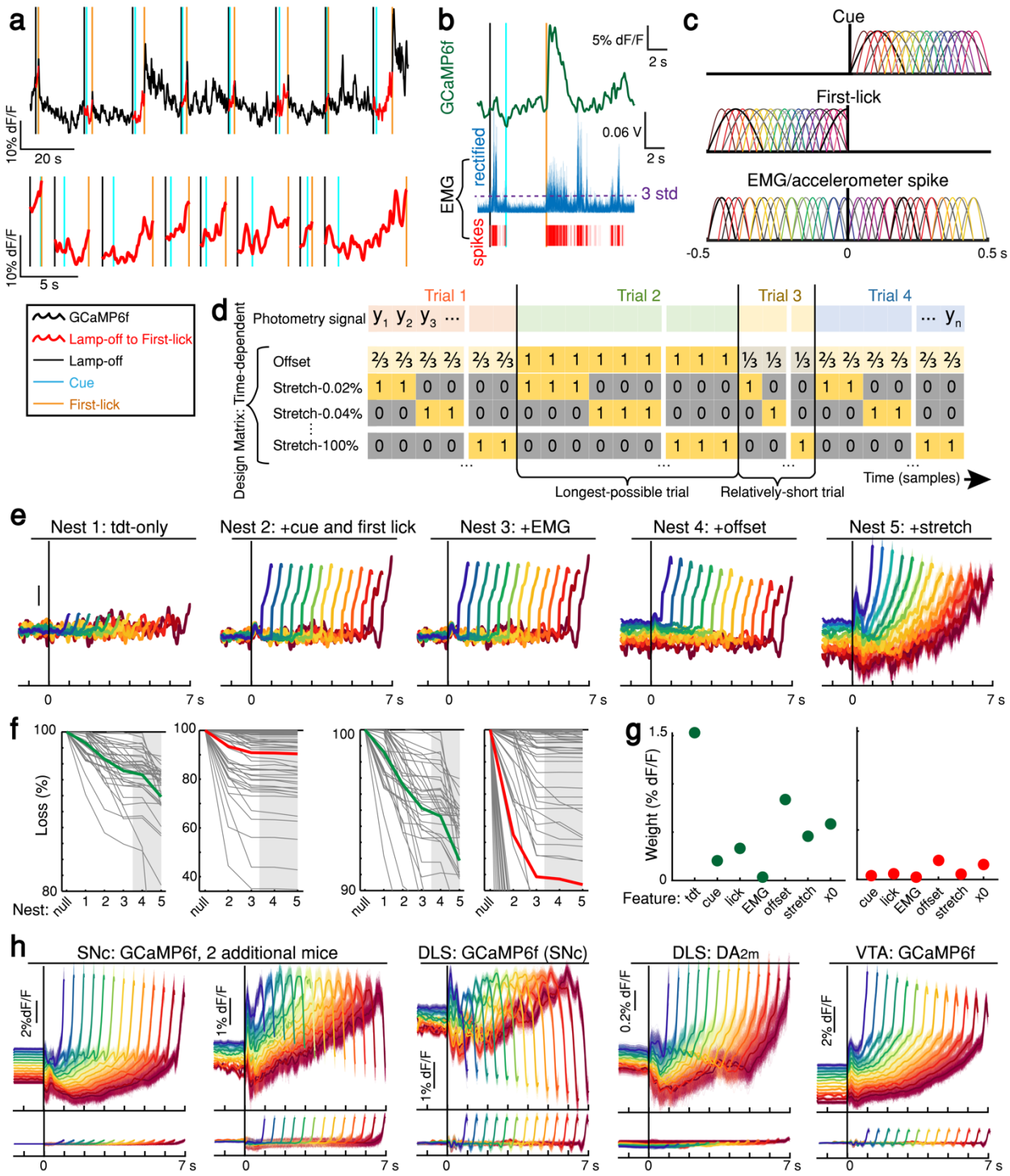
Extended Data Figure 4 | Cue-aligned average photometry signals showing reward-related responses: all mice, all fluorophores. a, DAN GCaMP6f signals at SNc cell bodies (12 mice). **b,** DAN GCaMP6f signals at axon terminals in DLS. The sharp, downward deflection immediately prior to movement onset was observed in every mouse (12/12) on every session and was not explained by movement artifacts. There appears to be a rapid "off" response. **c,** Striatal DA_{2m} signals at DLS (4 mice). **d,** DAN GCaMP6f signals at VTA cell bodies (4 mice). **e,** tdTomato signals (all sites, all sessions). **Insets**: Average signals for first-licks occurring between 3-3.25, aligned to cue (left of axis break) and aligned to first-lick (right of axis break). Traces plotted up till approximate movement onset (150 ms before first-lick).



Extended Data Figure 5 | Movement controls reliably detect movements, but there is no systematic difference in movement before first-lick during the timing interval. a, Schematic of movement controls. **b,** First-lick-aligned average movement signals on rewarded (red) and unrewarded (blue) trials. Pre-lick traces are truncated at the nearest cue-time for the averaged traces (dashed red, dashed blue). Left: one session; Right: all sessions. Dashed grey line: time of earliest-detected movement on most sessions (150 ms before first-lick). Average first-lick-locked tdt optical artifacts showed inconsistent directions even within the same session. Averages for all three types of artifact (consistently up, "Up"; consistently down, "Down"; and not consistent "NC") shown for all sessions. **c,** Breakdown of average tdt artifact direction by session at each recording site.

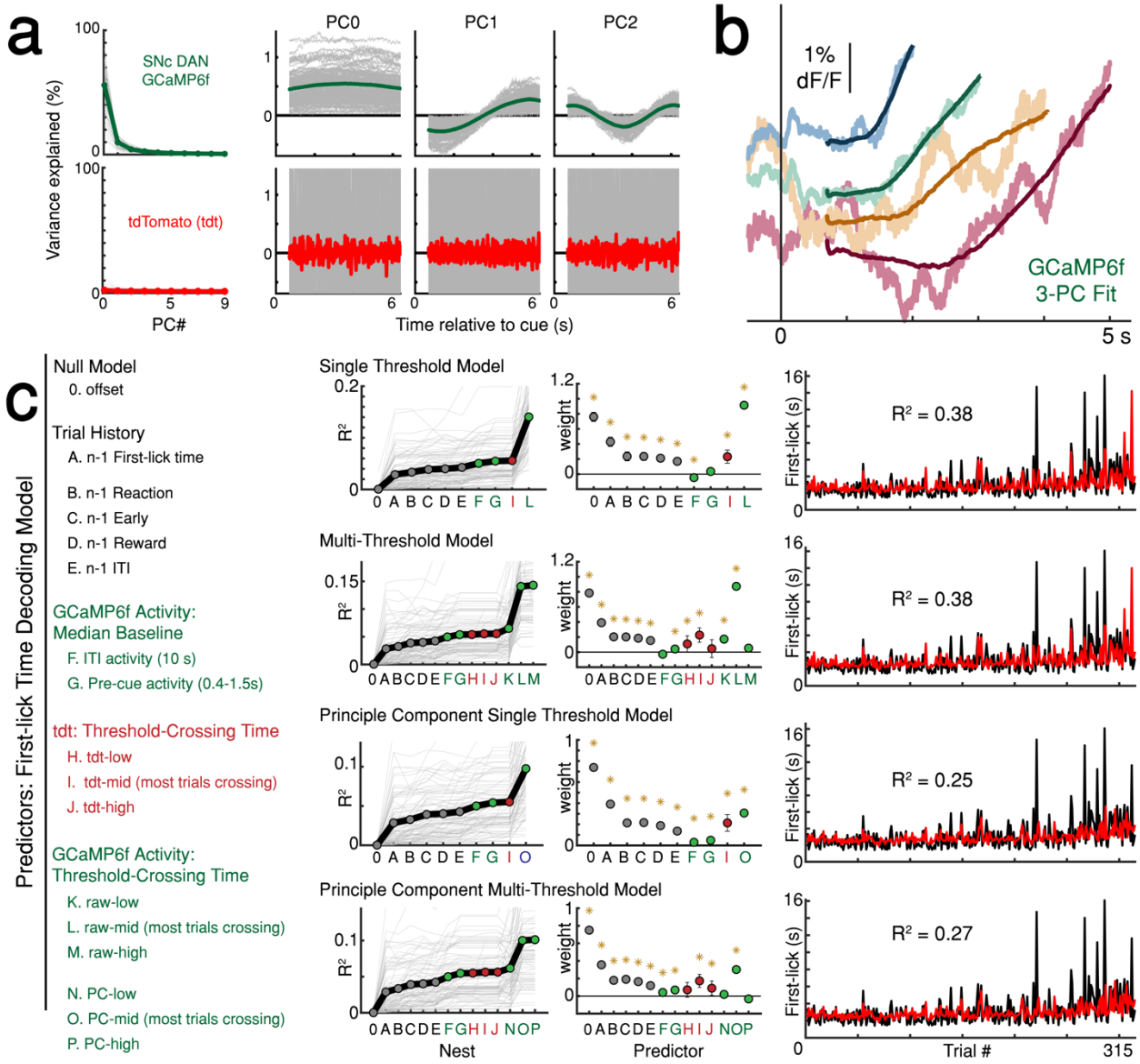


Extended Data Figure 6 | Baseline SNc DAN signals predict trial outcome, even when controlling for prior trial outcome and ongoing movement. Paired, consecutive trials were pooled into 4 categories based on the n-1th and nth trial outcomes (4 mice, the 17 sessions with highest signal to noise and number of trials). Categories: Early-Early (EE, where n-1th outcome is early, nth outcome is early: 2254 trials), Early-Reward (ER: 730 trials), Reward-Early (RE: 190 trials) and Reward-Reward (RR: 174 trials). To control for the contribution of movement to baseline signals, plots are shown only for trials with no licking during the last 5 seconds of the ITI before Lamp-off. **a,** Cue-aligned average DAN signals become more predictive of nth trial outcome as the cue time approaches. **b**, Lamp-off-aligned average DAN signals show "resetting" effect after the houselamp turns off. Before lamp-off, average DAN signals reflect the n-1th trial outcome; subsequently they reflect the nth trial outcome. **c**, Selectivity index taken on single trials quantifies the relative contribution of n-1th and nth trial outcomes to the prediction of the baseline signal. (Index calculated to exclude timepoints after the nth trial first-lick). **d,** Average ITI GCaMP6f signals aligned to most recent previous lick-time plotted up to onset of next spontaneous self-initiated lick during the ITI. (1 mouse, 5 sessions, truncated 150 ms before lick detection).

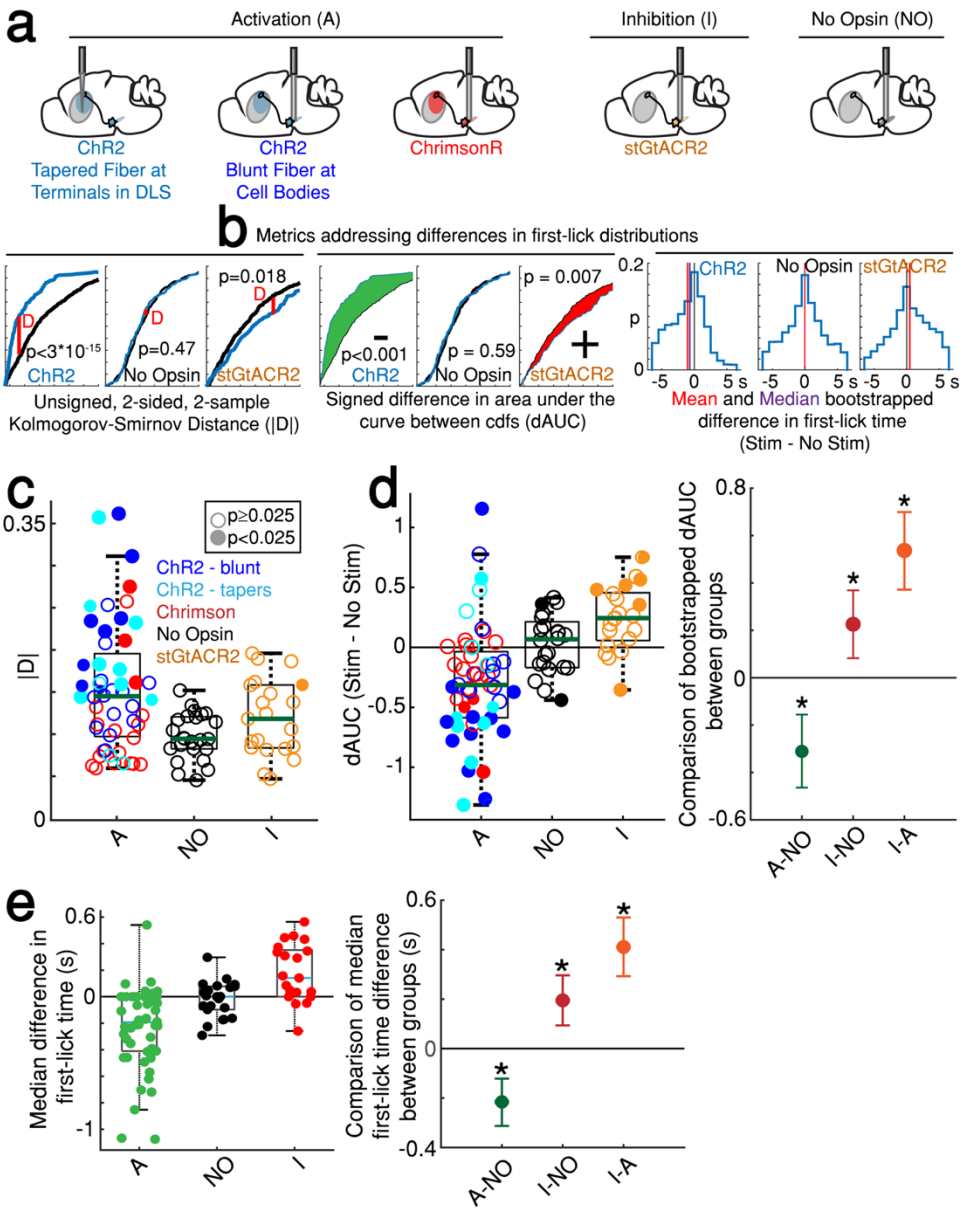


Extended Data Figure 7 | DAN GCaMP6f signal encoding model parameterization and model selection. a, Schematic of photometry timeseries fit by encoding model. The lamp-off to first-lick interval was isolated from each trial in a session (top) and concatenated to produce the timeseries fit by the model (bottom). **b,** Derivation of EMG spikes from raw signals. Thresholding of rectified EMG at 3 standard deviations (std) during an example trial. **c,** Optimized basis kernels for cue, first-lick, and EMG/accelerometer spikes to produce timing-independent features. **d,** Schematic of d *x* n Design Matrix for timing-dependent features. Note: timing-independent

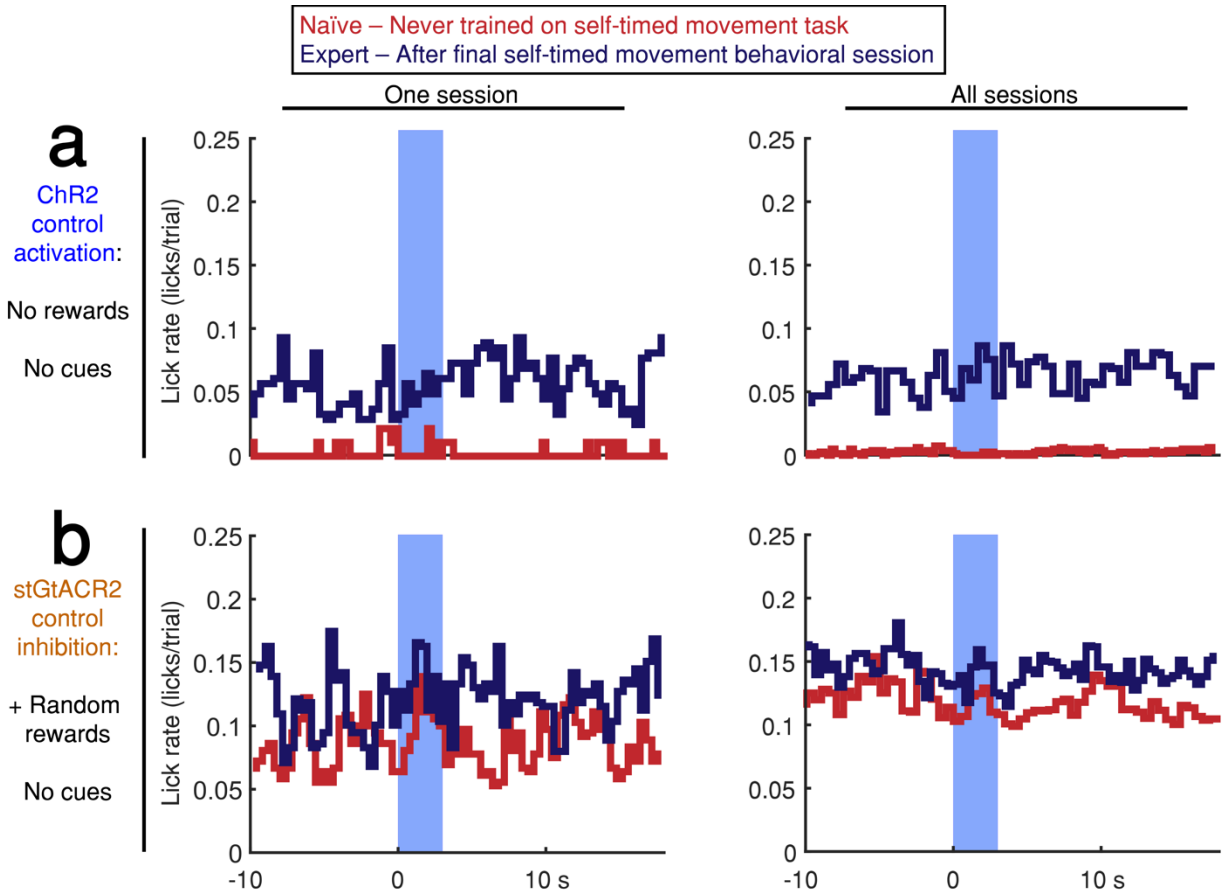
features not shown for clarity in schematic. **e,** GCaMP6f model fits by nest for example session. Model error simulated 300x (shading). **f**, Model loss by nest. Green: mean loss for SNc GCaMP6f; red: mean loss for tdTomato (tdt); grey lines: individual sessions; grey shading: timing-dependent nests. Left: full-scale view of all datasets. Right: mean GCaMP6f and tdt loss compared on same scale. **g,** Summary of feature weights across SNc GCaMP6f (left) and tdt (right) models (68 sessions each). Coefficient weights were rectified, summed, and divided by the number of predictors per feature. Error bars: 2*standard error (too small to see). All features were significant in both GCaMP6f and tdt models. **h,** Top: examples of the full timing-dependent model (nest 5) from additional mice for all recording conditions. Bottom: tdt control channel fit. Model errors simulated 300x. Some mice show downward-going movement-related spikes at SNc cell bodies (second panel). All mice showed downward-going movement-related spikes from SNc terminals in DLS (middle panel).



Extended Data Figure 8 | Variations of the first-lick time decoding model. a, Principle component analysis (PCA) of the ramping interval (0.7 s up to first-lick relative to cue). Green: mean GCaMP6f recorded at SNc; Red: mean tdTomato (tdt) recorded at SNc and VTA; Grey lines: single-session data. Left: Variance explained by first 10 principle components (PC). Right: Scores of first three principle components. X-axis shown for longest-possible interpolated trial duration; trials of shorter duration were interpolated to have the same number of samples for PCA. **b,** Example session data simulated with first 3 PCs. Light traces: actual averaged GCaMP6f activity truncated at first-lick onset; Dark traces: PC fits of the same trials. **c,** Decoding model variations. *: $p<0.05$, error bars: 95% confidence intervals. GCaMP6f threshold crossing time dominated every version of the model; n-1th trial first-lick time was consistently the second-best predictor.



Extended Data Fig. 9 | Variations on measurements of optogenetic effects. a, Strategy for optogenetic targeting of DANs. **b,** Comparison of four complementary metrics for addressing optogenetic effects. Left: unsigned Kolmogorov-Smirnov Distance (KS-D) assay of differences in first-lick time distribution. Center: signed, bootstrapped comparison of difference in area under the cdf curves (dAUC). Right: mean and median bootstrapped difference in first-lick time. **c,** KS-D Assay: all sessions. A: activation sessions; NO: no opsin sessions; I: inhibition sessions. Filled circles indicate significant difference between stimulated/unstimulated trials on single session ($p<0.025$, 2-sided, 2-sample KS test). **d,** Left: bootstrapped dAUC Assay: all sessions. Filled circles: significant difference on single session ($p<0.025$, 2-sided bootstrapped dAUC test, see Methods). Right: comparison of dAUC in first-lick distributions across all sessions between groups. Error bars denote bootstrapped 95% confidence interval (*: $p<0.05$). **e,** Left: median bootstrapped difference in first-lick time, stimulated-minus-unstimulated trials. Dots indicate single sessions. Comparison of median difference in first-lick time across all sessions. Error bars denote bootstrapped 95% confidence interval (*: $p < 0.05$).



Extended Data Figure 10 | Optogenetic DAN stimulation does not cause or prevent licking. **a,b** Stimulation-aligned lick-rate during stimulation-control sessions. Animals expressing ChR2 or stGtACR2 were tested in 1-3 control sessions both before exposure to the self-timed movement task (red) and in 1-2 control sessions after the end of behavioral training (red). Blue bar indicates stimulation period (3 s starting at time 0 s). Left: one control session, Right: all sessions. **a,** Activation control sessions (no cues or rewards). Animals were head-fixed on the behavioral platform in the absence of any cues or rewards and were stimulated randomly at the same pace as the standard 3.3 s self-timed movement task. Activation did not elicit immediate licking in any single session. **b,** Inhibition-control sessions (no cues, + random rewards). Animals were head-fixed on the behavioral platform in the absence of cues while receiving juice rewards at random times. Inhibition did not prevent licking in any single session.

Supplementary Information

1. Supplementary Methods

A. dF/F method characterization and validation

dF/F calculations are intended to reduce the contribution of slow fluorescence bleaching to fiber photometry signals, and many such methods have been described[1,20,26]. However, dF/F methods have the potential to introduce artifactual distortion when the wrong method is applied in the wrong setting. Thus, to derive an appropriate dF/F method for use in the context of the self-timed movement task, we characterized and quantified artifacts produced by 4 candidate dF/F techniques.

Detailed description of complementary dF/F methods.

1. Normalized baseline: a commonly used dF/F technique in which each trial's fluorescence is normalized to the mean fluorescence during the 5 s preceding the trial.
2. Low-pass digital filter: F_0 is the low-pass, digital infinite impulse response (IIR)-filtered raw fluorescence for the whole session (implemented in MATLAB with the built-in function *lowpass* with $f_c=5\cdot10^{-5}$ Hz, steepness=0.95).

3. Multiple baseline: a variation of Method 1, in which each trial's fluorescence is normalized by the mean fluorescence during the 5 s preceding the current trial, as well as 5 trials before the current trial and 5 trials after the current trial.
4. Moving average: F_0 is the 200 s moving average of the raw fluorescence at each point (100 s on either side of the measured timepoint).

Although *normalized baseline* (Method 1) is commonly used to correct raw fluorescence signals (F) for bleaching, this technique assumes that baseline activity has no bearing on the type of trial; however, because the mouse decides when to move in the self-timed movement task, it is possible that baseline activity may differ systematically with the mouse's choice on a given trial. Thus, normalizing F to the baseline period would obscure potentially physiologically-relevant signals. More insidiously, if baseline activity *does* vary systematically with the mouse's timing, normalization can also introduce substantial amplitude scaling and y-axis shifting artifacts when correcting F with this method (Extended Data Fig. 3c, middle panel). Thus, Methods 2-4 were designed and optimized to isolate photometry signals minimally distorted by bleaching signals and systematic baseline differences during the self-timed movement task. Methods 2-4 produced the same results in all statistical analyses, and the moving average method is shown in all figures.

Isolating minimally-distorted photometry signals with paired trial analyses of raw fluorescence. Although slow bleaching prevents comparison of raw photometry signals (F) at one time in a behavioral session with those at another time, the time-course of appreciable bleaching was slow enough in the reported behavioral sessions that minimal bleaching occurred over the course of 3 trials (~1 min, Extended Data Fig. 3a). Thus, F was comparable on sets of *paired* trials. To observe

the most minimally-distorted photometry signals possible, average F on paired trials was compared (Extended Data Fig. 3b,c). Because dF/F baseline DAN signals were systematically related to lick timing, we compared F baseline signals between all paired trials in which an Early (first-lick between 0.7-2.9s, unrewarded) trial was followed by a Late (first-lick between 3.4-7s, rewarded) trial ("ER" comparison). To ensure systematic differences did not result from subtle bleaching in the paired-trial interval, we reversed the ordering contingency and also compared all Late trials preceding Early trials ("RE comparison"). The same systematic relationship between baseline signals and first-lick time was found for paired trials analyzed by raw F (Extended Data Fig. 3c, left panel).

Quantification of artifactual amplitude scaling/baseline shifts introduced by dF/F processing. Each Candidate dF/F Method was applied to the same Paired Trial datasets described above. The resulting paired-fluorescence datasets were normalized after processing (minimum dF/F=0, maximum=1). The amount of distortion introduced by dF/F was quantified with a Distortion Index (DI), which was calculated as:

$$\text{Distortion Index, } DI(t) = abs(F(t)-dF/F(t))$$

where F(t) and dF/F(t) are the normalized, paired-trial raw fluorescence signal or dF/F signal at time *t*, respectively. *t* spanned from the beginning of the n-1th trial (-20 s) to the end of the nth trial (20s), aligned to the cue of the nth trial (Extended Data Fig. 3c, bottom panels). The DI shown in plots has been smoothed with a 200 ms moving average kernel for clarity.

As expected, normalizing fluorescence to the baseline period (*normalized baseline*) erased the correlation of baseline dF/F signals with first-lick time (Extended Data Fig. 3c, middle panels).

More insidiously, this also resulted in distortion of GCaMP6f dynamics *during* the timing interval, evident in the diminished difference between E-signals compared to R-signals relative to the shapes observed in the raw fluorescence paired-trial comparison (Extended Data Fig. 3c, middle-bottom panel). However, dF/F Methods 2-4 visually and quantitatively recapitulated the dynamics observed in the raw fluorescence comparison (Extended Data Fig. 3c, right panels).

These results were corroborated by time-in-session permutation tests in which datasets for single sessions were divided into thirds (beginning of session, middle of session, and end of session). The differences between baseline and ramping dynamics observed in whole-session averages were present even within these shorter blocks of time within the session (i.e., faster ramping and elevated baseline signals on trials with earlier self-timed licks). Furthermore, permutation tests in which the block identity (begin, middle, end) was shuffled showed that this pattern held when trials with earlier first-licks from the end of the session were compared with trials with later first-licks from the beginning of the session (and vice versa).

B. <u>Derivation of threshold and alternative decoding models</u>

Derivation of threshold models

As a metric of the predictive power of ramping DAN signals on first-lick time, we derived a threshold-crossing model. A threshold-crossing event was defined as the first time after the cue when the photometry signal exceeded and remained above a threshold level up until the time of first-lick on each trial. Importantly, while the analysis approach is reminiscent of pacemaker-accumulator models for timing, we make no claims that the analysis is evidence for pacemaker-

accumulator models. Rather threshold-crossing times provided a convenient metric to compare the rate of increase in signals between trials.

Photometry timeseries for GCaMP6f and tdt were de-noised by smoothing with a 100ms Gaussian kernel (kernel was optimized by grid screen of kernels ranging from 0, 30, 50, 80, 100, 150, 200 ms to minimize noise without signal distortion). To completely exclude the sensory- and motor-related transients locked to the cue and the first-lick events, the ramping interval was conservatively defined as 0.7 s post-cue up until 0.6 s before the first-lick. To eliminate chance crossings due to noise, we imposed a stiff, debounced threshold condition: to be considered a threshold crossing event, the photometry signal had to cross the threshold from low-to-high and remain above this level until the end of the ramping interval.

To derive an unbiased threshold for each session, we tested 100 evenly-spaced candidate threshold levels spanning the minimum-to-maximum photometry signal during the ramping interval for each session. Depending on threshold level, some trials never crossed, i.e., signal always remained below threshold or started and ended above threshold. Thus, the lowest candidate threshold for which there was a maximum number of trials crossing during the timing interval was selected as the “mid-level” threshold-crossing point. This threshold was specific to each photometry signal tested on each session. Threshold-crossing time was included in the decoding model as the normalized time on the ramping interval (0,1). If a trial never crossed threshold, it was encoded as a zero. If no trials ever crossed threshold, the threshold predictor was encoded as a vector of ones, thus penalizing the model for an additional predictor but providing no new information.

Multi-threshold Model

An alternative model employed 3 unbiased thresholds: 1) the lowest threshold with ≥50 trials crossing ("min"); 2) the lowest threshold with the most crossings ("mid," described above); and 3) the highest threshold with ≥50 trials crossing ("max"). For tdt datasets, trials rarely met the monotonic threshold constraint (usually the signals oscillated above and below the threshold throughout the ramping interval, failing to meet the debouncing constraint). Thus, to include tdt signals as conservatively as possible, we relaxed the 50-trial minimum constraint, taking the threshold with the most trials crossing, which was usually around 10 or fewer. The addition of more thresholds did not substantially improve the cross-validated model compared to the single-threshold model (Extended Data Fig. 8c).

Principle component analysis (PCA) threshold-crossing models

In another version of the decoding model, the threshold-crossing procedures were applied to ramping intervals fit with the first three PCs (as described in Methods) to derive a PCA version of the single-threshold and multi-threshold models. PCA analysis on tdt datasets showed no consistent PCs, and thus these PCs were not included in the decoding model. Instead, the actual tdt data was employed in the threshold model as in the other models described.

2. Supplementary Discussion

A unifying Reward-Prediction-Error based framework underlying dynamic DAN activity in timing tasks.

A framework that has explained many disparate experimental results from the dopaminergic system is temporal difference (TD) learning with reward-prediction errors (RPE)[27,33]. In this framework, DAN activity is thought to reflect the moment-to-moment difference in the animal's expectation versus its perception of the value of its current state, where value is defined as the temporally-discounted expectation of total future reward. In classical trace-conditioning paradigms, DANs fire in transient bursts to unexpected rewards and reward-predicting cues, whereas they pause their firing when expected reward is omitted. Indeed, we observed RPE-like signals in the cue-related transient, dips in activity after unrewarded first-licks, and surges in activity following rewarded first-licks (Fig. 1c-e, Extended Data Fig. 4a-d). Persistence of RPE-like signals in well-trained animals has been suggested to arise from the inherent imprecision in neural timing[2], which may reflect the animal's moment-to-moment uncertainty of its current state—i.e., its position in time–and, by extension to our task, uncertainty about its accuracy for a given self-timed lick[30]. Indeed, positive-going RPE-like signals were strongest for first-licks closest to the reward-boundary, presumably when the mouse's "confidence" of reward was lowest, consistent with the greatest RPE occurring when the mice were least certain of success (Extended Data Fig. 4a-d).

Whereas RPE-frameworks have explained *transient* bursts and pauses in DAN activity during trace conditioning and other types of learning experiments, DAN activity can also change more

slowly[30]. For example, "ramping" signals build up over seconds during goal-directed navigation[19], bandit tasks in which animals must complete multiple goals to receive reward[25,26], and tasks with visual cues of proximity to reward[20]. It has been suggested that DANs could signal different information via slow changes in activity (e.g., motivation, ongoing value, vigor) compared to fast-timescale activity (e.g., post-hoc RPE signals for learning), and a number of proposals have suggested that DANs multiplex different kinds of information over different timescales and contexts[21,26].

However, recent models have proposed RPE-based explanations that may be able to reconcile these seemingly disparate dopamine signals[20,27,30]. While these models do not refute the possibility that DANs could encode other types of information (e.g., value, vigor, etc.), they are attractive for their parsimonious explanation of how fast time-scale phenomena and slowly-evolving ramps could arise from the same underlying RPE-based calculation. In short, these models employ principles from TD learning to show how certain shapes of the value function (i.e., the assignment of values to the series of behavioral states comprising a task) can give rise to a *continuously changing* RPE, even in well-trained animals[20,27,30,34].

We were interested in whether an RPE-based framework could explain the results found in our self-timed movement task as well as results from other timing tasks[1]. To approach this question, we applied a key feature of TD learning algorithms to determine what an RPE-like signal would look like in different kinds of timing tasks. Specifically, we took advantage of the fact that *RPE is proportional to the derivative of the subjective value function under conditions of state*

uncertainty[27,30], as is the case during timing tasks in which the animal must rely on its own internal representation of time to guide behavior[27].

Thus, if the value landscape for a given behavioral task is known, and if DAN activity encodes RPE, the RPE-based framework makes predictions about the expected shape of dynamic DAN activity during the task. In a recent study, similar applications of this principle predicted the ramping DAN signals that were observed in virtual reality (VR) tasks in which animals were moved passively through VR spaces, as well as when the animals passively viewed abstract, dynamic visual cues indicating proximity to reward[20], suggesting the ramping in our task could be explained from similar principles.

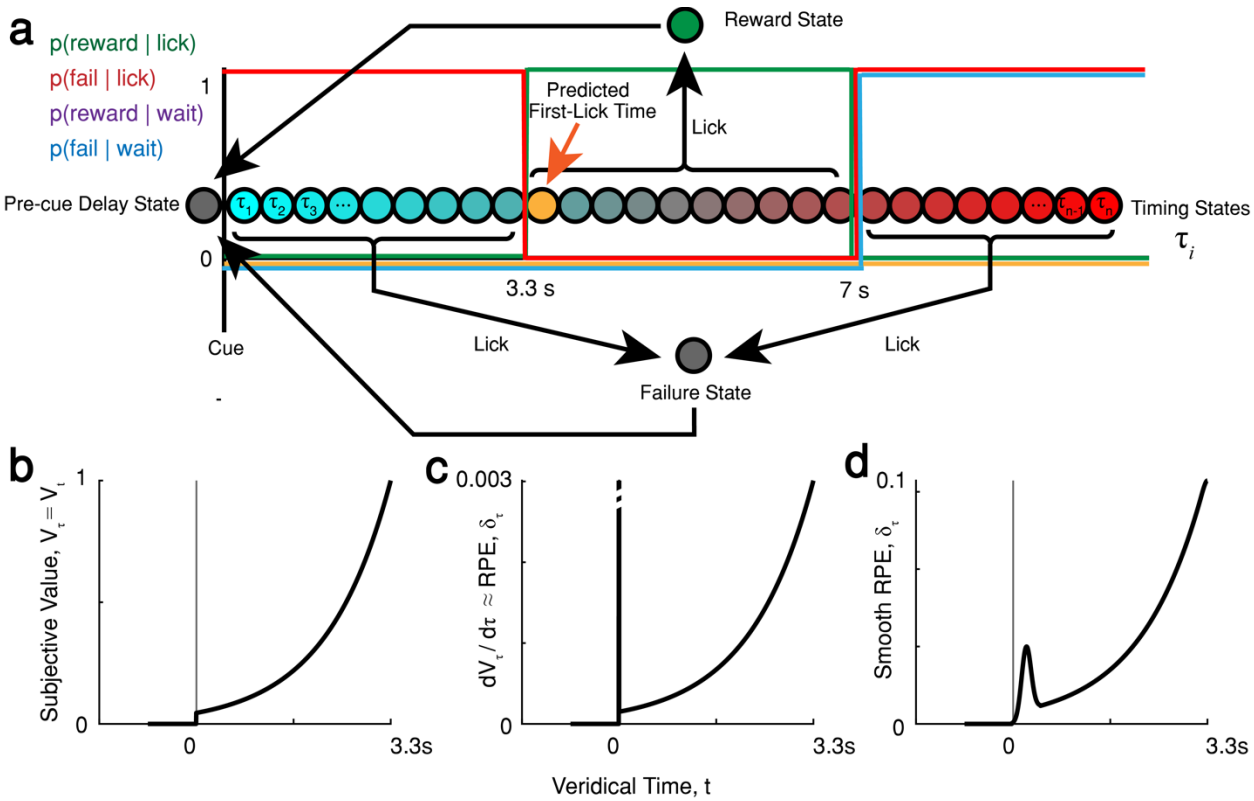
RPE-predictions for DAN responses during self-timed movement.

In a simple TD learning model of self-timed movement, time may be modeled as a continuous set of states through which a Markov agent must traverse to receive reward[35] (Supplementary Fig. 1a). At each state transition (timestep), the agent must decide whether to move (lick) or to wait based on the probability of transitioning to a reward or failure state. If the agent is an optimal timer, its subjective approximation of its current state, τ, accurately tracks veridical time, t, and it will thus withhold movement until the first moment at which reward will be available in response to licking (3.3 s in our experiment).

The value landscape of this model can be understood intuitively. When the cue event occurs, a well-trained agent can expect an increased possibility of reward in the next few seconds; thus, at this moment, value increases. However, reward never occurs within the first 3.3 s of the standard

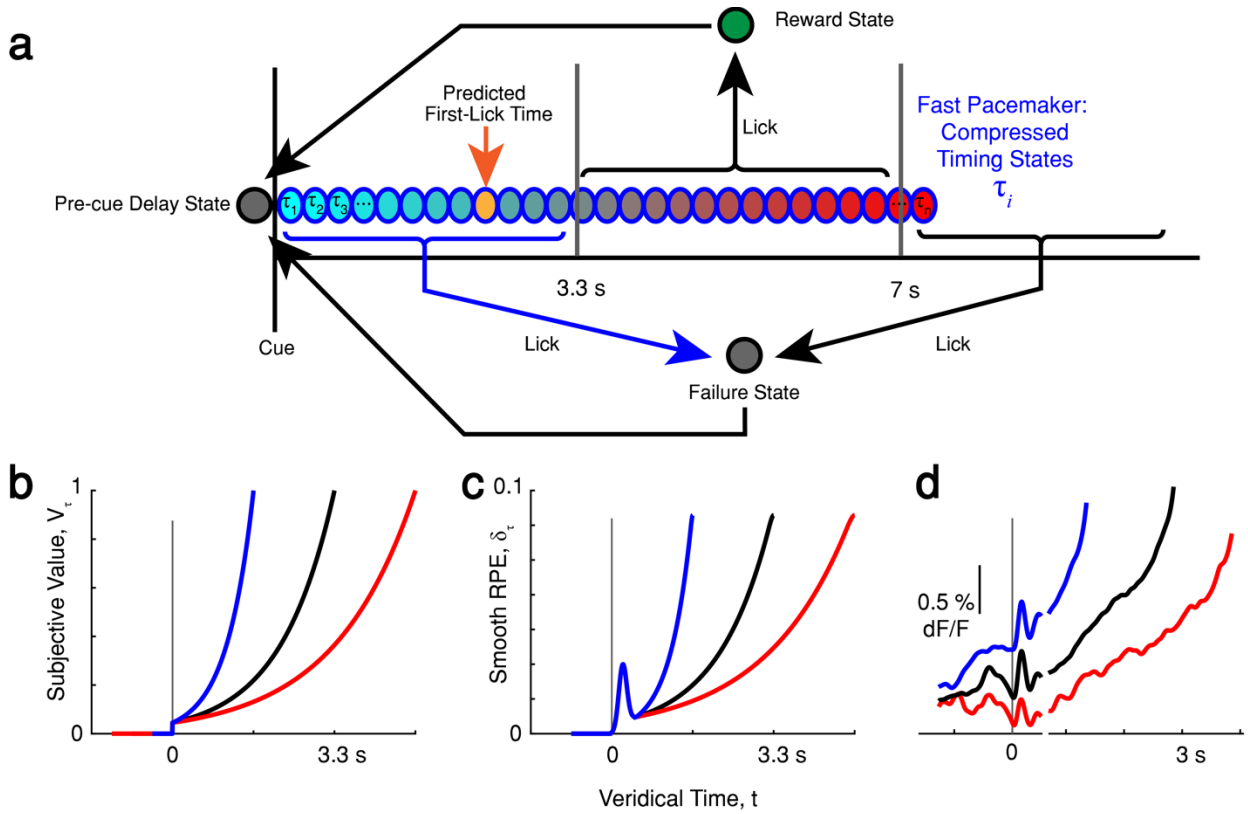
timing task we implemented; thus, value at the cue is necessarily lower than value at 3.3 s. In fact, value will constantly increase as time approaches 3.3 s. Thus, as long as the agent withholds licks, the value landscape, V_t, during the first few seconds is a monotonically increasing, convex function[36] (Supplementary Fig. 1b). If the agent is an optimal timer, the subjective approximation of the value function, $\hat{V}_\tau$, matches the true value function, and $\hat{V}_\tau = V_t$.

However, we assume that, because the timer does not have access to the true state identity, t, it is never certain of its subjective approximation of its state, τ. Under conditions of state uncertainty, RPE is approximately the derivative of the subjective value function[20,27], $\delta_\tau \approx \hat{V}'_\tau$, where δ_τ is RPE at subjective time τ, and $\hat{V}'_\tau$ is the time-derivative of the subjective value function. Thus, the shape of the RPE function, δ_τ is also quite simple: a transient increase at the cue followed by a slowly-evolving ramp (Supplementary Fig. 1c). If the RPE function is measured by a calcium indicator such as GCaMP6f, the binding kinetics of the indicator would tend to blur the RPE function, which we approximated by smoothing (Supplementary Fig. 1d).



Supplementary Figure 1 | Value and RPE Landscapes for an optimal timer predict DAN responses during the self-timed movement task. a, State space and probability of state transition for an optimal timer. Gold-shaded state is the first state from which reward is available, and thus is when the first-lick is predicted to occur. **b,** Estimated value function $\hat{V}_t$, where $\hat{V}_\tau \approx V_t$ for an optimal timer. An exponential value landscape is shown, consistent with prior literature[27]. However, any sufficiently convex function could be implemented with the same result[27,30]. The agent is expected to first-lick at the peak of the trajectory. **c,** RPE function for an optimal timer, estimated as $\delta_\tau \approx \hat{V}'_\tau$, the derivative of the subjective value function. Y-axis scaled to show ramp. **d,** Predicted DAN GCaMP6f signals for an optimal timer. The RPE function was smoothed with a gaussian kernel spanning *ca.* 10% of the interval to approximate GCaMP6f off-dynamics.

The modeled RPE function mirrors the shape of the dynamics observed in DAN signals: a cue-related transient followed by a slow ramp up to the time of first-lick. However, unlike the optimal timer in this model, mice, like humans, exhibit suboptimal timing behavior with variability proportional to the duration of the timed interval[2]. It has been proposed that this variability in timing results from imprecision in an internal clock, referred to classically as the internal "pacemaker[37]." When the pacemaker is fast, self-timed movements occur relatively early, whereas when the pacemaker is slow, later movements occur. These changes in the pacemaker rate would correspond to the mouse traversing the set of subjective states, τ, at different rates than the passage of veridical time, t (Supplementary Fig. 2a), resulting in relative *compression* and *stretching*, respectively, in the subjective value function, $\hat{V}_\tau$ (Supplementary Fig. 2b), with corresponding compression/stretching of the RPE function (Supplementary Fig. 2c).



Supplementary Figure 2 | Compressed and stretched Value and RPE Landscapes for a sub-optimal timer predict dynamic DAN responses during the self-timed movement task, but do not capture baseline offsets. a, Simple state space of self-timed movement task for a suboptimal timer with a fast pacemaker. The fast pacemaker "compresses" state space[30,35], resulting in traversal of the timing states faster than veridical time. The mouse can only make a decision based on which state it believes itself in; thus first-lick is expected to occur early (gold-shaded state). **b,** A compressed subjective value function ($\hat{V}_\tau$, blue) reflects relatively fast traversal through the value landscape compared with that of veridical time (V_t, black). Conversely, stretched $\hat{V}_\tau$ (red) reflects slow traversal, consistent with a slow pacemaker. The animal is expected to lick at the peak of the trajectory. **c,** Smoothed estimated RPE function ($\hat{V}'_\tau \approx \delta_\tau$). Compression/stretching of the value function produces ramping dynamics similar to those observed in DANs (**d**) and striatal dopamine (Fig. 1f). However, this model alone

does not explain the more tonic baseline offsets that were anti-correlated with upcoming movement time.

Strikingly, as this simple RPE-based model predicts, DAN signals observed during our self-timed movement task show different ramping dynamics depending on when the animal actually moved (Supplementary Fig. 2d), consistent with compression/stretching of the subjective value and RPE functions. When the animal moved relatively early (perhaps corresponding to a fast pacemaker), DAN ramping unfolded with a steeper slope, as if the ramping period were *compressed*. Conversely, when the animal moved late (perhaps corresponding to a slow pacemaker), DAN ramping unfolded with a shallower slope, as if the ramping interval were *stretched*. The idea of compression/stretching of DAN ramps was supported by our encoding model (Fig. 2, Extended Data Fig. 7), for which we needed to add a timing-dependent "stretch factor" to best capture the variance in GCaMP6f signals during the timed interval. Together, these observations could be explained by DANs encoding an RPE-like signal related to the animal's "belief" of its position in objective time, τ, as derived from in its position along the subjective value trajectory during the timing interval of the task.

In fact, a recent model described how a timing mechanism instantiated by the nigrostriatal system could lead to (the well-known) variability in self-timed intervals by stretching or compressing of subjective value trajectories[30]. The model posits that dopamine modulates the pacemaker rate (consistent with pharmacological and lesion studies), with increased dopamine availability (or efficacy) speeding the pacemaker, and decreased dopamine slowing the pacemaker[4-5,8-11]. In turn, the pacemaker controls the encoding of subjective time, and thus the steepness of the value

function with respect to objective, veridical time. It follows that variation in dopamine availability would compress or stretch the value landscape to varying degrees from trial-to-trial. This model is consistent with our findings of variable ramping slope in DANs signals from trial-to-trial. It is also consistent with neural recordings from striatal spiny projection neurons and parietal cortical neurons during similar self-timed movement tasks, for which temporal sequences of striatal and cortical firing during timing were compressed for early movements and stretched for late movements[6,24].

While the RPE-based view of DAN activity captures the dynamic DAN signals we observed, our simple RPE model alone does not capture the *baseline offsets* in DAN signals that were predictive of movement timing even after controlling for previous trial outcome and ongoing nuisance movements (Fig. 3, Extended Data Fig. 6c). More complex RPE-based explanations for these *tonic* offsets in DAN signals could be imagined with further assumptions (e.g., states like the pre-cue delay could also contain timing states that create offsets before the trial begins, etc.), but a parsimonious explanation for how and why these offsets emerge requires further investigation. Mohebi *et al*. recently showed baseline differences in the amount of dopamine in the nucleus accumbens core that were correlated with the recent history of reward rate: higher recent reward rates were related to higher tonic dopamine. However, in our task, animals tended to move later toward the end of sessions, resulting in periods of relatively high reward rate when the average tonic baseline signal was *lower* (baseline preceding rewarded trials—by definition, later movements—was systematically lower in our task, Fig. 1d-f), suggesting a more complex relationship between tonic DAN activity and reward rate in our task. While the origin of offsets in DAN signals remains unclear, these offsets were nonetheless inversely related to the first-lick time,

and thus directly related to the (inferred) pacemaker rate, consistent with pharmacological and lesion studies positing a positive correlation between dopamine availability and pacemaker rate[4-5,8-11,30].

Ramping signals in our photometry experiments were measured from a population of DANs. An important future question is whether ramps are also present at the level of individual neurons, or rather represent a progressive recruitment of individual neurons, or some combination of both. Prior studies have reported ramping signals in individual neurons during tasks with visual feedback of distance to reward[20], whereas others have observed decoupling between DAN firing rates and downstream DA release[26], making it unclear whether electrophysiology would be capable of addressing this question. Observation of individual neurons expressing calcium indicators with GRIN-lens equipped endoscopes may be better suited to this question.

RPE-based predictions for DAN responses during a temporal bisection task.

Whereas DAN signals during our self-timed movement task were consistent with classic observations of the influence of dopamine on the speed of the pacemaker, a recent study employing a different timing task found more complex DAN dynamics during timing. Soares *et al*. recorded SNc DAN GCaMP6f signals with fiber photometry as mice executed a classic temporal bisection perceptual task[1] (Supplementary Fig. 3a). Trials began when mice entered a nose-poke port and received an auditory start-timing cue. Mice had to remain in the port throughout a variable timing interval, which was terminated with a stop-timing auditory cue. Mice then reported whether the interval was shorter or longer than a criterion time (1.5 s) by choosing a left or right nose-poke port corresponding to a “long” or “short” judgment. Mice were trained to categorize intervals

spanning 0.6-2.4 s. As expected, trials with more extreme intervals were easier for the mice, whereas trials with intervals closer to the 1.5 s criterion time elicited chance performance (Supplementary Fig. 3b).

Figure to be modified from Soares *et al*, 2016, Figure 2. Awaiting reprint permissions.

Supplementary Figure 3 | A temporal bisection task shows relatively high DAN signals during the timing interval when the inferred pacemaker rate is relatively fast. Figures adapted from Soares *et al.,* 2016[1]. **a,** Task schematic. **b,** Psychometric curve for timing intervals of different duration. Criterion time: 1.5 s. **c,** Start-timing cue-aligned average SNc DAN GCaMP6f signals. Second peak occurs just after the stop-timing cue (intervals: 0.6, 1.05, 1.26, 1.74, 1.95, 2.4 s). Figure recolored to indicate average inferred pacemaker rate. Red: slow; blue: fast. Relative dF/F amplitude during baseline and immediately prior to stop-timing cue shown left and right. dF/F amplitudes during timing are higher when the inferred pacemaker rate is fast. Left: Correct trials. Right: Incorrect trials show the same dF/F relationship with pacemaker rate.

DANs exhibited complex dynamics during the bisection task, starting with a sharp transient after the start-timing cue and ending with second transient after the stop-timing cue (Supplementary Fig. 3c). Between the start-timing and stop-timing cues, DAN signals exhibited a U-shape with increasing time, which was visible for trials with longer intervals but was truncated prematurely for the shorter intervals. The authors focused their analyses on the transient occurring *after the stop-timing* cue. Short judgments (suggesting a slow pacemaker) were accompanied by relatively high-amplitude transients after the stop-cue, whereas long judgments (suggesting a fast pacemaker) showed relatively low-amplitude transients. These results seemed to suggest that relatively *high* DAN activity reflected a *slow* pacemaker, the opposite of what is expected based on the bulk of pharmacological and lesion studies[30], as well as the trend we observed during our self-timed movement task.

This surprising finding could be a unique feature of the bisection task. Unlike self-timed movements, in which animals directly report elapsed time with a movement, the temporal bisection task requires an additional computational step, in which the timed interval must be categorized as "long" or "short." However, prior pharmacological studies employing the bisection task found results consistent with the classic view that higher dopamine availability is associated with a faster pacemaker[30,38]—opposite the interpretation of Soares *et al*., but consistent with the findings of our self-timed movement task.

The discrepancy between our results and those found by Soares *et al*. could perhaps be traced to differences in the way DAN signals were analyzed. We focused our attention on DAN signals

unfolding *during timing* in our self-timed movement task, whereas these signals were not explored by Soares *et al*. We thus asked two questions: 1. What correlations exist between DAN signals and pacemaker rate in the bisection task *before* the timing interval? And, 2. What correlations exist *during* the timing interval itself?

Before addressing these questions, we note that the relationship between pacemaker and bisection judgment is not as straightforward as in self-timed movement, and thus we recolored Supplementary Fig. 3c to clarify this, employing the following intuition: For a trial to be correct in the bisection task, on average, the pacemaker must be either accurate or "conservatively inaccurate." In other words, a correct "short" judgment requires either accurate timing or a *slow* pacemaker (Supplementary Fig. 3c, red curves). Conversely, a correct "long" judgment requires either accurate timing or a *fast* pacemaker (Supplementary Fig. 3c, blue curves).

When we considered DAN signals *before* the timing interval for correct trials in the Soares *et al*. study (Supplementary Fig. 3c, left), we noticed what appears to be two strata of signal levels. Trials with "long" judgments (fast pacemaker on average) had relatively high baseline signals, whereas trials with "short" judgments (slow pacemaker on average) had lower baseline signals, consistent with the relationship between baseline offsets and pacemaker rate that we observed in our self-timed movement task. As in our task, these baseline offsets remained present during the timing interval, resulting in the same stratification of dF/F signals immediately prior to the stop-timing cue (except for the very earliest time, 0.6s, which overlaps decaying GCaMP6f signals related to the start-timing cue, likely causing an artifactual inflation of the signal just prior to the stop-cue due to the off-kinetics of the calcium indicator or kinetics of calcium clearance more generally).

Thus, it generally appears that DAN activity was *higher* on trials with fast pacemaker rates, both during and before the interval in which the animal was actually timing. Intriguingly, *incorrect* trials (to the right in Supplementary Fig. 3c) showed a relative convergence of the baseline signals preceding the start-cue, but then signals diverged during the timing interval, resulting in relatively *high* signals at the time of the stop-cue for incorrect "long" choices (i.e., a fast pacemaker), but relatively low signals at the time of the stop-cue for incorrect "short" choices (i.e., a slow pacemaker). This is consistent with the patterns observed on correct trials. Interpreted thusly, the Soares *et al*. result is consistent both with our results and with classic pharmacological studies relating higher/lower dopamine availability to faster/slower pacemaker rates, respectively. Soares *et al*. presented their subsequent analyses with these baseline differences normalized-out in some way (Fig. 3 of Soares *et al*.). It is possible that this "zeroing out" of the baseline offset may have hindered efforts to detect consistent effects during the timing interval due to reordering of the traces.

Because baseline offsets in the bisection task appear similar to those in our self-timed movement task, we asked whether dynamic DAN signals in the bisection task could similarly be explained by the task's RPE landscape. In their investigation of the stop-timing cue-related transient, Soares *et al*. showed that its amplitude is well-explained by a combination of temporal surprise and behavioral performance, and we applied these parameters to derive a value landscape consistent with their bisection task.

The inferred value landscape of the bisection task for an optimal agent was built from a few assumptions (Supplementary Fig. 4a):

1. As in our self-timed movement task, value increases immediately at the start-cue and continues to rise toward the time of expected potential reward delivery.

2. Because the longest interval is 2.4 s, the time until potential reward is known to be no more than ~3 s (including the time to report judgment). However, due to temporal uncertainty and the fact that a false start (leaving the port before the stop-timing cue) results in an error and loss of reward, there is a second jump in the value function at the time of the stop-cue when the feedback of the tone reorients the value function and indicates the opportunity to collect reward within a few hundred milliseconds.

3. Because value is temporally discounted at the start-cue by the possibility of the longest-possible interval, any stop-cue occurring before 2.4 s results in a sudden "teleportation" through the value landscape to the final limb of the task that occurs just before the judgment and ascertainment of trial outcome, similar to the jump in the value function in a recently-reported, virtual reality, spatial teleportation task[20]. Thus, assuming the value function trends upwards steadily, the amplitude of RPE-related transients following the stop-cue would *decrease* as the interval duration increases, because the sudden jump in the value function becomes progressively smaller.

4. To capture aspects related to behavioral performance, we additionally included contours in the value function during the timing interval to reflect the probability of a correct choice for intervals of different lengths. Specifically, a relative minimum in the value function

occurs near 1.5 s, when predicted performance is worst. However, a stop-timing tone near the criterion time also results in a smaller jump in the value function because the probability of a correct decision is also lower. Thus, the increase in value at the moment of decision was adjusted by the probability of a correct choice.

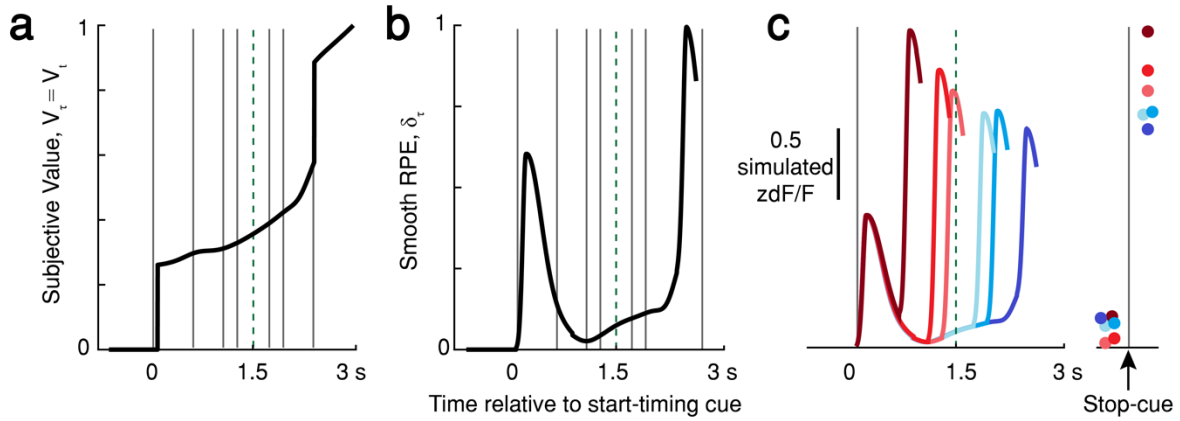
5. As in the simple RPE-model of our self-timed movement task, we modeled changes in pacemaker rate as compression/stretching of the subjective value landscape with respect to veridical time.

6. The agent traverses timing states during the timing interval, similar to the timing states in the self-timed movement task, but unlike our task, the bisection task does not require the agent to decide when to move. We assume the need to make a timed movement imposes a need for the agent to be relatively certain of its subjective timing state, τ, to make a decision, even though it is uncertain of its true state, t. The bisection task, on the other hand, is more similar to classical conditioning tasks in which the timing interval is not in the agent's control, and thus subjective state uncertainty increases with the distance from the last state-informative cue[30]. Thus, we took into account temporal blurring of the subjective state function, which would tend to reduce the convexity of the subjective value function and reduce the amplitude of ramping during the timing interval[30]. However, adding temporal blurring does not substantially change the fit-shape in our simplified model, and versions with or without blurring can reproduce the shape of the dynamic DAN signals.

Together, we arrived at a model of the RPE landscape for each of the six tested interval durations (Supplementary Fig. 4b,c). Importantly, this simple RPE-based model accurately captures the relative categorical amplitudes of the stop-timing cue-related transients, as follows: If the instantaneous DAN activity at the time of the stop-timing cue is relatively high, this would indicate that the animal is further along in the subjective value trajectory, resulting in 1) a *long* judgment, and 2) a relatively *smaller* RPE transient, because the underlying subjective value was higher at that moment. Conversely, if instantaneous DAN activity is relatively low at the stop-timing cue, this this would indicate that the animal is not very far along the subjective value trajectory, leading to 1) a *short* judgement and 2) a relatively *larger* stop-cue-related RPE transient, because the underlying subjective value was relatively low just before the stop-cue.

Now consider a particular (objective) time interval near the criterion time, for which the animal makes a mix of "long" and "short" choices (e.g., 1.74s; Supplementary Fig, 3b). Soares *et al*. found that the amplitude of the stop-timing cue-related GCaMP6f transient tended to be bigger when the animal made short choices, and this was taken as evidence that elevated DAN activity *slows* the internal clock. However, our model predicts that the size of the stop-cue-related transient will be inversely related to the amplitude of the underlying subjective value at that point, and thus inversely related to elapsed *subjective* time. It thus follows that if subjective time is more advanced on a given trial (i.e., faster pacemaker), the animal would tend to choose the long judgment on that trial, and the stop-timing RPE transient would be *smaller*. Conversely, if subjective time is less advanced on a trial (i.e., slower pacemaker), the animal would tend to choose the short judgment, and the stop-timing RPE transient would be *larger*.

Our RPE model accurately predicts the results of Soares *et al*.; however, our model holds that elevated DAN activity *speeds* the internal clock, consistent with most pharmacological studies but *opposite* the interpretation of Soares *et al*. Thus, our RPE-based model suggests a parsimonious explanation for DAN activity in both the self-timed movement and temporal bisection paradigms, with (1) relatively high DAN activity corresponding to a fast pacemaker; manifesting in (2) compression of the value landscape; thereby leading to (3) early movements (in the self-timed movement task) or long judgments (in the temporal bisection task).



Supplementary Figure 4 | Subjective Value and RPE Landscapes for the temporal bisection task predict dynamic DAN responses during the temporal bisection task, but do not capture baseline offsets. a, Estimated value function $\hat{V}_t$, where $\hat{V}_\tau \approx V_t$ for an optimal timer on a 2.4 s trial. Grey lines: test interval times. Green dashed line: criterion time (1.5 s). Value increases approaching the time when reward is available, increasing abruptly at the start- and stop-timing cues (0 and 2.4 s). **b,** Smoothed RPE function for an optimal timer, estimated as $\delta_\tau \approx \hat{V}'_\tau$, the derivative of the subjective value function. The RPE function was smoothed with an asymmetrical gaussian kernel spanning *ca*. 28% of the interval to approximate GCaMP6f off-dynamics. **c,** Predicted DAN GCaMP6f signals for an

optimal timer for the six test interval times. Traces truncated before reward collection for clarity. Right: relative simulated dF/F amplitude prior to the stop-timing cue and subsequent peak response. Amplitude before the stop-timing cue is directly proportional to clock speed; amplitude at the stop-timing cue-related peak is inversely proportional to clock speed.

Limitations of the RPE-based model.

The simple RPE-based models presented here explain dynamic DAN signals in both the bisection task and our self-timed movement task, but they do not explain the origin of baseline offsets. Mohebi *et al.*[26] recently-proposed that baseline offsets in ventral striatal dopamine levels could reflect the average recent reward rate, but we found that offset amplitude in DAN signals is at least partially independent of recent trial history during the self-timed movement task. It is possible that baseline variation arises from slow, random fluctuations in DAN activity, but further work is needed to explore the origins of these signals.

A second issue is the impact of optogenetic DAN activation and suppression on the rate of the pacemaker. In our self-timed movement task, DAN activation promoted early movements, consistent with increasing the pacemaker rate, whereas suppression promoted late movements, consistent with slowing the pacemaker rate (Fig. 4). However, Soares *et al.* reported an opposite effect for optogenetic manipulation during the bisection task, at least for DAN activation.

This difference between the tasks could be reconciled by a recent theoretical model proposed by Mikhael and Gershman to explain the behavior of the pacemaker in a wide range of classical conditioning and timing studies[30]. Their model shows that the pacemaker rate is expected to be

updated at the time of reinforcement by a Hebbian-like, bidirectional learning rule. If reward occurs exactly at the expected time, there is no update in the pacemaker rate. However, if reinforcement occurs before the expected time, this is interpreted as feedback that the pacemaker was running too slowly; thus, the update rule increases the pacemaker rate leading to expectation of reward at an earlier time on the next trial. Conversely, if reinforcement occurs after it was expected, this is interpreted as feedback indicating an overly fast pacemaker, resulting in an update that slows the pacemaker and expectation of a later reward on the next trial. The same principles apply to ongoing RPE during timing tasks.

In our self-timed movement task, we activated or inhibited DANs only *up to* the time of first-lick, which Mikhael and Gershman's model predicts will produce an effect on the pacemaker rate consistent with the sign of the manipulation (activate: increase, inhibit: decrease). However, Soares *et al*. continued optical stimulation *past* the end of the timing interval, until the end of the trial. When Mikhael and Gershman modeled stimulation in the Soares *et al*. task, they found that simulated DAN activation increased the pacemaker rate during the timing interval, but the continuing stimulation after the stop-timing cue rapidly counteracted this effect, resulting in *slowing* of the modeled pacemaker between the stop-cue and the judgment, leading an effect on pacemaker rate *inconsistent* with the sign of the manipulation, as observed in Soares *et al*. If this model is correct, the effect of stimulation on the animal's judgment in the Soares *et al*. task may have arisen due to continued manipulation of DAN activity *after* the timing interval had ended. A "retrospective" effect of this sort might seem counterintuitive, but such retrospective effects have long been observed in perceptual studies, in which recall of sensory stimuli can be enhanced by additional sensory cues presented shortly after stimulus offset, suggesting that sensory events are "buffered" briefly and can be altered by neural activity occurring between the sensory event and

the perceptual decision[39,40]. It is possible that a similar process could occur in the bisection task if DAN stimulation extends past the timing interval, although this is speculative. More work is needed to reconcile the optogenetic results in the self-timed movement and bisection tasks. To start, it would be informative to repeat the optogenetic experiments in the bisection task with optical stimulation limited to the period of the timed intervals only.

3. Supplementary References

33. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
34. Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. Dopamine reward prediction errors reflect hidden-state inference across time. *Nat. Neuro.* **20**, 581–589 (2017).
35. Sutton, R. S., & Barto, A. G. *Reinforcement Learning*. (2000).
36. Gershman, S. J. Dopamine Ramps Are a Consequence of Reward Prediction Errors. *Neural Comp.* **26**, 467–471 (2014).
37. Gibbon, J., Malapani, C., Dale, C. L., & Gallistel, C. Toward a neurobiology of temporal cognition: advances and challenges. *Curr. Opin. Neurobiol.* **7**, 170–184 (1997).
38. Morgan, L., Killeen, P.R., Fetterman, J.G. Changing rates of reinforcement perturbs the flow of time. *Behav. Processes* **30**, 259–271 (1993).
39. Gegenfurtner, K. R., & Sperling, G. Information Transfer in Iconic Memory Experiments. *J. Exp. Psych: Human Perception and Performance* 19, 845–866 (1993).

40. Herrington, T. M., & Assad, J. A. Neural activity in the middle temporal area and lateral intraparietal area during endogenously cued shifts of attention. *J. Neurosci.* **29**, 14160–14176 (2009).