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

Background: Mesolimbic dopamine (DA) neurons are central to cue guided reward seeking and action sequence learning. Yet, the mechanisms by which cue-induced DA neural activity drives goal-directed or habitual sequence execution remain unknown.
Methods: We designed two novel tasks to isolate the effect of sequence-delineating cues on DA-driven behavioral strategies and learning. In the lever insertion fixed-ratio 5 task (LI5), the lever insertion marked sequence initiation. In the lever retraction fixed-ratio 5 task (LR5), the lever retraction served as both sequence termination and reward-predictive cue.
Results: 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, to some extent, behavioral chunking.
Conclusion: 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: 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 are deviations from expected outcome that 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 to sequence initiation[9–12]. Mesolimbic DA activity predicts action sequences performance[9,10] and scale with distance to reward to sustain motivation toward goals[13,14]. Although DA has been related to habit learning and response automaticity[15–19], it remains unclear how DA responses at action and reward-predictive cues evolve over sequence learning when behavior is goal-directed, i.e. driven by outcome representation, and when it becomes automated and habitual. Indeed, studies of DA dynamics during sequence learning have not characterized whether behavior was habitual or goal-directed [9–12], with one exception [20], which lacked within-session resolution during early habit formation.

Habits are commonly defined as stimulus-response behaviors[21], and are predominantly identified as an absence of goal-directed control (i.e. insensitivity to outcome devaluation or contingency degradation)[22,23]. This description has been extended to include chunked action sequences executed automatically without outcome expectation until completion[24–26]. Several studies reveal the importance of sequence-delineating cues in habit learning and behavioral chunking[27–32], either by making stimulus-response associations more explicit and/or by reducing attentional demand during execution of behavior. We previously showed that in a discrete-trials task, lever insertion, which signals trial onset and sequence initiation, and retraction, which signals reward delivery and sequence termination, promote habit, automaticity and behavioral chunking[27,28,33]. Yet, it remains unclear whether these cues differentially shape the development of these behaviors. It is also unknown whether their presentation differentially influence reward expectation and is associated with different distinct mesolimbic DA activity patterns reflecting specific response strategies.

To assess how sequence initiation and termination cues influence habitual or goal-directed response strategy, we designed two isomorphic fixed-ratio-5 tasks in which only one sequence-delineating cue was present, either lever insertion in LI5 task, or lever retraction in LR5 task. Based on prior work[28], we hypothesized that the sequence termination cue in the LR5 task would serve as response feedback, alleviating requirements for reward monitoring, thereby promoting automatic and chunked responses. We further predicted that this proximal reward-predictive cue would drive core features of habits such as insensitivity to both outcome devaluation and reversed instrumental contingency[27,28,33].

We propose that DA-mediated credit assignment[34–36] may link the sequence termination cue to habit formation by shifting the expectation of reinforcement from individual presses to the entire sequence as a chunk executed automatically. To test this, we tracked mesolimbic DA neuron activity along the action sequence in the LI5 and LR5 task, and across training. Finally, we used optogenetic manipulation to demonstrate the role of DA-mediated credit assignment in the development of automaticity and behavioral chunking.

MATERIALS & METHODS

Wild-type (WT) (n = 48) or TH-Cre Long-Evans rats (n = 36) were used in this study (details in supplement methods). We designed two fixed-ratio 5 tasks to discern the respective role of initiation (lever insertion) and termination (lever retraction) cues in action sequence learning. In the lever insertion fixed-ratio 5 (LI5) task, trials began with lever insertion. Rats had to complete a sequence of five lever presses to obtain the reward (0.1 ml 20% sucrose), with the lever retracting randomly during reward retrieval and thus not predicting sequence completion (**Fig. 1a**). In

contrast, in the lever retraction fixed-ratio 5 (LR5) task, the lever retracted immediately after the fifth lever press, serving as the termination cue to signal reward delivery, while the lever insertion occurred during reward retrieval and did not signal trial onset (**Fig. 1b**).

The first experiment aimed at characterizing behavior in the two tasks. To measure habit learning, sensitivity to outcome devaluation was assessed after 12 training sessions. In the second and third experiments, we used fiber photometry to respectively record ventral tegmental area (VTA) DA neural activity and DA release in the NAc Core. Habit learning was assessed by testing response sensitivity to an omission schedule. DA signals were recorded throughout training and during the omission test. In the fourth experiment, TH-Cre$^+$ rats received optogenetic VTA DA neurons stimulation at sequence completion and in absence of lever retraction cue for 12 consecutive sessions. Virus expression and optic fiber placement were verified with histological procedures. For full methodological details, see supplement methods.

RESULTS

Sequence termination cues promote habit-like behavior

Task differences rapidly emerged during training with increased response rate for rats trained in the LR5 task, compared to the LI5 task (**Fig. 1c-d**). While both groups had identical inter-press intervals (IPI) in the first session, with a similar proportion of IPIs shorter than 0.5s (**Fig. 1e**), LR5 rats showed a greater proportion of shorter IPIs on the last training session (**Fig. 1f**), indicating greater action automaticity. Consistent with these observations, LR5 rats rapidly learned to continuously press on the lever until its retraction, leading to a drastic decrease in the likelihood of port checking during sequence execution (**Fig. 1g**). In contrast, LI5 rats maintained a high probability of port-checking across sessions (**Fig. 1g-h**). This pattern suggests that LR5 rats learned to concatenate individual lever presses into a single behavioral sequence executed without intermediate reward expectation, a hallmark of behavioral chunking. Early in training, LR5 rats often check the port after 1-3 lever presses (**Fig. 1i**), but by training end, they reliably completed full lever press sequences without checking (**Fig. 1j**). In contrast, LI5 rats showed no evidence of chunking, as they persisted in port checking after each response throughout training (**Fig. 1i** and **1j**).

Task differences in action automaticity and behavioral chunking were associated with differences in sensitivity to sensory-specific satiety-induced outcome devaluation (**Suppl. Fig. 1**). Indeed, we observed a higher proportion of rats insensitive to outcome devaluation in the LR5 task compared to the LI5 task (devaluation ratio ≤ 0; 8/15 in LR5 task vs 1/15 in LI5 task). We found no effect of sex in these tests (**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 a reduced sensitivity 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 sequence execution depend on cue-mediated reward expectation and prediction, we used fiber photometry to record VTA DA neuron activity across training in the LR5 and LI5 tasks. GCaMP6f was expressed in (TH)-Cre$^+$ rats to selectively target VTA DA neurons with 95% specificity (**Fig. 2a-c**). We replicated the task differences in behavior (**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 DA neural activity at lever insertion, lever retraction, during lever presses, and at reward retrieval. 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 lever insertion signals opportunity to respond for reward, DA neural activity increased at lever insertion (**Fig. 2d**) and at reward retrieval (**Fig. 2f**), with no change across trials and no response at sequence completion (**Fig. 2e**). In contrast, in the LR5 task, where lever retraction signals sequence completion and reward delivery, we observed rapid training-related changes in DA activity shifting from reward to the retraction cue as learning progressed (**Fig. 2h-i**), and no training-related changes at lever insertion (**Fig. 2g**). With extended training in the LR5 task, both the lever retraction and reward-related DA responses declined (**Fig. 2h-i**; last 15 trials).

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

DA neural activity patterns during the lever press sequence differed across tasks (**Fig. 3**). To compare activity at lever presses, we examined the first 4 responses as the fifth press coincides with lever retraction in the LR5 task (**Fig. 2**). In the LI5 task, DA neural activity associated with sequence initiation changed over training, initially peaking after the first lever press (first and next 15 trials) and progressively shifting to before the lever press (last 15 trials; **Fig. 3a**). DA response decreased as rats progressed in the sequence from the first to fourth response (**Fig. 3e-f**), with an overall suppression immediately post-press during the last 15 trials (**Fig. 3 a-d and g**). Similarly, phasic DA responses evoked by premature port-checking also declined along the sequence (**Suppl. Fig. 3a-d**). Notably, 1-2s after each port entry, we observed an increased inhibition in DA neural activity along the sequence (**Suppl. Fig. 3e-g**). This inhibition in LI5-trained rats might reflect a negative reward prediction error (RPE) signal, indicating a growing violation of reward expectation as rats approached sequence completion.

In contrast, LR5-trained rats showed minimal DA modulations around lever presses (**Fig. 3h-k**), with modest decreases across training (**Fig. 3l-n**). Deconvolution analysis, isolating neural responses to specific events while avoiding contamination from adjacent events (supplement methods), confirmed similar patterns of DA neural activity between the first and last training sessions in both tasks (**Suppl. Fig. 4)**.

Patterns of DA release in the Nucleus Accumbens Core in response to cues, lever presses and reward are similar to VTA

We next asked if the distinct patterns of VTA DA neuron activity were similarly observed in DA release in the NAc Core, a primary target for VTA DA neurons implicated in DA-mediated cue processing and reward learning[37–40]. To this end, we infused the DA sensor dLight1.2 into NAc Core in male and female rats (**Fig. 4a-b**).
DA release patterns tended to mirror VTA DA neuron activity across training. 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, LI5 rats showed a reward-evoked DA release that somewhat diminished by training end (**Fig. 4e**), whereas in LR5 rats, the reward-related peak rapidly shifted to the lever retraction cue (**Fig. 4h**).
Consistent with VTA recordings, LI5 rats exhibited DA release time-locked to lever presses (**Suppl. Fig. 5a-g**), but the amplitude of the DA signal remained stable across training. No lever press-elicited DA release was observed in LR5-trained rats (**Suppl. Fig 5h-n**).

Combining VTA and NAc Core recordings, we found that in the LR5 task, where lever retraction signals sequence completion, DA responses rapidly shift within a single session, from reward retrieval to sequence termination. In contrast, in the LI5 task, where no cue signals reward

delivery, DA signals remain anchored to the reward, even after extended training. Importantly, recurrent port checking in LI5 task likely reflects progress monitoring. Together, these results illustrate how animals adapt distinct behavioral strategies to achieve their goals based on the sequence-delineating cue presented.

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 underwent a single session under an omission schedule to evaluate behavioral and dopaminergic responses to reversed instrumental contingencies. The omission procedures mirrored the task structures experienced during training, but required rats to refrain from pressing the lever for 20 seconds to receive a reward (**Fig. 5a-b, supplement methods**).

During the omission test, LR5-trained rats responded more than LI5 rats (**Fig. 5c**), indicating reduced ability to adapt to the new contingencies. LR5 rats completed the 5-lever press sequence faster than LI5 rats (**Fig. 5d**). LI5 rats completed fewer sequences in the omission test compared to the last training session whereas LR5 rats completed more sequences (**Fig. 5e**). While LI5 rats frequently checked the magazine between responses, LR5 rats remained cue-driven and completed most sequences in chunks of 5 presses, even in absence of reward (**Fig. 5f**). This strategy difference is further confirmed by the smaller port-checking probability in LR5 rats (**Fig. 5g).**

Task-related differences in omission sensitivity were mirrored by distinct DA dynamics in VTA and NAc Core. LI5 rats exhibited a robust decrease in VTA DA neuron activity following reward omission (**Fig. 5j-k**) that was absent in LR5 rats. While VTA DA activity increased in both groups following unexpected reward delivery, this response was greater in LI5 rats (**Fig. 5l-m**). Similarly, NAc Core DA release more strongly decreased after omitted reward in LI5 rats and the response was attenuated in LR5 rats (**Fig. 5n-o**), along with no group differences in response to unexpected reward (**Fig. 5p-q**). Port entry latency after reward omission (**Fig. 5h**) or delivery (**Fig. 5i**) did not differ between groups, ruling out behavioral difference at these events as a source of mesolimbic DA activity discrepancy. These results suggest that LR5 training impairs adaptation to reward omission, by attenuating both negative RPEs following reward omission and positive RPEs following unexpected reward delivery.

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

Given that lever retraction paired with reward delivery evokes phasic VTA DA activity and promotes automaticity and behavioral chunking in LR5 rats, we tested whether mimicking this DA activity could replicate the behavior. TH-Cre^{+} rats received unilateral VTA infusions of DIO-ChR2 or DIO-eYFP virus and were implanted with an optic fiber (**Fig. 6a**). We modified the LI5 task (**Fig. 6b**) to incorporate brief optogenetic stimulation (0.5s; 20mW; 20Hz) to simulate the DA burst observed upon sequence completion and lever retraction in the LR5 task. As in the original LI5 task, the lever remained extended until rats retrieved the reward, preventing them from using lever retraction as a reward-predictive cue.

Optogenetic stimulation at sequence completion progressively increased response rates in ChR2 rats across training compared to eYFP rats, also receiving laser presentations (**Fig. 6c-d**). While both groups showed similar IPI distribution in the first session (**Fig. 6e**), by the last session, ChR2 rats exhibited a higher proportion of very short IPIs (**Fig. 6f**). Optogenetic stimulation also reduced port-checking in ChR2 rats (**Fig. 6g**), with all individuals decreasing port entry likelihood between first and last sessions (**Fig. 6h**). Finally, sequence length analysis revealed no group difference

during the first session, with mostly single presses before port check (**Fig. 6i**, ~60%). In the last session, eYFP rats maintained frequent port checking, whereas ChR2 rats made short bouts of 2 or 3 presses before checking the port (**Fig. 6j**), consistent with emerging behavioral chunking.

Although both groups obtained a similar number of rewards, ChR2 rats exceeded the response requirement and retrieved rewards faster than eYFP rats (**Suppl. Fig. 6a-c**). With the parameters used in this experiment (0.5s; 20mW; 20Hz, no cue), ChR2 rats did not show robust intra-cranial self-stimulation (ICSS) in subsequent tests, except for one rat (**Suppl. Fig. 6d-f**). This indicates that VTA DA neuron ICSS does not explain the behavioral effects, as DA-driven responses would be expected to delay, not accelerate, reward retrieval. Importantly, robust ICSS was observed when laser stimulation was increased to 2s (**Suppl. Fig. 6g-i**). In summary, brief unilateral VTA DA neuron activation at sequence completion is sufficient to increase automaticity and, to a lesser extent, behavioral chunking.

DISCUSSION

We studied how sequence initiation and termination cues differentially influence reward expectation along the sequence and drive habitual and goal-directed strategies. We found that the lever retraction cue marking sequence completion in the LR5 task facilitated habit-like behavior. Conversely, in the LI5 task, in which lever insertion signals trial onset, rats exhibited greater goal-directedness, repeatedly checking for reward between lever presses. These behavioral strategies were associated with distinct mesolimbic DA dynamics, with LR5 rats exhibiting a shift in DA activity from the reward to its predictive cue, while LI5 rats maintained reward-related DA activity with expectation along sequence execution. 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, in absence of lever retraction cue, was sufficient to increase automaticity and behavioral chunking.

We previously showed that lever cues at sequence boundaries promote habit and behavioral chunking[27,28,33]. These findings have been replicated using other cue types such as auditory tones[30,31,41]. Here, we extended these findings by demonstrating that the lever retraction cue, but not the insertion, is sufficient to promote habit-like behavior. Indeed, LR5-trained rats showed automatic, chunked responses without reward expectation until sequence completion[42,43]. They were also less sensitive to outcome devaluation and negative contingencies in the omission schedule, suggesting reduced behavioral flexibility. These results support the hypothesis that the proximal reward-predictive cue provides response feedback, reducing the need for action-outcome monitoring and promoting habit formation[28,41]. Our prior work[28,44] showed that rats required to complete five consecutive lever presses without port checking reached LR5-like levels of efficiency, but their behavior relied on internal monitoring and remained goal-directed. Thus, the feedback cue is the critical feature driving the shift toward habitual strategy in the LR5 task.

While DA has been implicated in both sequence[8–10,45] and habit learning[18,46,47], these constructs are often studied separately. Yet, in a recent action sequence study, Van Elzelingen et al.[20] linked shifts in DMS DA release to habit development over extended training, though early learning dynamics were not examined. In contrast, our findings in VTA and NAc show that DA activity is highly dynamic from the very first session, closely tracking changes in reward expectation throughout sequence execution, and reflecting key differences in habitual versus goal-directed strategies. In LR5-trained rats, DA signals rapidly shifted from reward retrieval to the proximal reward predictive stimulus (i.e lever retraction) as previously observed[9,10]. During lever presses, there was no reward expectation, as reflected by the lack of within-sequence port entry and the absence of DA neural activity.

In the LI5 task, DA responses during early training peaked after the first lever press, but with extended training, this peak preceded the press, indicating that rats learned to expect reward after the lever insertion cue and at the lever approach. Additionally, each lever press evoked a phasic VTA DA burst that decayed across the sequence, while no DA release was observed in NAc. This contrasts with prior studies where accumbal DA progressively ramps up during a motivated approach in an action sequence task[10], decision-making task[48,49] or when rats have to navigate a maze to obtain a reward[13,50]. In our LI5 task, the absence of ramping DA might be caused by the absence of sensory feedback along the chain of action[51]. Instead, in LI5 rats, inter-press port checking triggered a biphasic VTA DA neuron response, with a brief peak synchronous to the port entry, followed by inhibition of DA activity one second later. This biphasic pattern likely reflects reward expectation when rats entered the port and a negative RPE when reward expectation was violated. The progressive increase in negative RPE magnitude along the sequence suggests that LI5 rats use an internal model of the task, as their reward expectation builds while approaching reward delivery.

We found a training-related decline in DA signals in both VTA and NAc Core in the LR5 task but not in the LI5 task. Similar reductions in cue-evoked DA release have been reported as animals became experts in the task[9,52,53]. This decline is unlikely due to signal bleaching, as DA activity fully recovered following unexpected reward delivery in the omission test. Instead, it may reflect learned expectation of the sequence termination cue[52,53]. Alternatively, with habit formation, termination cue encoding may shift from the mesolimbic to the nigrostriatal pathway[17,19,54–57]. This pathway shift during learning has been suggested as a neural basis for the development of automaticity and the transition from goal-directed to habitual control[18,19,54].

Dickinson et al.[58] first showed that extensive reinforcement history and habit learning drives behavioral persistence under an omission schedule. Accordingly, training in the LI5 and LR5 tasks lead to distinct behavioral and dopaminergic adaptations during the omission test. LI5-trained rats maintained goal-directed control and were able to flexibly reduce their response in absence of reward. Their neural signature showed negative RPE following reward omission and positive RPE following unexpected reward delivery. Conversely, LR5 rats' resistance to reward omission was associated with attenuated positive RPE and blunted negative RPE, suggesting reduced sensitivity to negative contingency. Studies have shown that DA responses to conditioned reward cues and reward omission scale with uncertainty of reinforcement[6,59]. Under high uncertainty, VTA DA neurons are activated by reward omission, potentially driving adaptive learning and behavioral switching to obtain reward[60]. After training under reward uncertainty in the LI5 task, repeated negative RPEs after reward omission likely drives behavioral extinction and adaptation to the reversal in instrumental contingencies. In contrast, when reward delivery is predicted with certainty by the lever retraction in the LR5 task, rats may persist to respond, leading to habitual behavior that is less sensitive to outcome omission and relies more on cues. These results suggest that the lack of negative RPE signal is linked to the reduced behavioral flexibility that characterizes habits.

Dopamine-dependent synaptic plasticity in corticostriatal circuits is essential for habit learning[18,47]. In the striatum, sensorimotor events such as lever presses, trigger synaptic activity that leaves behind eligibility traces lasting approximatively 1-2 seconds[36,61]. If a reward is delivered during this window, DA signaling can induce synaptic plasticity, facilitating credit assignment between an action and its outcome[36,61–63]. Our lab showed that VTA DA neuron ICSS in an action sequence task results in the backpropagation of the reinforcement signal, promoting sequence completion[8]. Similarly, another study showed that optogenetic stimulation of the DMS at the first lever press promotes habit[20]. Interestingly, both studies involved seeking-taking chain schedules in which pressing on a lever triggered extension of another lever and pressing on this second lever resulted

in reward delivery. This type of task may facilitate the backpropagation of eligibility traces from proximal cues to distal cues. Here, the sequence termination cue elicited a transient DA response as rats learned to complete the LR5 sequence. Mimicking this DA signal with optogenetics, without the actual lever retraction cue, promoted partial automaticity and behavioral chunking. A possible mechanism is that optogenetic DA stimulation time-locked to the final lever press facilitates credit assignment between sequence completion and the reward. With repetition, this may support the backward spread of credit to earlier action in the sequence, ultimately reinforcing the sequence as a whole unit[61].

Notably, DA-stimulated rats often exceeded the required number of lever presses to receive the reward. While this might resemble ICSS, the brief (0.5s) VTA stimulation used here lacks the strong reinforcing effect seen with longer (1s) stimulation durations reported in other studies[2,3,64,65] (but see[66]). Despite this, ChR2 rats remained engaged in the task, as evidenced by the number of rewards obtained and a short retrieval latency. Thus, the additional lever presses likely indicate the difficulty in inhibiting a prepotent behavior due to increased speed in sequence execution. Likewise, LR5 trained rats often pressed on the lever while retracting, leading to sequences of 6 or 7 lever presses (**Fig. 1**).

To conclude, we found sequence initiation and termination cues differentially affect reward expectation along the sequence and drive distinct response strategies that are encoded by specific mesolimbic DA patterns. The goal-directed strategy was associated with recurrent and increasing DA-encoded reward expectation throughout the sequence whereas the habitual strategy was marked by DA signals confined to the sequence termination and attenuated RPEs following expectation violations. These findings underscore the crucial role of mesolimbic DA signals at sequence termination in assigning credit to the whole action sequence to drive automated and habitual behavior.

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.

Data and code availability

Data and python code used for analysis and figures have been deposited at G-node and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

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. van Elzelingen, W. *et al.* Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Current Biology* **32**, 1163-1174.e6 (2022).
21. Dickinson, A. Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society* **308**, 67–78 (1985).
22. Watson, P. & de Wit, S. Current limits of experimental research into habits and future directions. *Curr Opin Behav Sci* **20**, 33–39 (2018).
23. Vandaele, Y. & Janak, P. H. Defining the place of habit in substance use disorders. *Prog Neuropsychopharmacol Biol Psychiatry* (2017) doi:10.1016/j.pnpbp.2017.06.029.
24. 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).
25. Du, Y., Krakauer, J. W. & Haith, A. M. The relationship between habits and motor skills in humans. *Trends Cogn Sci* **26**, 371–387 (2022).
26. 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).
27. 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).
28. 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).
29. Bouton, M. E. Context, attention, and the switch between habit and goal-direction in behavior. *Learn Behav* **49**, 349–362 (2021).
30. Thrailkill, E. A., Michaud, N. L. & Bouton, M. E. Reinforcer predictability and stimulus salience promote discriminated habit learning. *J Exp Psychol Anim Learn Cogn* **47**, 183–199 (2021).
31. Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A. & Bouton, M. E. Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *J Exp Psychol Anim Learn Cogn* **44**, 370–384 (2018).
32. Turner, K. M. & Balleine, B. W. Stimulus control of habits: Evidence for both stimulus specificity and devaluation insensitivity in a dual-response task. *J Exp Anal Behav* (2023) doi:10.1002/jeab.898.
33. Vandaele, Y. *et al.* Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife* **8**, 1–29 (2019).
34. Tang, J. C. Y. *et al.* Dynamic behaviour restructuring mediates dopamine-dependent credit assignment. *Nature* **626**, 583–592 (2024).
35. 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).
36. Yagishita, S. *et al.* A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science (1979)* **345**, 1616–1620 (2014).
37. Di Chiara, G. Nucleus accumbens shell and core dopamine: differential role in behavior and addiction. *Behavioural Brain Research* **137**, 75–114 (2002).
38. 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).
39. 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).
40. 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).

41. Lin, Z. *et al.* Sustaining temporal attention prevents habit expression during operant learning in rats. *Sci Rep* **10**, 1–7 (2020).
42. Graybiel, A. M. The Basal Ganglia and Chunking of Action Repertoires. *Neurobiol Learn Mem* **136**, 119–136 (1998).
43. Graybiel, A. M. & Grafton, S. T. The striatum: Where skills and habits meet. *Cold Spring Harb Perspect Biol* **7**, 1–14 (2015).
44. 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.
45. Hollon, N. G., Arnold, M. M., Gan, J. O., Walton, M. E. & Phillips, P. E. M. Dopamine-associated cached values are not sufficient as the basis for action selection. *Proc Natl Acad Sci U S A* **111**, 18357–18362 (2014).
46. Yin, H. H. *et al.* Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nat Neurosci* **12**, 333–341 (2009).
47. Wang, L. P. *et al.* NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron* **72**, 1055–1066 (2011).
48. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat Neurosci* **19**, 117–126 (2015).
49. Mohebi, A. *et al.* Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).
50. Krausz, T. A. *et al.* Dual credit assignment processes underlie dopamine signals in a complex spatial environment. *Neuron* **111**, 3465-3478.e7 (2023).
51. Mikhael, J. G., Kim, H. G. R., Uchida, N. & Gershman, S. J. The role of state uncertainty in the dynamics of dopamine. *Current Biology* **32**, 1077-1087.e9 (2022).
52. 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).
53. 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).
54. Keramati, M. & Gutkin, B. Imbalanced Decision Hierarchy in Addicts Emerging from Drug-Hijacked Dopamine Spiraling Circuit. *PLoS One* **8**, (2013).
55. 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).
56. Ambrosi, P. & Lerner, T. N. Striatonigrostriatal circuit architecture for disinhibition of dopamine signaling. *Cell Rep* **40**, (2022).
57. 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).
58. Dickinson, A., Squire, S., Varga, Z. & Smith, J.W. Omission Learning after Instrumental Pretraining. *The Quarterly Journal of Experimental Psychology Section B* **51**, 271–286 (1998).
59. Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science (1979)* **299**, 1898–1902 (2003).
60. Ishino, S. *et al.* Dopamine error signal to actively cope with lack of expected reward. *Sci Adv* **9**, (2023).
61. Yamaguchi, K. *et al.* A behavioural correlate of the synaptic eligibility trace in the nucleus accumbens. *Sci Rep* **12**, (2022).

62. Shindou, T., Shindou, M., Watanabe, S. & Wickens, J. A silent eligibility trace enables dopamine-dependent synaptic plasticity for reinforcement learning in the mouse striatum. *European Journal of Neuroscience* **49**, 726–736 (2019).
63. Fisher, S. D. *et al.* Reinforcement determines the timing dependence of corticostriatal synaptic plasticity in vivo. *Nat Commun* **8**, (2017).
64. Steinberg, E. E. *et al.* A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* **16**, 966–973 (2013).
65. Pascoli, V., Terrier, J., Hiver, A. & Lüscher, C. Sufficiency of Mesolimbic Dopamine Neuron Stimulation for the Progression to Addiction. *Neuron* **88**, 1054–1066 (2015).
66. 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).

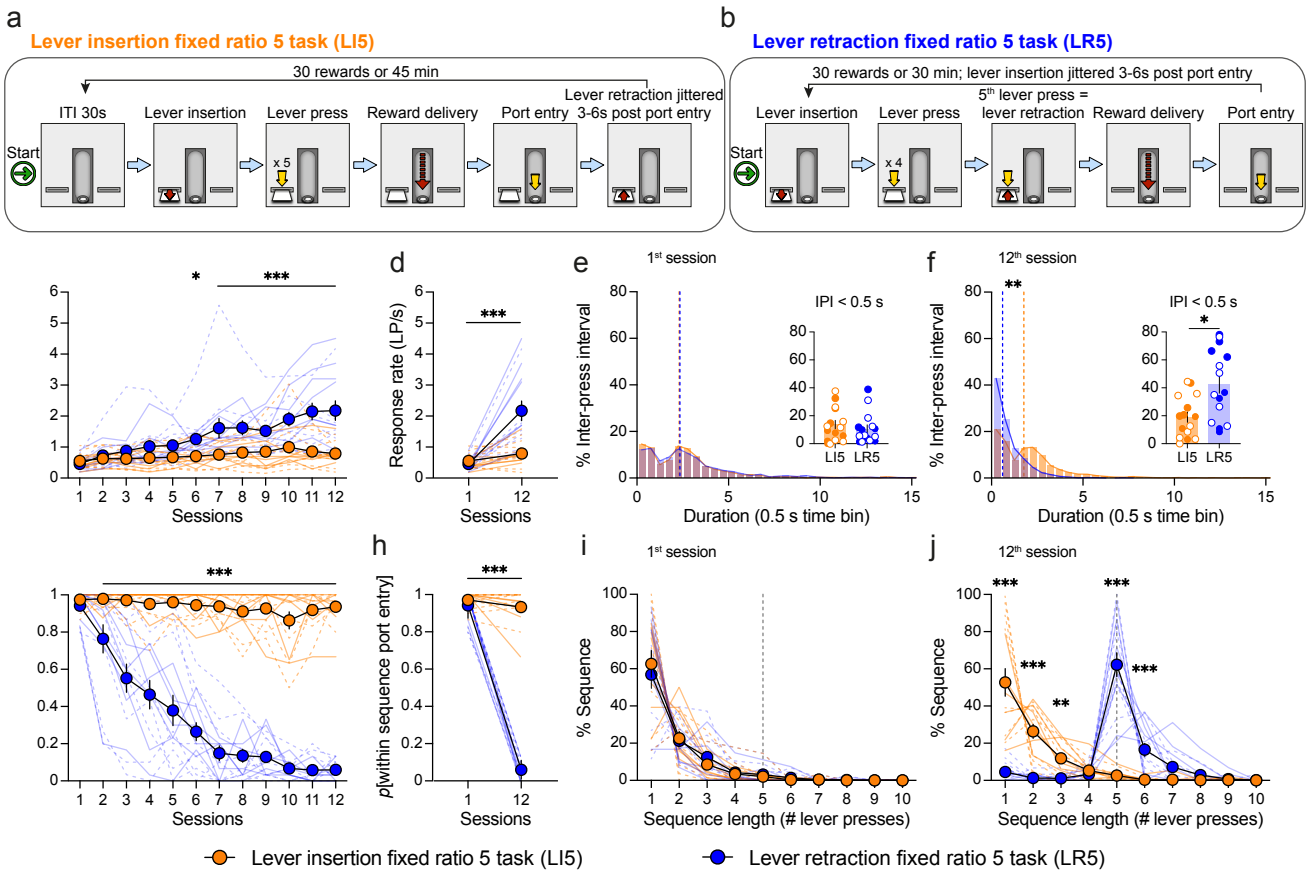

Figure 1. The lever retraction cue, but not the lever insertion cue, promotes action automaticity and behavioral chunking. (**a**) Lever insertion fixed-ratio 5 (LI5) task design. (**b**) Lever retraction fixed-ratio 5 (LR5) task design. (**c**) Response rate in lever press per second (LP/s), within 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.

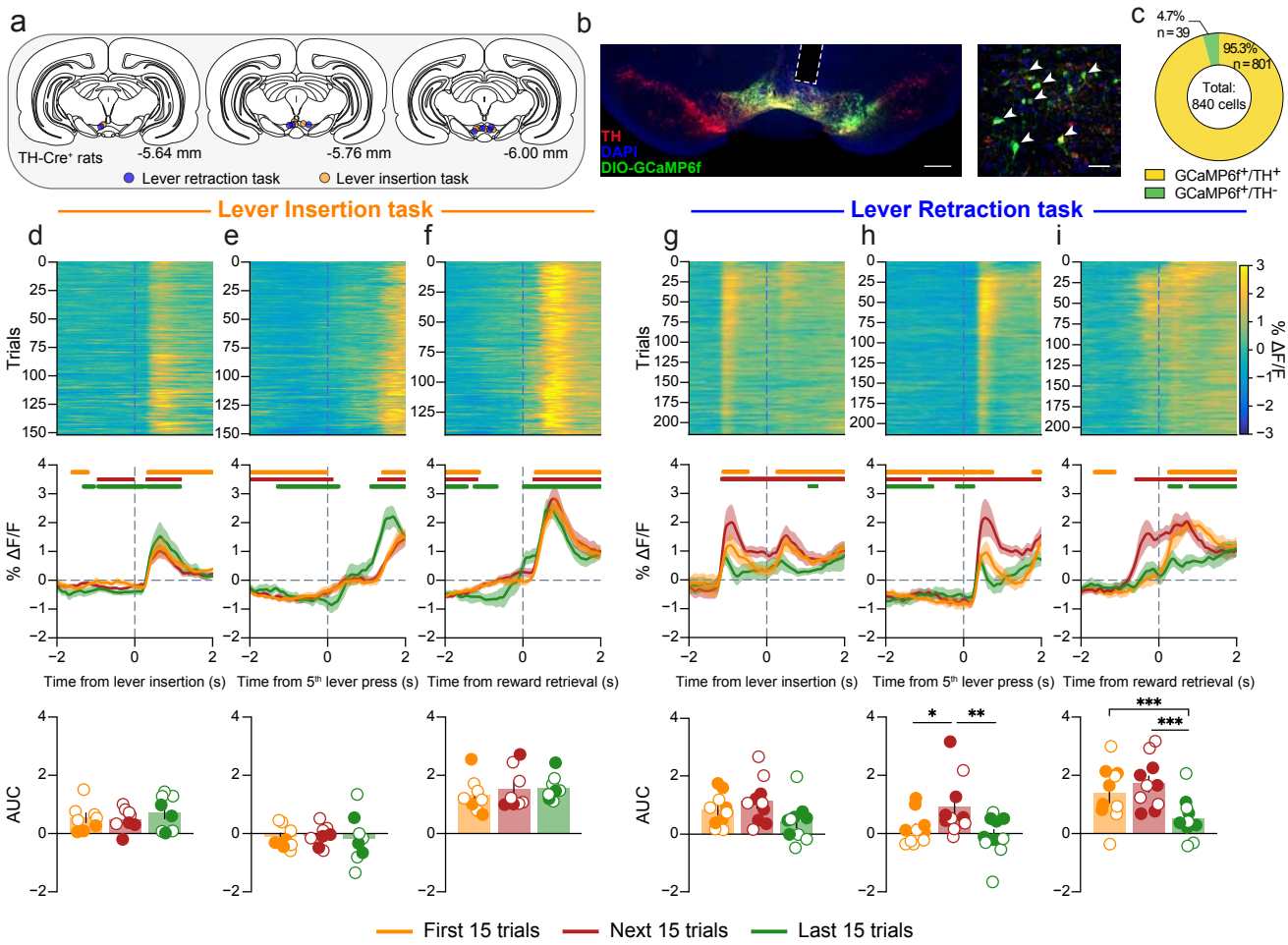

Figure 2. Training-dependent changes in VTA dopamine neuronal phasic activity in response to cues and reward in the LR5 task, but not the LI5 task. (**a**) Coronal sections showing location of optic fiber tips relative to Bregma for TH-Cre+ rats in LI5 and LR5 tasks. (**b**) Left panel: DIO-GCaMP6f was 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.

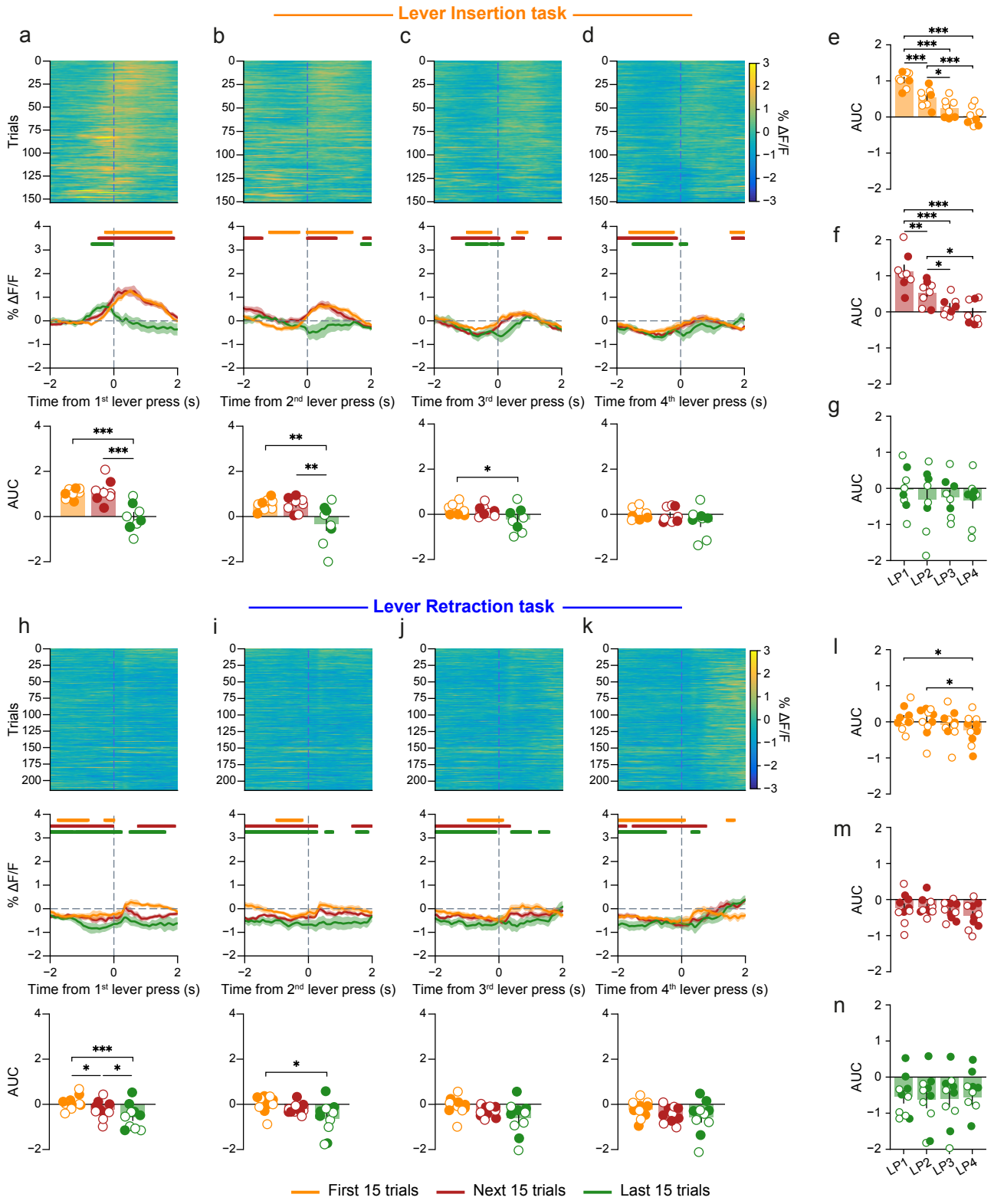

Figure 3. Distinct patterns of VTA DA neuronal activity during lever press sequences in the LI5 and LR5 tasks. (**a-d**) Sequence of the first four lever presses for the LI5 task. The upper panel heatmaps illustrate trial-by-trial dopamine neuron activity, averaged across rat. The middle panel represents 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. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

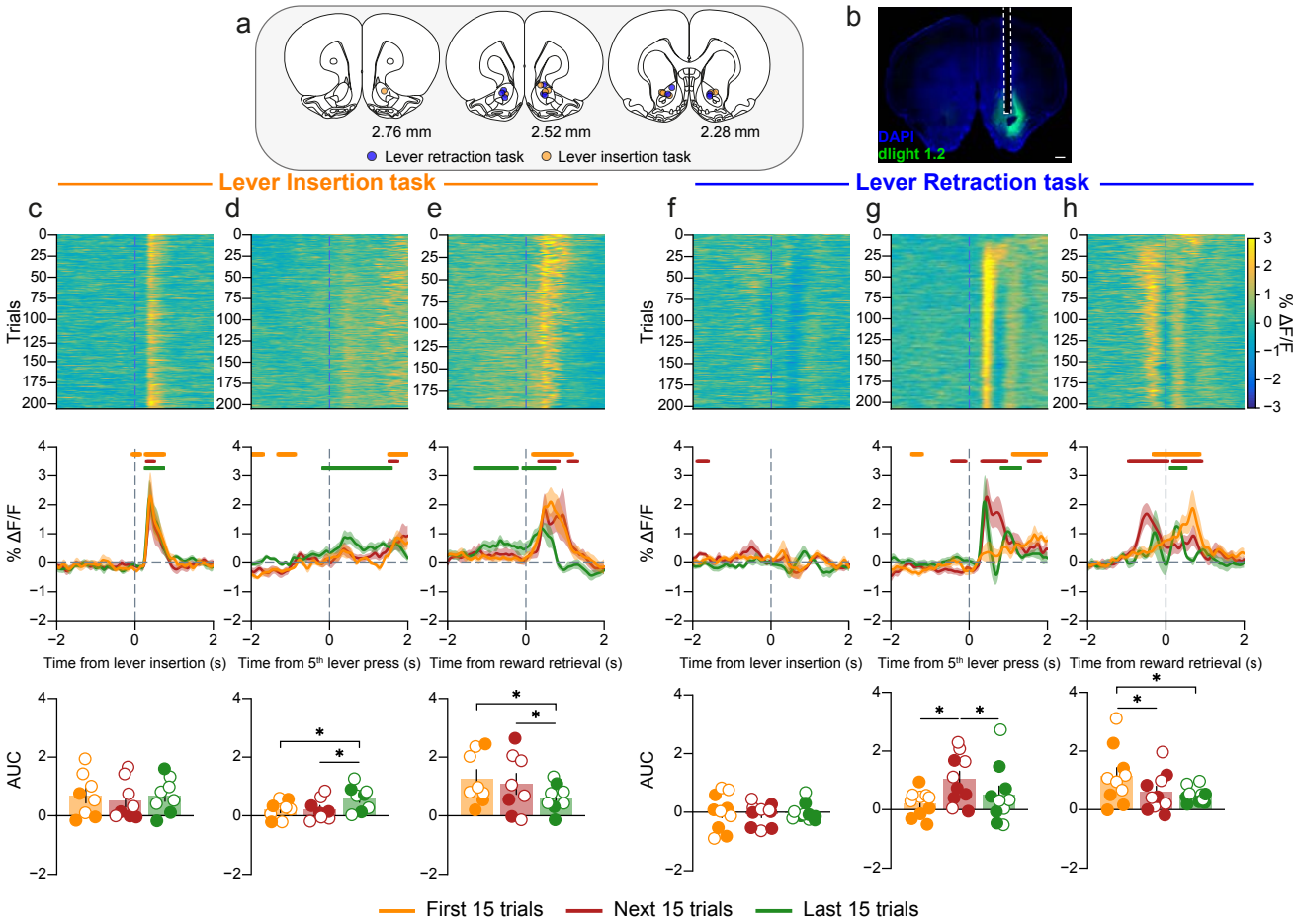

Figure 4. Dynamic changes in dopamine release in the NAc Core in response to cues and reward in the LI5 and LR5 tasks. (**a**) Coronal sections showing location of optic fiber tips relative to Bregma in LI5 and LR5 tasks. (**b**) dLight 1.2 was infused in the NAc Core, and the optic fiber, represented 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.

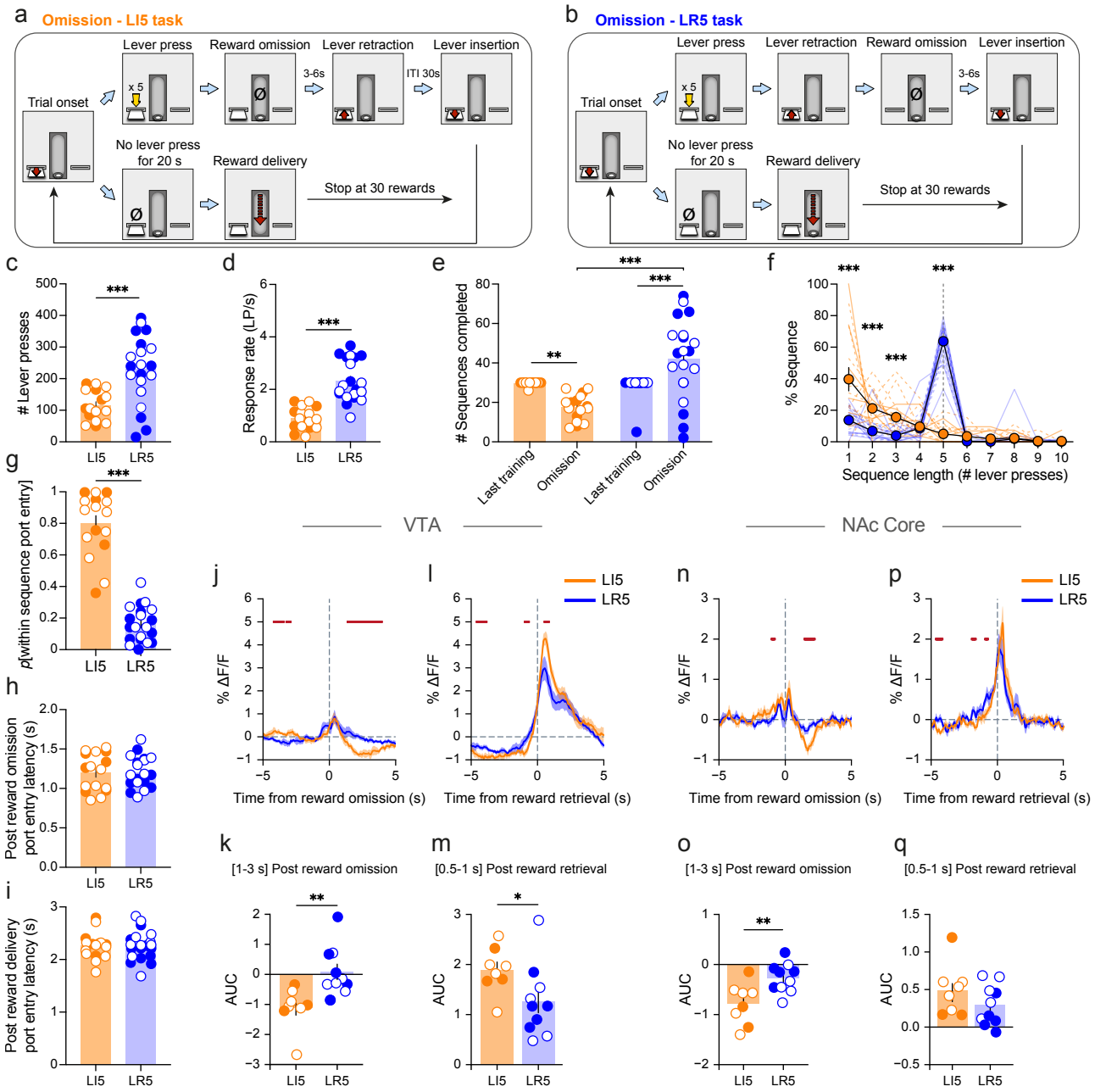

Figure 5. Impaired contingency learning induced by the lever retraction cue is associated with attenuated RPE DA signals in the mesolimbic pathway. Omission test design for the LI5 task (**a**) and the LR5 task (**b**). (**c**) Total number of lever presses during the omission test (two-way ANOVA, task effect: $F_{1,32} = 17.87$, $p < 0.001$; no task x region interaction: $F_{1,32} < 1$, $p = 0.529$). (**d**) Response rate in lever press per second (LP/s), within the lever press sequence (two-way ANOVA, task effect: $F_{1,32} = 40.49$, $p < 0.001$; no task x region interaction: $F_{1,32} < 1$, $p = 0.821$). (**e**) Number of sequences completed during the last LI5/LR5 training session and the omission test (three-way RM ANOVA, task effect: $F_{1,32} = 18.39$, $p < 0.001$; task x session interaction: $F_{1,32} = 34$, $p < 0.001$; no task x session x structure interaction $F_{1,32} = 1.43$, $p = 0.24$). (**f**) Frequency distribution of sequence lengths during the omission session. The grey dashed line represents the "optimal" sequence length of 5 lever presses (three-way RM ANOVA, no task effect: $F_{1,32} < 1$, $p = 0.585$; task x sequence length interaction: $F_{9,288} = 47.06$, $p < 0.001$; no task x sequence length x region interaction: $F_{9,288} < 1$, $p = 0.979$; Student Newman-Keuls post hoc pairwise comparison). (**g**) Probability to enter the port within the lever press sequence (within-sequence port entry) (two-way ANOVA, task effect: $F_{1,32} = 155.75$, $p < 0.001$; no task x region interaction: $F_{1,32} = 1.98$, $p = 0.169$). (**h**) 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: $F_{1,32} = 4.41$, $p = 0.044$). (**i**) 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: F1,32 < 1, p = 0.361). (**j**) Phasic VTA dopamine neuron activity 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, $t_{16} = -3.30$, p = 0,004). (**l**) Phasic VTA neuron activity around reward retrieval. (**m**) Area under the curves for 0.5 to 1 seconds after reward retrieval (unpaired two-tailed t-test comparison, t16= 2.15, p = 0.047). (**n**) Phasic NAc Core DA signal around port entries following reward omission. (**o**) Area under the curves for 1 to 3 seconds after reward omission (unpaired two-tailed t-test comparison, t16= -3.07, p = 0,007). (**p**) Phasic NAc Core DA signal around reward retrieval. (**q**) 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 = 16 (6 males; 10 females); LR5 n = 20 (10 males; 10 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.

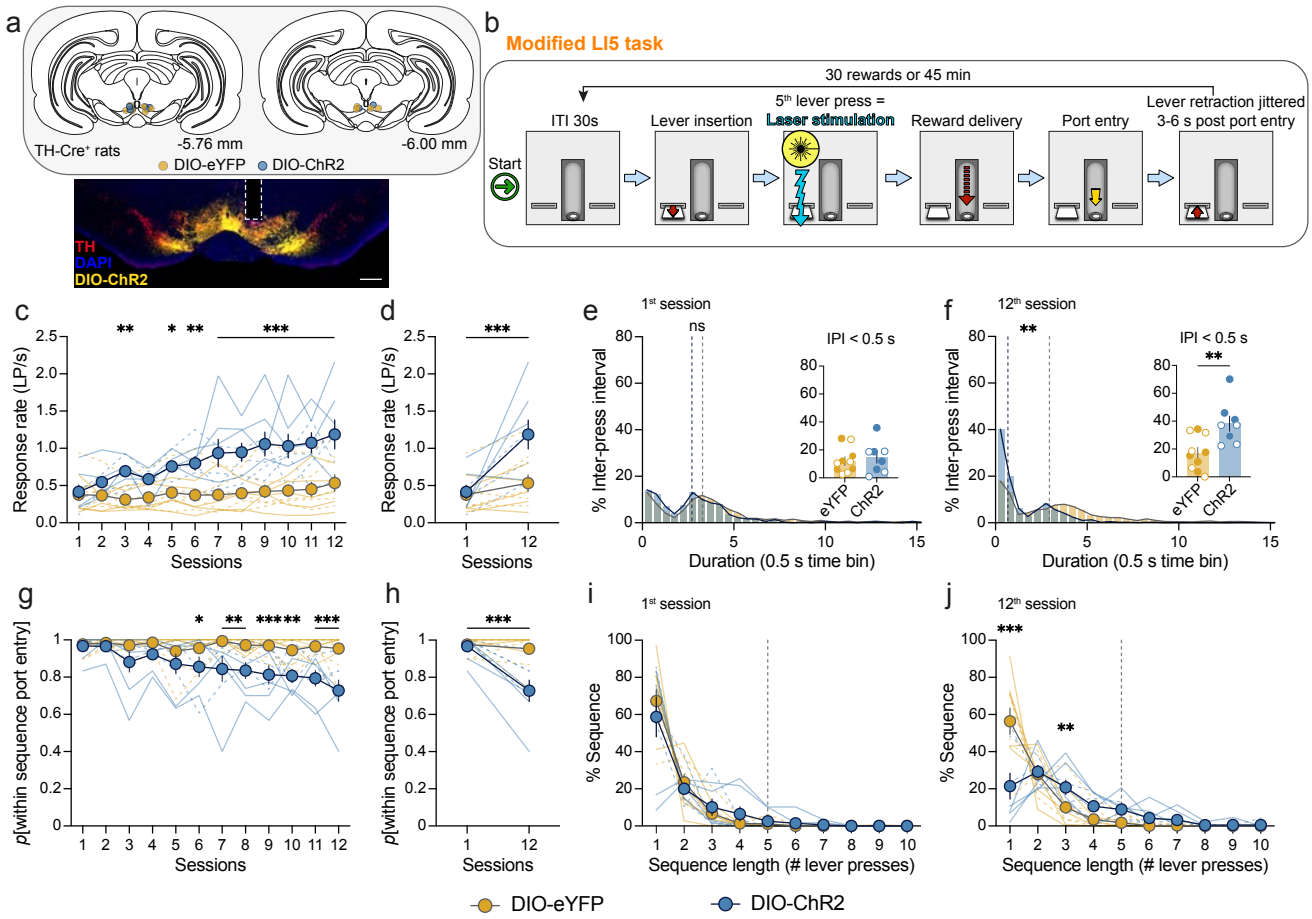

Figure 6. Optogenetic VTA DA neuron stimulation at the time of sequence completion drives automaticity and behavioral chunking. (**a**) Targeting VTA DA neurons with DIO-eYFP or DIO-ChR2 virus in TH-Cre+ rats. Scale bar: 500µm. (**b**) Task design for optogenetic stimulation of VTA DA neurons at sequence 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.