



Published in final edited form as:

Nature. 2021 February ; 590(7846): 451–456. doi:10.1038/s41586-020-03050-5.

Cell-type specific asynchronous modulation of PKA by dopamine in learning

Suk Joon Lee¹, Bart Lodder¹, Yao Chen^{1,2}, Tommaso Patriarchi^{3,4}, Lin Tian³, Bernardo L. Sabatini¹

¹Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA ²Department of Neuroscience, Washington University School of Medicine, St. Louis, MO 63110, USA ³Department of Biochemistry and Molecular Medicine, University of California, Davis, MA 95616, USA ⁴Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

Summary

Reinforcement learning models postulate that dopamine (DA) releasing neurons (DANs) encode information about action and action outcome and provide a teaching signal to striatal spiny projection neurons (SPNs) in the form of DA release¹. DA is thought to guide learning via dynamic and differential modulation of protein kinase A (PKA) in each class of SPN². However, the real-time relationship between DA and SPN PKA remains untested in behaving animals. Here, we monitor the activity of DANs, extracellular DA levels, and net PKA activity in SPNs in the nucleus accumbens in mice during learning. We find positive and negative modulation of DA that evolves across training and is both necessary and sufficient to explain concurrent fluctuations in SPN PKA activity. The modulations of PKA in SPNs that express type-1 and type-2 DA receptors are dichotomous such that they are selectively sensitive to increases and decreases in DA, respectively, which occur at different phases of learning. Thus, PKA-dependent pathways in each class of SPNs are asynchronously engaged by positive or negative DA signals during learning.

Introduction

Across phylogeny, dopamine (DA) release in the brain induces cellular plasticity that promotes behavioral adaptation³. In mammals, DA action in the nucleus accumbens (NAc), a striatal region heavily innervated by ventral tegmental area (VTA) DA neurons (DANs), mediates the association of motor actions with action outcomes, as necessary for individuals to learn to repeat behaviors that led to good outcomes^{1,2,4,5}. Manipulations of VTA DANs

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Author contributions

S.L. designed the study, developed the photometry system, collected and analyzed data, and wrote the manuscript. B.L. collected and analyzed data, and edited the manuscript. Y.C. conceived the study and edited the manuscript. T.P. and L.T. developed the dLight sensor and shared expertise. B.L.S. designed and supervised the study and wrote the manuscript.

Competing interests

The authors declare no competing interests.

and NAc DA levels established the sufficiency of DA release in NAc for action reinforcement^{6,7,8,9}. Furthermore, VTA DAN activity and NAc DA levels encode reward prediction error (RPE) - the difference between the actual and expected values of the outcome of an action^{10,11,12,13}.

The anatomical and molecular organizations of spiny projection neurons (SPNs), the principle cells of NAc, suggest an antagonistic model of DA action on SPNs^{2,14}. NAc SPNs (analogous to the direct and indirect pathway SPNs of dorsal striatum) consist of striatomesencephalic SPNs, which innervate midbrain regions, and striatopallidal SPNs, which innervate the ventral pallidum^{15,16}. This anatomic division correlates with molecular differences: striatomesencephalic SPNs express G_{α_s} -coupled type-1 DA receptors (D1Rs) by which DA enhances cAMP production and protein kinase A (PKA) activity whereas striatopallidal SPNs express $G_{\alpha_i/o}$ -coupled type-2 DA receptors (D2Rs) which inhibit cAMP production and suppress PKA^{15,16}. Models of reinforcement learning² incorporate these differences and link RPE-encoding DA transients to PKA-dependent modulation of excitability^{17,18}, synaptic plasticity^{19,20}, and transcription^{21,22} in SPNs.

Here, we investigate the relationship between DA and net SPN PKA activity (i.e., the balance between PKA and phosphatase activities) in freely-behaving mice undergoing reward-based learning with multichannel fiber photometry and fluorescence lifetime photometry (FLiP)²³. Our results support a model of DA- and basal ganglia-dependent learning in which PKA in D1R- and D2R-SPNs is asynchronously engaged to mediate the action of positive and negative DA signals during learning.

DRs can regulate SPN PKA *in vivo*

To examine the ability of each DA receptor (DR) class to modulate PKA in awake animals, we monitored net SPN PKA activity following intraperitoneal delivery of DA receptor agonists and antagonists. We expressed FLIM-AKAR, previously validated *in vitro*^{24,25} and *in vivo*²³, in either D1R- or D2R-SPNs in NAc by injecting an adeno-associated virus (AAV) that expressed the sensor in a Cre-dependent manner into transgenic mice (Fig. 1a). Fluorescence was detected using an optical fiber, and fluorescence lifetimes were measured with time-correlated single-photon counting at 1Hz (Fig. 1a, Extended Data Fig. 1).

In *Drd1a-Cre* mice, FLIM-AKAR expressed in NAc D1R-SPNs reported an increase in net PKA activity (~160ps change) following administration of D1R agonist (SKF81297 hydrobromide), consistent with D1R-mediated activation of PKA (Fig. 1b, replotted from 23). The D1R-agonist induced change was not observed in D1R-SPNs expressing FLIM-AKAR with a mutated PKA phosphorylation site (FLIM-AKAR^{T391A}). In contrast, D1R antagonist (SKF83566 hydrobromide) slightly but significantly increased (~17ps) FLIM-AKAR lifetime (Fig. 1b), indicating a reduction in D1R-SPNs' PKA activity. Pre-administration of D1R antagonist largely blocked D1R agonist response (Fig. 1b and d), confirming the specificity of the agonist.

In *Adora2a-Cre* mice, D2R agonist (sumanirole maleate) increased (~22ps) FLIM-AKAR lifetime, indicating suppression of D2R-SPNs' net PKA activity (Fig. 1c). Conversely, D2R

antagonist (eticlopride hydrochloride) decreased the lifetime (~16ps), demonstrating an increase in D2R-SPNs' net PKA activity (Fig.1c). Furthermore, pre-injection of D2R antagonist blocked the effect of D2R agonist (Fig.1d). There was a slight change in the lifetime of phosphorylation-site mutant AKAR in both D1R-antagonist and D2R-agonist experiments (Fig.1b-c), potentially arising from injection-induced hemodynamic changes.

Changes in SPN PKA in response to pharmacological manipulation of DA receptors revealed bidirectional DA receptor-dependent regulation of SPN PKA activity *in vivo*. Furthermore, FLiP's ability to detect behaviorally-induced changes in SPNs PKA activity is suggested by the small (~10ps) reduction in FLIM-AKAR lifetime observed in D1R-SPNs at injection, which was independent of the drug administered (Fig.1b).

Plasticity of DA signals across learning

Since fluorescence lifetime changes of FLIM-AKAR in response to reward can last 40-60s²³, we designed a slow-time-scale, food-reward reinforced task (Fig.2a) to investigate how DAN activity, NAc DA, and SPN PKA are modulated during learning. Mice were habituated to the arena for 1 day and then trained on the full task for 11 days (days 1-11). On day 12, the reward was omitted from 25% of successful trials to collect 'reward-omission' trials. On day 13, the LED was turned off for the session in order to collect 'LED-omission' trials in which mice occasionally managed to perform correct movements despite the absence of the LED cue and received 'unexpected' rewards. Animals learned three key components of the task: 1) staying in the trigger zone to initiate a new trial; 2) running toward the receptacle zone after the LED cue; and 3) waiting in the receptacle zone once having entered it (Extended Data Fig.2a).

Given that NAc DA release can be dissociated from increases in VTA DAN spiking²⁶, we simultaneously monitored somatic DAN activity in VTA, activity of VTA DAN axons in NAc, and ensuing DA transients in NAc to compare their patterns across training within the same mice. To simultaneously monitor both DAN activity and NAc DA release, we expressed jRCaMP1b²⁷ in VTA DANs and dLight²⁸ in NAc (Fig.2b).

In beginner animals, NAc DA levels increased robustly following reward delivery but only minimally at the time of LED cue (Fig.2c,e,f). In expert animals, the magnitude of DA release after reward was lower than in beginner animals, whereas that after the LED cue was larger (Fig.2c,e,f). LED-evoked DA release occurred in unrewarded failure trials and also increased across training, and there was no significant difference in LED-evoked DA release in success versus failure trials across training (data not shown). In LED-omission trials, there was no significant increase in DA level at the time of an omitted LED cue ('LO-LED', Fig.2d,g), indicating that the LED-evoked DA response requires the cue, which likely drives reward expectation due to its learned association with reward. Furthermore, as evidenced by intermediate performance states, the shift in DA release from reward to cue occurred gradually across training (Extended Data Fig.2i) and correlated with success rate (Fig.2e-f).

We tested if positive and negative modulation of DA release occurs in the trained states by examining reward omission (day 12) and LED omission (day 13) trials. Consistent with

previous studies, DA levels dipped below baseline at the time of expected reward delivery when the reward was omitted ('RO-dip', Fig.2d,g). In contrast, the DA peak after reward was larger in rewarded LED-omission trials than in regular rewarded trials (Fig.2d,g). The DA dip following reward omission and the larger DA response in unexpectedly rewarded LED-omission trials are consistent with bidirectional modulation of DA by reward expectation.

The patterns of activity in DAN soma and terminals were similar to those of DA levels both across learning and within the trained state (Fig.2h, Extended Data Fig.2), and the contributions of behavioral events to each signal were not statistically different in generalized linear models relating the two (Extended Data Fig.3). In addition, we calculated the correlations among deconvolved fluorescence signals (to account for different kinetics of the signals) or signal peaks, and found that substantial (50-60%) variance in DAN terminal activity and DA release in NAc are explained by DAN soma activity (Fig.2i). To test the causal relationships between DAN soma activity and terminal activity, we optogenetically manipulated the former during the behavior while monitoring the latter. Bidirectional modulation of soma activity blunted the terminal activity response to LED, reward, and reward omission (Extended Data Fig.4).

In summary, the task induced both positive and negative modulation of DAN activity and DA release by reward expectation across training stages. These bidirectional changes in DA signals during the slow-time scale behavior permit examination of how SPN PKA activity responds to evolving DA signaling during learning. We performed additional photometry controls such as recordings in mice expressing eGFP or a DA-binding-site-mutant dLight to examine for potential movement and hemodynamic artifacts, checking for optical crosstalk between red and green channels, and photo-bleaching across days (Extended Data Fig.5).

DA asynchronously modulates SPN PKA

DA release patterns in the two hemispheres were similar in this non-lateralized behavior (Extended Data Fig.6a-b). Therefore, we performed simultaneous measurements of DA levels and net SPN PKA activity by expressing two green fluorescent sensors, dLight and FLIM-AKAR, in different hemispheres (Fig.3a). Consistent with the above results, dLight responses evoked by the cue and by reward increased and decreased, respectively, across training (Fig.3b-c). Similarly, net PKA activity in D1R-SPNs increased at the time of reward consumption (Fig.3b, *right*, red aligned to R) in beginner animals, consistent with the activation of PKA by G_{α_s} -coupled D1 receptors. At expert stages, the D1R-SPN PKA activity increase shifted to LED onset (Fig.3b-c). For rewarded trials in trained animals, net PKA activity detected by the sensor likely increased due to both cue- and reward-evoked DA release (Fig.3b) as 1) the LED cue increased PKA activity in unrewarded failure trials and in reward-omission trials; and, 2) the magnitude of net PKA activation was larger in rewarded compared to unrewarded trials. Furthermore, D1R-SPN PKA activation was larger in rewarded LED-omission trials of expert animals than in the regular rewarded trials, consistent with larger DA release evoked by unexpected reward triggering greater PKA activation (Fig.3d-e).

In contrast, net PKA activity in D2R-SPNs was not strongly modulated in beginner animals (Fig.3b). Furthermore, although a slight reduction in D2R-SPN net PKA activity following the LED cue is consistent with PKA inhibition by $G_{\alpha i/o}$ -coupled D2 receptors activated by DA, this pattern did not change across learning (Fig.3b-c). Therefore, this dip in PKA activity was likely not caused by DA, of which the release pattern changed markedly across learning. In contrast, failure trials of intermediate and expert animals, in which DA levels significantly dropped below baseline, significantly increased net PKA activity in D2R-SPNs (Fig.3b-c). A similar pattern of D2R-SPNs FLIM-AKAR lifetime occurred in reward-omission trials, consistent with this signal reflecting activation of PKA by DA decreasing below baseline (Fig.3d-e). Furthermore, FLIM-AKAR lifetime modulations observed during behavior occurred in individual trials (Extended Data Fig.7) and did not reflect movement artifacts (Extended Data Fig.6c). Thus, PKA activities in D1R- and D2R-SPNs respond to different DA dynamics and are not strongly modulated at the same time: D1R-SPN PKA is activated at early learning stages by rewards, which increase DA levels, as well as by each reward-predictive cue and reward after learning whereas D2R-SPN PKA is only activated at late learning stages by failure to achieve expected rewards, which decreases DA below baseline.

With the exception of the small reduction in D2R-SPN net PKA activity in success trials, the net PKA activity patterns in SPNs could be explained by the evolution of DA dynamics across learning. To test causality in this relationship, we exploited DA-receptor targeting pharmacology established above (Fig.3f-g). Both LED- and reward-driven increases of D1R-SPNs net PKA activity were largely blocked by D1R antagonist (SKF83566 hydrobromide). This demonstrates that D1R-SPN PKA activation that is correlated with DA release is, indeed, DA-receptor dependent. Similarly, the activation of D2R-SPN PKA by reward omission was blocked by D2R antagonist (eticlopride hydrochloride), indicating that this PKA activation is mediated by D2Rs and requires basal DA binding to D2Rs. In contrast, the small reduction in D2R-SPNs net PKA activity after reward consumption was not significantly blocked by D2R antagonist.

To further test the sufficiency in the DA-SPN PKA activity relationship, we investigated the effect of optogenetically activating (ChrimsonR²⁹) or inactivating (stGtACR2³⁰) DANs to modulate SPN PKA (Fig.4a). We confirmed the ability of the two opsins to bi-directionally modulate DA levels in a graded manner (Fig.4b,e). An activation protocol that achieved peak DA release similar a food-reward response increased D1R-SPN net PKA activity in D1R-dependent manner to a similar extent as the reward (Fig.4b,c,h). Similar effects were achieved by stimulating DAN axons in NAc and by ramping DA to mimic patterns observed during goal-directed movements³¹ (Extended Data Fig.8b-c). Furthermore, the relationship between PKA activation and DA release was non-linear (Extended Data Fig.8d), suggesting that PKA activation depends on both the amount and kinetics of DA release.

In contrast, D2R-SPN net PKA activity was minimally modulated by DAN activation (Fig.4d), supporting that, consistent with its insensitivity to D2R antagonism (Fig.3g), the small reduction in D2R-SPN PKA activity induced by reward is DA independent. Furthermore, the insensitivity of D2R-SPN PKA to DAN activation and increasing DA suggests that D2R-dependent inhibition of PKA is close to maximal at basal DA. Non-

physiological levels of DA did modulate D2R-SPN net PKA activity (Extended Data Fig.8a).

DAN inactivation had converse effects. D1R-SPN net PKA activity was not significantly changed in response to DAN inactivation (Fig.4f), indicating minimal engagement of D1Rs by basal DA, which makes D1Rs unresponsive to DA dips. Importantly, optogenetic inhibition of DANs isolated the effect of DA decreases on D1R-SPN PKA, unlike in reward-omission trials in which the cue-evoked DA increase precedes the dip. Conversely, D2R-SPN net PKA activity significantly increased in a D2R-dependent manner during DAN inactivation that decreases DA to a similar extent as reward omission in trained animals (Fig.4e,g,h), consistent with basal engagement of D2Rs by DA that allows unbinding of DA from D2Rs to disinhibit D2R-SPN PKA.

There were two results from the optogenetic experiments that were not consistent with the pharmacological manipulations. First, inactivation of DANs, which dipped DA below baseline to a similar extent as reward omission, did not affect D1R-SPN PKA activity (Fig.4f), but D1R antagonism suppressed D1R-SPN PKA (Fig.1b). Second, activation of DANs to increase DA levels to a similar extent as food rewards did not affect D2R-SPN PKA activity (Fig.4d), but D2R agonism reduced D2R-SPN PKA activity (Fig.1c). Besides possible non-specific and circuit-wide drug effects, these discrepancies can result from stronger and longer-lasting effects of D1R antagonist and D2R agonist compared to physiologically-induced short-term changes in DA.

PKA inhibition in SPNs slows learning

We investigated if selective inhibition of PKA activity in each neuron class affects learning by virally overexpressing PKI $_{\alpha}$, an endogenous inhibitory peptide that blocks PKA activation *in vitro*²⁴ and *in vivo*²³. PKI expression in D1R-SPNs altered learning, as evident in daily averages of time in trigger zone, speed after LED, and entering failure rate (Extended Data Fig.9b-d). The effect of D1R-SPN PKA inhibition appeared in early training (day 1-3), consistent with significant D1R-SPN PKA activation by reward in beginner animals. Additional differences in trigger zone time after day 3 suggests that D1R-SPN PKA activation continues to contribute to learning in intermediate animals (Extended Data Fig.9b). D1R-SPN PKA inhibition also reduced overall speed (Extended Data Fig.9i). However, this effect cannot fully explain the above effects because the average speed after LED onset was reduced for the PKI group only in the first half of training (Extended Data Fig.9c), indicating that mice were physically capable of fulfilling task criteria.

The effects of D2R-SPN PKA inhibition was significant at later stages – speed after LED and entering failure rate in days 4-7 were altered (Extended Data Fig.9f-g). This effect is consistent with D2R-SPN PKA activation by failure to receive rewards in intermediate and expert mice (Fig.3b). On the other hand, there was no significant difference in time in trigger zone (Extended Data Fig.9e), consistent with the lack of DA modulation by failure to initiate LED (LED omission). Interestingly, although D2R-SPN PKA was strongly activated during reward-omission trials, PKA inhibition in D2R-SPNs did not alter extinction of task-related behaviors (data not shown). This is consistent with the recent demonstration that extinction

involves neither DA dips nor D2R-SPNs in NAc²⁰ and depends on different parts of the striatum³².

Cell-type specific PKA inhibition produced mild effects on behavior, and not a complete impairment of learning. This may be because we designed a focal manipulation – i.e., inhibition of PKA in specific cells in a subregion of NAc. In addition, the subtlety of the effects could also be due to 1) the incomplete blockade of PKA in PKI-expressing cells; 2) the engagement of PKA-independent cellular plasticity in SPNs³³; and 3) the contribution of other brain regions that act in parallel to NAc or compensate for the loss of PKA signaling in NAc³⁴.

Discussion

Directly measuring the downstream effects of positive and negative DA transients on SPNs had been difficult to achieve due to the challenges of monitoring intracellular signaling in behaving animals (see pioneering work^{35,36,37}). Employing FLiP, we monitored net PKA activity in SPNs during behavior with ~1s temporal resolution and showed, for the first time, that PKA activity in D1R- and D2R-SPNs dynamically follow the patterns of DA release evoked by conditioned and unconditioned events. Furthermore, we established the causality of the relationship between DA and SPN-PKA activity by combining pharmacology and optogenetics with FLiP. Despite the potential contributions of many GPCRs and Ca²⁺-dependent adenylyl cyclases, DA, at times of its phasic modulation, largely dictates the state of PKA in SPNs. Furthermore, our findings provide *in vivo* evidence for a mechanistic model of SPN PKA modulation by DA (Extended Data Fig.10) that explains the selective sensitivity of D1R- and D2R-SPN PKA activity to increase and decrease of DA, respectively, due to the different basal occupancy of the receptors. Demonstration of D2R-SPN PKA response to DA dips during behavior is especially valuable in light of the recent finding that DA dips induce PKA-dependent spine growth in D2R-SPNs²⁰.

As a consequence of their temporally-dissociated DA-dependent modulation of PKA, SPNs of the direct (striatomesencephalic) and indirect (striatopallidal) pathways may have independent functions during learning. Based on our results, we endorse that the former promotes an initial association between an action and an outcome and the latter refines the learned behavior once the association is established. This model is in line with previous theoretical^{38,39} and conceptual^{1,40} models as well as recent experimental results²⁰ that support the same core concept of opponent reinforcement learning.

Methods

Animals

Experimental manipulations were performed in accordance with protocols approved by the Harvard Standing Committee on Animal Care following guidelines described in the US National Institutes of Health Guide for the Care and Use of Laboratory Animals. *Drd1a-Cre* (B6.FVB(Cg)-Tg(Drd1-cre)EY262Gsat/Mmucd, 030989-UCD) and *Adora2a-Cre* (B6.FVB(Cg)-Tg(Adora2a-cre)KG139Gsat/Mmucd, 036158-UCD) mice on *C57BL/6J* backgrounds⁴¹ were acquired from MMRRC UC Davis. *DAT-IRES-Cre* (B6.SJL-

Slc6a3tm1.1(cre)Bkmm/J, 006660)⁴² and *C57BL/6J*(000664) mice were acquired from the Jackson Laboratory. *DAT-IRES-Cre; Drd1a-Cre* and *DAT-IRES-Cre; Adora2a-Cre* mice were bred in house by crossing heterozygous parent lines. All transgenic animals used for experiments were heterozygous for the relevant Cre allele. Mouse ages were 2-4 months, and male and female mice were used in approximately equal proportion. Mice were housed on a 12hr dark/12hr light reversed cycle. No sample size precalculation was performed. For PKI experiments, we randomly assigned mice to PKI and GFP group in a manner that each of a pair matched in age and sex was randomly assigned to a different group. For PKI experiments, no blinding was attempted.

Viruses

Recombinant adeno-associated viruses (AAVs of serotype 1, 8, 9) were used to express transgenes of interest in either Cre-recombinase dependent or independent manner. AAVs were packaged by commercial vector core facilities (Addgene, Boston Children's Hospital Vector Core, Janelia Vector Core, Penn Vector Core, UNC Vector Core) and stored at -80°C upon arrival. Viruses were used at a working concentration of 10^{12} to 10^{14} genomic copies per ml. 300 nl of virus was used for all experiments except for PKI experiments for which we used 600 nl of virus at 1.2×10^{14} genomic copies per ml and 300 nl of virus at 4×10^{13} genomic copies per ml for *Drd1a-Cre* and *Adora2a-Cre*, respectively (we used a lower volume and titer for the latter due to greater toxicity in this line). Most viral plasmids are available in Addgene: AAV-FLEX-FLIM-AKAR (#60445), AAV-FLEX-FLIM-AKART391A (#60446), AAV-CAG-dLight1.1 (#111067, hSyn construct available upon request to Dr. Lin Tian), AAV-hSyn-FLEX-NES-jRCaMP1b (#100850), AAV-hSyn-FLEX-ChrimsonR-tdTomato (#62723), AAV-hSyn-SIO-stGtACR2-FusionRed (#105677), AAV-FLEX-PKIalpha-IRES-nls-mRuby2 (#63059), AAV-Cag-FLEX-eGFP (#59331), AAV-Syn-Flex-GCaMP6f (#100833). AAV-hSyn-dLight^{D103A} is available upon request to Dr. Lin Tian.

Surgery

Inhaled isoflurane was used as anesthesia. Virus was stereotactically injected into either NAc core (anteroposterior (AP) +1.2 mm, medio-lateral (ML) \pm 1.3 mm relative to bregma; dorsoventral (DV) 4.1mm below brain surface) or VTA (AP -3.3 mm, ML +0.48 mm, DV 4.5 mm). For fiber photometry or optogenetic experiments, an optical fiber (MFC_200/230-0.37_4.5mm_MF1.25_FLT mono fiber optic cannula, Doric Lenses) was implanted 200 μm above the injection site.

Behavior

To motivate animals to perform the visual cue guided task, we food restricted mice such that they remained at 80~90% of their initial weight. Mice were given 2~3 g of regular chow daily in addition to the variable number of 20mg dustless precision chocolate flavor pellets (F05301, Bio Serv) consumed during the task. Food restriction was started at least one day before commencing the behavioral training. During the task, a mouse (sometimes connected to a patch cord) was allowed to freely move inside an 8x16 inch box, which contained a pellet receptacle and a white LED on one of the short walls. Receptacle entry was detected by an infrared sensor installed inside the receptacle. Animal movements were captured by

cameras (WV-CP504, Panasonic or FL3-U3-13E4M, PointGrey) connected to Ethovision 11.5 or Bonsai 2.3 software that controlled all behavioral apparatuses and were synchronized with MATLAB 2012a software used to acquire photometry data.

The behavioral task structure is as follows. After the enforced 120 s inter-trial-interval (ITI), the mouse can self-initiate a trial by entering the trigger zone (a zone opposite to the wall containing the pellet receptacle indicated in green in Fig.2a but not marked in the actual behavior box) and staying in the zone for 3 consecutive seconds. If the mouse enters the trigger zone during the enforced ITI or exits the trigger zone before 3s after entering the zone, the trial is not initiated. If the mouse succeeds in initiating the trial, the LED above the receptacle turns on, signaling the start of the trial. Once the LED turns on, the mouse must enter the LED zone (a zone near the wall containing the pellet receptacle indicated in blue in Fig.2a but not marked in the box) within 5s. If the mouse fails to enter the LED zone within 5s, the trial is terminated, and the LED turns off. If the mouse enters the LED zone in time, it has to stay in the LED zone for additional 3 consecutive seconds. After this enforced waiting period, a single 20mg pellet is dispensed, and the LED turns off. If the mouse prematurely exits the LED zone during the waiting period, the trial is terminated, and the LED turns off. After termination of the trial either by a success or a failure, the next enforced ITI of 120s starts. Mice were trained for 11 days starting after 1 day of a 40-minute habituation session during which they were given 10 free pellets in the receptacle. On day 12, the reward was omitted from 25% of successful trials to collect 'reward omission' data. On day 13, the LED was turned off for the entire session to collect 'LED omission data'. Most mice reached 90% success rate (# of success trials/# of total trials) in 9 days (Extended Data Fig.2a).

Each behavior session was run until a mouse initiated 20 trials (including success and failure trials) or a 2hr time limit was reached, except on reward-omission days (day 12) when the session continued until 5 reward-omission trials were collected. Because mice had to self-initiate trials, session durations were variable (50min - 2hr). Most sessions included 20 trials except the first few sessions at the start of training and the extinction sessions of PKI experiments (9 out of 806 sessions for photometry experiments, 15 out of 611 sessions for PKI experiments).

To compare signals across different stages of learning across all animals, we defined a 'beginner' stage as days in which the success rate was below 50% of the maximum (plateau) success rate, an 'intermediate' stage as days in which the success rate was between 50~90% of the maximum success rate, and an 'expert' stage as days in which the success rate was above 90% of the maximum success rate. A highly-sensitive optical system allowed us to use dim excitation (5~16 μ W) and perform fiber photometry on all days to track progressive changes in neural signals.

Fluorescence lifetime photometry (FLiP)

FLiP was carried out with the optical system described in Extended Data Fig.1 and using the method described in detail in Lee et al. (2019). Briefly. All filters in the system were purchased from Semrock. A pulsed laser (BDS-473-SM-FBE Becker and Hickl – BH – operating at 50Mhz) was used as the light source to excite FLiM-AKAR. For fluorescence

detection, a high-speed hybrid photomultiplier tube (PMT, HPM-100-07-Cooled, BH) controlled by DCC-100-PCI (BH) was used. The hybrid PMT was connected to an SPC-830 (BH), a time correlated single photon counting (TCSPC) board, which detects the time delay between the pulsed excitation and the photon detection by the PMT. The data was collected by custom software written in MATLAB 2012a, which calculated the average lifetime of detected photons at 1s intervals. This interval for average lifetime measurements was empirically determined to have enough photons to accurately estimate the lifetime (>200,000 photons/measurement) without running into a photon count limit (~1,000,000 photons/s) of the TCSPC board. The typical excitation power needed to generate the appropriate rate of photons (~400 kHz) for TCSPC was 0.6~1 μ W (measured at the output end of the patch cord).

To estimate the change in lifetime of FLIM-AKAR, we measured the average lifetime for each 1s time bin by calculating the mean photon arrival time (the population mean of the delay between the pulsed excitation and the fluorescence photon arrival as described) using the following equation⁴³:

$$\tau = t - t_0 = \frac{\int tF(t) \cdot dt}{\int F(t) \cdot dt} - t_0$$

in which F(t) is the photon count of a fluorescence lifetime decay curve at time bin t, and t_0 is the offset of the lifetime histogram, which can be estimated by fitting a double exponential curve to a lifetime histogram. We performed this calculation for the 8ns time range in a lifetime histogram as this interval was minimally contaminated by a secondary fluorescence peak resulting from the autofluorescence of a fiber. The length of the patch cord was chosen to maximize the time separation between the sensor fluorescence peak and the fiber autofluorescence peak (~10ns time delay for light to travel from the one end of the patch cord to the other end and back). The lifetime of FLIM-AKAR was reported as a change in lifetime (delta lifetime), which was calculated by subtracting the average lifetime of a baseline period (a period before the event of interest) from the average lifetime transient. Delta lifetime was plotted in 1s time bin for behavioral experiments and in 10s average for pharmacology experiments.

Fiber photometry and optogenetics

Both fiber photometry (fluorescence intensity fiber photometry) and optogenetics were carried out with the optical system described in Extended Data Fig.1. All filters in the system were purchased from Semrock except for a Doric Minicube (FMC5_E1(465-480)_F1(500-540)_E2(555-570)_F2(580-680)_S, Doric Lenses) that houses multiple dichroic mirrors and filters.

For fiber photometry, 470nm (M470F3, Thorlabs) and 565nm (M565F3, Thorlabs) LEDs were used to excite dLight (or GCaMP) and jRCaMP, respectively. Both LEDs were frequency modulated by a digital signal processor system (RX8-5-12, Tucker-Davis Technology) to carry out locked in amplification of PMT outputs. The average power levels of LEDs (measured at the output end of the patch cord) were 9.3 μ W, 5.4 μ W, 16.2 μ W, 5.6 μ W, and 14.5 μ W for dLight, soma jRCaMP, terminal jRCaMP, soma GCaMP, and

terminal GCaMP excitation. For fluorescence detection, H7422-40 (Hamamatsu) and H10770(P)A-40 (Hamamatsu) PMTs were used. PMTs were connected to a low-noise current preamplifier (SR570, Stanford Research Systems). The signal generated from SR570 was locked-in-amplified by RX8-5-12 using the frequency of the LED used to excite the sensor that fluoresces in the light spectrum assigned to a corresponding PMT. The locked-in-amplified signal was collected by a data acquisition board (PCI-6115, National Instruments), which was controlled by the custom software written in MATLAB 2012a. The raw fluorescence data was collected at 1kHz. It was subsequently smoothed by a moving average filter (width of 200ms) and down-sampled to 100Hz. The relative change in fluorescence $df(f)/f_0 = (f(t)-f_0)/f_0$ was calculated using f_0 equal to the average of the baseline period (20s before the trigger zone entry). For comparisons and averaging across mice, df/f of an individual mouse was normalized to the 99 percentile of df/f value across all sessions for the mouse.

For activation of Chrimson, a 593.5nm laser (SKU: YL-593-00100-CWM-SD-03-LED-F, Optoengine) was used with 2ms laser pulses delivered at 20Hz with exception of DA ramping experiments (Extended Data Fig.8). For DA ramping experiments, we increased the frequency of the light pulses (2ms width) gradually from 24Hz to 34Hz for 3s, 16Hz to 34Hz for 5s, and 4Hz to 30Hz for 7s by changing the frequency every 500ms. Reported laser powers were measured by a digital optical power meter (PM100D, Thorlabs) at the end of the patch cable, while the laser was operating in a continuous mode. For activation of stGtACR2, a 473nm laser (MBL-III-473, Optoengine) was used. For all stGtACR2 activation experiments, the laser was operated in a continuous mode (without pulsing). Reported laser powers were measured in the same manner as in the Chrimson experiments.

Pharmacology

The following concentrations of drugs were used for intraperitoneal (IP) injection: D1R agonist (SKF 81297 hydrobromide, 10mg/kg), D1R antagonist (SKF 83566 hydrobromide, 3mg/kg), D2R agonist (Sumanriole maleate, 4mg/kg), D2R antagonist (Eticlopride hydrochloride, 0.5mg/kg). All drugs were dissolved in sterile saline. Drug solutions were IP injected in 0.1ml solution/10g of mouse using a fine needle insulin syringe (BD 324911, BD Bioscience). On average, it took 30~60s from the beginning of scruffing a mouse to the end of IP injection.

Deconvolution of fluorescence intensity photometry data

To gain a better understanding of the underlying neural activity generating the measured fluorescence transients, we deconvolved fluorescence signals into population events using the one-dimensional constrained deconvolution algorithm developed by Pnevmatikakis et al. (2016)⁴⁴, which assumes that the sensor fluorescence follows an autoregressive process. Before deconvolution, we smoothed the raw fluorescence data collected at 1kHz by a moving average filter (200ms width) and down-sampled the data to 10Hz. We subsequently calculated df/f using the bottom 5th percentile value of a rolling time window of 300s as f_0 . This allowed us to perform deconvolution on fluorescence signal for an entire session without introducing a trial structure to the data. The deconvolution of df/f was performed using a conic programming method and a second order autoregressive process. Correlations

between deconvolved signals were calculated by computing a correlation between the number of population events in each 1s time interval.

Encoding model

To more objectively reveal the relationships between behavior and fluorescence signals, we created a generalized linear model to predict the deconvolved fluorescence signal (dLight and jRCaMP) from observed stimuli and behavioral parameters. We processed both observed stimuli and behavioral parameters into time bin of 100ms to match the time bin of the deconvolved fluorescence signal. We constructed additional explanatory variables by introducing multiple time shifts to the observed stimulus and behavioral parameter variables. For continuous variables (speed, acceleration, rotation, position), $-0.3 \sim +0.3$ s time shifts were introduced; for event variables (movement initiation, cue, reward delivery, receptacle entry), $-0.3 \sim +1.0$ s time shifts were introduced. We reorganized the data from a session into a trial structure in order to have a fixed distribution of the number of trials when we split the data into fitting and testing sets. Model parameters (coefficients) were learned from fitting sets and evaluated using testing sets (5 fold cross validation). We performed Lasso on the data from all sessions of each learning stage and each mouse to find a minimum set of explanatory variables that allows the mean squared error (MSE) to be within 1 standard error from the minimum MSE (a standard error of MSE was calculated from 5 cross validation sets). We decided on the final common set of explanatory variables by selecting all variables selected by Lasso across different learning stages. With the final set of explanatory variables, we performed a linear regression. The contribution of each variable category (kinematics, movement initiation, position, cue, reward, accuracy, previous trial) was calculated by setting the coefficients of all the variables assigned to each category as zero and computing the correlation between the actual fluorescence signal and the signal predicted by the model: $\text{contribution} = (R^2_{\text{full}} - R^2_{\text{partial}}) / R^2_{\text{full}}$.

Histology

Virus injected mice were euthanized and perfused transcardially with PBS followed by 4% PFA (in PBS). After >24 hr post-fix in 4% PFA, brains were sliced at 50 μ m thick using a vibrating blade microtome (Leica Biosystems VT1000S), mounted on glass slides with a DAPI mounting medium (Vectashield, H-1200), and imaged under a wide-field microscope with a 10x objective (VS120, Olympus). Images were acquired through OlyVIA 2.9 and processed via ImageJ 1.52i, which was also used for cell counting with local contrast enhancement (CLAHE) and particle analysis.

Statistics

For all data presentation, we first averaged trials for individual mice, and, then, averaged across mouse averages. All error bars were SEM across mouse averages, unless stated otherwise in the legend. All pair-wise comparisons were two-sided. All multiple comparisons were Bonferroni corrected. Exact p-values are reported in Table S1. R^2 represents a Pearson's correlation coefficient. For analysis of a trend of a single group across multiple time periods, one-way repeated measure (RM) ANOVA were used to take into account of repeated measurements from the same subjects, (if needed) followed by Bonferroni post-hoc comparisons. Similarly, for analysis of a trend of multiple groups across

multiple time periods, two-way RM ANOVA was used, followed by Bonferroni post-hoc comparisons. For the analyses in dLight-jRCaMP (Fig.2), dLight of dLight-AKAR (Fig.3), D1R-SPN AKAR of dLight-AKAR (Fig.3) experiments, 1 out of 10, 1 out of 16, and 1 out of 14 subjects, respectively, were dropped from RM ANOVA due to a missing value in serial measurements (no measurement made). In the case of data sets with multiple missing values, we analyzed the data instead by fitting a mixed model (one-way) as implemented in GraphPad Prism 8.0⁴⁵ (Extended Data Fig.2i). This mixed model uses a compound symmetry covariance matrix and is fit using Restricted Maximum Likelihood (REML). In the absence of missing values, this method gives the same p-values and multiple comparisons tests as RM ANOVA. In the presence of missing values (missing completely at random), the results can be interpreted like repeated measures ANOVA. Both RM ANOVA and mixed-effect analysis did not assume sphericity and were corrected by Geisser and Greenhouse epsilon hat method.

Figure experimental and analysis details

Fig.1a: Schematic of a coronal section at 1.2mm depicting viral injection of AAV1-FLEX-FLIM-AKAR into NAc of *Drd1a-Cre* and *Adora2a-Cre* mice for D1R-SPN and D2R-SPN FLIM-AKAR measurements, respectively (left). An optical fiber was implanted 200 μ m above the injection site (middle). A hybrid PMT and time correlated single photon counting were used to measure the fluorescence lifetime of FLIM-AKAR, which provides an estimate of the ratio between phosphorylated and un-phosphorylated FLIM-AKAR (right). Higher PKA activity results in a faster fluorescence decay (lower fluorescence lifetime) of the sensor (red).

Fig.1d: Peak lifetime change was calculated from 100s average around the peak.

Fig.2b: Schematic of a sagittal section depicting viral expression and fiber implantation in *DAT-IRE5-Cre* mice. AAV9-hSyn-dLight1.1 and AAV1-hSyn-FLEX-NES-jRCaMP1b were injected into NAc and VTA, respectively. Fibers were implanted 200 μ m above the two injections sites. NAc fiber was used to collect both dLight and DAN terminal jRCaMP signals. The VTA fiber was used to collect the DAN somatic jRCaMP signal.

Fig.2c-h: df/f was normalized across all sessions for each mouse such that the 99 percentile response=1.

Fig.2e: Mean of the normalized signal was calculated over 3s starting at LED onset.

Fig.2f: Mean of the normalized signal was calculated over 3s around the peak after the receptacle entry.

Fig.2g: Mean of the normalized signal was calculated over 0s-3s after time of expected pellet dispensing or 0s-3s after time of expected LED onset.

Fig.2i: R^2 was calculated from linear fit of each 20s interval (-5 to +15s with respect to the trigger zone entry) separately, averaged for each session, then averaged for individual mouse. R^2 between signal peaks were calculated from peaks above 1SD of session data.

Fig.3a: Schematic of a coronal section at 1.2 mm depicting injection of AAV9-hSyn-dLight1.1 and AAV1-FLEX-FLIM-AKAR into NAc of *Drd1a-Cre* or *Adora2a-Cre* mice (left). Schematic of a coronal section depicting dual fiber photometry in which one fiber is used for intensity measurements of dLight fluorescence and the other for fluorescence lifetime measurements of FLIM-AKAR (right). The fibers were implanted 200 μ m above injection sites.

Fig.3c top: Mean of the normalized signal was calculated over 0s-3s after LED onset, 5.8-15.8 after LED onset, and 3s around the peak after receptacle entry.

Fig.3c middle: Mean amplitude (ps) calculated over 0s-40s after LED.

Fig.3c bottom: Mean amplitude (ps) calculated over 0s-40s after LED for success trials and 5.8s-45.8s after LED for failure trials.

Fig.3e: Mean of the normalized signal for dLight was calculated over 0s-3s after dispensing or time of expected reward. Average response of FLIM-AKAR (ps) was calculated from peak amplitudes after dispensing or time of expected reward.

Fig.3f-g: Drug was IP delivered 10mins before recording began.

Fig.4a: Schematic of a sagittal section depicting viral expression and fiber implantation. *left*, AAV9-hSyn-dLight1.1 was injected into NAc of *DAT-IRES-Cre* mice. AAV1-hSyn-FLEX-ChrimsonR-tdTomato or AAV8-hSyn-SIO-stGtACR2-FusionRed was injected into VTA for DAN activation or inactivation, respectively. Fibers were implanted 200 μ m above the injection sites. *right*, As on the left except AAV1-FLEX-FLIM-AKAR was injected to NAc of *DAT-IRES-Cre; Drd1a-Cre* and *DAT-IRES-Cre; Adora2a-Cre* mice for, respectively, D1R-SPN and D2R-SPN FLIM-AKAR measurements.

Fig.4b-h: Statistics for dLight and AKAR signal was performed on mean df/f (0-3s for activation, 0-10s for inactivation) and mean lifetime (0-20s for activation, 5-25s for inactivation experiments), respectively. Average per mouse was calculated from 10 trials for optogenetic illumination and reward responses and 5 trials for reward omission responses. Reward responses of SPN FLIM-AKAR were acquired in separate cohorts of untrained animals. Reward-omission responses of dLight and SPN FLIM-AKAR were acquired from separate cohorts that were trained on the behavioral task.

Fig.4b: Stimulation parameters were 20Hz train of 2ms pulses at 14.3mW. For food rewards, the response of dLight was aligned to its peak after receptacle entry and time shifted for maximum overlap with an optogenetic response.

Fig.4e: Stimulation parameters were continuous illumination at 5mW.

Fig.4h: IP injection of drug was at least 10min before recording. Stimulation parameters were 1s illumination/20Hz train of 2ms pulses at 14.3mW (left) and 5s continuous illumination at 2.5mW (right).

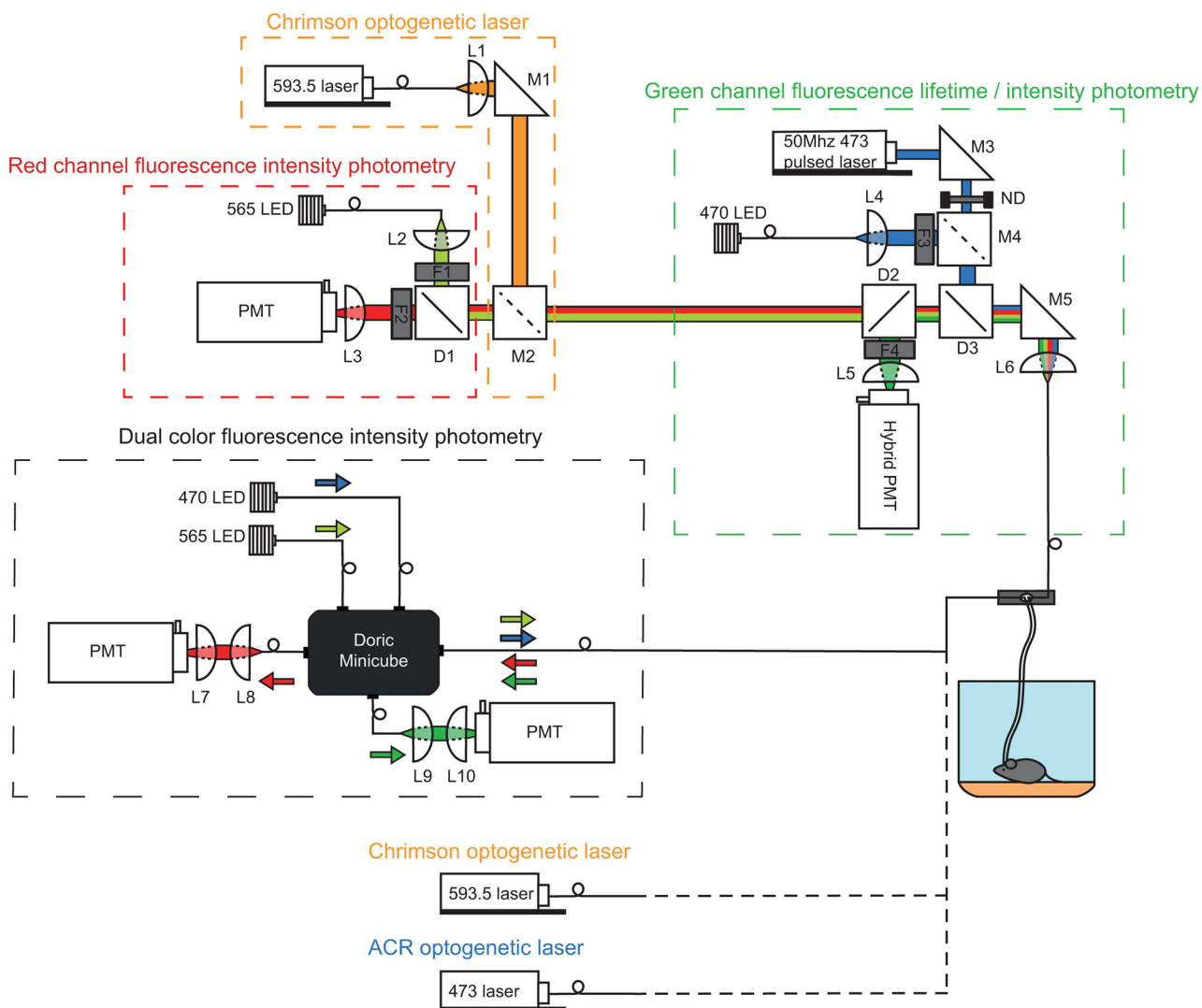
Data availability

All data (MATLAB data files) are available online via the public repository managed by Harvard Medical School (https://sharehost.hms.harvard.edu/neurobiology/?sabatini/DA_PKA).

Code availability

Custom MATLAB codes are available online via the public repository managed by Harvard Medical School (https://sharehost.hms.harvard.edu/neurobiology/?sabatini/DA_PKA).

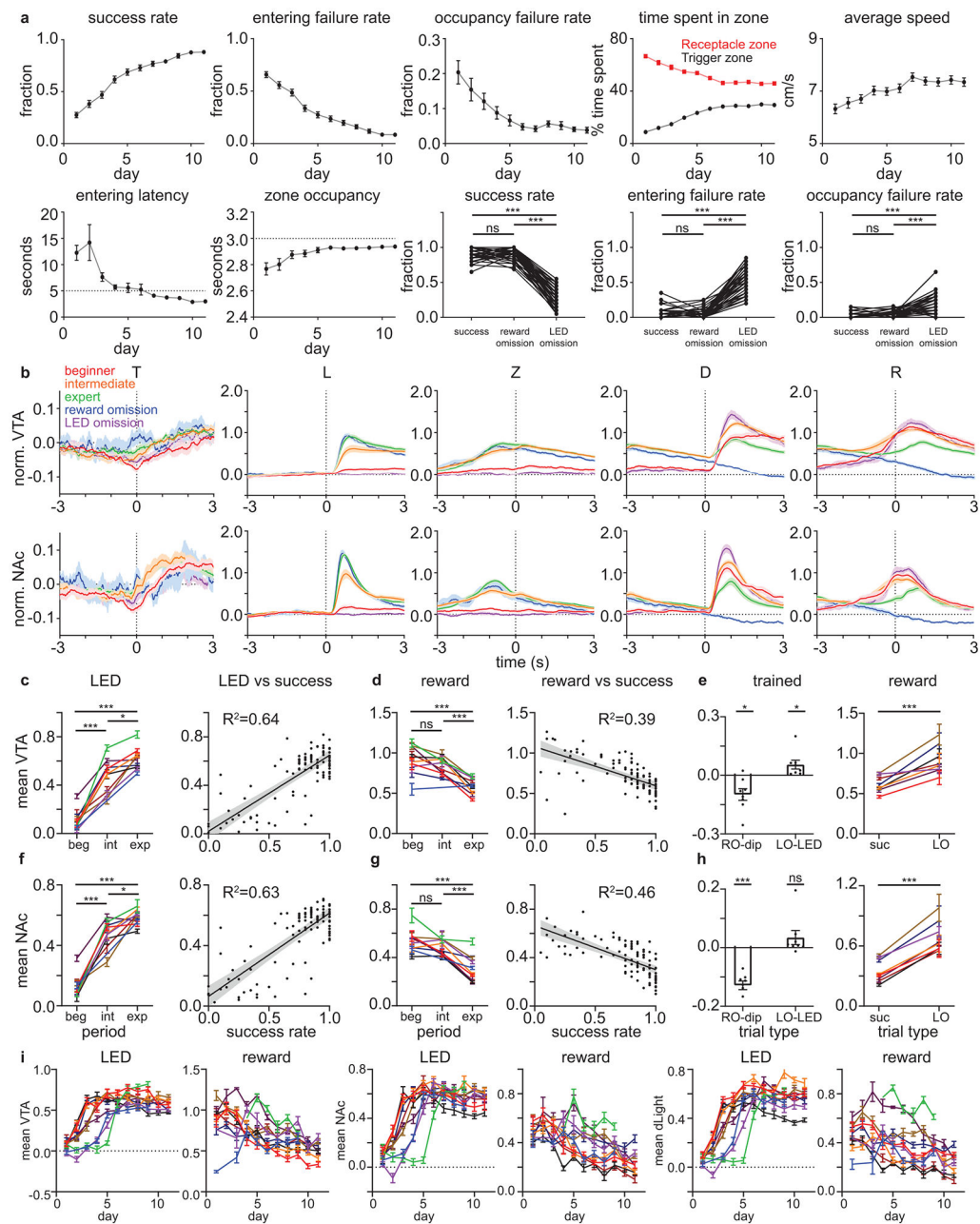
Extended Data



Extended Data Fig. 1 | Multi-purpose photometry system for FLiP, fiber photometry, and optogenetics

The system consists of two independent multi-color photometry units. The top photometry unit consists of three sub-components used for: (1) red channel fluorescence photometry, (2) Chromson optogenetic laser activation, and (3) green channel fluorescence lifetime and

intensity photometry. For (1), red channel photometry was accomplished using a fiber coupled 565nm LED (M565F3, Thorlabs) for excitation whose output was collimated in free-space by L2 and filtered by F1 (554/23, Semrock). Red fluorescence was separated from the excitation light by dichroic D1 (573LP, Semrock), filtered by F2 (630/60, Semrock), and focused onto a PMT (H10770(P)A-40, Hamamatsu) by L3. For (2), Chrimson optogenetic light was provided by a fiber coupled 593.5nm laser (SKU: YL-593-00100-CWM-SD-03-LED-F, Optoengine) whose output was collimated by L1 and combined with the red photometry path via M2, a mirror that can be inserted or removed, respectively, for Chrimson optogenetic stimulation or red channel photometry. For (3)'s green channel fluorescence lifetime measurement mode, a 50Mhz 473nm pulsed laser (BDS-473-SM-FBE, Becker & Hickl) was fed through a rotating neutral density filter for power adjustment, reflected by D3 (488LP dichroic, Semrock), and focused onto a patch cable by L6. Emission light was passed through D3, reflected by D2 (532LP dichroic, Semrock), filtered by F4 (517/22, Semrock), and focused by L5 to a high-speed hybrid PMT (HPM-100-07-Cooled, Becker and Hickl). The hybrid PMT was connected to a time correlated single photon counting board (SPC-830, Becker and Hickl) for fluorescence lifetime measurements. For (3)'s green channel fluorescence intensity measurement mode, a fiber coupled 470nm LED (M470F3, Thorlabs) was collimated by L4, filtered by F3 (482/18, Semrock), and reflected by a removable mirror (M4); emission light was detected by a PMT (H7422-40, Hamamatsu). Alternatively, when fluorescence lifetime measurements were not needed, the bottom photometry unit was used. This simple "dual color fluorescence intensity photometry" unit consists of 470nm and 565nm LEDs (Thorlabs), two PMTs (H10770(P)A-40, Hamamatsu), and a Doric Minicube (FMC5_E1(465-480)_F1(500-540)_E2(555-570)_F2(580-680)_S, Doric Lenses) that are connected by patch cables. For both photometry units, LEDs were driven by a digital signal processor system (RX8-5-12, Tucker-Davis Technology) for frequency modulation to carry out locked in amplification of sensor signals detected by PMTs. In addition to the two main photometry units, a 593.5nm laser (SKU: YL-593-00100-CWM-SD-03-LED-F, Optoengine) and a 473nm laser (MBL-III-473, Optoengine) with independent patch cable connections were installed for Chrimson optogenetics and stGtACR2 optogenetics, respectively, for VTA DAN activity manipulation while monitoring NAc.



Extended Data Fig. 2 I. Plasticity in DA release and DAN activity dynamics

a, Behavioral parameters demonstrating that mice are able to learn the visual cue guided operant conditioning described in Fig. 2a. Top: from the left, success rate (number of rewarded trials / total number of trials, one-way RM ANOVA, $F(6.298,384.2)=107.8$, $p<0.0001$), entering failure rate (number of receptacle zone entering failure trials / total number of trials, one-way RM ANOVA, $F(5.782,352.7)=103.8$, $p<0.0001$), occupancy failure rate (number of premature receptacle zone exit trials / number of receptacle zone entering success trials, one-way RM ANOVA, $F(4.253,225.4)=7.324$, $p<0.0001$), time spent in zone (time spent in a zone / total session time, receptacle: one-way RM ANOVA, $F(3.171,193.4)=51.12$, $p<0.0001$; trigger: $F(4.544,277.2)=110.9$, $p<0.0001$), average speed

(one-way RM ANOVA, $F(3.969,242.1)=15.26$, $p<0.0001$). Bottom: from the left, entering latency (delay to enter the receptacle zone after the LED cue, one-way RM ANOVA, $F(1.760,107.4)=9.652$, $p=0.0003$), zone occupancy (time spent in the receptacle zone after entering the zone during a trial, $3s$ =maximum, one-way RM ANOVA, $F(5.229,277.1)=3.420$, $p=0.0045$). Last three graphs depict success rate (one-way RM ANOVA, $F(1.727,72.54)=668.6$, $p<0.0001$), entering failure rate (one-way RM ANOVA, $F(1.328, 55.79)=244.5$, $p<0.0001$), and occupancy failure rate (one-way RM ANOVA, $F(1.295,54.40)=61.04$, $p<0.0001$) comparisons for regular, reward omission, and rewarded LED omission sessions of expert mice. $n=64$ mice from all photometry behavior experiments. Plotted as mean \pm SEM across mice and dots=mouse averages. *** $p<0.001$ for Bonferroni corrected post-hoc comparisons.

b, DAN activity across learning. The average responses for beginner, intermediate, expert, reward omission (of expert mice), and rewarded LED omission (of expert mice) trials are shown in red, orange, green, blue, and purple, respectively. Dashed vertical lines indicate the behavioral time stamps (T=trigger zone entry, L=LED on, Z=receptacle zone entry, D=pellet dispensing, R=receptacle entry). Top: normalized df/f of VTA jRCaMP signal showing VTA DAN soma activity. Bottom: normalized df/f of NAc jRCaMP signal showing VTA DAN terminal activity. $n=10$ mice. Plotted as mean \pm SEM across mice.

c, VTA jRCaMP response (mean of normalized signal) to LED (mean of 0s-3s after LED cue) during training. *left*, Individual mouse average plotted for different training periods (beginner, intermediate, and expert) plotted as mean \pm SEM across trials (one-way RM ANOVA, $F(1.870,14.96)=73.76$, $p<0.0001$). *right*, Daily average of LED response vs. success rate where each dot represents a daily (session) measurement of a mouse plotted as linear regression fit \pm its 95% CI.

d, VTA jRCaMP response to reward (mean of 3s around the peak after receptacle entry) during training plotted as c (one-way RM ANOVA, $F(1.835,14.68)=40.40$, $p<0.0001$).

e, VTA jRCaMP response in trained mice. *left*, response to reward omission (RO dip, mean of 3s-6s (shifted for slow soma jRCaMP signal) after expected time of pellet dispensing, one-sample t-test, $p=0.013$) and to LED omission (LO-LED, mean of 0s-3s after expected time of LED onset, one-sample t-test, $p=0.045$) plotted as mean \pm SEM across mice. *right*, response to reward (mean of 0s-3s after pellet dispensing) in regular success (suc) and LED omission (LO) trials (paired t-test, $p=0.001$) plotted as c.

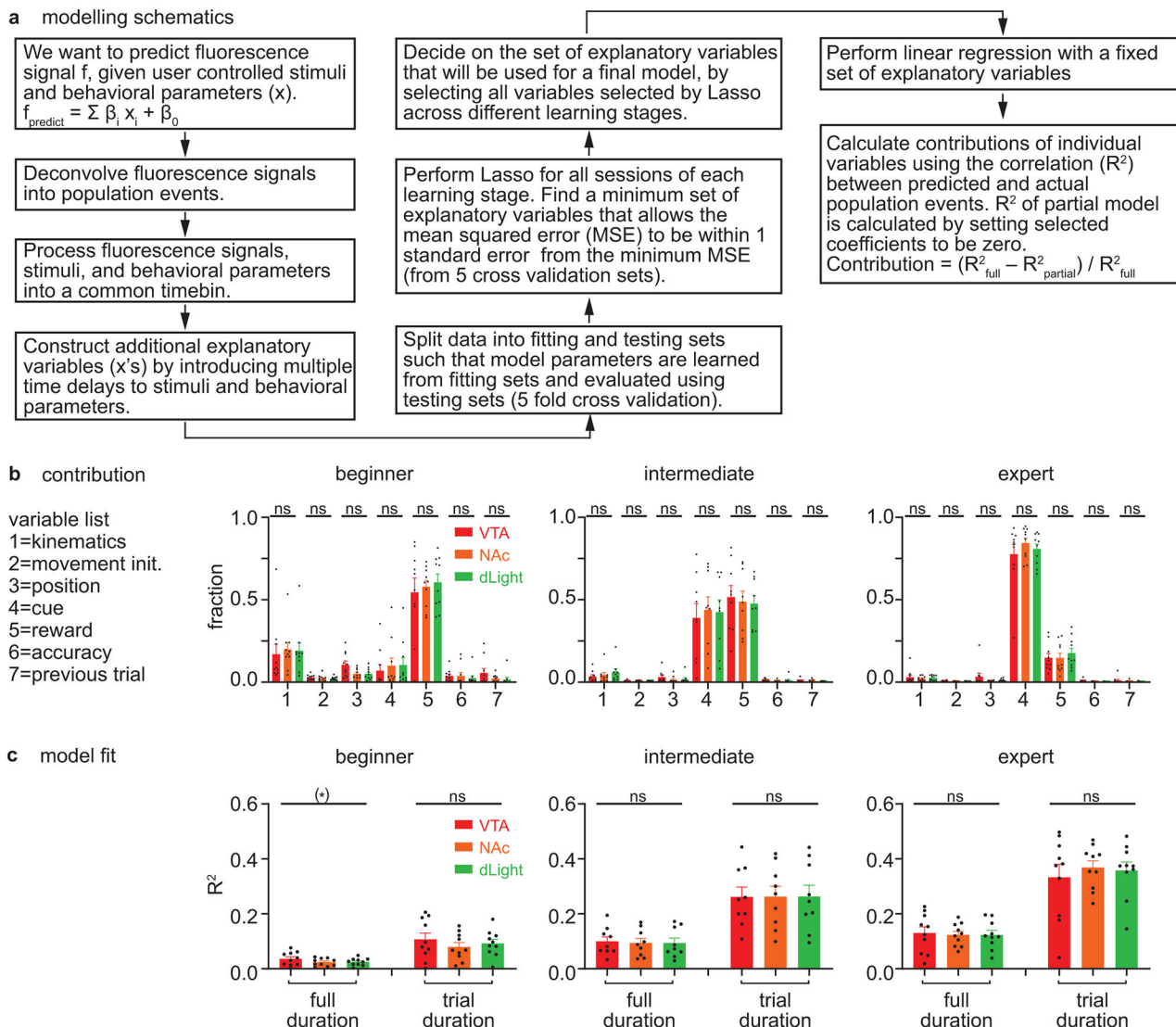
f, As in c for NAc jRCaMP (one-way RM ANOVA, $F(1.986,15.89)=98.52$, $p<0.0001$).

g, As in d for NAc jRCaMP (one-way RM ANOVA, $F(1.939,15.51)=33.95$, $p<0.0001$).

h, As in e NAc jRCaMP response in trained mice. *left*, response to reward omission (RO dip, mean of 0s-3s after expected time of pellet dispensing, one-sample t-test, $p<0.0001$) and to LED omission (LO-LED, mean of 0s-3s after expected time of LED onset, one-sample t-test, $p=0.189$). *right*, response to reward (mean of 0s-3s after pellet dispensing) in regular success (suc) and LED omission (LO) trials (paired t-test, $p<0.0001$).

i, Daily average response of VTA jRCaMP, NAc jRCaMP, and dLight response to LED and reward across training for individual mouse. (from the left, one-way mixed-effects analysis, $F(1.589,13.35)=25.07$, $p<0.0001$; $F(2.339,18.71)=11.70$, $p=0.0003$; $F(1.720,14.45)=25.75$, $p<0.0001$; $F(2.984,23.87)=15.69$, $p<0.0001$; $F(1.806,15.17)=35.41$, $p<0.0001$; $F(2.835,22.68)=12.73$, $p<0.0001$). Plotted as mean \pm SEM across trials for each mouse.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ for one-sample t-tests, and paired t-tests, Bonferroni corrected post-hoc comparisons. All t-tests are two-sided.



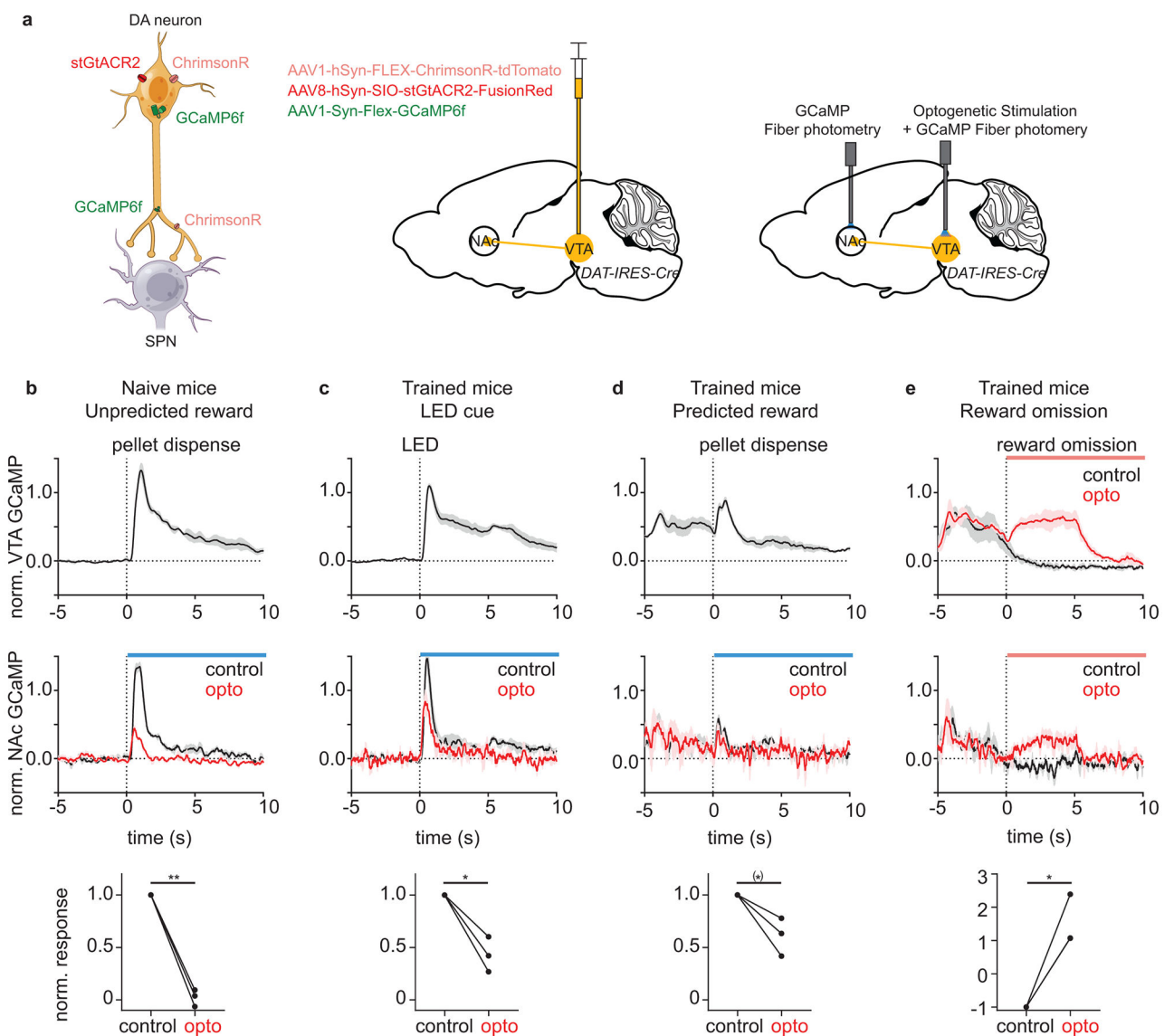
Extended Data Fig. 3 | Relationship between DA signal and behavioral parameters

a, Schematics of building a generalized linear model that relates user controlled stimuli and behavioral parameters to fluorescence signals. Briefly, there are 3 types of explanatory (independent) variables in the model. Continuous variables (speed, acceleration, rotation, position) continuously change their values as time passes. Event variables (movement initiation, cue, reward delivery, receptacle entry) are 0 except at a time point of an event when they temporarily change their value to 1. Whole trial variables (accuracy=0 for current trial failure, 1 for current trial success; previous trial=0 for previous trial failure, 1 for previous trial success) change their values in the beginning of a trial and stay constant until the next trial.

b, Comparison of average variable contributions for VTA jRCaMP (red), NAc jRCaMP (orange), and dLight (green) for beginner (*left*), intermediate (*middle*), and expert (*right*)

sessions. Contribution of each category was calculated by a method described in a. Kinematic variables include speed, acceleration, and rotation variables. Other categories are assigned to an individual variable (a set of time shifted variables). Mean contributions to 3 signals were compared by one-way RM ANOVA for each variable category. Plotted as mean \pm SEM across mice (n=10 mice).

c, Comparison of model fits for VTA jRCaMP (red), NAc jRCaMP (orange), and dLight (green) for beginner (*left*), intermediate (*middle*), and expert (*right*) sessions plotted as mean \pm SEM across mice (n=10 mice). Model fit was estimated by the correlation between actual and predicted signals from the model. Left set of bars represent correlations during a full duration (-40~+80s respect to the trigger zone entry). Right set of bars represent correlations during a trial duration (-5+15s respect to the trigger zone entry). Model fits for 3 signals were compared by one-way RM ANOVA. (*) $p < 0.10$ for one-way RM ANOVA (Bonferroni corrected).



Extended Data Fig. 4 l. Bidirectional modulation of DAN activity during behavior

a. Schematic describing the experimental procedure. *left*, Expression of stGtACR2, ChrimsonR, and GCaMP6f in DANs. *middle*, Injection of three viral vectors into VTA of a *DAT-IRES-CRE* mouse. *right*, The NAc fiber was used to collect terminal GCaMP signal. The VTA fiber was used to collect somatic GCaMP signal and to optogenetically activate (ChrimsonR) or inactivate (stGtACR2) DANs.

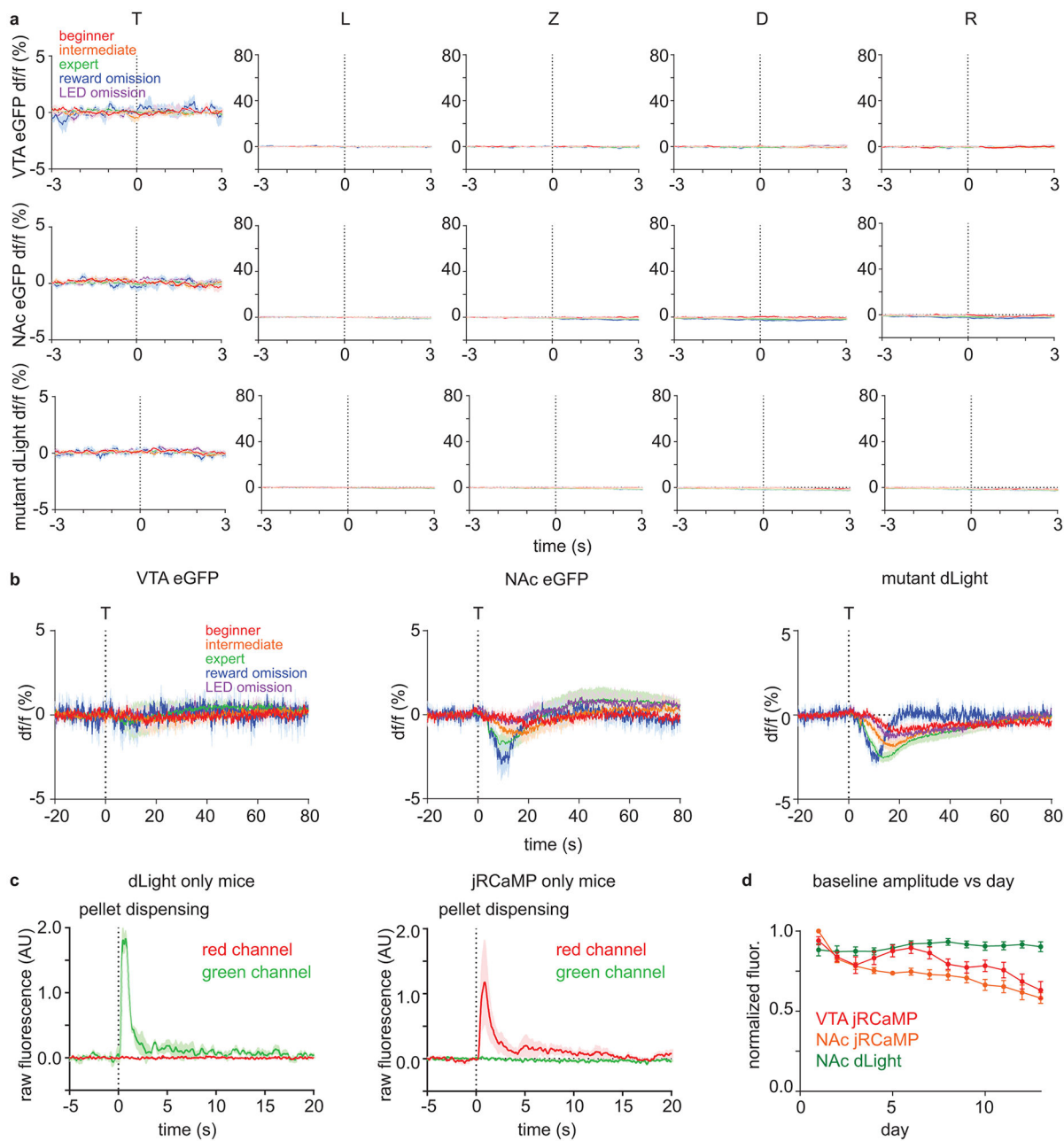
b. VTA GCaMP (*top*) and NAc GCaMP (*middle*) response (normalized to 99 percentile of the single session) to unpredicted rewards in naïve (untrained) mice. Mean NAc GCaMP signal normalized to the control response was compared between control and inactivation trials (*bottom*, paired t-test, $p=0.002$). Dotted line represents time of pellet dispensing.

c. VTA GCaMP (*top*) and NAc GCaMP (*middle*) response (normalized to 99 percentile of trained animal sessions) to reward predictive LED cue in trained mice. Comparison of mean NAc GCaMP signal during control and inactivation trials (*bottom*, paired t-test, $p=0.027$). Dotted line represents time of LED onset.

d. VTA GCaMP (*top*) and NAc GCaMP (*middle*) response (normalized to 99 percentile of trained animal sessions) to predicted reward (reward following LED cue) in trained mice. Comparison of mean NAc GCaMP signal during control and inactivation trials (*bottom*, paired t-test, $p=0.065$). Dotted line represents time of pellet dispensing.

e. VTA GCaMP (*top*) and NAc GCaMP (*middle*) response (normalized to 99 percentile of trained animal sessions) to reward omission in trained mice. Comparison of mean NAc GCaMP signal during control and activation trials (*bottom*, 2 of 3 lines overlapping, paired t-test, $p=0.029$). Dotted line represents time of expected reward delivery.

(*) $p<0.10$, * $p<0.05$, ** $p<0.01$ for paired t-test. All graphs are plotted as mean \pm SEM across mice ($n=3$ mice) and dot=mouse average. Average of signal (0s-10s) normalized to the control response was used for the comparison between control and optogenetic trials. The average response of each mouse was calculated from 5-6 trials. Blue and red bars indicate the periods of blue laser illumination for stGtACR2 and red laser illumination for ChrimsonR, respectively. VTA GCaMP signal could not be collected for blue laser illumination period due to optical crosstalk. All t-tests are two-sided.



Extended Data Fig. 5 l. Movement and optical artifacts cannot explain dLight and jRCaMP signal patterns

a. df/f (%) of different controls. The average signals for beginner, intermediate, expert, reward omission (of expert mice), and rewarded LED omission (of expert mice) trials are shown in red, orange, green, blue, and purple, respectively. Dashed vertical lines indicate the behavioral time stamps (T=trigger zone entry, L=LED on, Z=receptacle zone entry, D=pellet dispensing, R=receptacle entry). Top: df/f (%) of eGFP signal from VTA of *DAT-IRES-Cre* mice (n=4 mice) that were injected with AAV1-Cag-FLEX-eGFP into VTA. Middle: df/f (%) of eGFP signal from NAc of *DAT-IRES-Cre* mice (n=4 mice) that were injected with AAV1-Cag-FLEX-eGFP into VTA. Bottom: df/f (%) of DA binding mutant dLight (D103A

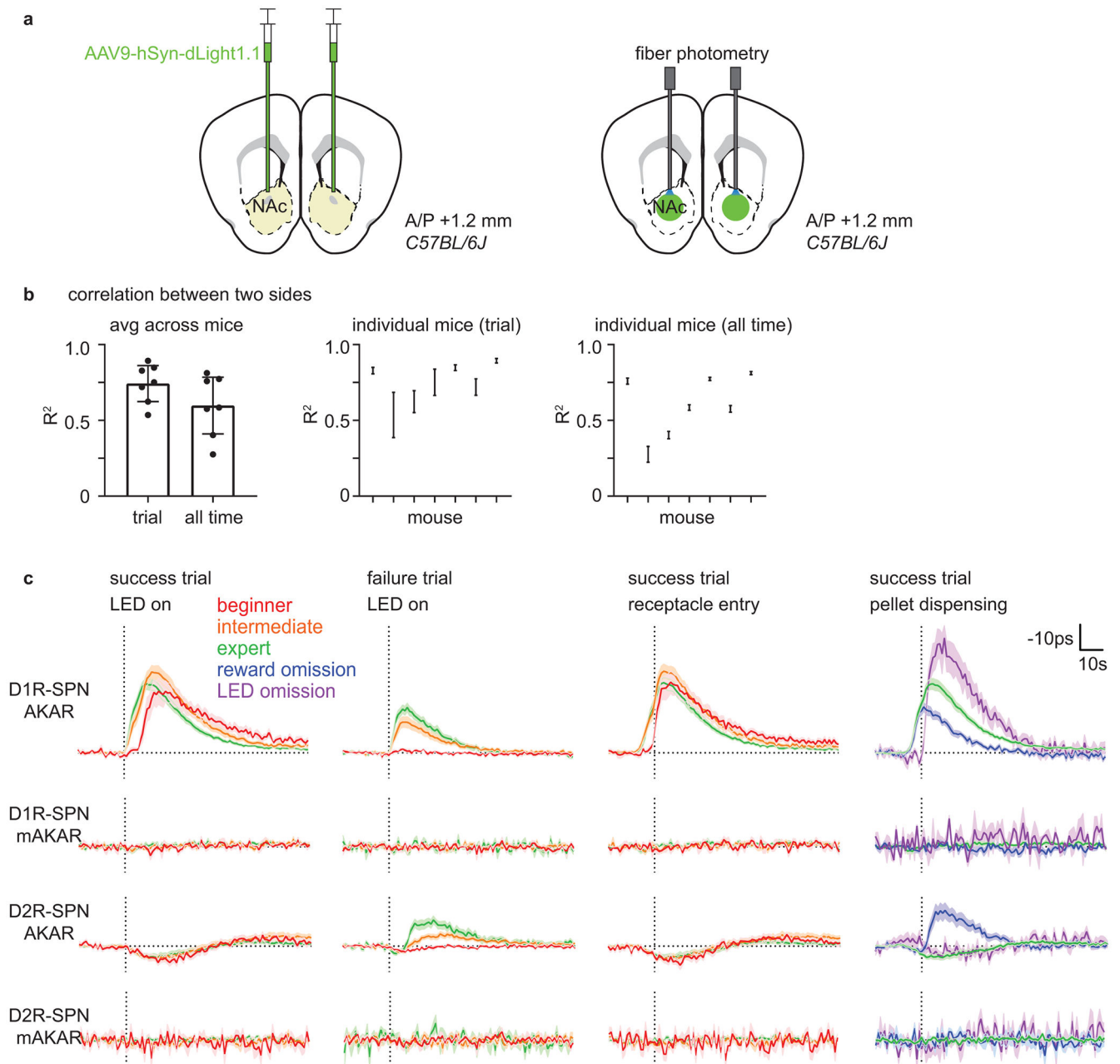
mutation) signal from NAc of *C57BL/6J* mice (n=8 mice) that were injected with AAV9-hSyn-dLight^{D103A} into NAc.

b, df/f (%) of different controls that are magnified in df/f axis and de-magnified in time axis. VTA eGFP (*left*, n=4 mice), NAc eGFP (*middle*, n=4 mice), and NAc mutant dLight (*right*, n=8 mice) signal aligned to the time of trigger zone entry (dashed vertical line). There was a minor (compared to sensor responses) change in NAc eGFP and mutant dLight signal that develops across learning (possibly due to hemodynamic effects).

c, Test for the optical crosstalk between green and red spectrum for simultaneous dual color photometry for dLight and jRCaMP. Mice were given unexpected free food pellets, and signal was aligned to the time of pellet dispensing. *left*, Raw fluorescence signal in red and green spectrum from NAc of *C57BL/6J* mice (n=3 mice, 10 trials/mouse) injected with AAV9-hSyn-dLight1.1 into NAc. *right*, Raw fluorescence signal in red and green spectrum from NAc of *DAT-IRES-Cre* mice (n=3 mice, 10 trials/mouse) injected with AAV1-hSyn-FLEX-NES-jRCaMP1b into VTA.

d, Baseline (pre-trial) raw fluorescence estimating the change in a signal strength due to photo-bleaching and viral expression change across days. Raw fluorescence were normalized by the maximum value of each mouse across all sessions. Error bar=SEM of mouse averages. n=10 mice.

All graphs are plotted as mean \pm SEM across mice (n=10 mice).

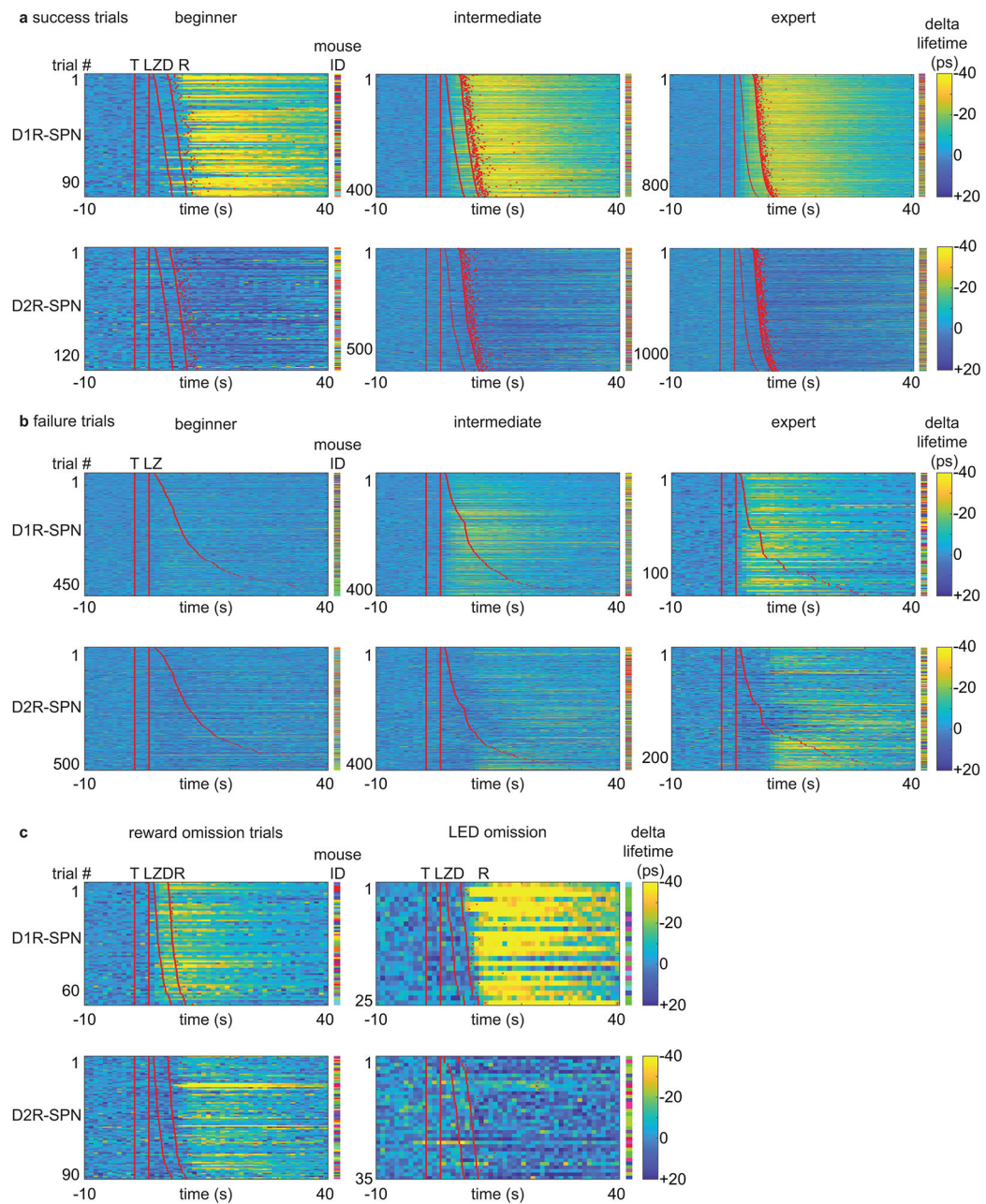


Extended Data Fig. 6 | Bilateral dLight measurement and mutant FLIM-AKAR control experiments

a, Schematic describing a strategy to measure DA level in both hemispheres. AAV9-hSyn-dLight1.1 was bilaterally injected into NAc of *C57BL/6J* mice. Then, two optical fibers were implanted 200µm above the injection sites in two hemispheres.

b, Relationship between dLight signals from two hemispheres. *left*, Average correlation between two dLight signals (df/f) during trial duration ($-5+15s$ respect to the trigger zone entry) and all time (analyzed in 20s time window for each linear fit) plotted as mean \pm 95% CI across mouse averages (two-sided paired t-test, $p=0.002$). *middle*, Correlation of

individual mouse for trial duration plotted as mean \pm 95% CI across session averages for each mouse. *right*, Correlation of individual mouse for all time plotted as middle. $n=7$ mice. **c**, Comparison between FLIM-AKAR and FLIM-AKAR^{T391A}, which has a point mutation at the PKA phosphorylation site, signals. AAV1-FLEX-FLIM-AKAR or AAV1-FLEX-FLIM-FLIM-AKAR^{T391A} was injected into NAc of *Drd1a-Cre* or *Adora2a-Cre* mice for these experiments. From the top, D1R-SPN FLIM-AKAR (D1R-SPN AKAR), D1R-SPN FLIM-AKAR^{T391A} (D1R-SPN mAKAR), D2R-SPN FLIM-AKAR (D2R-SPN AKAR), D2R-SPN FLIM-AKAR^{T391A} (D2R-SPN mAKAR). From the left, signals were aligned to the time (dashed vertical line) of “LED on” for success and failure trials separately, “receptacle entry” for success trials, and “pellet dispensing” for success, reward omission, and LED omission trials. Signals for beginner, intermediate, expert, reward omission (of expert mice), rewarded LED omission (of expert mice) trials are shown in red, orange, green, blue, and purple, respectively. Plotted as mean \pm SEM across mice. $n=14$ mice (D1R-SPN AKAR), 7 mice (D1R-SPN mAKAR), 18 mice (D2R-SPN AKAR), 6 mice (D2R-SPN mAKAR).

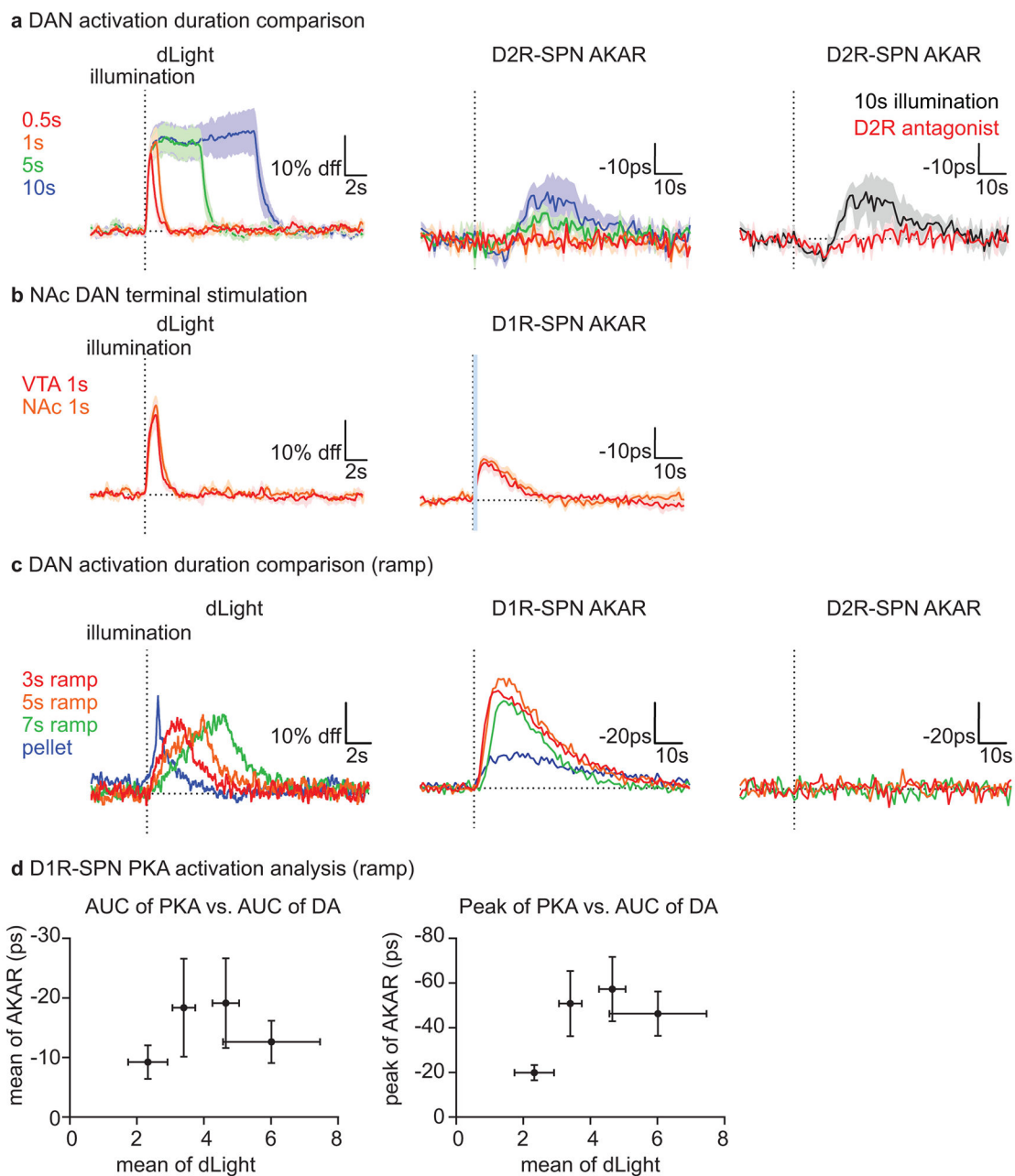


Extended Data Fig. 7 l. Plasticity in SPN PKA activity patterns during learning

a, Heatmaps of SPN FLIM-AKAR response for success trials during learning. Each row represents an individual trial. Red lines or dots indicate behavioral time stamps (T=trigger zone entry, L=LED on, Z=receptacle zone entry, D=pellet dispensing, R=receptacle entry). Different colors in mouse ID columns represent different mice for an individual row. Top: D1R-SPN FLIM-AKAR responses of *Drd1a-Cre* mice. n=98 trials (beginner), 418 trials (intermediate), 873 trials (expert) from 14 mice. Bottom: D2R-SPN FLIM-AKAR responses of *Adora2a-Cre* mice. n=134 trials (beginner), 596 trials (intermediate), 1152 trials (expert) from 18 mice.

b, Heatmaps of SPN FLIM-AKAR response for failure trials during learning. Plotted as a. Top: D1R-SPN FLIM-AKAR responses of *Drd1a-Cre* mice. n=497 trials (beginner), 402 trials (intermediate), 122 trials (expert) from 14 mice. Bottom: D2R-SPN FLIM-AKAR responses of *Adora2a-Cre* mice. n=528 trials (beginner), 416 trials (intermediate), 218 trials (expert) from 18 mice.

c, Heatmaps of SPN FLIM-AKAR response for reward omission trials (*left*) and rewarded LED omission trials (*right*). Plotted as a. Top: D1R-SPN FLIM-AKAR responses of *Drd1a-Cre* mice. n=69 trials (reward omission) from 14 mice, 25 trials (rewarded LED omission trials) from 6 mice. Bottom: D2R-SPN FLIM-AKAR responses of *Adora2a-Cre* mice. n=91 trials (reward omission) from 18 mice, 35 trials (rewarded LED omission trials) from 10 mice.



Extended Data Fig. 8 I. Transient change in DAN activity is sufficient to modulate SPN PKA activity in NAc

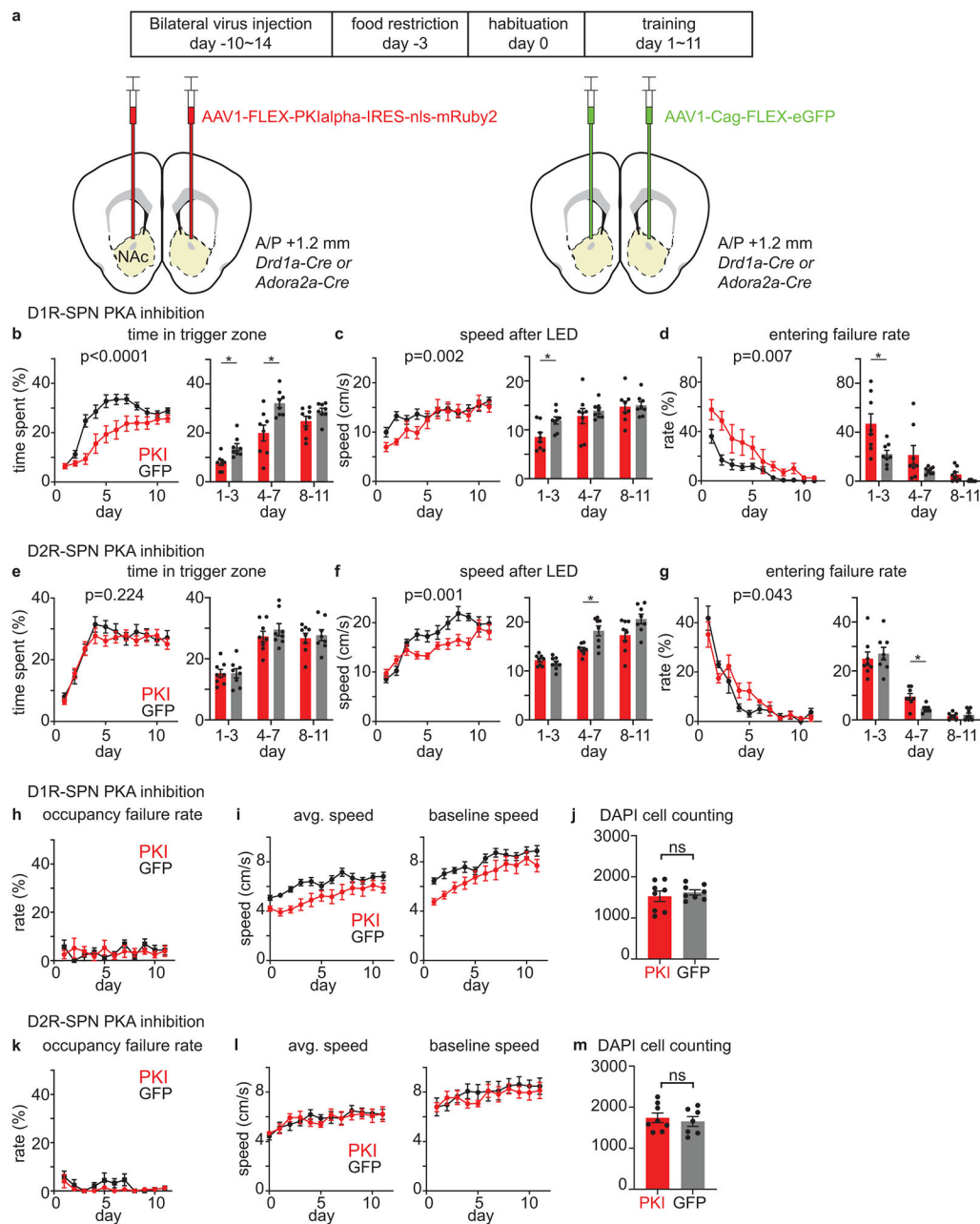
a, *left*, dLight responses in *DAT-IRES-Cre* mice ($n=3$ mice) to DAN activation (20Hz, 2ms pulse width, 14.3mW illumination) for different durations of illumination (red=0.5s, orange=1s, green=5s, blue=10s, one-way RM ANOVA, $F(1.064,2.127)=35.89$, $p=0.023$). *middle*, D2R-SPN FLIM-AKAR responses in *DAT-IRES-Cre; Adora2a-Cre* mice ($n=4$ mice) plotted in the same way as the left (one-way RM ANOVA, $F(1.353,4.059)=4.140$, $p=0.109$). *right*, D2R-SPN FLIM-AKAR responses in *DAT-IRES-Cre; Adora2a-Cre* mice ($n=4$ mice) to 10s illumination without (black) and with (red) IP injection D2R antagonist at least 10mins before recording (paired t-test, valley: $p=0.338$, peak: $p=0.033$). Statistics

performed on mean dLight (0-10s) and AKAR (end of illumination to +20s except for valley estimation (0-10s) for D2R antagonist experiment) signal. Note: To test if D2R-SPN PKA activity can respond to DAN activation at all, we activated DAN for 10s, which increases DA levels far more than does a natural food reward response (b, *left*). This non-physiological level of DA release results in a bidirectional modulation of D2R-SPN net PKA activity (b, *right*) with net PKA activity slightly decreasing and then increasing. However, given that D2R antagonist does not significantly affect the initial reduction in PKA activity, this reduction is unlikely to be D2R mediated. On the other hand, the delayed activation of PKA was blunted by D2R antagonist, suggesting a contribution of indirect circuit mechanisms, such as modulation of the activity of D2R-expressing cholinergic interneurons in the NAc. b, dLight and D1R-SPN FLIM-AKAR responses to DAN terminal stimulation (20Hz, 2ms pulse width) in NAc (red=VTA DAN stimulation for 1s/10.5mW, orange=DAN terminal stimulation for 1s/7.7mW). *left*, dLight responses in *DAT-IRES-Cre* mice (n=3 mice, paired t-test, p=0.429). *right*, D1R-SPN FLIM-AKAR responses in *DAT-IRES-cre; Drd1a-Cre* mice (n=6 mice, paired t-test, p=0.597). Statistics performed on mean dLight (0-3s) and AKAR (end of illumination to +20s) signal.

c, Optogenetic induction of ramping DA level change and consequent PKA activity change in SPNs to different illumination duration. *left*, dLight responses in *DAT-IRES-Cre* mice (n=3 mice) to ramping DAN activation for different ramping durations (red=3s, orange=5s, green=7s ramping activation) and to food reward (blue) (one-way RM ANOVA, $F(1.123, 2.246)=4.260$, p=0.163). To induce ramping DA level change, the frequency of stimulation was gradually increased from 24Hz to 34Hz for 3s (at 10.5mW), 16Hz to 34Hz for 5s (at 6.1mW), and 4Hz to 30Hz for 7s (at 10.5mW). *middle*, D1R-SPN FLIM-AKAR responses in *DAT-IRES-Cre; Drd1a-Cre* mice (n=4 mice) plotted in the same way as the left (one sample t-test on 7s ramp, p=0.038). *right*, D2R-SPN FLIM-AKAR responses in *DAT-IRES-Cre; Adora2a-Cre* mice (n=4 mice) plotted in the same way as the left (one sample t-test on 7s ramp, p=0.779). All t-tests are two-sided.

d, D1R-SPN PKA activation vs. DA release analysis. *left*, mean of D1R-SPN AKAR (0s-80s) vs. mean of dLight (0s-20s) for different stimulations (reward, 3s ramp, 5s ramp, and 7s ramp). *right*, peak of D1R-SPN AKAR (0s-80s) vs. mean of dLight (0s-20s) for different stimulations. Each data point represents the average and the standard error across mice (n=3 mice for dLight, n=4 mice for AKAR).

All t-tests are two-sided. All graphs are plotted as mean \pm SEM (if shaded) across mice. Dashed vertical line=illumination onset. The average response of each mouse was calculated from 10 trials. Blue bars indicate the periods of laser illumination (NAc) for ChrimsonR during which accurate FLIM-AKAR measurements were not possible. Pellet response for dLight was aligned to the peak after receptacle entry and time shifted so that the upward slope starts near 0s. Pellet response for D1R-SPN AKAR was aligned to the receptacle entry. Statistics for c were performed on mean dLight signal (0-20s) and mean AKAR signal (0-80s).



Extended Data Fig. 9 l. Selective PKA inhibition in SPNs slows learning

a, Schematic describing a strategy to investigate the effect of D1R-SPN or D2R-SPN PKA inhibition on behavior. AAV1-FLEX-PKIalpha-IRES-nls-mRuby2 was injected into NAc of *Drd1a-Cre* or *Adora2a-Cre* mice to selectively inhibit PKA in D1R-SPN or D2R-SPN, respectively. For control groups, AAV1-Cag-FLEX-eGFP was injected, instead. 10-14 days after surgery, mice were started on a behavior schedule that includes 1 day of habituation (day 0) and 11 days of training (day 1~11).

b, Effect of D1R-SPN PKA inhibition on the fraction of time spent in trigger zone (time spent in trigger zone/total session time). Daily average time (*left*, two-way RM ANOVA day

x group interaction, $F(10,140)=5.565$, $p<0.0001$) and multiday average (*right*, unpaired t-test, $p=0.012$, 0.017 , 0.323).

c, Effect of D1R-SPN PKA inhibition on the speed after LED (average speed during 0-1.2s after LED onset. 1.2s is the minimum latency to enter the receptacle zone after LED cue). Daily average speed after LED onset (*left*, two-way RM ANOVA day x group interaction, $F(10,140)=2.923$, $p=0.002$) and multiday average (*right*, unpaired t-test, $p=0.046$, >0.999 , >0.999).

d, Effect of D1R-SPN PKA inhibition on the entering failure rate (# of entering failure trials/ total # of trials). Daily average entering failure rate (*left*, two-way RM ANOVA day x group interaction, $F(10,140)=2.591$, $p=0.007$) and multiday average (*right*, unpaired t-test, $p=0.043$, 0.367 , 0.073).

e, Effect of D2R-SPN PKA inhibition on the time spent in trigger zone. Daily average time (*left*, two-way RM ANOVA day x group interaction, $F(10,140)=1.322$, $p=0.224$) and multiday average (*right*, unpaired t-test, $p>0.999$, >0.999 , >0.999).

f, Effect of D2R-SPN PKA inhibition on the speed after LED. Daily average speed after LED onset (*left*, two-way RM ANOVA day x group interaction, $F(10,140)=3.124$, $p=0.001$) and multiday average (*right*, unpaired t-test, $p=0.938$, 0.011 , 0.145).

g, Effect of D2R-SPN PKA inhibition on the entering failure rate. Daily average entering failure rate (*left*, two-way RM ANOVA day x group interaction, $F(10,140)=1.951$, $p=0.043$) and multiday average (*right*, unpaired t-test, $p>0.999$, $=0.011$, >0.999).

h, Occupancy failure rate (number of premature receptacle zone exit trials / number of receptacle zone entering success trials) of D1R-SPN PKA inhibition experiments (one-way RM ANOVA on GFP group, $F(3.213,22.49)=1.918$, $p=0.153$; one-way RM ANOVA on PKI group, $F(2.424,16.97)=0.352$, $p=0.747$).

i, *left*, Average speed (total distance travelled / total session time, two-way RM ANOVA group effect, $F(1,14)=6.506$, $p=0.023$). *right*, Average speed during baseline period (-20s before trigger zone entry to trigger zone entry, two-way RM ANOVA group effect, $F(1,14)=4.304$, $p=0.057$).

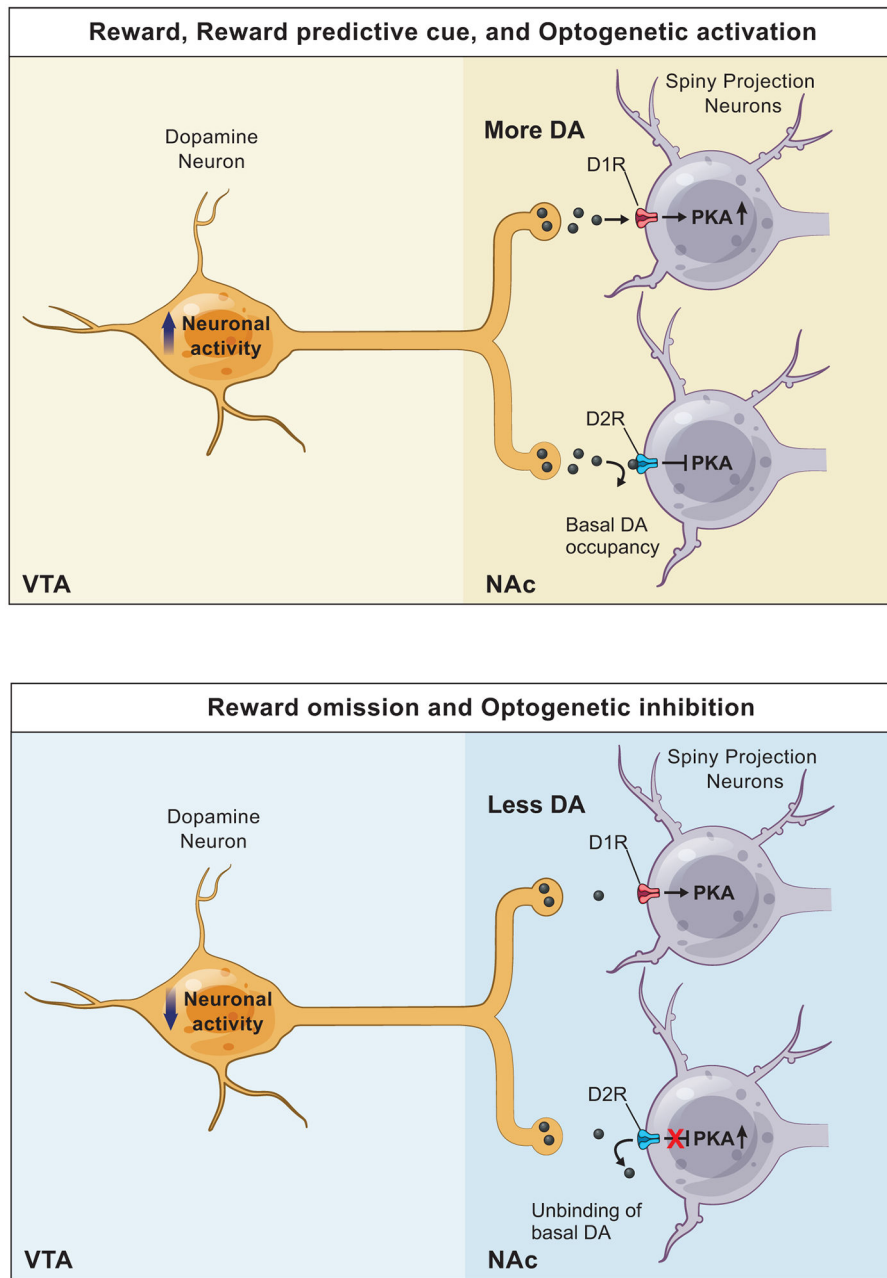
j, DAPI cell counting for D1R SPN PKA inhibition (unpaired t-test, $p=0.517$).

k, Occupancy failure rate of D2R-SPN PKA inhibition experiments (one-way RM ANOVA on GFP group, $F(2.293,16.05)=2.062$, $p=0.155$; one-way RM ANOVA on PKI group, $F(1.598,11.19)=1.482$, $p=0.264$).

l, *left*, Average speed (two-way RM ANOVA group effect, $F(1,14)=0.011$, $p=0.916$). *right*, Average speed during baseline period (two-way RM ANOVA group effect, $F(1,14)=0.211$, $p=0.653$).

m, DAPI cell counting for D2R SPN PKA inhibition (unpaired t-test, $p=0.598$).

* $p<0.05$ for Bonferroni corrected unpaired t-tests. All t-tests are two-sided. Individual dots on the bar graph=a mouse. All graphs are plotted as mean \pm SEM across mice ($n=8$ for each group).



Extended Data Fig. 10 | Model of DA action on SPN PKA activity

Overview of DA action on SPN PKA activity.

top, When DAN activity increases in response to a reward, reward predictive cue or optogenetic activation of DANs, more DA is released from DAN terminals in the NAc. This increase in DA level allows DA to bind to D1R, which increases the activity of adenylyl cyclase, the level of cAMP, and ultimately the activity of PKA in D1R-SPNs. In contrast, the increase in DA level has a minimal impact on D2R, which is occupied by the basal level of DA.

bottom, When DAN activity decreases in response to a reward omission or optogenetic inhibition of DANs, DA release from DAN terminals in the NAc decreases below the

baseline. This decrease in DA level has a minimal impact on D1R, which is not occupied by the basal DA. In contrast, the decrease in DA level allows the basal DA to unbind from D2R, which disinhibits PKA activity in D2R-SPNs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by NIH (B.L.S.: U19NS113201 and R35NS105107; Y.C.: F32DA035543; L.T.: U01NS013522 and U01NS090604), Howard Hughes Medical Institute (B.L.S.), Sackler Scholar Programme in Psychology (S.L.), Schuurman Schimmel van Outeren Stichting (B.L.) and Hendrik Muller fonds (B.L.). Graphical illustration was provided by Sigrid Knemeyer (sigrid@scistories.com).

References

1. Bromberg-Martin ES, Matsumoto M & Hikosaka O Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* 68, 815–834 (2010). [PubMed: 21144997]
2. Kravitz AV & Kreitzer AC Striatal mechanisms underlying movement, reinforcement, and punishment. *Physiology* 27, 167–177 (2012). [PubMed: 22689792]
3. Vidal-Gadea AG & Pierce-Shimomura JT Conserved role of dopamine in the modulation of behavior. *Commun. Integr* 5, 440–447 (2012).
4. Steinberg EE et al. Positive reinforcement mediated by midbrain dopamine neurons requires D1 and D2 receptor activation in the nucleus accumbens. *PLoS One* 9, doi: 10.1371/journal.pone.0094771 (2014).
5. Hikida T, Kimura K, Wada N, Funabiki K & Nakanishi S Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron* 66, 896–907 (2010). [PubMed: 20620875]
6. Tsai HC et al. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* 324, 1080–1084 (2009). [PubMed: 19389999]
7. Steinberg EE et al. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci* 16, 966–973 (2013). [PubMed: 23708143]
8. Saunders BT, Richard JM, Margolis EB & Janak PH Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat. Neurosci* 21, 1072–1083 (2018). [PubMed: 30038277]
9. Coddington LT & Dudman JT The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci* 21, 1563–1573 (2018). [PubMed: 30323275]
10. Schultz W, Dayan P & Montague PR A neural substrate of prediction and reward. *Science* 275, 1593–1600 (1997). [PubMed: 9054347]
11. Cohen JY, Haesler S, Vong L, Lowell BB & Uchida N Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482, 85–88 (2012). [PubMed: 22258508]
12. Eshel N, Tian J, Bukwich M & Uchida N Dopamine neurons share common response function for reward prediction error. *Nat. Neurosci* 19, 479–486 (2016). [PubMed: 26854803]
13. Day JJ, Roitman MF, Wightman RM & Carelli RM Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci* 10, 1020–1028 (2007). [PubMed: 17603481]
14. Shen W, Flajolet M, Greengard P & Surmeier DJ Dichotomous control dopaminergic of striatal synaptic plasticity. *Science* 321, 848–851 (2008). [PubMed: 18687967]
15. Gerfen CR et al. D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* 250, 1429–1432 (1990). [PubMed: 2147780]
16. Kupchik YM et al. Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. *Nat. Neurosci* 18, 1230–1232 (2015). [PubMed: 26214370]

17. Skeberdis VA et al. Protein kinase A regulates calcium permeability of NMDA receptors. *Nat. Neurosci* 9, 501–510 (2006). [PubMed: 16531999]
18. Lee HK et al. Phosphorylation of the AMPA receptor GluR1 subunit is required for synaptic plasticity and retention of spatial memory. *Cell* 112, 631–643 (2003). [PubMed: 12628184]
19. Yagishita S et al. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345, 1616–1620 (2014). [PubMed: 25258080]
20. Iino Y et al. Dopamine D2 receptors in discrimination learning and spine enlargement. *Nature* 579, (2020).
21. Lau GC, Saha S, Faris R & Russek SJ Up-regulation of NMDAR1 subunit gene expression in cortical neurons via a PKA-dependent pathway. *J. Neurochem* 88, 564–575 (2004). [PubMed: 14720206]
22. Nayak A, Zastrow DJ, Lickteig R, Zahniser NR, Browning MD, Maintenance of late-phase LTP is accompanied by PKA-dependent increase in AMPA receptor synthesis. *Nature* 396, 482–486 (1998). [PubMed: 9853758]
23. Lee SJ, Chen Y, Lodder B & Sabatini BL Monitoring behaviorally induced biochemical changes using fluorescence lifetime photometry. *Front. Neurosci* 13, doi: 10.3389/fnins.2019.00766 (2019).
24. Chen Y, Saulnier JL, Yellen G & Sabatini BL A PKA activity sensor for quantitative analysis of endogenous GPCR signaling via 2-photon FRET-FLIM imaging. *Front. Pharmacol* 5, doi: 10.3389/fphar.2014.00056 (2014).
25. Chen Y et al. Endogenous Gαq-coupled neuromodulator receptors activate protein kinase A. *Neuron* 96, 1070–1083 (2017). [PubMed: 29154125]
26. Mohebi A et al. Dissociable dopamine dynamics for learning and motivation. *Nature* 570, 65–70 (2019). [PubMed: 31118513]
27. Dana H et al. Sensitive red protein calcium indicators for imaging neural activity. *ELife* 5, 1–24 (2016).
28. Patriarchi T et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* 360, 1420–1428 (2018).
29. Klapoetke NC et al. Independent optical excitation of distinct neural populations. *Nat. Methods* 11, 338–346 (2014). [PubMed: 24509633]
30. Mahn M et al. High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nat. Commun* 9, doi:10.1038/s41467-018-06511-8 (2018).
31. Howe MW, Tierney PL, Sandberg SG, Phillips PEM & Graybiel AM Prolonged dopamine signaling in striatum signals proximity and value of distant rewards. *Nature* 500, 575–579 (2013). [PubMed: 23913271]
32. Matamales M et al. Local D2- To D1-neuron transmodulation updates goal-directed learning in the striatum. *Science* 367, 549–555 (2020). [PubMed: 32001651]
33. Jiang SZ et al. NCS-Rapgef2, the protein product of the neuronal Rapgef2 gene, is a specific activator of D1 dopamine receptor-dependent ERK phosphorylation in mouse brain. *eNeuro* 4, doi: 10.1523/ENEURO.0248-17.2017 (2017)
34. Ilango A et al. Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci* 34, 817–822 (2014). [PubMed: 24431440]
35. Goto A et al. Circuit-dependent striatal PKA and ERK signaling underlies rapid behavioral shift in mating reaction of male mice. *PNAS* 112, 6718–6723 (2015). [PubMed: 25964359]
36. Yamaguchi T et al. Role of PKA signaling in D2 receptor-expressing neurons in the core of the nucleus accumbens in aversive learning. *PNAS* 112, 11383–11388 (2015). [PubMed: 26305972]
37. Ma L et al. A highly sensitive A-kinase activity reporter for imaging neuromodulatory events in awake mice. *Neuron* 99, 1–15 (2018). [PubMed: 30001504]
38. Collins AGE, & Frank MJ Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev* 121, 337–366 (2014). [PubMed: 25090423]
39. Gurney KN, Humphries MD, & Redgrave P A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interface. *PLoS Bio.* 13, doi: 10.1371/journal.pbio.1002034 (2015)

40. Gerfen CR, & Surmeier DJ Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci* 34, 441–466 (2011) [PubMed: 21469956]
41. Gerfen CR, Paletzki R & Heintz N GENSAT BAC Cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* 80, 1368–1383 (2013). [PubMed: 24360541]
42. Backman CM et al. Characterization of a mouse strain expressing Cre recombinase from the 3' untranslated region of the dopamine transporter locus. *Genesis* 44, 383–90 (2006). [PubMed: 16865686]
43. Lee SR, Escobedo-lozoya Y, Szatmari EM, & Yasuda R Activation of CaMKII in single dendritic spines during long-term potentiation. *Nature* 458, 299–304 (2009). [PubMed: 19295602]
44. Pnevmatikakis EA et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* 89, 285–299 (2016). [PubMed: 26774160]
45. Motulsky HJ How to report the methods used for the mixed model analysis. https://www.graphpad.com/guides/prism/8/statistics/stat_how-to-report-the-methods-used.htm (2020).

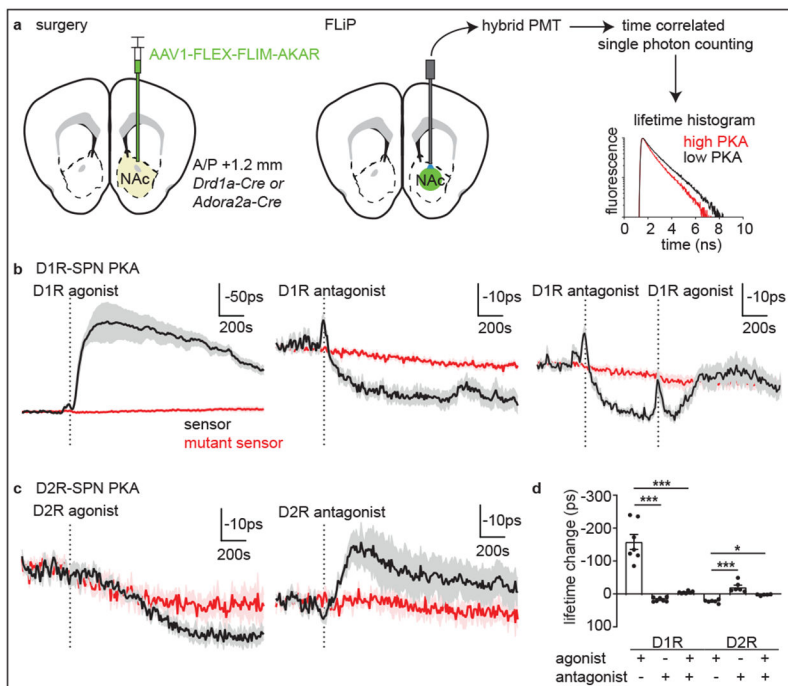


Fig. 1 | Fluorescence lifetime photometry (FLiP) reveals bidirectional changes in SPN PKA activity *in-vivo*

a, Schematic describing viral injection and optical fiber implantation (*left*). A hybrid PMT and time-correlated single-photon counting were used to measure the fluorescence lifetime of FLiM-AKAR (*right*) which shortens upon PKA phosphorylation.

b, Average lifetime changes of FLiM-AKAR (black) and FLiM-AKAR^{T391A} (red) expressed in D1R-SPNs in response to D1R agonist (*left*), antagonist (*middle*), and antagonist + agonist (*right*) ($n=7$ mice). Dashed lines indicate the end of the injection. Plotted as mean \pm SEM across mice.

c, Average FLiM-AKAR lifetime changes in D2R-SPNs in response to D2R agonist (*left*) and antagonist (*right*) ($n=6$ mice for FLiM-AKAR, 5 for FLiM-AKAR^{T391A}) plotted as b.

d, FLiM-AKAR lifetime changes (in b-c) in D1R-SPNs following D1R agonist and/or antagonist injections or those in D2R-SPNs following injection of D2R agonists and/or antagonists plotted as mean \pm SEM across mice. * $p<0.05$, *** $p<0.001$, Bonferroni-corrected post-hoc comparisons for one-way ANOVA.

For all figures, refer to Table S1 for exact p-values.

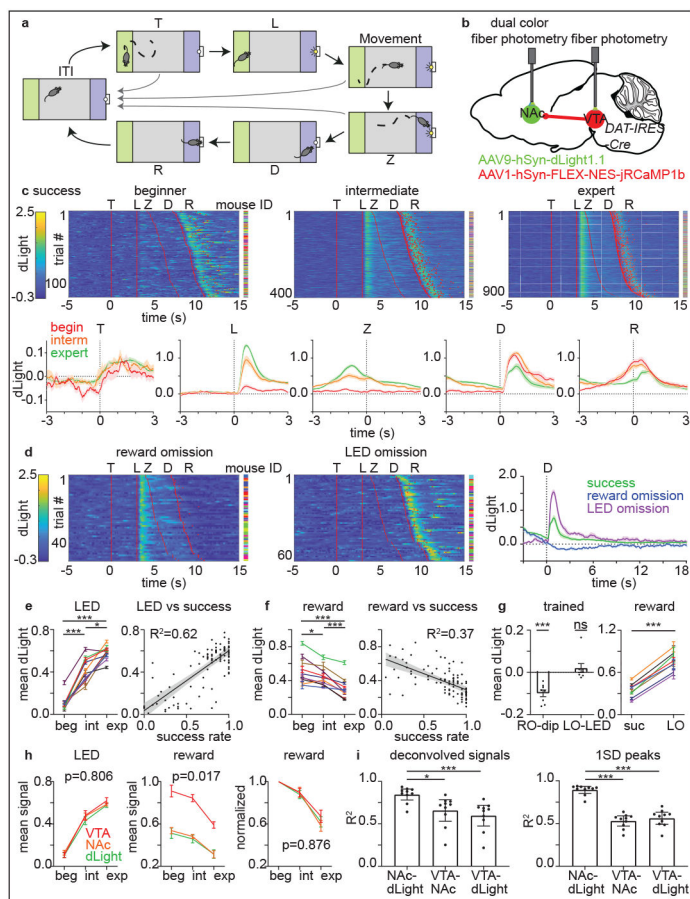


Fig. 2 |. Plasticity of DA release and DAN activity dynamics across learning

a, After the inter-trial interval (≥ 120 s), a mouse can initiate a trial by entering (T) and staying in the trigger zone (green) for ≥ 3 s, which activates the LED (L). The mouse must, within 5s, subsequently enter (Z) and stay ≥ 3 s in the receptacle zone (blue) to trigger food reward dispensing (D), after which the mouse enters the receptacle (R) to collect the pellet. The LED turns off at trial-end whether due to success (food delivery) or failure (return to the ITI).

b, Strategy for simultaneous measurement of DA levels and DAN activity.

c, *top*, Heatmaps of dLight responses in success trials across learning (each row shows trial). Red lines or dots indicate behavioral timestamps. The mouse ID for each row is represented by a different color. *bottom*, Average responses across mice aligned to behavioral timestamps plotted as mean \pm SEM across $n=10$ mice.

d, dLight signals as above in reward-omission trials (*left*) and rewarded LED-omission trials (*middle*) for expert mice. *right*, Average signals. Plotted as c ($n=8$ mice).

e, LED-evoked dLight signals on average for individual mice (plotted as mean \pm SEM across trials) at different learning stages (*left*) and for daily sessions plotted versus success rate (plotted as linear regression fit \pm its 95% CI) (*right*). $n=9$ mice.

f, Reward-evoked dLight signals plotted as e. $n=9$ mice.

g, Responses of individual expert mice to reward omission (RO dip) and LED omission (LO-LED) plotted as mean \pm SEM across mice (*left*) and to reward in regular-success (suc) and LED-omission (LO) trials plotted as e. n=8 mice.

h, Average signals to LED (*left*), reward (*middle*), and reward normalized to beginner response (*right*) plotted as mean \pm SEM across mice (n=9) compared by two-way RM ANOVA period x signal interaction.

i, Correlation between pairs of fluorescence channels using deconvolved signals (*left*) or signal peaks (*right*) during trials plotted as mean \pm 95% CI across n=10 mice.

*p<0.05, ***p<0.001 for Bonferroni-corrected post-hoc comparisons for one-way RM ANOVA, one-sample t-tests, and paired t-tests. All t-tests are two-sided.

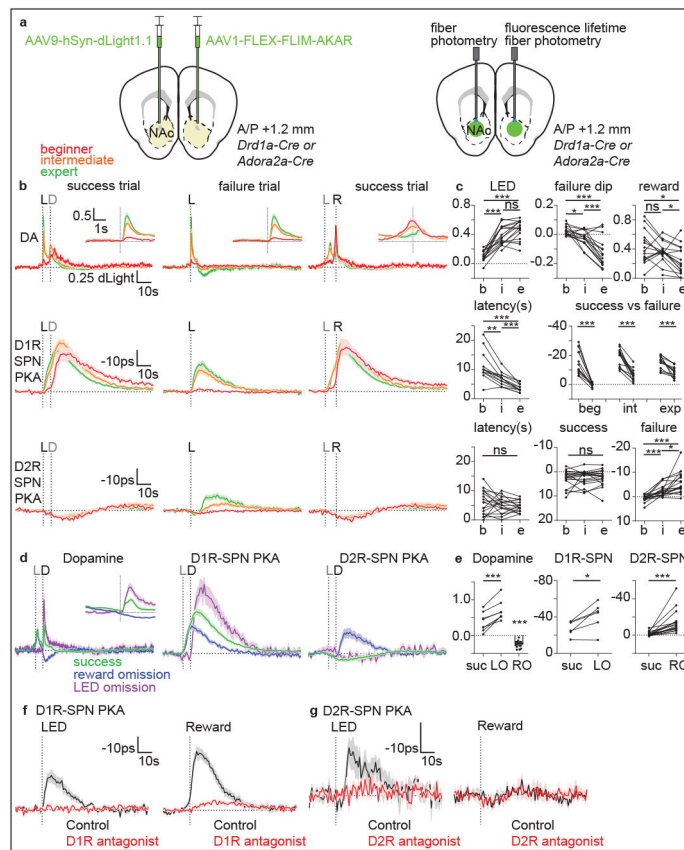


Fig. 3 | D1R- and D2R-SPN PKA activities are dynamically modulated and follow patterns of DA during learning

a. Strategy to measure SPN PKA activity and DA release.

b. Average responses of dLight of *Drd1a-Cre* and *Adora2a-Cre* mice ($n=16$) (top), D1R-SPN FLIM-AKAR in *Drd1a-Cre* mice ($n=14$) (middle), and D2R-SPN FLIM-AKAR in *Adora2a-Cre* mice ($n=18$) (bottom). Alignments are to (dashed vertical line) LED onset (L, black) for success (left) and failure (middle) trials or receptacle entry (R, black) for success trials (right). The average times across mice of pellet dispensing (D) and LED onset (L) are in gray.

c. top, dLight signal amplitudes after LED onset (left), in failure trials at typical time of reward delivery (middle), and to reward (right). middle, D1R-SPN PKA response latency (to 50% peak) and mean amplitude for success vs. failure trials. bottom, D2R-SPN PKA response latency (to 50% valley) and mean amplitudes for success and failure trials.

d. dLight and FLIM-AKAR responses in reward-omission and rewarded LED-omission trials aligned to pellet dispensing (D, black). Average time of LED onset (L) is in gray.

e. Average responses to rewards in regular-success (suc) and rewarded LED-omission (LO) trials, or at time of expected reward delivery in reward-omission (RO) trials for dLight (left), D1R-SPN PKA (middle), and D2R-SPN PKA (right).

f. Effects of D1R antagonism on D1R-SPN PKA activity transients ($n=4$) aligned to onset of an LED cue previously associated with food delivery but with no food delivered (left, 5 trials/mouse) or to a free (i.e. no action required) reward aligned to receptacle entry (right, 10 trials/mouse).

g, As in f for D2R-SPN FLIM-AKAR (n=4 mice) D2R antagonism.

All graphs are plotted as mean \pm SEM across mice except c and e where dots are mouse averages; *p<0.05, **p<0.01, ***p<0.001 for Bonferroni corrected post-hoc comparisons for one-way RM ANOVA, one sample t-tests, and paired t-tests. All t-tests are two-sided. Numbers of trials and mice per condition given in Extended Data Fig.7.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

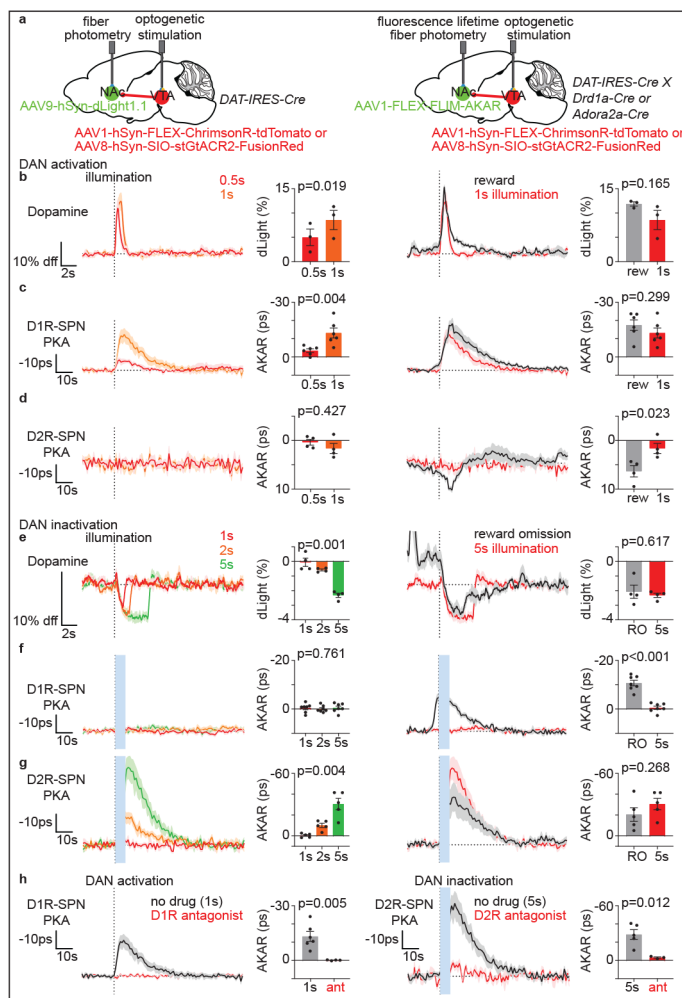


Fig. 4 | Transient changes in DAN activity are sufficient to modulate SPN PKA activity
a, Strategy for modulating DAN activity and measuring DA release (*left*) or SPN PKA activity (*right*).
b, dLight responses of *DAT-IRES-Cre* mice ($n=3$) to DAN activation (20Hz) for different durations (*left*) and of naïve animals to food reward compared to 1s illumination (*right*). Dashed line=time of illumination onset or reward delivery.
c, As in **b** showing D1R-SPN FLIM-AKAR responses of *DAT-IRES-Cre; Drd1a-Cre* mice ($n=6$).
d, As in **b** showing D2R-SPN FLIM-AKAR responses of *DAT-IRES-Cre; Adora2a-Cre* mice ($n=4$).
e, dLight responses of *DAT-IRES-Cre* mice ($n=4$) to DAN inactivation (stGtACR2) continuous illumination for different durations (*left*) and of trained mice to reward omission compared 5s inactivation (*right*). Dashed line=time of illumination onset or expected reward delivery. Baseline dLight signal in reward omission is due to the response to antecedent LED cue.
f, As in **e** for D1R-SPN FLIM-AKAR responses of *DAT-IRES-Cre; Drd1a-Cre* mice ($n=7$).

g, As in e for D2R-SPN FLIM-AKAR responses of *DAT-IRES-Cre; Adora2a-Cre* mice (n=5).

h, Responses of FLIM-AKAR responses in D1R-SPNs to DAN activation (1s) without (n=6) or with (n=4) injection of D1R antagonist (*left*) and in D2R-SPNs to DAN inactivation (5s) without (n=5) and with (n=3) injection of D2R antagonist (*right*).

All graphs are plotted as mean \pm SEM across mice. Dot=average for each mouse. Statistics: paired t-test, unpaired t-test, one-way RM ANOVA. All t-tests are two-sided. Blue bars=laser illumination.