Sequence termination cues drive habits via dopamine-mediated credit assignment

AUTHORS

Robin Magnard[1]; Yifeng Cheng[1,3]; Joanna Zhou[1]; Haley Province[1]; Nathalie Thiriet[4]; Patricia H. Janak[1,2,3]; Youna Vandaele[1,4]

AFFILIATIONS

[1]Department of Psychological and Brain Sciences, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD.
[2]Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD.
[3]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD.
[4]Université de Poitiers, INSERM, U-1084, Laboratoire des Neurosciences Expérimentales et Cliniques, Poitiers, France.

CORRESPONDENCE

Dr. Youna Vandaele, Université de Poitiers, INSERM, U-1084, Laboratoire des Neurosciences Expérimentales et Cliniques, Poitiers, France.

Phone: +33782507731

Email address: youna.vandaele@univ-poitiers.fr

Dr. Robin Magnard, Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA.

Phone: 410-516-7994

Email address: robin.magnard@jhu.edu

ABSTRACT

Mesolimbic dopamine (DA) neurons are central to sequence learning and habit formation. Yet, the mechanisms by which cue-induced DA neural activity drives goal-directed or habitual sequence execution remain unknown. We designed two novel tasks to investigate how sequence initiation and termination cues influence DA-driven behavioral strategies and learning. We found that sequence initiation and termination cues differentially affect reward expectation during action sequences, with only the termination cue contributing to greater outcome devaluation insensitivity, automaticity and behavioral chunking. Mesolimbic fiber photometry recording revealed that this habit-like behavior was associated with a rapid backpropagation in DA signals from the reward to the immediately preceding cue and with attenuated DA reward prediction error signals, which reflected greater behavioral inflexibility. Finally, in absence of external cues, brief optogenetic stimulation of VTA DA neurons at sequence termination was sufficient to drive automaticity and behavioral chunking. Our results highlight the critical role of cue-evoked DA signals at sequence termination in mediating credit assignment and driving the development of habitual action sequence execution.

KEYWORDS: Dopamine; Nucleus Accumbens; Ventral Tegmental Area; Habit; Behavioral Chunking; Credit assignment; Inflexibility; Reward Prediction Error

INTRODUCTION

Mesolimbic dopamine (DA) neurons play a critical role in cue-mediated reward expectation and sequence learning [1–5]. Notably, midbrain DA neurons encode reward prediction errors (RPE) [1–8], which reflect deviations from outcome expectation and drive learning by conferring motivational value to actions and reward-predictive cues [3,8]. Across sequence learning, DA signals in the nucleus accumbens (NAc) backpropagate from the reward and the action to the most distal reward-predictive cue [9–12]. Furthermore, it has been shown that mesolimbic DA neural activity predict performance of action sequences [9,10] and flexibly scale with distance to reward to sustain motivation toward a goal [13,14]. Although DA has been related to habit learning and response automaticity [15–19], it remains unclear how DA activity at action and reward-predictive cues changes over sequence learning when behavior is goal-directed, i.e. driven by outcome representation, and when it becomes automated and habitual.

Although habits are commonly identified as an absence of goal-directed control (i.e. insensitivity to both outcome devaluation and contingency degradation)[20], their description has been extended to include the formation of chunked action sequences that are automatically executed in absence of outcome expectation until their completion [21–23]. Previous findings indicate that the presence of sequence-delineating cues, such as those marking the beginning or the end of action sequences, plays a critical role in the development of habit, characterized by an insensitivity to outcome devaluation, a strong automaticity and behavioral chunking [24–26].

Here, we designed two new tasks to distinguish the role of sequence initiation and termination cues in this response strategy. We hypothesized that the sequence termination cue would favor habit by serving as response feedback and alleviating requirements for action monitoring. We propose that credit assignment, mediated by DA activity [27–29], might be a key mechanism linking the sequence termination cue to habit formation. Specifically, we hypothesize that DA signals assign value to the most salient element of the sequence. As learning progresses, DA-driven reinforcement may evolve from individual actions to the sequence as a whole. Such a shift may drive the transition from goal-directed to habitual behavior by consolidating and automating chunked action sequences. Thus, we used fiber photometry to determine how mesolimbic DA signal changes over training during cue-mediated sequence learning depending on the type of cue (sequence initiation vs. termination cue) and the nature of behavioral control (goal-directed versus habit-like). Finally, we used optogenetic manipulation to demonstrate the role of DA-mediated credit assignment in the development of automaticity and behavioral chunking.

RESULTS

Sequence termination cues promote habit-like behavior

To discern the respective role of sequence initiation cues and sequence termination cues in instrumental behavior, we designed two tasks in which the lever could either serve as the cue that triggers the initiation of the sequence of actions (i.e. lever insertion) or as the cue that signals its completion (i.e. lever retraction). We assessed different features of habit, namely, action automaticity, behavioral chunking, and insensitivity to outcome devaluation.

After initial training under a free-operant un-cued fixed-ratio 5 (FR5) schedule, rats were trained in one of two procedures designed to isolate the impact of the sequence initiation and termination cues (**Fig. 1a-b**). The first group of rats was trained in the lever insertion fixed-ratio 5 (LI5) task in which trials began with insertion of the lever. Rats were required to complete a sequence of 5 lever presses to obtain the reward (0.1 ml of sucrose 20%). In this LI5 task, reward delivery was

not signaled by lever retraction, which occurred at a random interval during reward retrieval (**Fig. 1a**). The second group of rats was trained in the lever retraction fixed-ratio 5 (LR5) task in which reward delivery was directly signaled by lever retraction, occurring at completion of five lever presses. However, the lever insertion cue was not relevant for signaling trial onset since it occurred during reward retrieval (**Fig. 1b**). These procedures allowed us to isolate and evaluate the behavioral impact of the lever insertion cue in the LI5 task and the lever retraction cue in the LR5 task.

Task differences rapidly emerged during training with a greater increase in response rate for rats trained in the LR5 task (**Fig. 1c**). Rats trained in the LR5 task exhibited a higher response rate in the last LR5 session compared to the first LR5 session and compared to the last LI5 session (**Fig. 1d**). The distribution of inter-press intervals (IPI) was comparable between tasks on the first session, with a similar proportion of IPIs shorter than 0.5-s (**Fig. 1e**), LR5 rats had a greater proportion of shorter IPIs on the last training session as compared with LI5 rats (**Fig. 1f**). This reduced IPI duration suggests greater action automaticity in the LR5 task that emphasizes the lever retraction cue at the end of sequence completion, compared to the LI5 task that emphasizes the lever insertion cue prior to sequence initiation.

Congruent with these observations, in the LR5 task, rats rapidly learned to complete the required ratio by pressing on the lever until its retraction. This resulted in a drastic decrease across LR5 sessions in the likelihood of checking the port during sequence execution (**Fig. 1g**). In contrast, in the LI5 task where there was no sequence termination cue, rats checked the reward port regularly, with port checking probability remaining near the maximum possible across sessions (**Fig. 1g-h**). The rapid increase in response rate concomitant with suppression of port checking during sequence execution in the LR5 task suggests that rats learned to concatenate individual lever presses into a single behavioral sequence, a process referred to as behavioral chunking, executing the sequence without reward expectation until its completion. Indeed, rats in the LR5 group made a majority of short sequences (1-3 lever presses before port checking) during the first session (**Fig. 1i**) but learned to complete the full lever press sequence without checking the port by the end of training (**Fig. 1j**). In contrast, rats trained in the LI5 task showed no evidence of behavioral chunking across training, as they persisted in port checking after every few lever presses from the first (**Fig. 1i**) to the last (**Fig. 1j**) training session.

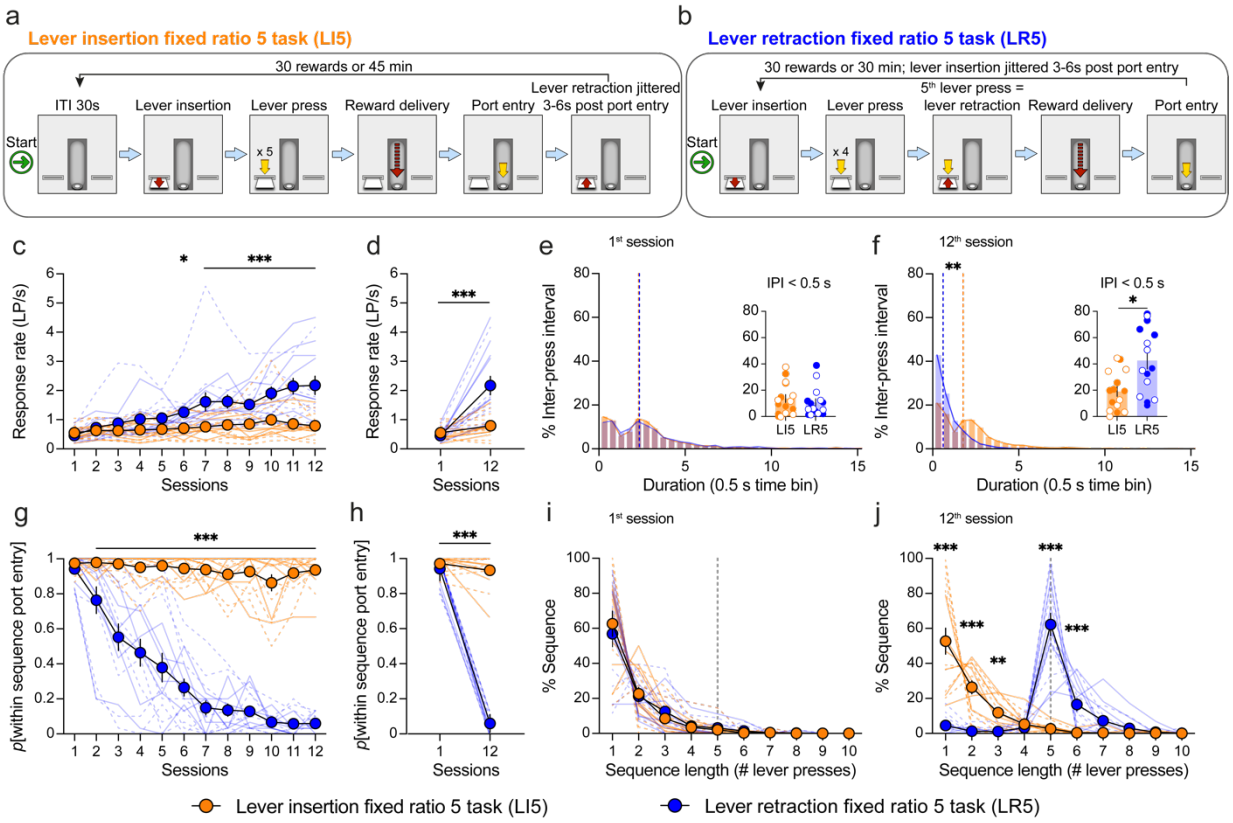

Figure 1. The lever retraction cue, but not the lever insertion cue, promotes action automaticity and behavioral chunking. (**a**) *Lever insertion fixed-ratio 5 (LI5) task design.* (**b**) *Lever retraction fixed-ratio 5 (LR5) task design.* (**c**) *Response rate in lever press per second (LP/s), within the lever press sequence during LI5/LR5 training sessions (two-way RM ANOVA, task effect:* $F_{1,28} = 10.6$, $p = 0.003$; *task x session interaction:* $F_{11,308} = 10.07$, $p < 0.001$; *Student Newman-Keuls post hoc pairwise comparison).* (**d**) *Response rate on the first versus last session in the LI5/LR5 tasks (two-way RM ANOVA, task effect:* $F_{1,28} = 11.54$, $p = 0.002$; *task x session interaction:* $F_{1,28} = 23.88$, $p < 0.001$; *Student Newman-Keuls post hoc pairwise comparison).* (**e**) *Frequency distribution of inter-press intervals (IPI) for the 1st LI5/LR5 session (Mann-Whitney Rank Sum t-test median comparison U = 455, p = 0.725). Insert panel: individual proportion of IPI under 0.5 s (Mann-Whitney Rank Sum t-test, U = 94, p = 0.455).* (**f**) *Frequency distribution of IPI for the 12th LI5/LR5 session (Mann-Whitney Rank Sum t-test median comparison U = 290, p = 0.007). Insert panel: individual proportion of IPI under 0.5 s (Mann-Whitney Rank Sum t-test, U = 51, p = 0.011).* (**g**) *Probability to enter the port within the lever press sequence (within-sequence port entry) during training in LI5/LR5 tasks (two-way RM ANOVA, task effect:* $F_{1,28} = 247.71$, $p < 0.001$; *task x session interaction:* $F_{11,308} = 39.95$, $p < 0.001$; *Student Newman-Keuls post-hoc pairwise comparison).* (**h**) *Probability of within sequence port entry on the first versus last training session in the LI5/LR5 tasks (two-way RM ANOVA, task effect:* $F_{1,28} = 483.56$, $p < 0.001$; *task x session interaction:* $F_{1,28} = 687.4$, $p < 0.001$; *Student Newman-Keuls post hoc pairwise comparison).* (**i**) *Frequency distribution of sequence lengths during the first training session in the LI5/LR5 tasks. The grey dashed line represents the optimal sequence length of 5 lever presses (two-way RM ANOVA, no task effect:* $F_{1,28} < 1$, $p = 0.808$; *no task x sequence length interaction:* $F_{9,252} < 1$, $p = 0.95$). (**j**) *Frequency distribution of sequence length during the last training session in the LI5/LR5 tasks. The grey dashed line represents the optimal sequence length of 5 lever presses (three-way RM ANOVA, no task effect:* $F_{1,28} < 1$, $p = 0.919$; *task x sequence length interaction:* $F_{9,252} = 49.1$, $p < 0.001$; *Student Newman-Keuls post-hoc pairwise comparison). Data*

*shown as mean ± SEM, superimposed with individual data point. LI5: n = 15 (7 males; 8 females); LR5: n = 15 (7 males; 8 females). Solid lines and filled circles: males; Dashed lines and empty circles: females. ITI: inter-trial interval. *p < 0.05; **p < 0.01; ***p < 0.001.*

Task differences in action automaticity and behavioral chunking were associated with differences in sensitivity to sensory-specific satiety-induced outcome devaluation (**Suppl. Fig. 1**). Although habitual responding is unexpected after the relatively short training used in these studies, we detected habit-like behavior, with a higher proportion of rats with insensitivity to devaluation (devaluation ratio ≤ 0) for the LR5 task compared to the LI5-task (**Suppl. Fig. 1**; 8/15 for LR5 task vs 1/15). We found no effect of sex on any of these performance variables (**Suppl. table 1**). Taken together, these results show that the lever retraction cue, rather than the lever insertion cue, promotes habit-like behavior, characterized by action automaticity, behavioral chunking, and greater propensity to show insensitivity to outcome devaluation.

Distinct patterns of VTA DA neuron activity in response to cues and reward depending on response strategy

To determine if DA dynamics during behavioral sequence execution depend on cue-mediated reward expectation and prediction, we next used fiber photometry to examine ventral tegmental area (VTA) DA neuron activity across training in the LR5 and LI5 tasks. For selective recording of DA neurons, we expressed GCaMP6f in the VTA of tyrosine hydroxylase (TH)-Cre$^+$ rats, which allowed targeting of TH$^+$ DA neurons with 95% specificity (**Fig. 2a-c**). We replicated the differences in behavior observed in the two tasks (**Suppl. Fig. 2**). We observed no sex differences in behavior or neural activity and combined both sexes in the following analyses (**Suppl. table 1**).

We monitored changes in DA neuron activity across trials in the LI5 and LR5 tasks at the lever insertion and retraction cues, at reward retrieval, and during lever presses. Changes in action automaticity and behavioral chunking occurred within the very first session in the LR5 task, with an increase in response rate and a decrease in port checking from the first 15 trials to the next 15 trials of the first session (**Suppl. Fig. 2b and f**). Thus, for both tasks, we analyzed changes in activity across 3 sets of 15 trials: the first 15 trials of the first session (trials 0-15), the second 15 trials of the first session (trials 16-30) and the last 15 trials of the last training session.

In the LI5 task, where the lever insertion cue signals opportunity to lever press for reward, there was an increase in DA neuron activity associated with the lever insertion cue (**Fig. 2d**), and after reward retrieval (**Fig. 2f**), with no change in either of these over repeated trials, and no obvious change at completion of the ratio (**Fig. 2e**). In contrast, in the LR5 task, where lever retraction serves as the cue for sequence completion and reward delivery, we observed rapid training-related changes in DA neuron activity. While changes in DA neuron activity associated with the lever insertion cue were not statistically significant (**Fig. 2g**), we observed a shift from reward retrieval to the lever retraction cue as learning progressed (**Fig. 2h-i**). Specifically, from the first to the second 15 trials, the DA response to the lever retraction cue increased. In parallel, during the first 15 trials, DA neuron activity peaked during reward retrieval (**Fig. 2i**), with onset of this peak shifting over subsequent trials to before the port entry (**Fig. 2i**), when the lever retracts (**Fig. 2h)**. Both the lever retraction cue and reward peaks of activity decreased with extended training in the LR5 task (**Fig 2h-i**; last 15 trials).

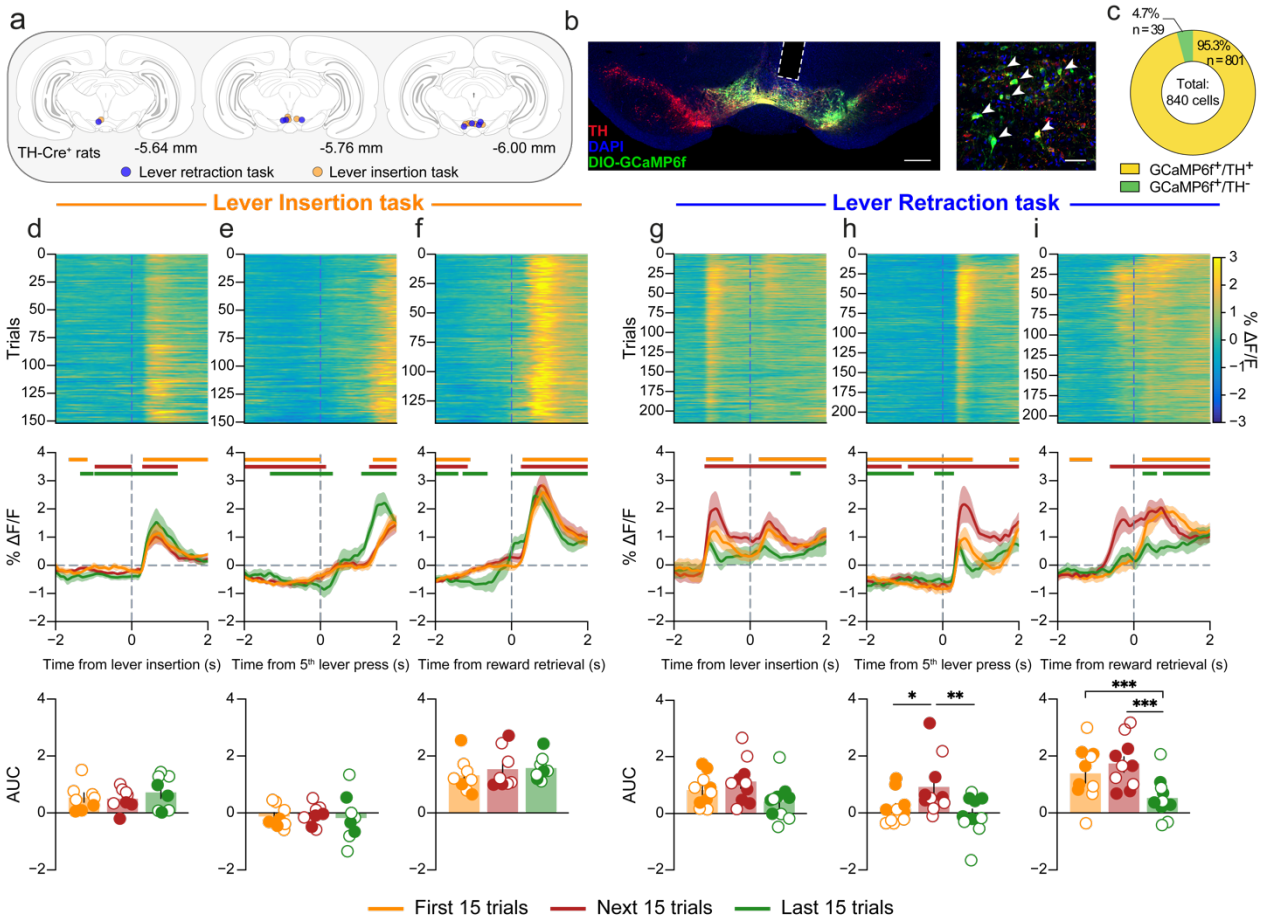

Figure 2. Training-dependent changes in VTA dopamine neuronal phasic activity in response to cues and reward in the LR5 task, but not the LI5 task. *(**a**) Coronal sections showing location of optic fiber tips relative to Bregma for TH-Cre$^+$ rats in LI5 and LR5 tasks. (**b**) Left panel: DIO-GCaMP6f was infused in the VTA of TH-Cre$^+$ rats, and the optic fiber, represented with the white dash line, was implanted above the infusion site. Scale bar: 500 µm. Right panel: histology confirmation of DIO-GCaMP6f (green) and TH (red) co-expression (indicated by white arrows). Scale bar: 50 µm. (**c**) GCaMP6f virus is predominantly expressed in TH+ neurons. (**d**) Top panel: LI5 task heatmap displaying trial-by-trial dopamine neuron activity, averaged across rats, around lever insertion. Middle panel: phasic dopamine neuron activity around lever insertion averaged across rats and trials for the first 15 trials (orange line), the next 15 trials (red line) and the final 15 trials (green line) of training. Bottom panel: area under the curves (AUCs) for 0 to 1 s following lever insertion (one-way RM ANOVA, no trial effect: $F_{2,14} < 1$, $p = 0.43$). (**e**) Same as (**d**) for the 5th lever press event in the LI5 task (one-way RM ANOVA, no trial effect: $F_{2,14} < 1$, $p = 0.863$). (**f**) Same as (**d**) for reward retrieval in the LI5 task (one-way RM ANOVA, no trial effect: $F_{2,14} = 2.29$, $p = 0.147$). (**g**) Same as (**d**) for lever insertion in the LR5 task (one-way RM ANOVA, no trial effect: $F_{2,18} = 2.71$, $p = 0.09$). (**h**) Same as (**d**) for the 5th lever press in the LR5 task (one-way RM ANOVA, trial effect: $F_{2,18} = 7.26$, $p = 0.005$; Student Newman-Keuls post hoc pairwise comparison). (**i**) Same as (**d**) for reward retrieval in the LR5 task (one-way RM ANOVA, trial effect: $F_{2,18} = 21.04$, $p < 0.001$; Student Newman-Keuls post hoc pairwise comparison). Data shown as means ± SEM, superimposed with individual data point. LI5 n = 8 (3 males; 5 females); LR5 n = 10 (5 males; 5 females). Orange, red and green horizontal bars indicate significant deviation from baseline at 99% confidence level. Filled circles: males; Empty circles: females. *p < 0.05; **p < 0.01; ***p < 0.001.*

Distinct patterns of VTA DA neuron activity during lever press sequences between the LI5 and LR5 tasks

Analysis of changes in DA neuron activity across lever presses revealed intriguing task differences (**Fig. 3**). To compare activity after the lever press action, we examined the first 4 responses only as the 5th response is coupled to lever retraction in the LR5 task as shown above (**Fig. 2**). In the LI5 task, DA neuron activity associated with the initiation of the lever press sequence changed over time, with a peak of activity at the initiation of the sequence initially following the first lever press (first and next 15 trials) and progressively shifting to before the lever press (last 15 trials; **Fig. 3a**). Furthermore, DA neuron activity decreased as rats progressed in the execution of the sequence from the first to the fourth lever press (**Fig. 3e-f**), with an overall suppression immediately post-press during the last 15 trials (**Fig. 3a-d and g**).

In contrast, in the LR5 task, large modulations of DA activity around lever presses were not observed (**Fig. 3h-k**). Instead, a modest decrease in signal occurred across training (**Fig. 3l-n**). Similar patterns of DA neuron activity were observed during the first and last training sessions in the LI5 and LR5 tasks when the recorded signals were deconvoluted to isolate the neural responses associated with individual behavioral events (lever press, reward retrieval) while avoiding contamination from neighboring events (**Suppl. Fig. 3**; details in Online Methods).

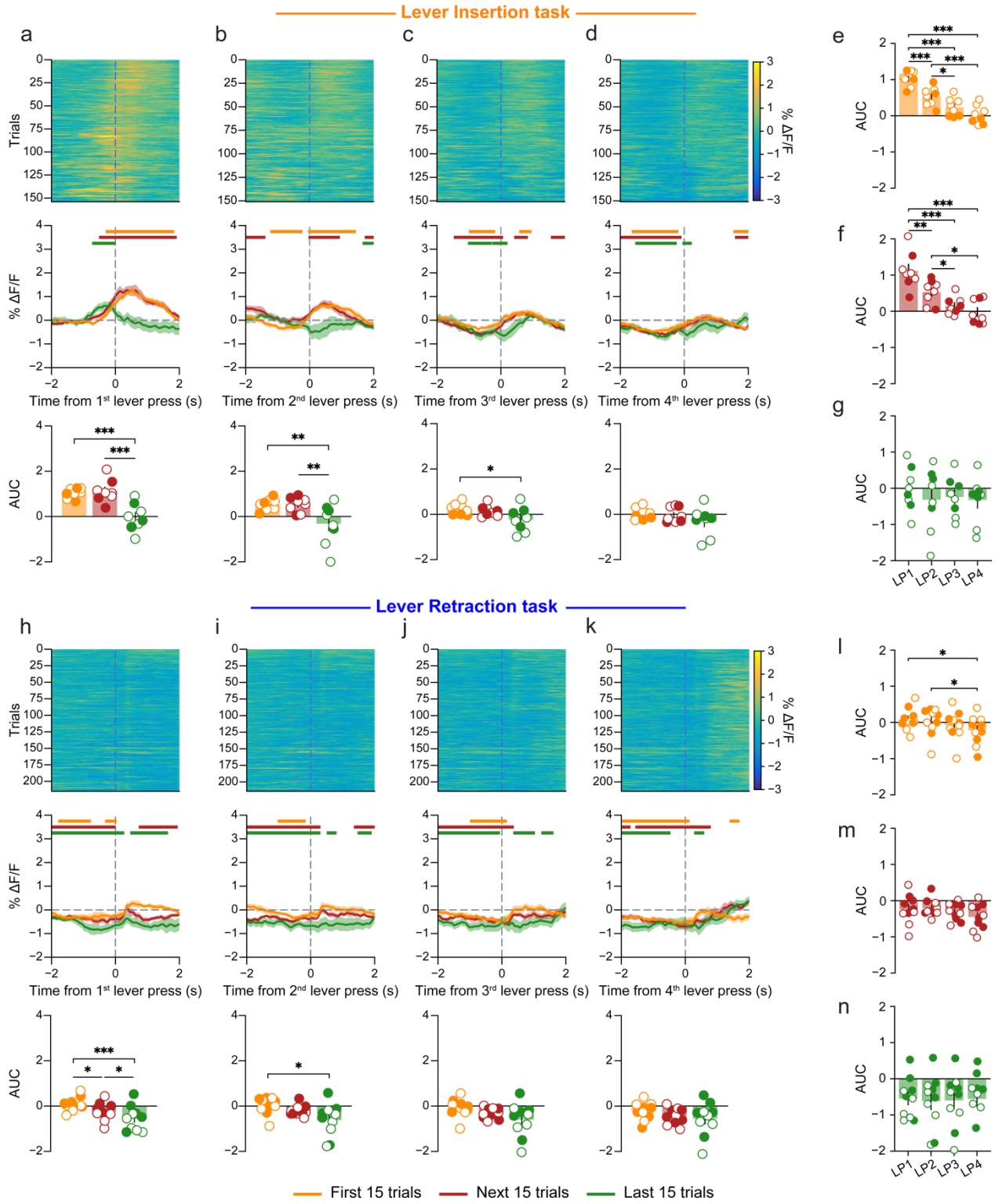

Figure 3. Distinct patterns of VTA DA neuronal activity during lever press sequences in the LI5 and LR5 tasks. *(**a-d**) Sequence of the first four lever presses for the LI5 task. The upper panel heatmaps illustrate trial-by-trial dopamine neuron activity, averaged across rat. The middle panel represents dopamine neuron activity averaged across rats and trials, delineated for the first 15 trials (orange line), the next 15 trials (red line), and the final 15 trials (green line) of the training*

*period. The bottom panel displays the area under the curves (AUCs) for the 0 to 1 seconds following the event. (**a**) 1st lever press (LP1; one-way RM ANOVA, trial effect: $F_{2,14}$ = 14.71, p < 0.001. Student Newman-Keuls post hoc pairwise comparison). (**b**) 2nd lever press (LP2; one-way RM ANOVA, trial effect: $F_{2,14}$ = 9.55, p = 0.002. Student Newman-Keuls post hoc pairwise comparison). (**c**) 3rd lever press (LP3; one-way RM ANOVA, trial effect: $F_{2,14}$ = 4.54, p = 0.03. Student Newman-Keuls post hoc pairwise comparison). (**d**) 4th lever press (LP4; one-way RM ANOVA, no trial effect: $F_{2,14}$ = 2.09, p = 0.160). (**e-g**) Area under the curves (AUCs) for the 0 to 1 s window following the first four lever presses of the sequence in the LI5 task, measured across the three trial periods (two-way RM ANOVA, trial effect: $F_{2,14}$=10.12, p = 0.002; lever press effect: $F_{3,21}$ = 28.63, p < 0.001; lever press x trial interaction $F_{6,42}$ = 5.77, p < 0.001). (**e**) First 15 trials (one-way RM ANOVA, lever press effect: $F_{3,21}$ = 33.05, p < 0.001. Student Newman-Keuls post hoc pairwise comparison). (**f**) Next 15 trials (one-way RM ANOVA, lever press effect: $F_{3,21}$ = 17.71, p < 0.001. Student Newman-Keuls post hoc pairwise comparison). (**g**) Final 15 trials (one-way RM ANOVA, lever press effect: $F_{3,21}$ = 2.22, p = 0.116). (**h-k**) same as (**a-d**) for the LR5 task. (**h**) 1st lever press (LP1; one-way RM ANOVA, trial effect: $F_{2,18}$ = 9.95, p < 0.001. Student Newman-Keuls post hoc pairwise comparison). (**i**) 2nd lever press (LP2; one-way RM ANOVA, trial effect: $F_{2,18}$ = 4.60, p = 0.024. Student Newman-Keuls post hoc pairwise comparison). (**j**) 3rd lever press (LP3; one-way RM ANOVA, no trial effect: $F_{2,18}$ = 2.95, p = 0.08). (**k**) 4th lever press (LP4; one-way RM ANOVA, no trial effect: $F_{2,18}$ = 1.57, p = 0.234). (**l-n**) same as (**e-g**) for the LR5 task (two-way RM ANOVA, trial effect: $F_{2,18}$ = 4.77, p = 0.02; lever press effect: $F_{3,27}$ = 3.23, p = 0.04; no lever press x trial interaction $F_{6,54}$ = 1.25, p = 0.29). (**l**) First 15 trials (one-way RM ANOVA, lever press effect: $F_{3,27}$ = 4.12, p = 0.015. Student Newman-Keuls post hoc pairwise comparison). (**m**) Next 15 trials (one-way RM ANOVA, lever press effect: $F_{3,27}$ = 1.55, p = 0.225). (**n**) Final 15 trials (one-way RM ANOVA, lever press effect: $F_{3,27}$ < 1, p = 0.824). Data shown as means ± SEM, superimposed with individual data point. LI5 n = 8 (3 males; 5 females); LR5 n = 10 (5 males; 5 females). Orange, red and green horizontal bars indicate significant deviation from baseline at 99% confidence level. Filled circles: males; Empty circles: females. *p < 0.05; **p <0.01; ***p < 0.001.*

Patterns of DA release in the nucleus accumbens core are similar to VTA in response to cues, lever presses and reward

We next asked if the distinct patterns of VTA DA neuron activity we observed would be reflected in similar patterns of DA release in the nucleus accumbens core (NAc core), a primary target for VTA DA neurons that is implicated in DA-mediated cue processing and reward learning [30–33]. To this end, we infused the dopamine sensor dLight1.2 into NAc core in male and female rats (**Fig. 4a-b**).

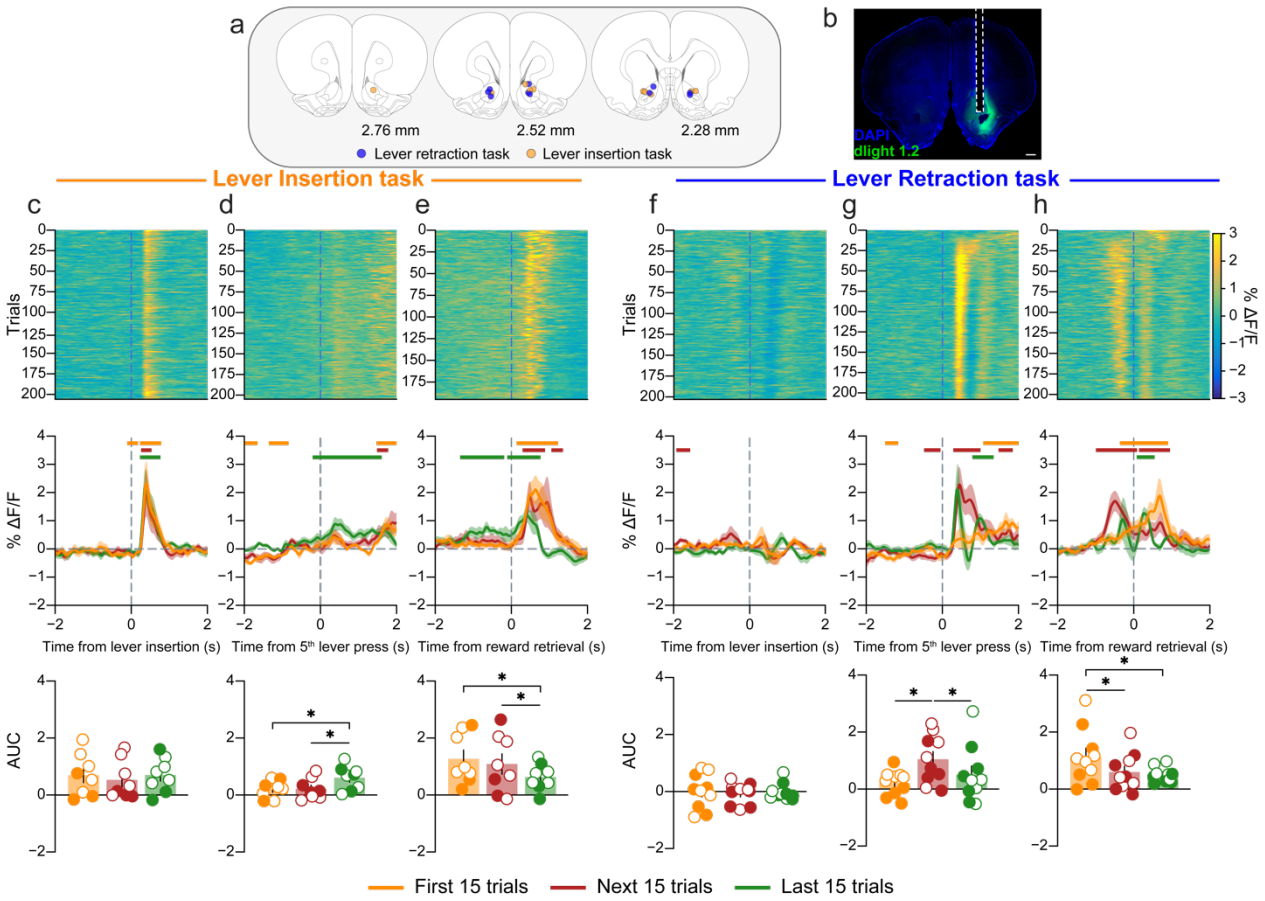

Figure 4. Dynamic changes in dopamine release in the NAc core in response to cues and reward in the LI5 and LR5 tasks. *(**a**) Coronal sections showing location of optic fiber tips relative to Bregma in LI5 and LR5 tasks. (**b**) dLight 1.2 was infused in the NAc core, and the optic fiber, represented with the white dash line, was implanted above the injection site. Scale bar: 500 µm. (**c**) Top panel: LI5 task heatmap displaying trial-by-trial dopamine release, averaged across rats, around lever insertion. Middle panel: phasic dopamine release around lever insertion averaged across rats and trials for the first 15 trials (orange line), the next 15 trials (red line) and the last 15 trials (green line) of training. Bottom panel: area under the curves (AUCs) for 0 to 1 seconds following lever insertion (one-way RM ANOVA, no trial effect: $F_{2,14} < 1$, $p = 0.59$). (**d**) Same as (**c**) for the 5th lever press event in the LI5 task (one-way RM ANOVA, trial effect: $F_{2,14} = 5.29$, $p = 0.02$; Student Newman-Keuls post hoc pairwise comparison). (**e**) Same as (**c**) for reward retrieval in the LI5 task (one-way RM ANOVA, trial effect: $F_{2,14} = 5.42$, $p = 0.02$; Student Newman-Keuls post hoc pairwise comparison). (**f**) Same as (**c**) for lever insertion in the LR5 task (one-way RM ANOVA, no trial effect: $F_{2,18} < 1$, $p = 0.86$). (**g**) Same as (**c**) for the 5th lever press in the LR5 task (one-way RM ANOVA, trial effect: $F_{2,18} = 5.53$, $p = 0.013$; Student Newman-Keuls post hoc pairwise comparison). (**h**) Same as (**c**) for reward retrieval in the LR5 task (one-way RM ANOVA, trial effect: $F_{2,18} = 6.23$, $p = 0.009$; Student Newman-Keuls post hoc pairwise comparison). Data shown as means ± SEM, superimposed with individual data point. LI5 n = 8 (3 males; 5 females); LR5 n = 10 (5 males; 5 females). Orange, red and green horizontal bars indicate significant deviation from baseline at 99% confidence level. Filled circles: males; Empty circles: females. *p < 0.05.*

Changes in patterns of DA release tended to mirror the observed changes in VTA DA neuron activity across training in the LI5 and LR5 tasks. Specifically, the lever insertion cue elicited a peak

in DA release in the LI5 task (**Fig. 4c**) but not in the LR5 task (**Fig. 4f**). Conversely, the fifth lever press did not evoke DA activity in the LI5 task (**Fig. 4d**), while it caused a significant DA release in the LR5 task (**Fig. 4g**). In addition, we observed a prominent reward peak that somewhat diminished by the end of the training in the LI5 task (**Fig. 4e**), and, in contrast, a reward peak that rapidly shifted to the lever retraction cue in the LR5 task (**Fig. 4h**).

Consistent with measures of VTA DA neuron activity, we observed peaks of DA release during lever presses in the LI5 task but not in the LR5 task. These peaks did not change across training (**Suppl. Fig. 4**).

Together with the findings from VTA-GCaMP recordings, these results suggest that a shift in phasic DA signals from reward retrieval to the termination of the sequence occurs within the first session in the LR5 task (from the first 15 trials to the next 15 trials) in which lever retraction explicitly cues sequence completion and reward delivery, but not after extended training in the LI5 task in which no cue signals imminent reward delivery.

Impaired contingency learning induced by the lever retraction cue is associated with attenuated RPE DA signals in the VTA and NAc Core

After training in the LR5 and LI5 tasks, rats were tested in a single session under an omission schedule to evaluate their behavioral and dopaminergic responses to reversed instrumental contingencies. The omission procedures mirrored the task structures experienced during training. During the omission tests, all rats had to refrain from pressing the lever for 20 seconds to receive a reward. For the LI5 task, five presses on the lever resulted in omission of the reward and lever retraction was delayed until the port entry (**Fig. 5a**), while for the LR5 task, the fifth lever press resulted in omission of the reward and lever retraction (**Fig. 5b**).

During the omission tests, rats trained in the LR5 task exhibited a greater number of lever presses compared to rats trained in the LI5 task (**Fig. 5c**), indicating a reduced ability to adapt to the new contingencies relative to LI5-trained rats. The mean latency to enter the port after reward omission (**Fig. 5d**) or delivery (**Fig. 5e**) did not differ in LR5 and LI5 rats. The difference in behavioral flexibility was paralleled by differences in DA neuron activity in the VTA and DA release in the NAc Core. LI5 rats exhibited a robust decrease in VTA DA neuron activity following omission of the reward (**Fig. 5f-g**) that was absent in LR5 rats. While VTA DA activity increased in both groups following unexpected reward delivery, this was greater in LI5 rats (**Fig. 5h-i**). We also observed an attenuated decrease in DA signals in the NAc Core when expected reward was omitted in rats trained in the LR5 task compared to rats trained in the LI5 task (**Fig. 5j-k**). However, we did not observe a significant difference in the DA responses to unexpected reward delivery (**Fig. 5l-m**). Task differences in mesolimbic DA neuron activity cannot be attributed to a difference in port checking behavior since response latencies after reward omission or delivery were similar between the tasks (**Fig 5d-e**). These results suggest that, following training in the LR5 task, rats are deficient in adapting to reward omission and that this deficiency is associated with attenuation of negative reward prediction error (RPE) signals following reward omission and positive RPE signals following unexpected reward delivery.

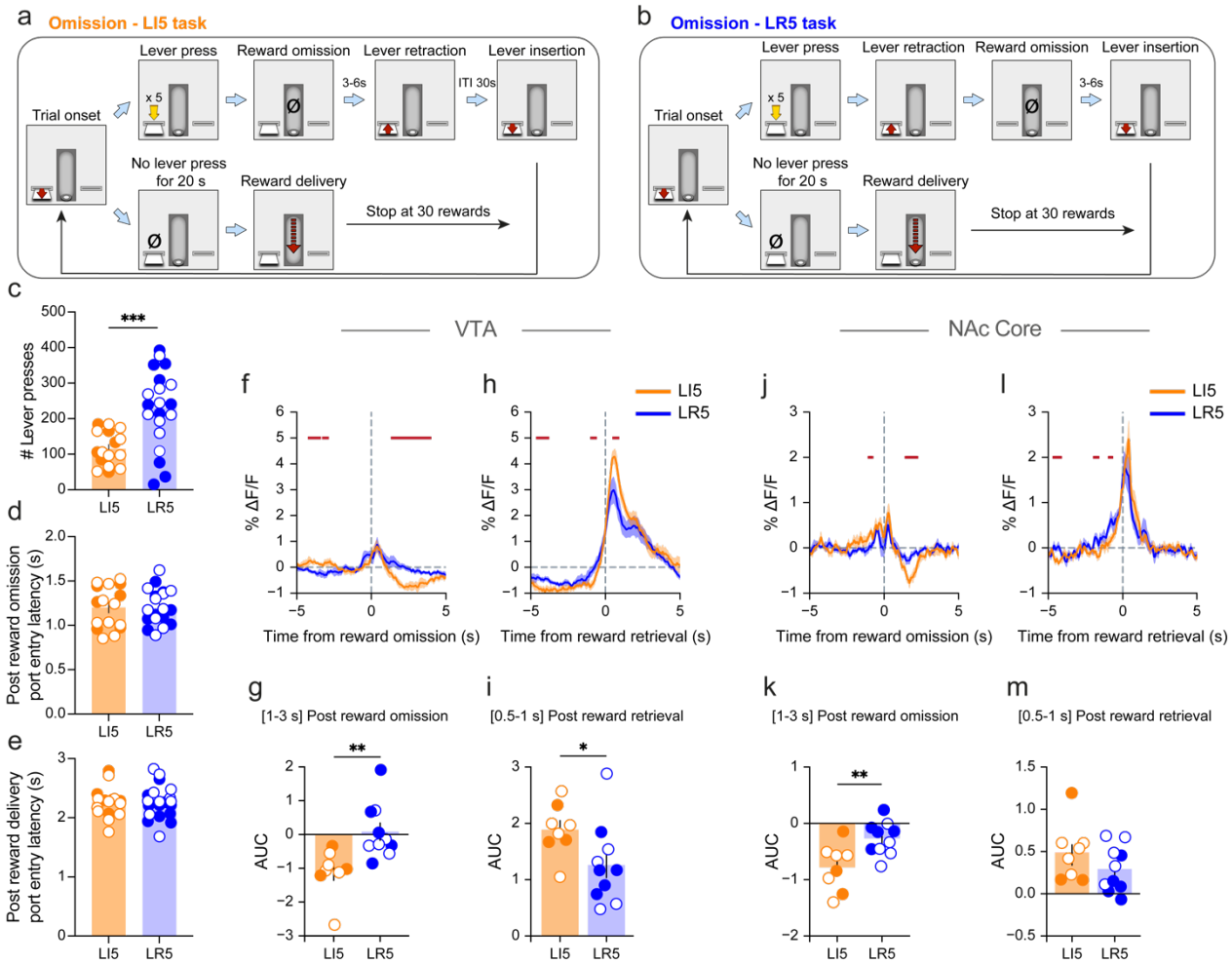

Figure 5. Impaired contingency learning induced by the lever retraction cue is associated with attenuated RPE DA signals in the mesolimbic pathway. *Omission test design for the LI5 task (**a**) and the LR5 task (**b**). (**c**) Total number of lever presses during the omission test (two-way ANOVA, task effect: $F_{1,32} = 17.87$, $p < 0.001$; no task x region interaction: $F_{1,32} < 1$, $p = 0.529$; Student Newman-Keuls post hoc pairwise comparison). (**d**) Port entry latency following omission of the reward (two-way ANOVA, no task effect: $F_{1,32} < 1$, $p = 0.95$; no region effect: $F_{1,32} < 1$, $p = 0.74$; task x region interaction: $F_{1,32} = 4.41$, $p = 0.044$). (**e**) Port entry latency following unexpected reward delivery (two-way ANOVA, no task effect: $F_{1,32} < 1$, $p = 0.940$; no task x region interaction: $F_{1,32} < 1$, $p = 0.361$). (**f**) Phasic VTA dopamine neuron activity around port entries following reward* omission. *(**g**) Area under the curves for 1 to 3 seconds after reward omission (unpaired two-tailed t-test comparison, $t_{16} = -3.30$, $p = 0{,}004$). (**h**) Phasic VTA neuron activity around reward retrieval. (**i**) Area under the curves for 0.5 to 1 seconds after reward retrieval (unpaired two-tailed t-test comparison, $t_{16} = 2.15$, $p = 0.047$). (**j**) Phasic NAc Core DA signal around port entries following reward* omission. *(**k**) Area under the curves for 1 to 3 seconds after reward omission (unpaired two-tailed t-test comparison, $t_{16} = -3.07$, $p = 0{,}007$). (**l**) Phasic NAc Core DA signal around reward retrieval. (**m**) Area under the curves for 0.5 to 1 seconds after reward retrieval (unpaired two-tailed t-test comparison, $t_{16} = 1.36$, $p = 0.191$). Data shown as means ± SEM, superimposed with individual data point. LI5 n = 8 (3 males; 5 females); LR5 n = 10 (5 males; 5 females). Red horizontal bars indicate significant difference between LI5 and LR5 tasks at 95% confidence level. Filled circles: males; Empty circles: females. *p < 0.05; **p < 0.01; ***p < 0.001.*

Optogenetic VTA DA neuron stimulation at the time of sequence completion drives automaticity and behavioral chunking

Given our prior observation that lever retraction cues paired with reward delivery evoke phasic activity of VTA dopamine neurons and are associated with the development of both automaticity and behavioral chunking in LR5 rats, we sought to replicate this behavioral pattern by mimicking cue-elicited DA in an optogenetic experiment. TH-Cre^{+} rats were unilaterally infused in the VTA with DIO-ChR2 or DIO-eYFP virus and implanted with an optic fiber (**Fig. 6a**). We modified the LI5 task (**Fig. 6b**) to incorporate brief optogenetic stimulation (0.5 s; 20 mW; 20 Hz) to artificially recreate the activity peak observed upon completion of the fifth lever press and lever retraction in the LR5 task. As with the original LI5 task, the lever remained extended until rats entered the port to consume the reward, preventing them from using the lever retraction as a reward-predictive cue.

Optogenetic stimulation at sequence completion resulted in a progressive increase in response rates in ChR2 rats over training sessions as compared to eYFP rats receiving equal presentations of laser (**Fig. 6c-d**). The distribution of inter-press intervals (IPI) was comparable between ChR2 and eYFP rats on the first session (**Fig. 6e**) but differed between groups by the last session (**Fig. 6f),** with a remarkable increase in the proportion of very short IPIs in ChR2 rats (**Fig. 6f insert**). These findings suggest that stimulating VTA DA neurons at sequence completion resulted in progressively faster sequence execution across training. In addition, optogenetic stimulation of VTA DA neurons resulted in a moderate but significant decrease in port checking across training sessions (**Fig. 6g**). Thus, all ChR2 rats exhibited a decrease in port entry likelihood between the first and last session (**Fig. 6h**). Finally, analysis of lever press sequence length revealed no difference between the two groups during the first session, with a majority of single presses in both eYFP and ChR2 rats (**Fig. 6i**, ~60%). However, on the last training session, eYFP rats persisted with recurrent port checking, whereas ChR2 rats reduced the proportion of single presses, tending to respond in bouts of 2 or 3 presses before checking the port (**Fig. 6j**).

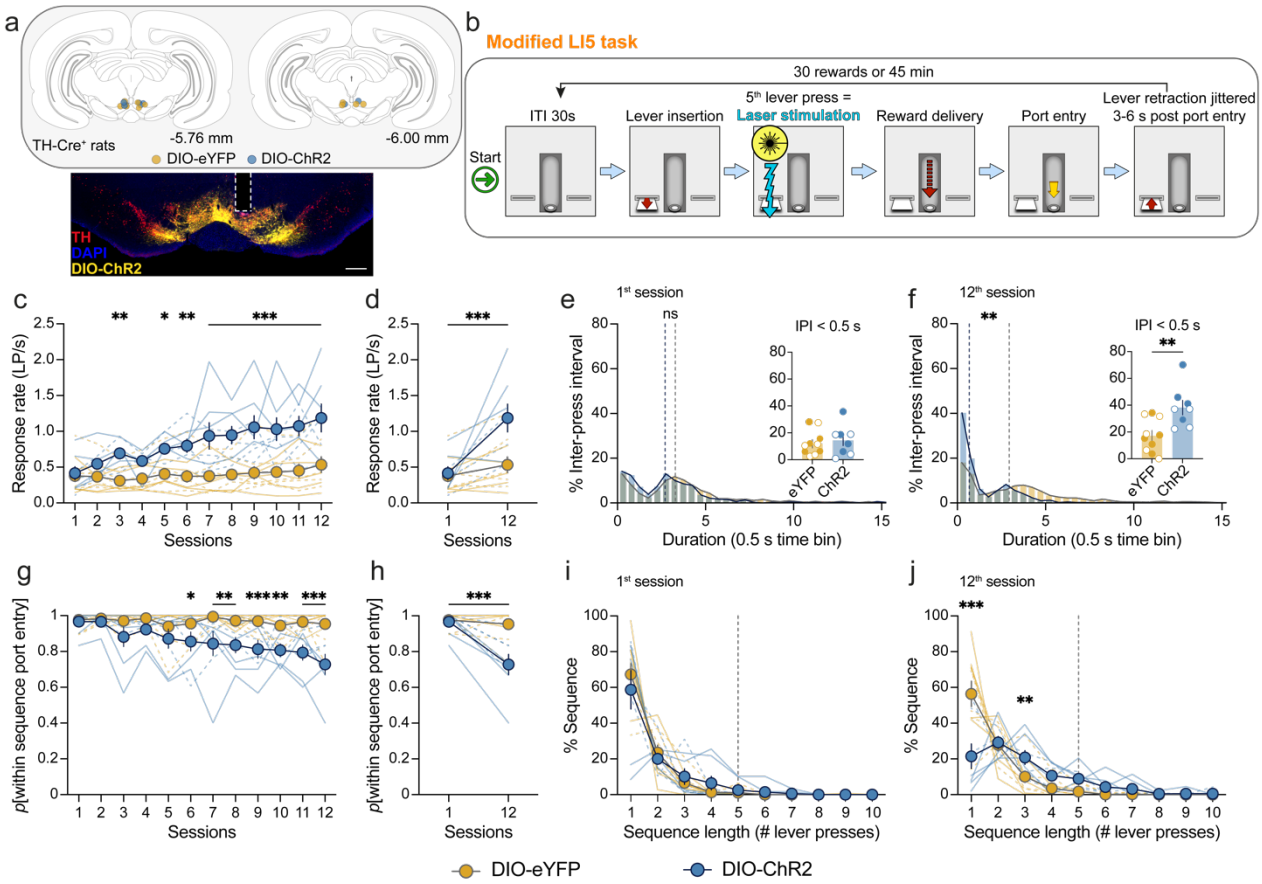

Figure 6. Optogenetic VTA DA neuron stimulation at the time of sequence completion drives automaticity and behavioral chunking. *(**a**) Targeting VTA DA neurons with DIO-eYFP or DIO-ChR2 virus in TH-Cre$^+$ rats. Scale bar: 500µm. (**b**) Task design for optogenetic stimulation of VTA DA neurons at sequence completion. (**c**) Response rate in lever press per second (LP/s) within the lever press sequence (two-way RM ANOVA, group effect: $F_{1,16}$ = 15.97, p = 0.001; group x session interaction: $F_{11,176}$ = 4.5, p < 0.001; Student Newman-Keuls post hoc pairwise comparison). (**d**) Response rate on the first versus last session (two-way RM ANOVA, group effect: $F_{1,16}$ = 6.58, p = 0.021; group x session interaction: $F_{1,16}$ = 8.17, p = 0.011; Student Newman-Keuls post hoc pairwise comparison). (**e**) Frequency distribution of inter-press intervals (IPI) for the first session in DIO-eYFP and DIO-ChR2 rats (Mann-Whitney Rank Sum t test median comparison U = 395, p = 0.23). Insert panel: individual proportion of IPI under 0.5 s (t test, no group effect t_{16} = 0.476, p = 0.64). (**f**) Frequency distribution of inter-press intervals (IPI) for the last session in DIO-eYFP and DIO-ChR2 rats (Mann-Whitney Rank Sum t test median comparison U = 274, p = 0.004). Insert panel: individual proportion of IPI under 0.5 s (t test, group effect t_{16} = 3.22, p = 0.005). (**g**) Probability to enter the port within the sequence (within-sequence port entry) (two-way RM ANOVA, group effect: $F_{1,16}$ = 13.84, p = 0.005; group x session interaction: $F_{11,176}$ = 3.95, p < 0.001; Student Newman-Keuls post hoc pairwise comparison). (**h**) Probability of within sequence port entry on the first versus last session in DIO-eYFP and DIO-ChR2 rats (two-way RM ANOVA, group effect: $F_{1,16}$ = 9.41, p = 0.007; group x session interaction: $F_{1,16}$ = 18.06, p < 0.001; Student Newman-Keuls post hoc pairwise comparison). (**i**) Frequency distribution of sequence length during the first training session. The grey dashed line represents the optimal sequence length of 5 lever presses (two-way RM ANOVA, no group effect: $F_{1,16}$ < 1, p = 0.53; no group x sequence length interaction: $F_{9,144}$ < 1, p = 0.77). (**j**) Frequency distribution of sequence length during the last training session. The grey dashed line represents the optimal sequence*

*length of 5 lever presses (two-way RM ANOVA, group effect: $F_{1,16} = 5.16$, $p = 0.037$; group x sequence length interaction: $F_{9,144} = 8.64$, $p < 0.001$; Student Newman-Keuls post hoc pairwise comparison). Data shown as mean ± SEM, superimposed with individual data point. DIO-eYFP n = 10 (5 males; 5 females); DIO-ChR2 n = 8 (4 males; 4 females). Solid line and filled circles: males; Dashed line and empty circles: females. *p < 0.05; **p < 0.01; ***p < 0.001.*

Although both groups obtained a similar number of rewards throughout training (**Suppl. Fig. 5a**), ChR2 rats exceeded the response requirement (**Suppl. Fig. 5b**) and retrieved the reward with a shorter latency than eYFP rats (**Suppl. Fig. 5c**). The tendency to exceed the response requirement could be interpreted as resulting from the reinforcing effect of VTA DA neuron stimulation [2,3,34]. However, with the exception of one rat, optogenetic stimulation of VTA DA neurons in ChR2 rats failed to support robust intra-cranial self-stimulation (ICSS) with the stimulation parameters used in this experiment (0.5 s; 20 mW; 20 Hz, in absence of cue) (**Suppl. Fig. 5d to f**), indicating that reinforcing effects of DA neuron stimulation do not account for the results reported here. Indeed, if rats were responding for DA neuron stimulation, a longer, not shorter, latency to retrieve reward would be expected. Importantly, robust ICSS was observed when laser stimulation was increased to 2s and was paired with a cue light, as shown previously [2,3,34] (**Suppl. Fig. 5g to i**). In sum, these findings show that brief unilateral VTA DA neuron activation is sufficient to promote behavioral chunking of action sequences.

DISCUSSION

In this study, we examined the role of phasic DA signals during performance of action sequences that were differentially impacted by sequence initiation and termination cues. We found that a lever retraction cue marking sequence completion in the LR5 task facilitated habit-like behavior. Conversely, when the sequence completion cue was absent in the LI5 task, in which lever insertion acted as a cue to indicate the opportunity to perform actions for reward access, rats exhibited greater goal-directedness, without behavioral chunking, as indicated by repeated reward checking between lever presses. These distinct response strategies were associated with differential changes in mesolimbic DA signals, which captured variations in reward expectation along the execution of action sequences in the LR5 and LI5 tasks. We also found that behavioral inflexibility in rats trained in the LR5 task was associated with blunted positive and negative RPE-like signals in an omission test. Finally, optogenetic stimulation of VTA DA neurons at sequence termination, immediately prior to reward delivery and in absence of the lever retraction cue, was sufficient to promote automaticity and behavioral chunking.

We have previously shown that lever cues at the boundaries of action sequences promote habit formation and behavioral chunking [24–26]. Here, we extended these findings by demonstrating that the lever retraction cue, but not the lever insertion cue, is sufficient to promote habit-like behavior. Rats trained in the LR5 task were more likely to show an insensitivity to outcome devaluation compared to those trained in the LI5 task. These LR5-trained rats also displayed higher automaticity in lever pressing, without reward expectation until completion of the sequence, which is indicative of behavioral chunking [35,36]. Finally, LR5 rats were less sensitive than LI5 rats to a reversal of instrumental contingency in the omission test, suggesting some form of behavioral inflexibility. These results support the hypothesis that the reward predictive cue (i.e. lever retraction) serves as response feedback that favors habit formation and behavioral chunking by alleviating the need for action monitoring [25]. In contrast, rats trained in the LI5 task need to build and use an action model to monitor progress during sequence execution. Similar action monitoring was observed in prior work[25,37], when rats where required to respond in a sequence of five consecutive lever presses without port checking to earn the reward. Although this resulted in highly

efficient performance similar to that seen here in the LR group, action monitoring was required in absence of the feedback cue, which maintained goal-directed control. Thus, the feedback cue is the critical feature for habitual strategy adoption in the LR5 task.

Task differences in reward expectation along the behavioral sequence were reflected in distinct DA signal patterns that were similar in the VTA and in the NAc core. In the LR5 task, DA signals rapidly shifted from reward retrieval to the lever retraction cue, aligning with the certain reward expectation at action sequence termination. Conversely, DA neurons activity was absent around lever presses, when there is no reward expectation due to behavioral chunking [23,38]. In the LI5 task, where reward delivery remained unpredictable, DA signals showed a sustained peak at the reward. The greater goal-directedness and regular port checking between individual lever presses suggest some form of reward expectation at initiation and during sequence execution, which was associated with DA peaks after individual lever presses, with a decay in activity from sequence initiation to sequence termination. Furthermore, the peak at sequence initiation followed the first lever press during early training, but preceded it after extended training. These results suggest that across training, rats learned to expect the reward after the lever insertion cue and at the approach of the lever, when they initiated the action sequence.

Task differences in DA signal patterns could reflect credit assignment to distinct events in the sequence. DA signaling has been proposed to mediate credit assignment [29], i.e., the appropriate connection of a given action or stimulus with its outcome. In the LI5 task, DA signal backpropagate to the beginning of the action sequence and thus, could assign credit to the action policy. This is in agreement with prior studies showing a shift in DA signal from the proximal to the distal lever press across training in an action sequence task [9,10]. In contrast, in the LR5 task, DA signals backpropagate to the immediately preceding sequence termination cue and assign credit to this cue.

We found an attenuation of DA signals across training in the LR5 task but not in the LI5 task in VTA and NAc core. Prior studies reported a similar phenomenon, where cue-induced DA release within the ventral striatum decreased as animals became expert in the task [9,39,40]. This decrease in activity cannot be attributed to an extinction of fluorescent signal over time since VTA DA neuron activity fully recovered following unexpected reward delivery in the omission test. Instead, decreased DA signals over training may be due to the learned expectation of the sequence termination cue [39,40]. It is also possible that with habit learning, encoding of the motivational properties of the sequence termination cue shift from the mesolimbic pathway to the nigrostriatal pathway [17,19,41–44]. Indeed, both pathways support reinforcement learning but through different mechanisms[8]. While DA neurons in the VTA confer motivational value to action and reward-associated cues, SNc DA neurons support more inflexible action-specific learning[8]. Relatedly, the sequential shift in DA control from the mesolimbic to the nigrostriatal pathway has been suggested as a neural basis for the development of automaticity and the transition from goal-directed to habitual control [18,19,41].

Training in the LI5 and LR5 tasks contributed to differences in rats' ability to adapt to a change in instrumental contingency during the omission test. When rats responded under goal-directed control in the LI5 task, they were able to flexibly learn the new contingencies and adjusted their behavior accordingly. This learning was accompanied by the decreases in of DA signals following reward omission (negative RPE) and increases in DA signals following unexpected reward delivery (positive RPE). Conversely, LR5 rats displayed attenuated positive RPE signals and absent negative RPE signals compared to LI5 rats, which may explain their persistence to press the lever despite reward omission. This persistence may result from the presence of the sequence

termination cue and, since this cue is still presented at the end of the sequence, they keep pressing despite reward omission.

The sequence termination cue elicited a large DA transient in both VTA and NAc as rats learned to complete the sequence in the LR5 task. Optogenetically mimicking this DA feedback cue in rats without actual lever retraction biased behavior towards automaticity. A possible mechanism underlying the effect of optogenetic manipulation on our automaticity measures is that optogenetic DA stimulation reinforces actions preceding the stimulation, initially the final lever press of the sequence, and, over time via back propagation links together the relevant preceding presses[45]. However, we found no evidence of DA signal backpropagation across presses in the LR5 task. Thus, another possible mechanism is that peaks of DA induced by optogenetic stimulations assign credit to the state at the end of sequence completion such that rats complete the sequence to reach that state without encoding the action policy. Interestingly, rats that received DA stimulation often exceeded the required number of lever presses to receive the reward. While this behavior could be mistaken for intra-cranial self-stimulation (ICSS), the brief (0.5s) stimulation of VTA DA neurons in our study does not exhibit the strong reinforcing properties typically associated with longer (1s) stimulation durations, as reported in other studies [2,3,34,46] (but see [47]). Despite this, the ChR2 rats remained engaged in the task, as evidenced by the number of rewards obtained throughout training and their short latency to retrieve rewards. Therefore, the additional lever presses likely indicate the difficulty in inhibiting a prepotent behavior due to an increased speed in sequence execution. Likewise, rats trained in the LR5 task tended to keep pressing on the lever while retracting, leading to sequences of 6 or 7 lever presses (**Fig. 1**)

To conclude, we have shown that sequence initiation and termination cues differentially affect reward expectation during action sequences, which drive different response strategies that are encoded by distinct mesolimbic DA patterns. Our study provides critical new insights into the role of DA in shaping motivated behaviors, offering a new understanding of its function in behavioral control and response strategies during sequence learning. We show that DA encodes both habit-like and goal-directed response strategies but through differential credit assignment to either the sequence termination cue or the action policy. This dual role highlights dopamine's pivotal influence on learning processes, bridging gaps in our understanding of how reward signals shape complex behaviors.

MATERIALS & METHODS

Subjects

Wild-type (WT) (n = 48; 23 males and 25 females) or TH-Cre^{+} Long-Evans rats (n = 36; 17 males and 19 females) were used in this study. These rats express Cre recombinase under the control of the tyrosine hydroxylase (TH) promoter [48]. Rats were individually housed under a 12 h/light/dark cycle (lights ON at 7 am). They were food restricted to 90% of free feeding weight, starting one week before training, but had *ad libitum* access to water throughout the experiment. All rats were 8 to 10 weeks old at the beginning of the experiment. These studies were carried out in accordance with the recommendations of the Guide for the Care and Use of Laboratory Animals (Institute of Instrumental Training Laboratory Animal Resources, Commission of Life Sciences, National Research Council, 1996). The protocol was approved by the animal care and use committee of Johns Hopkins University.

Behavioral training

Apparatus

The apparatus consisted of operant chambers (Med associates) each equipped with a retractable lever on the left side of the right wall. The reward (0.1mL of 20% sucrose) was delivered over 3 s in a magazine equipped with an infra-red beam, and located next to the lever. The house-light, located on the back wall of the operant chamber remained illuminated for the length of the session.

Initial training

Rats experienced one magazine training session in which 30 sucrose rewards were delivered under a random interval 60 s schedule. The following days, they were trained to press the lever to earn reward on a fixed-ratio 1 (FR1) schedule. Sessions were limited to 1 hour or 30 reward deliveries, whichever occurred first. Rats were trained one session per day, for 3 to 5 sessions, until they earned the maximum 30 rewards in less than an hour. Rats were then trained on a FR5 schedule (five responses per reward) for 5 additional sessions, limited to 30 min or 30 reward deliveries, whichever occurred first. During initial instrumental training, the lever was continuously presented and no cue indicated when reward was delivered.

LI5 and LR5 task design

The lever insertion fixed-ratio 5 (LI5) task and the lever retraction fixed-ratio 5 (LR5) task were designed such that only one of the lever cues (lever insertion or retraction, respectively) is relevant to signal reward while the other cue, is made irrelevant. Both groups of rats were trained in parallel in the LI5 task or LR5 task. Two rats were excluded from analyses because they omitted the reward more than 70% of the trials on average across the 12 sessions, leaving a total of 30 rats (LI5: n=15, 8 females, 7 males; LR5: n=15, 8 females, 7 males).

LI5 task

As detailed in **Fig 1a**, each trial begun with a 30 s inter-trial interval prior to the insertion of the lever in the cage. Rats were then required to press the lever five times to receive the reward. The fifth lever press triggered reward delivery, which was not signaled by the lever retraction cue. To ensure that rats did not pay attention to the retraction of the lever, this event occurred during reward consumption, at a random interval ranging from 3 to 6 s after rats' entry into the magazine or at the exit of the magazine, whichever occurred first. The session ended after 45 min or 30 reward deliveries. For each trial, the lever remained extended until the rats completed the ratio and entered the magazine to consume the sucrose.

LR5 task

As detailed in **Fig 1b**, each trial begun with the insertion of the lever in the cage. The completion of the FR5 requirement led to the retraction of the lever which signaled reward delivery, occurring at the fifth lever press. To ensure that rats did not pay attention to the insertion of the lever, this event occurred at a random interval ranging from 3 to 6 s after the rats entered the magazine to consume the sucrose or at magazine exit, whichever occurred first. The session ended after 30 min or 30 reward deliveries. For each trial, the lever remained extended until completion of the ratio.

Outcome devaluation by sensory-specific satiety

Each rat received 2 days of testing, separated by one reinforced training session. Rats were given 1 hour free access to their training reward (sucrose 20%; devalued condition) or to a control reward, which never served as a reinforcer (grain-based pellet; valued condition). Pre-feeding occurred in feeding cages in the experimental room. Immediately after pre-feeding, rats were placed in the operant chambers for a test session conducted under extinction with the same

LI5/LR5 task design. Test sessions were limited to 10 trials or 10 minutes of lever presentation, whichever occurred first. Experimental valued/devalued conditions were reversed for the second test session and the order of conditions was counterbalanced across rats.

Omission test

As illustrated in **Fig 5a-b** the omission schedule consisted on a single test session in which animals had to refrain from pressing the lever for 20 s in order to obtain the reward. In contrast, completion of the FR5 resulted in omission of the reward. Every lever press reset the 20 s timer. The fundamental structures of the LI5 and LR5 tasks were maintained: for rats trained in the LI5 task, lever insertion indicated the initiation of a trial, while for those trained in the LR5 task, lever retraction denoted the fulfillment of the response requirement. The session only ended after 30 rewards were delivered, with no time limit.

Surgical procedure

Surgeries for viral infusions and optic fiber implants were carried out as previously described [3,49]. Rats were anesthetized with 5% isoflurane and placed in a stereotaxic frame, after which anesthesia was maintained at 1–3%. Rats were administered carprofen anesthetic (5 mg/kg), and cefazolin antibiotic (70 mg/kg) subcutaneously. The top of the skull was exposed and holes were drilled for viral infusion, optic fiber implant, and four skull screws. Viral injections were made using a microsyringe pump at a rate of 0.1 µl/min. Syringes were left in place for 5 min, then raised 200 µm dorsal to the injection site, left in place for another 10 min, then removed slowly. Implants were secured to the skull with dental cement. Optogenetic manipulations commenced at least 4 weeks after surgery. A total of 5 to 6 weeks was allowed for virus expression prior to photometry recording.

Fiber photometry

For fiber photometry recording of VTA dopamine neurons bulk calcium imaging, 1 µl Cre-dependent GCaMP6f (AAVDJ-EF1a-DIO-GCaMP6f; titer 5.25×10^{12} particles/ml, Stanford University) was infused unilaterally into the VTA (AP -5.8 mm, ML ±0.7 mm, DV -7.9 mm relative to Bregma) of TH-Cre^{+} rats. In order to avoid residual florescence recording from SNc, low-auto-fluorescence optic fibers (400 µm diameter; 0.57 numerical aperture, Doric Lenses) were inserted with a ±10° angle, directed toward the VTA midline, just dorsal to the injection site (AP -5.8 mm, ML ±1.7 mm, DV -7.4 mm relative to Bregma). Fiber photometry recording of dopamine signaling in the NAc was achieved via infusion of the dopamine sensor dLight1.2 virus (pAAV5-hSyn-dLight1.2: titer 5-5.75 $\times 10^{12}$ particles/ml, addgene) in the NAc Core (AP +1.8 mm, ML ±1.7 mm, DV -7 mm relative to Bregma) of WT rats. Optic fibers in the NAc Core were implanted at AP +1.8 mm, ML ±1.7 mm, DV -6.8 mm relative to Bregma.

Optogenetics

AAV5-Ef1a-DIO-ChR2-eYFP (titer $4–4.2 \times 10^{12}$ particles/ml, University of North Carolina) or AAV5-EF1a-DIO-eYFP (titer $4–6 \times 10^{12}$ particles/ml, University of North Carolina) were infused unilaterally (0.7 µl at each target site, for a total of 2.8 µl per rat, as previously described [2,3]) at the following coordinates from Bregma for targeting VTA cell bodies: AP -6.2 and -5.4 mm, ML ±0.7 mm, DV -8.5 and -7.5. Custom-made optic fiber implants (300-µm glass diameter) were inserted unilaterally in the VTA, just above and between viral injection sites at the following coordinates relative to Bregma: AP -5.8 mm, ML ±0.7 mm, DV -7.5.

Fiber photometry recording

We assessed VTA dopamine neuron activity in TH-Cre$^+$ rats during training in the LI5 task (n = 8; 5 females, 3 males) or the LR5 task (n = 10; 5 females, 5 males). Dopamine release in the NAc Core was assessed in WT rats trained in the LI5 task (n = 8, 5 females, 3 males) or the LR5 task (n = 10, n= 5 females, 5 males). For the group of rat recorded in the VTA, 3 LI5 trained and 6 LR5-trained did not experienced a jittered insertion of the lever in the cage. Instead, for the LI5 trained rats, the lever was retracted in the cage 0.5 s after the port entry following ratio completion. For the LR5-trained rats, the lever was inserted in the cage 1.5 s after the port entry following ratio completion.

Operant chambers were equipped with a real-time fiber photometry recording system (Tucker-Davis Technologies). A fluorescence mini-cube (Doric Lenses) was utilized to transmit light streams from a 465 nm LED modulated at 211 Hz and a 405 nm LED modulated at 531 Hz. The 465 nm LED passed through a GFP excitation filter, while the 405 nm LED passed through a 405 nm bandpass filter. The power of the LEDs was set at approximately 100 µW. These light sources were connected to an optic fiber implanted in the rat's brain. Power output (~30 µW and ~70 µW for the 405 nm and the 465 nm channel respectively) was daily measured at the tip of the patch cord, before and after each rat's session to ensure a stable power across rats and recording days. Both GCaMP6f and dLight fluorescence emitted by neurons below the fiber tip were captured by the mini-cube and passed through a GFP emission filter. The fluorescence signal was then amplified and focused onto a high sensitivity photoreceiver (Newport, Model 2151). To account for bleaching and movement artifacts, the brightness produced by the 465 nm excitation, which stimulates calcium-dependent GCaMP6f fluorescence, was demodulated and compared to the isosbestic 405 nm excitation, which stimulates GCaMP6f or dLight in a calcium and dopamine-independent manner.
Rats received initial operant training under the FR1 schedule before surgery Five to six weeks after virus infusion, rats were given two FR1 reminder sessions to acclimate rats to patch cord tethering and ensure they were able to move easily. Photometry recording was conducted throughout 8 training sessions in the LI5 and LR5 tasks and during the omission test.

Optogenetic stimulation

Optogenetic stimulation in the modified-LI5 task

Illumination was provided by 473 nm lasers (OptoEngine), adjusted to read 20 mW from the tip of the patch cord at constant illumination. Light power was measured before and after every behavioral session to ensure that laser power was constant within session. Black tape was wrapped around the ceramic sleeve used to connect the patch cord and the implanted optic fiber to block visible light transmission that could be used as a cue during the task. A total of 8 ChR2 rats (4 females, 4 males) and 10 eYFP (5 females, 5 males) were run in this experiment. As shown in **Fig 6b**, the LI5 task was modified to include a brief laser stimulation (0.5 s; 20 mW; 20 Hz, 5 ms pulse duration) at the 5th lever press. Rats were trained for 12 consecutive sessions that terminated after 45 minutes or 30 rewards obtained, whichever occurred first.

Intra-cranial self-stimulation (ICSS)

At the end of the optogenetic experiment, all rats were tested for ICSS with the stimulation parameters used in the modified-LI5 task to assess the possible reinforcing effect of the laser stimulation. During 8 1 h daily sessions, rats had access to two nosepoke ports; a response at the active nosepoke resulted in delivery of a 0.5 s train of light pulses (20 Hz, 5 ms pulse duration). Inactive nosepokes were recorded but had no consequence. Eight additional 1 h ICSS sessions were conducted in which a response in the active nosepoke resulted in a 2 s laser stimulation (20

Hz, 5 ms duration) paired with a cue light in the active nosepoke. Correct ICSS in these conditions (above 150 responses at the active nosepoke) was used as criteria for rats' inclusion in the optogenetic experiment, in addition to correct virus expression and optic fiber placement in the VTA.

Tissue collection

Rats were deeply anesthetized with sodium pentobarbital and transcardially perfused with cold phosphate buffered (PB) saline (PBS) followed by 4% paraformaldehyde (PFA). Brains were removed and post-fixed overnight in 4% PFA. Brains were then cryoprotected in a PB solution containing 30% sucrose for at least 48 h, frozen into liquid isopentane and store at -80°C until processed for histology.

Histology

Brains were processed into 50 µm-thick coronal sections on a cryostat (Leica Microsystems). To verify viral expression of dLight and optic fiber placement, brain slices were directly mounted on microscope slide and coverslipped with Vectashield mounting medium containing DAPI. Brain sections were then imaged with a Zeiss Axio 2 microscope. To verify viral GCaMP, ChR2 and eYFP expression in midbrain dopamine neurons, we performed immunohistochemistry for tyrosine hydroxylase (TH) and GFP (for GCaMP only). Sections were washed in PBS and incubated with bovine serum albumin (BSA) and Triton X-100 (each 0.2%) for 20 min. 10% normal donkey serum (NDS) was added for a 30 min incubation, before primary antibody incubation (mouse anti-GFP, 1:1500, Invitrogen; rabbit anti-TH, 1:500, Fisher Scientific) overnight at 4°C in PBS with BSA and Triton X-100 (each 0.2%). Sections were then washed with PBS and incubated with 2% NDS in PBS for 10 min. Secondary antibodies were added (1:200 Alexa Fluor 488 donkey anti-mouse; 1:200 594 donkey anti-rabbit, Invitrogen) for 2 h at room temperature. Sections were then washed in PBS and in PB and mounted with Vectashield mounting medium containing DAPI. Fluorescence as well as optic fiber tracks were then visualized. In order to determine the specific targeting of TH-Cre$^+$ neurons by DIO-viruses, 20 x three-channel images along the medial-lateral and anterior-posterior gradients of the midbrain were taken, using equivalent exposure and threshold settings. With the TH channel turned off, GFP$^+$ or eYFP$^+$ cells were first identified by a clear ring around DAPI-stained nuclei. The TH channel was then overlaid, and the proportion of GFP$^+$ or eYFP$^+$ cells co-expressing TH below optic fiber placements was counted.

Statistical analysis

Behavioral analysis

Data were subjected to three-way repeated measures analyses of variance (session as within-factor and task, sex and region as between-subject factors), followed by post-hoc comparisons when indicated, using Student Newman-Keuls pairwise comparison. Significance was assessed against a type I error rate of 0.05. Behavioral data that did not follow a normal distribution were analyzed using non parametric tests (Mann-Whitney Rank Sum t test for group comparison). Statistical tests were conducted on SigmaStat (Systat software Inc., San Jose, USA) and SPSS (IBM, Amorak, NY).

Analysis of photometry recordings

- ***Signal acquisition***

A real-time signal processor (RP2.1, Tucker-Davis Technologies) running Synapse software was used to modulate the output of each LED and record the photometry signals. The signals were sampled from the photodetector at a rate of 6.1 kHz and demodulated and decimated to 1017 Hz before being saved to disk. For analysis purposes, both signals were downsampled to 102 Hz,

and a least-squares linear fit was applied to the 405 nm signal to align it with the 465 nm signal. This fitted 405 nm signal served as the reference for normalizing the 465 nm signal using the formula % ΔF/F =100 * (465 nm signal – fitted 405 nm signal)/(fitted 405 nm signal).

- ***Main analysis across trials during training***

For LI5/LR5 tasks, all ΔF/F photometry traces were analyzed from -2 to 2 s around each event of the behavioral sequence comprising the lever insertion (LI), the sequence of lever presses (LP1 to LP5) and the port entry (PE; reward retrieval). To analyze rapid changes in activity across training sessions, trials from all recording sessions were concatenated and the mean activity across the first 15 trials, the next 15 trials (trials 16-30) and the last 15 trials was compared for each task. Unless otherwise specified, the area under the curve (AUC) was computed from 0 to 1 s post-event. For statistical analysis, two-way repeated measure ANOVAs (trial set as within-subject factor and sex as between-subject factor) were conducted. Deviation from baseline (%dF/F ≠ 0) was determined via bootstrapped confidence intervals (99% or 95% CIs) as described in [50]. A consecutive threshold of 17 consecutive 10 ms time bins was applied for statistical evaluation.

- ***Deconvolution analysis***

For analysis purposes, both signals were downsampled to 25 Hz. In order to isolate response to each behavioral event (lever presses and reward retrieval), kernels were calculated as previously described [51]: time-dependent GCaMP6f signal was modeled as the sum of the response to each behavioral event. Calcium response to each event was computed as the convolution of a time series representing the time of the event (series of 0 s and 1 s, in which 1 s correspond to the event timestamp) and to the kernel corresponding to the response profile to that event. The model can be written as follows, where g(t) is the GCaMP6f signal, where a, b, etc., are example of behavioral event time series, and where k^a and k^b are the kernels for the corresponding event:

$$g(t) = g0 + \sum_{t'=-1s}^{2s} a(t-t')k^a(t') + \sum_{t'=-1s}^{2s} b(t-t')k^b(t') + \cdots + error$$

For each recording session, the coefficients of the kernels were solved using the method of least-squares in Python (LinearRegression from sklearn package). **Suppl. Fig. 3** displays individual rats' kernel across the first and last recording sessions in the LI5 and LR5 tasks.

To identify when the signal significantly deviated from baseline, we used bootstrapping to obtain the 99% confidence interval and applied the consecutive thresholds method [50]. More specifically, the lower and upper bound of the confidence interval must be above or below 0 for a least 5 consecutive 40 ms time bins to detect statistical deviation from baseline.

- ***Analysis of GCaMP6f and dLight signal under omission schedule***

Photometry traces of % ΔF/F were analyzed within a time window of -5 to 5 seconds around two key events: the port entry that followed a reward omission or unexpected reward delivery. We computed the AUC from 0.5 to 1 second after the reward retrieval. To capture the delayed signal depression following the omission of the reward, we also calculated the AUC from 1 to 3 seconds after omission. To identify significant differences between the LI5 and LR5 tasks as shown in **Fig. 5**, we conducted a comparison of peri-event waveforms between the two tasks, using the two-sample t-test and bootstrap technique (bootstrap difference distribution of randomly resampled means) as described in [50]. Deviation from baseline (%dF/F ≠ 0) was determined via bootstrapped confidence intervals (95% CIs) as described in [50]. A consecutive threshold of 17 consecutive 10-ms time bins was applied for statistical evaluation of differences between tasks and from baseline.

ACKNOWLEDGEMENTS

The authors thank all members of the Janak lab for their helpful discussions on the experimental findings.

FUNDING AND DISCLOSURES

This work was supported by NIH grant R01DA035943 (PHJ) and IRESP grant AAPSPA2021-V1-07 (NT). The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

RM, YV and PHJ designed the experiments; RM, YV, collected the data with technical assistance from JZ, YC and HP; RM, YV analyzed the data with assistance from YC; PHJ and NT provided financial support; RM, YV and PHJ interpreted the data and wrote the manuscript. RM, YC, JZ, HP, NT, PHJ and YV provided inputs, made edits and approved the final version of the manuscript.

REFERENCES

1. Keiflin, R. & Janak, P. H. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron* **88**, 247–263 (2015).
2. Keiflin, R., Pribut, H. J., Shah, N. B. & Janak, P. H. Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. *Current Biology* **29**, 93-103.e3 (2019).
3. Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat Neurosci* **21**, 1072–1083 (2018).
4. Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural Circuitry of Reward Prediction Error. *Annu Rev Neurosci* **40**, 373–394 (2017).
5. Engelhard, B. *et al.* Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).
6. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science (1979)* **275**, 1593–1599 (1997).
7. Fiorillo, C. D., Newsome, W. T. & Schultz, W. The temporal precision of reward prediction in dopamine neurons. *Nat Neurosci* **11**, 966–973 (2008).
8. Fraser, K. M., Pribut, H. J., Janak, P. H. & Keiflin, R. From Prediction to Action: Dissociable Roles of Ventral Tegmental Area and Substantia Nigra Dopamine Neurons in Instrumental Reinforcement. *Journal of Neuroscience* **43**, 3895–3908 (2023).
9. Collins, A. L. *et al.* Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci Rep* **6**, 1–15 (2016).
10. Wassum, K. M., Ostlund, S. B. & Maidment, N. T. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biol Psychiatry* **71**, 846–854 (2012).
11. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat Neurosci* **19**, 117–126 (2015).

12. Phillips, C. D., Hodge, A. T., Myers, C. C., Leventhal, D. K. & Burgess, C. R. Striatal dopamine contributions to skilled motor learning. *J Neurosci* e0240242024 (2024) doi:10.1523/JNEUROSCI.0240-24.2024.
13. 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).
14. Kim, H. R. *et al.* A Unified Framework for Dopamine Signals across Timescales. *Cell* **183**, 1600-1616.e25 (2020).
15. Faure, A., Haberland, U., Condé, F. & El Massioui, N. Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. *Journal of Neuroscience* **25**, 2771–2780 (2005).
16. Faure, A., Leblanc-Veyrac, P. & El Massioui, N. Dopamine agonists increase perseverative instrumental responses but do not restore habit formation in a rat model of Parkinsonism. *Neuroscience* **168**, 477–486 (2010).
17. Belin, D. & Everitt, B. J. Cocaine Seeking Habits Depend upon Dopamine-Dependent Serial Connectivity Linking the Ventral with the Dorsal Striatum. *Neuron* **57**, 432–441 (2008).
18. Wickens, J. R., Horvitz, J. C., Costa, R. M. & Killcross, S. Dopaminergic mechanisms in actions and habits. *Journal of Neuroscience* **27**, 8181–8183 (2007).
19. Lerner, T. N. Interfacing behavioral and neural circuit models for habit formation. *J Neurosci Res* **98**, 1031–1045 (2020).
20. Dickinson, A. Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society* **308**, 67–78 (1985).
21. Haith, A. M. & Krakauer, J. W. The multiple effects of practice: skill, habit and reduced cognitive load. *Curr Opin Behav Sci* **20**, 196–201 (2018).
22. Du, Y., Krakauer, J. W. & Haith, A. M. The relationship between habits and motor skills in humans. *Trends Cogn Sci* **26**, 371–387 (2022).
23. Dezfouli, A. & Balleine, B. W. Habits, action sequences and reinforcement learning. *Eur J Neurosci. 2012 Apr;35(7):1036-51. doi: 10.1111/j.1460-9568.2012.08050.x.* (2012).
24. Vandaele, Y. *et al.* Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife* **8**, 1–29 (2019).
25. Vandaele, Y. & Janak, P. H. Lack of action monitoring as a prerequisite for habitual and chunked behavior: Behavioral and neural correlates. *iScience* **26**, 105818 (2023).
26. Vandaele, Y., Pribut, H. J. & Janak, P. H. Lever insertion as a salient stimulus promoting insensitivity to outcome devaluation. *Front Integr Neurosci* **11**, 1–13 (2017).
27. Tang, J. C. Y. *et al.* Dynamic behaviour restructuring mediates dopamine-dependent credit assignment. *Nature* **626**, 583–592 (2024).
28. Hamid, A. A., Frank, M. J. & Moore, C. I. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733-2749.e16 (2021).
29. Yagishita, S. *et al.* A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science (1979)* **345**, 1616–1620 (2014).
30. Di Chiara, G. Nucleus accumbens shell and core dopamine: differential role in behavior and addiction. *Behavioural Brain Research* **137**, 75–114 (2002).
31. Fraser, K. M. & Janak, P. H. Long-lasting contribution of dopamine in the nucleus accumbens core, but not dorsal lateral striatum, to sign-tracking. *Eur J Neurosci* **46**, 2047–2055 (2017).
32. Wassum, K. M., Ostlund, S. B., Loewinger, G. C. & Maidment, N. T. Phasic mesolimbic dopamine release tracks reward seeking during expression of Pavlovian-to-instrumental transfer. *Biol Psychiatry* **73**, 747–755 (2013).
33. Saunders, B. T. & Robinson, T. E. The role of dopamine in the accumbens core in the expression of Pavlovian-conditioned responses. *Eur J Neurosci* **36**, 2521–2532 (2012).

34. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* **16**, 966–973 (2013).
35. Graybiel, A. M. The Basal Ganglia and Chunking of Action Repertoires. *Neurobiol Learn Mem* **136**, 119–136 (1998).
36. Graybiel, A. M. & Grafton, S. T. The striatum: Where skills and habits meet. *Cold Spring Harb Perspect Biol* **7**, 1–14 (2015).
37. Vandaele, Y., Ottenheimer, D. J. & Janak, P. H. Dorsomedial striatal activity tracks completion of behavioral sequences. *eNeuro* 2021.04.01.437899 (2021) doi:10.1101/2021.04.01.437899.
38. Dezfouli, A. & Balleine, B. W. Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLoS Comput Biol* **9**, e1003364 (2013).
39. Clark, J. J., Collins, A. L., Sanford, C. A. & Phillips, P. E. M. Dopamine encoding of pavlovian incentive stimuli diminishes with extended training. *Journal of Neuroscience* **33**, 3526–3532 (2013).
40. Goedhoop, J., Arbab, T. & Willuhn, I. Anticipation of Appetitive Operant Action Induces Sustained Dopamine Release in the Nucleus Accumbens. *Journal of Neuroscience* **43**, 3922–3932 (2023).
41. Keramati, M. & Gutkin, B. Imbalanced Decision Hierarchy in Addicts Emerging from Drug-Hijacked Dopamine Spiraling Circuit. *PLoS One* **8**, (2013).
42. Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W. & Everitt, B. J. Parallel and interactive learning processes within the basal ganglia: relevance for the understanding of addiction. *Behav Brain Res* **199**, 89–102 (2009).
43. Ambrosi, P. & Lerner, T. N. Striatonigrostriatal circuit architecture for disinhibition of dopamine signaling. *Cell Rep* **40**, (2022).
44. Yang, H. *et al.* Nucleus Accumbens Subnuclei Regulate Motivated Behavior via Direct Inhibition and Disinhibition of VTA Dopamine Subpopulations. *Neuron* **97**, 434-449.e4 (2018).
45. Tang, J. C. Y. *et al.* Dynamic behaviour restructuring mediates dopamine-dependent credit assignment. *Nature 2023 626:7999* **626**, 583–592 (2023).
46. Pascoli, V., Terrier, J., Hiver, A. & Lüscher, C. Sufficiency of Mesolimbic Dopamine Neuron Stimulation for the Progression to Addiction. *Neuron* **88**, 1054–1066 (2015).
47. Millard, S. J. *et al.* Cognitive representations of intracranial self-stimulation of midbrain dopamine neurons depend on stimulation frequency. *Nature Neuroscience 2024 27:7* **27**, 1253–1259 (2024).
48. Witten, I. B. *et al.* Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* **72**, 721–733 (2011).
49. Steinberg, E. E. *et al.* Positive reinforcement mediated by midbrain dopamine neurons requires D1 and D2 receptor activation in the nucleus accumbens. *PLoS One* **9**, (2014).
50. Jean-Richard-dit-Bressel, P., Clifford, C. W. G. & McNally, G. P. Analyzing Event-Related Transients: Confidence Intervals, Permutation Tests, and Consecutive Thresholds. *Front Mol Neurosci* **13**, (2020).
51. Parker, N. F. *et al.* Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat Neurosci* **19**, 845–854 (2016).