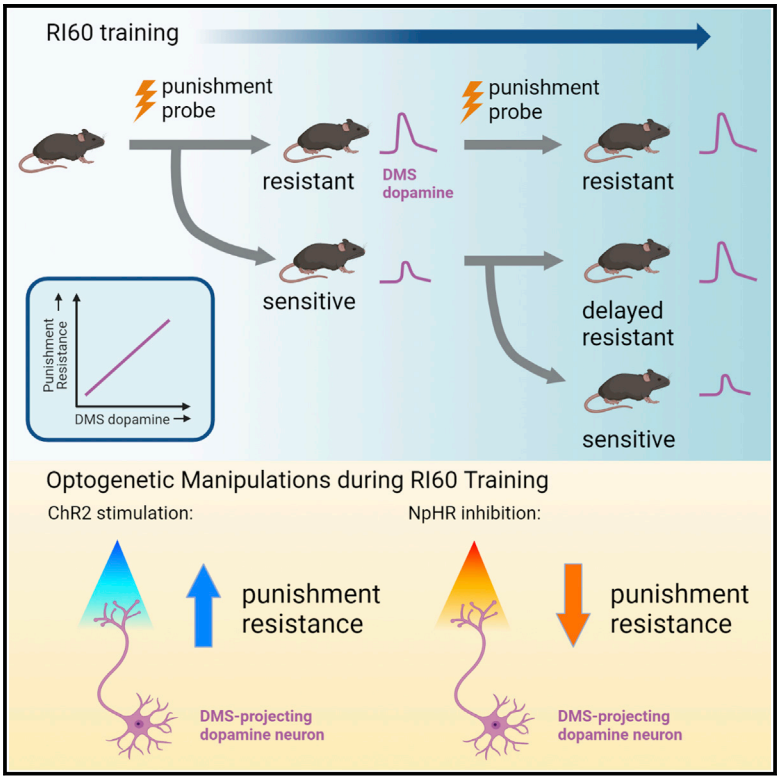
Article

Current Biology

Dopamine signaling in the dorsomedial striatum promotes compulsive behavior

Graphical abstract



Highlights

- Random interval training causes punishment-resistant reward seeking in some mice
- DMS dopamine signals predict which mice will develop punishment resistance
- Enhancing DMS dopamine signals accelerates the development of punishment resistance
- Inhibiting DMS dopamine signals slows the development of punishment resistance

Authors

Jillian L. Seiler, Caitlin V. Cosme, Venus N. Sherathiya, Michael D. Schaid, Joseph M. Bianco, Abigael S. Bridgemohan, Talia N. Lerner

Correspondence

talia.lerner@northwestern.edu

In brief

Seiler et al. investigate how dorsal striatal dopamine circuits control the development of punishment-resistant reward seeking, a mouse model of compulsive behavior. They identify DMS dopamine signaling as a key part of the circuitry that drives the emergence of punishment resistance in male and female mice.

Seiler et al., 2022, Current Biology *32*, 1175–1188
March 14, 2022 © 2022 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2022.01.055

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Dopamine signaling in the dorsomedial striatum promotes compulsive behavior

Jillian L. Seiler,[1,2] Caitlin V. Cosme,[1,4] Venus N. Sherathiya,[1] Michael D. Schaid,[1] Joseph M. Bianco,[1] Abigael S. Bridgemohan,[3] and Talia N. Lerner[1,5,6,*]

[1]Department of Neuroscience, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA
[2]Department of Psychology, University of Illinois at Chicago, Chicago, IL 60607, USA
[3]Department of Biology, Northwestern University Weinberg College of Arts & Sciences, Evanston, IL 60208, USA
[4]Present address: Alkermes Inc., Waltham, MA, USA
[5]Twitter: @LernerLab
[6]Lead contact
*Correspondence: talia.lerner@northwestern.edu
https://doi.org/10.1016/j.cub.2022.01.055

SUMMARY

Compulsive behavior is a defining feature of disorders such as substance use disorders. Current evidence suggests that corticostriatal circuits control the expression of established compulsions, but little is known about the mechanisms regulating the development of compulsions. We hypothesized that dopamine, a critical modulator of striatal synaptic plasticity, could control alterations in corticostriatal circuits leading to the development of compulsions (defined here as continued reward seeking in the face of punishment). We used dual-site fiber photometry to measure dopamine axon activity in the dorsomedial striatum (DMS) and the dorsolateral striatum (DLS) as compulsions emerged. Individual variability in the speed with which compulsions emerged was predicted by DMS dopamine axon activity. Amplifying this dopamine signal accelerated animals' transitions to compulsion, whereas inhibition delayed it. In contrast, amplifying DLS dopamine signaling had no effect on the emergence of compulsions. These results establish DMS dopamine signaling as a key controller of the development of compulsive reward seeking.

INTRODUCTION

Animals learn about the consequences of their actions through reinforcement. Positive or negative outcomes lead to the formation of action-outcome associations, which allow an animal to predict consequences and act purposefully. Action-outcome learning relies on the dorsomedial striatum (DMS) and supports goal-directed behavior.[1,2] The chief benefit of goal-directed behavior is that it permits flexibility when outcomes change. However, excessive flexibility might lead animals to prematurely abandon strategies that would be productive in the long run, as when action-outcome associations fluctuate or are probabilistic. How should we choose when we should continue historically good reward-seeking strategies and when we should abandon old strategies that are no longer beneficial? Miscalculations in the answer to this question are a defining feature of disorders such as substance use disorders (SUDs).

We set out to examine this question using a mouse model of punishment-resistant reward seeking (often termed compulsion in the animal literature). Punishment resistance was measured as the tendency of mice to continue reward seeking when faced with a possible aversive shock outcome. Punishment-resistant reward seeking appears to depend on dorsal striatal brain regions and their cortical inputs,[3–5] but little is known about how it emerges.[5]

One hypothesis is that punishment-resistant reward seeking results from habit formation.[3,4,6,7] Habits, which require the dorsolateral striatum (DLS),[8] decouple actions from outcomes and promote the use of stimulus-response associations to drive behavior.[9] Under this hypothesis, the development of punishment-resistant reward seeking would require DLS, and habit formation would precede punishment resistance. Indeed, the dependence of reward-seeking behavior on DLS dopamine precedes the development of punishment-resistant reward seeking for addictive drugs.[7,10] However, habit formation is not *required* for punishment-resistant reward seeking to emerge: when rats were trained to perform new action sequences each day to get cocaine, punishment-resistant drug seeking developed that was independent of habit and DLS dopamine.[11]

A second hypothesis is that punishment-resistant reward seeking arises due to strengthened action-outcome associations in DMS. Enhanced activity and plasticity in orbitofrontal cortex to DMS projections have been observed in animals that compulsively self-stimulate their ventral tegmental area (VTA) dopamine neurons and in animals that compulsively self-administer methamphetamine.[12–15]

To distinguish between these hypotheses, we trained mice to work for sucrose on a random interval schedule (RI60) known to promote habit and tested whether they became punishment resistant. We used fiber photometry to record dopamine axon activity in DMS and DLS during this behavior. Dopamine regulates reward learning and is a critical neuromodulator in both DMS and DLS.[16–20] Dopamine receptor blockade in DMS can inhibit action-outcome learning,[21] while dopamine signaling in

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

DLS is required for habit formation.[22] We found that DMS, but not DLS, dopamine axon signals predicted which individual mice would become punishment resistant. Optogenetic manipulation of DMS dopamine confirmed its causal and temporally specific role in the development of punishment-resistant reward seeking.

RESULTS

A random interval schedule of reinforcement promotes punishment-resistant reward seeking

We first determined whether a random interval schedule of reinforcement (RI60), previously shown to promote habit,[1,23–25] elicited punishment-resistant reward seeking. We compared RI60 training to training on a random ratio schedule (RR20). Although other training paradigms have been used to elicit habits,[26] RI60 and RR20 schedules are an established comparison due to their tendency to elicit similar rates of action.[24] After initial fixed ratio (FR1) training, mice advanced to RI30 or RR10, then to RI60 or RR20 (Figure 1A). To assess punishment-resistant reward seeking, we performed shock probes after 1 or 2 days and 13 or 14 days of RI60/RR20 training. During the shock probes, nosepokes were accompanied by a $^1/_3$ risk of shock (0.2 mA, 1 s; Figure 1B). The shock intensity was chosen based on previous studies of punishment-resistant reward seeking[12] and was aversive to male and female mice in a fear-conditioning paradigm with 12 tone-shock pairings (two-way ANOVA, $F_{1,50} = 16.46$, $p < 0.001$; Sidak's multiple comparison male shock paired versus no shock, $p < 0.01$; female shock paired versus no shock, $p < 0.05$; Figure S1A). To test whether punishment-resistant reward seeking developed in tandem with another test of habit-like behavioral inflexibility, a subset of mice were given an omission probe at the end of training (Figure 1C).[23,27,28] During the omission probe, mice were required to withhold nosepokes to receive rewards, reversing the previously learned contingency.

We observed a significant main effect of schedule (RI60 versus RR20) on the number of shocks mice were willing to receive on the shock probes (two-way ANOVA, $F_{1,43} = 6.37$, $p < 0.05$) and an interaction of schedule and training time ($F_{1,43} = 4.54$, $p < 0.05$). Furthermore, after extended RI60 training, mice increased the number of shocks they were willing to receive (Bonferroni, $p < 0.01$; Figure 1D). During the second shock probe, both RI60- and RR20-trained mice initially continued to nosepoke at the same rates relative to their training baseline, but RR20-trained mice more rapidly reduced their responding (two-way ANOVA, main effect of training time, $F_{4.426,181.5} = 4.99$, $p < 0.001$; Bonferroni, $p < 0.05$ bins after 50 min; Figure S1B). RR20-trained mice were also more willing than RI60-trained mice to explore alternative actions during the second shock probe session, indicated by a higher fraction of responses at the inactive nosepoke (unpaired t test, $p < 0.05$; Figures S1D–S1F).

RI60-trained mice took longer to complete the omission probe than RR20-trained mice (unpaired t test, $p < 0.05$; Figure 1E). RR20-trained mice almost immediately stopped nosepoking, whereas RI60-trained mice continued well into the session (two-way ANOVA, interaction of schedule and time, $F_{11,275} = 2.22$, $p < 0.05$; Figure S1C). RR20-trained mice, therefore, maintain a higher level of flexibility than RI60-trained mice when presented with either punishments or a reversed contingency.

What differences between RI60 and RR20 might elicit differences in punishment-resistant or omission-resistant reward seeking? As previously reported,[24,25] RI60 and RR20 training schedules provoked approximately equivalent rates of nosepoking (Figure S1G). However, RI60-trained mice made fewer nosepokes per reward (mixed-effects analysis, $F_{1,42} = 21.70$, $p < 0.0001$; Figure S1H) and earned significantly more rewards per training session than RR20-trained mice (unpaired t test, $p < 0.0001$; Figure S1I). Therefore, RI60 and RR20 have different effort demands and incur different reward histories, which could influence learning trajectories.

RI60-trained mice showed significant individual variability in punishment resistance, which was not due to variation in body weight (Figure S1J). We wondered whether the same individuals who withstood a high number of shocks also took longer to learn the omission contingency, reflecting generalized inflexibility. We found that these 2 measures were not significantly correlated ($r = 0.42$; not significant, ns; Figure 1F). Thus, although RI60 training promotes inflexibility in the form of both punishment-resistant and omission-resistant reward seeking, these two phenomena do not always occur in the same individuals and their development may rely on different brain circuits.

Three punishment-related phenotypes emerge with extended RI60 training

We wondered whether individual mice were taking different strategies to "solve" the RI60 task. We divided the RI60-trained mice into 3 groups based on a post hoc evaluation of shock probe performance: "punishment resistant" (PR) mice tolerated many shocks on both probes, "delayed punishment resistant" (DPR) mice increased the number of shocks tolerated from the first to the second probe, and "punishment sensitive" (PS) mice tolerated few shocks on both probes (Figure 2A STAR Methods; Methods S1A). There was a significant interaction of phenotype and training time ($F_{3,\ 40} = 24.18$, $p < 0.0001$); only DPR mice showed a significant increase in shocks received across the probes ($p < 0.0001$;Figure 2B).

We also analyzed our data by sex and found that PR mice were more likely to be male while PS mice were more likely to be female (Figure S2A). Male mice tolerated more shocks than females (two-way ANOVA; main effect of training time $F_{1,35} = 417.7$, $p < 0.001$; main effect of sex $F_{1,35} = 26.19$, $p < 0.0001$; interaction $F_{1,35} = 4.94$, $p < 0.05$; Bonferroni shock 1, $p < 0.01$, and shock 2, $p < 0.0001$; Figure S2B) and had higher nosepoke rates during RI60 (unpaired t test of male versus female across days of training, $p < 0.0001$; Figure S2C). Variance in punishment-resistant reward seeking was not explained by differences in body weight (Figure S1J), and sex could not fully account for large individual variance. Nevertheless, given these sex differences, we were careful to include a balance of male and female mice going forward in all our experiments.

An analysis of RI60 behavior in PR, DPR, and PS mice showed interesting differences (Figures 2 and S2). PR and DPR mice had higher rates of nosepoking than PS mice (mixed-effects analysis; main effect of training time $F_{5.236,156.3} = 9.79$, $p < 0.0001$; main effect of phenotype $F_{2,33} = 18.59$, $p < 0.0001$; interaction $F_{26,388} = 3.28$, $p < 0.0001$; Figure 2C). PR mice took significantly

CellPress
OPEN ACCESS

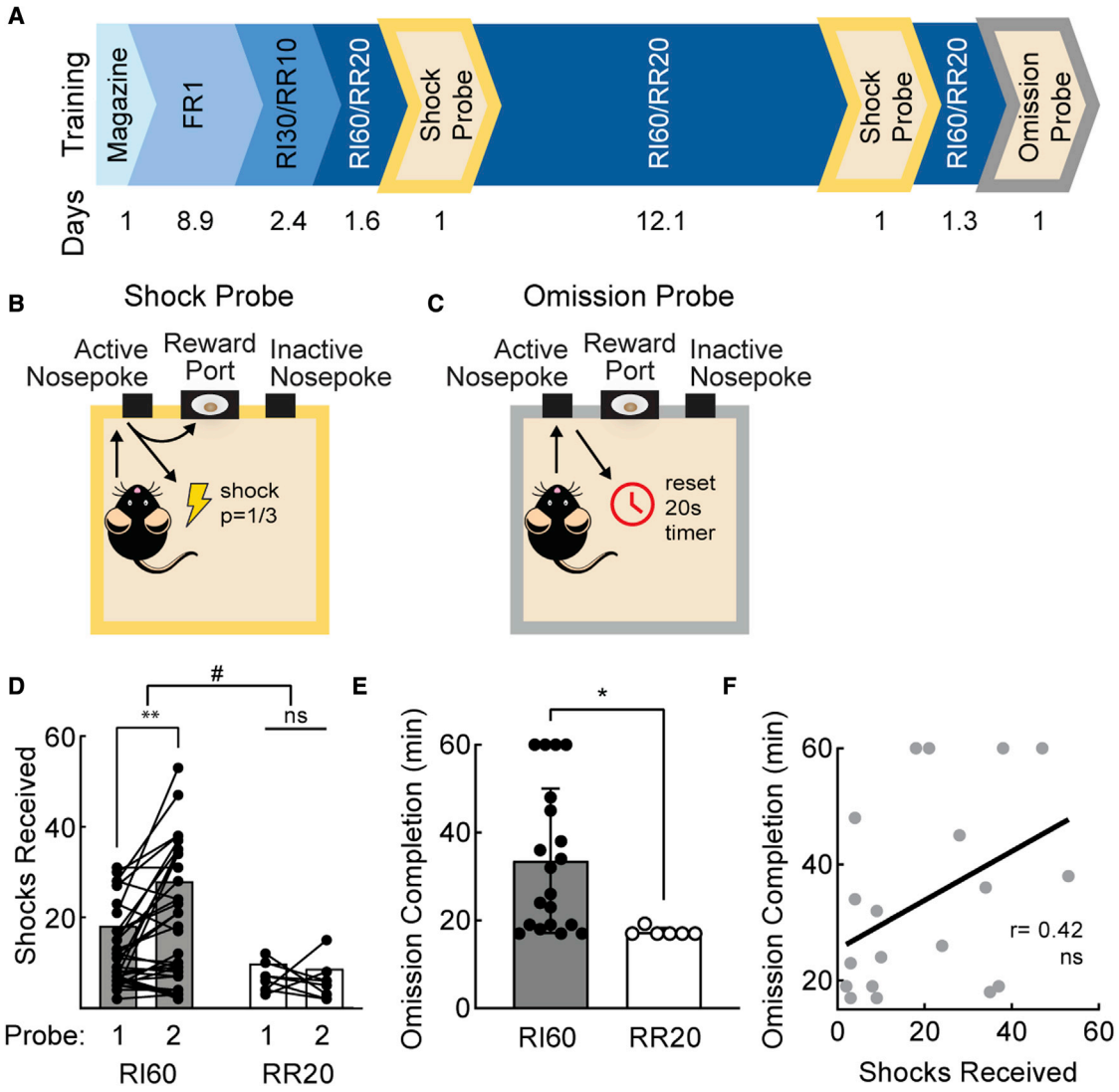


Figure 1. A random interval schedule of reinforcement promotes punishment-resistant reward seeking

(A) Timeline of operant training and probes. Average days per stage of training below.

(B) Schematic of shock probe: active nosepokes incurred a $^1/_3$ probability of shock.

(C) Schematic of omission probe: active nosepokes delayed reward by 20 s.

(D) Shocks received on early and late shock probes for RI60-trained (black; n = 36) and RR20-trained (white; n = 9) mice. Bars represent mean; points represent individuals. Main effect $^{\#}p < 0.05$, multiple comparisons, $^{**}p < 0.01$.

(E) Average time to complete omission probe for RI60-trained (black; n = 20) and RR20-trained (white; n = 7) mice. Error bars represent SD. $^{*}p < 0.05$

(F) Correlation between shocks received on late shock probe and omission completion time for RI60-trained mice tested in both probes (r = 0.42, ns).

See also Figure S1.

longer than PS mice to complete the omission probe (one-way ANOVA, $F_{2,18} = 4.36$, $p < 0.05$; Tukey's multiple comparison, PR = 51.50 ± 17 s versus PS = 27.92 ± 13.65 s, $p < 0.05$; Figure 2D). Meanwhile, PS mice were more likely to explore the inactive nosepoke during the late shock probe (mixed-effects analysis; main effect of phenotype, $F_{2,33} = 3.79$, $p < 0.05$; Tukey's multiple comparisons test, RI60 days 13 and 14 for PS versus DPR, $p < 0.05$; one-way ANOVA for shock 2, $F_{2,18} = 5.36$, $p < 0.05$; Tukey's multiple comparison PS versus DPR, $p < 0.05$; second shock for PS versus PR, $p < 0.05$; Figures 2E–2G). Port entry and reward rates did not differ among PR, DPR, and PS mice (Figures 2H and 2I). As a result, PS mice are more "efficient," making fewer nosepokes per reward (mixed-effects analysis; main effect of training time $F_{4.65,87.86} = 2.42$, $p < 0.05$; main effect of phenotype $F_{2,21} = 22.77$, $p < 0.0001$; interaction $F_{24,227} = 1.76$, $p < 0.05$; Figure 2J).

In RI60, a mouse is unlikely to receive a reward if it has recently received one. To be efficient, mice should wait to resume nosepoking after a reward. PS mice waited an average of 38 ± 15 s, significantly longer than the other groups (one-way ANOVA, $F_{2,33} = 8.57$, $p < 0.01$; Tukey's multiple comparison, PR = 21 ± 5, $p < 0.01$; DPR = 22 ± 4, $p < 0.01$; Figure 2K). We also looked

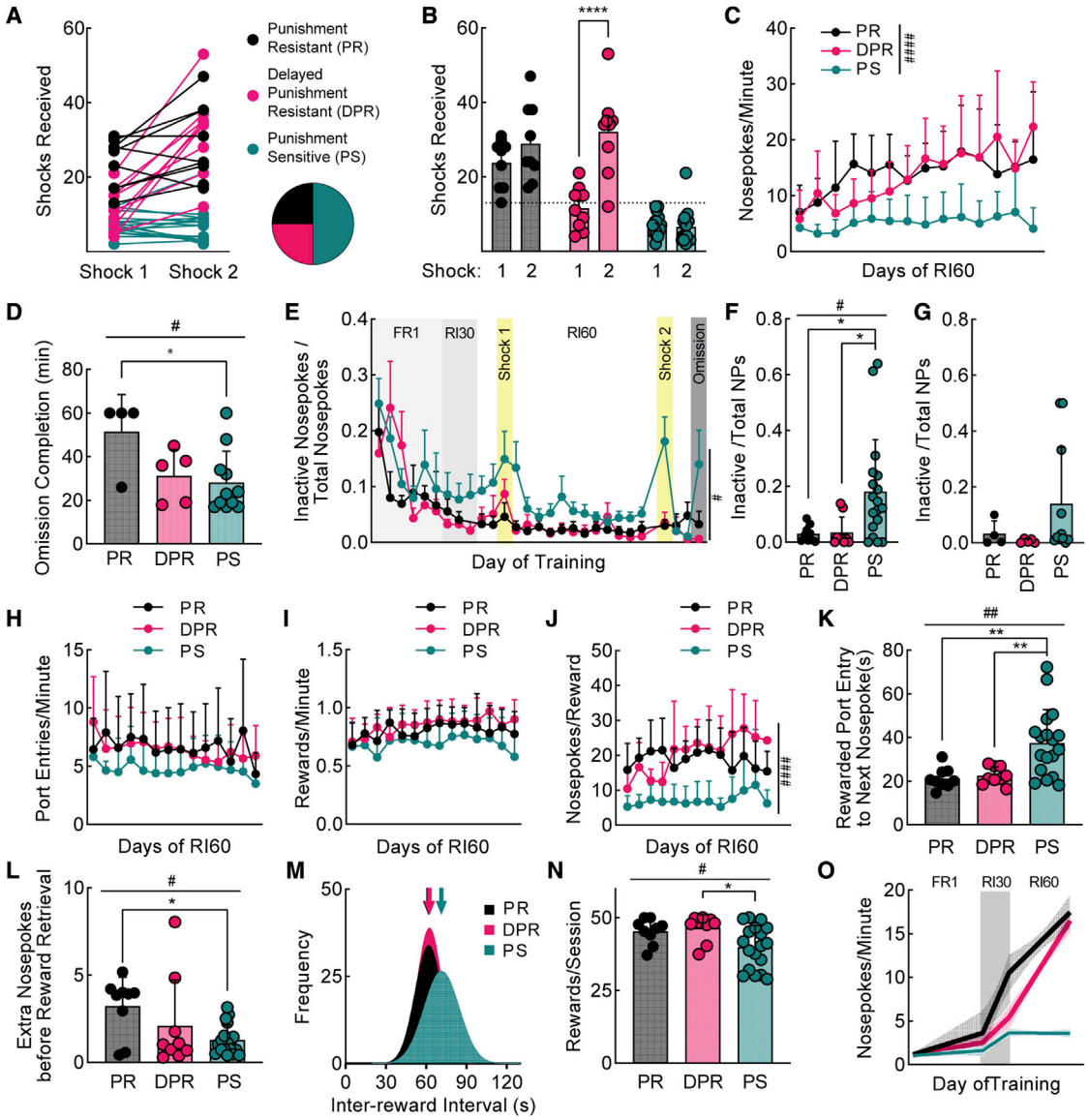


Figure 2. Three punishment-related phenotypes emerge with extended RI60 training

(A) Shocks received by RI60-trained mice (same data as Figure 1D). Individuals represented by points connected with lines. Mice were classified as punishment resistant (PR; black), delayed punishment resistant (DPR; pink), or punishment sensitive (PS; teal) based on number of shocks received during shock probes (see STAR Methods; PR n = 9, DPR n = 9, and PS n = 18 for all panels, unless specified). Pie chart shows proportions of PR, DPR, and PS mice.

(B) Average shocks received on early and late shock probes for each phenotype.

(C) Average nosepokes per minute across RI60 training by phenotype.

(D) Average time to complete omission probe (PR n = 4, DPR n = 5, and PS n = 11).

(E) Average nosepokes on inactive port as a fraction of total nosepokes across training.

(F) Average nosepokes on inactive port as a fraction of total nosepokes during second shock probe.

(G) Average nosepokes on inactive port as a fraction of total nosepokes during omission probe (PR n = 4, DPR n = 5, and PS n = 11).

(H) Average number of port entries per minute across RI60 training.

(I) Average number of rewards earned per minute across RI60 training.

(J) Average nosepokes per reward across RI60 training.

(K) Average time from rewarded port entry to next nosepoke.

(L) Average unrewarded "extra" nosepokes made following rewarded nosepoke, prior to rewarded port entry.

(M) Distribution of inter-reward interval times for each group. Arrows represent mean.

(N) Average rewards earned per RI60 session.

(O) Segmental linear regression showing slope of nosepokes made per minute in FR1, RI30, and RI60 schedules. Shaded region represents 95% confidence bands. All error bars represent SD. Main effects $^{\#}p < 0.05$, $^{\#\#}p < 0.01$, $^{\#\#\#}p < 0.001$, $^{\#\#\#\#}p < 0.0001$. Multiple comparisons $^{*}p < 0.05$, $^{**}p < 0.01$, $^{****}p < 0.0001$. See also Figure S2.

at whether mice made "extra" nosepokes after a rewarded nosepoke before going to collect their reward. PS mice made significantly fewer extra nosepokes than PR mice (one-way ANOVA, $F_{2,33} = 4.24$, $p < 0.05$; Tukey's multiple comparison, PS = 1.28 ± 0.85 versus PR = 3.23 ± 1.65, $p < 0.05$; DPR = 2.1 ± 2.63; Figure 2L), which also maximized their efficiency.

Why would PR and DPR mice expend more effort than necessary? On an RI60 schedule, rewards become available in a normal distribution around 60 s. Although unlikely, some intervals are much shorter than 60 s. To detect these shorter intervals and collect rewards as quickly as possible, mice must nosepoke constantly. PS mice pay for their efficiency with longer inter-reward intervals (Figure 2M). The average inter-reward interval for PS mice is 79.77 ± 26.54 s compared to 71.2 ± 32.3 s for PR and 69.58 ± 35.18 s for DPR (K-S Test, PR versus PS, $p < 0.0001$; DPR versus PS, $p < 0.0001$). Although the reward rates were not different on any particular day of RI60 training (Figure 2I), when the number of rewards per session was averaged over all days, PS mice received significantly fewer rewards (one-way ANOVA, $F_{2,33} = 4.45$, $p < 0.05$; Tukey's multiple comparison, PS = 39.91 ± 7.39 versus DPR = 46.56 ± 4.49, $p < 0.05$; PR = 45.33 ± 4.33; Figure 2N). The advantage of a high-effort strategy is the maximization of reward.

PR and DPR mice take similar reward-seeking strategies in the RI60 task, so to examine differences between them, we looked at earlier training data. PR mice escalate their nosepoking as soon as they enter RI30 (Figure 2O). DPR mice escalate their nosepoking later in RI60, concurrent with the development of punishment resistance. Thus, the timing of nose-poke escalation is a key behavioral predictor of punishment-resistant reward seeking. Importantly, nose-poke escalation emerges in PR mice before any experience of punishment, suggesting these mice have a predisposition toward developing punishment resistance related to their initial reward-seeking strategy.

Dopamine axon signals in the DMS predict punishment-resistant reward seeking

To understand the neural circuits underlying the development of punishment-resistant reward seeking, we recorded the activity of dopamine axons in the dorsal striatum. Dopaminergic projections to DMS and DLS are distinct, meaning dopamine-mediated reinforcement learning can be separately effectuated in these two areas.[29–31] To record the activity of dopamine axons in DMS and DLS in freely moving mice, we injected an adeno-associated virus (AAV) expressing cre-dependent GCaMP7b[32] (AAV5-CAG-FLEX-jGCaMP7b-WPRE) into substantia nigra pars compacta (SNc) in DAT-IRES-cre mice (Figures 3A and 4A). DAT-IRES-cre mice are reported to exhibit a 17% reduction of dopamine transporter (DAT)[33] and novelty-induced hyperactivity.[34] However, heterozygous mice in our colony do *not* have a significant reduction in DAT protein levels in DMS or DLS (Figures S3A and S3B) and are *not* hyperactive in a novel open field (Figures S3C and S3D). We implanted fiber optic probes above DMS and DLS and recorded from both areas simultaneously in all mice. Thus, we could be certain any observed differences between DMS and DLS signals were not due to differences in behavior between groups.

We began by examining DMS dopamine axon activity occurring during rewarded and unrewarded nosepokes. We verified GCaMP was expressed in dopaminergic (TH+) neurons in medial SNc (Figures 3B and 3C) and correct probe locations (Figures 3D and S3E). We compared DMS dopamine axon activity in RI60-trained (PR, DPR, PS; Figure 3E) and RR20-trained mice (RR20; Figure S3H). Peaks in DMS dopamine axon activity at the time of a rewarded nosepoke were clearer in PR and DPR mice than in PS or RR20 mice, an observation that was not true simply due to poor signal in PS or RR20 mice, as all mice had similar frequencies and amplitudes of GCaMP events across the full recordings (Figures S3F and S3G). A main effect of training time on the frequency of GCaMP events was observed but was the same across all groups (mixed-effects analysis; main effect of time $F_{1,27} = 7.9$, $p < 0.01$; Figure S3F). Peaks in response to rewarded nosepokes emerged during RI30 in PR mice, whereas peaks emerged more slowly across RI60 training in DPR mice (Figure 3E). Unrewarded nosepokes resulted in small positive deflections in all groups during RI30, however, negative deflections appeared during RI60 in PR and DPR mice. A positive deflection for rewarded nosepokes and a negative deflection for unrewarded nosepokes creates a difference in dopamine axon activity in response to the same motor action depending on the outcome. We calculated a rewarded-unrewarded peak score for each mouse by subtracting the minimum of the unrewarded-nosepoke peri-stimulus time histogram (PSTH) from the maximum of the rewarded-nosepoke PSTH. The rewarded-unrewarded peak score changed across training (mixed-effects analysis; $F_{1.61,29.8} = 13.83$, $p < 0.001$; Figure 3F) and was significantly different by phenotype ($F_{2,\ 33} = 8.16$, $p < 0.01$), with a significant interaction between the two ($F_{4,37} = 3.29$, $p < 0.05$).

DMS dopamine axon activity tracked with the development of punishment-resistant behavior in PR, DPR, and PS groups, but we also wanted to assess whether an individual's rewarded-unrewarded peak score (independent of group classification) could predict punishment resistance. Indeed, the shocks received were significantly correlated with DMS rewarded-unrewarded peak score on a mouse-by-mouse basis ($r = 0.53$, $p < 0.0001$; Figure 3G). Performance on the omission probe was not correlated with this score ($r = 0.1$, ns; Figure 3H).

We also examined DMS dopamine axon signals surrounding the time of rewarded and unrewarded port entries and noticed ramping activity preceding rewarded port entries (Figure S3I). Ramping toward reward has been described primarily in VTA-nucleus accumbens (NAc) dopamine circuits[35–39] but also occurs in DMS dopamine axons.[40] Here, we additionally note that ramping in DMS dopamine axons is more prominent in PR and DPR mice than in PS or RR20 mice (Figure S3I). To encapsulate ramping activity quantitatively, we measured the area under the curve (AUC) of our fiber photometry signal from −5 to 0 s relative to the rewarded port entry. DPR mice showed a significant increase from early to late RI60 training (mixed-effects analysis; interaction of time and phenotype, $F_{3,19} = 4.77$, $p < 0.05$; Bonferroni, $p < 0.05$; Figure S3J), indicating ramping in DMS dopamine axons could also be related to the development of punishment resistance.

Dopamine signals in the DLS do not predict punishment-resistant reward seeking

We next examined DLS dopamine axon activity (Figure 4A). We verified the expression of GCaMP across medial and lateral SNc, which both contain DLS-projecting dopamine neurons (89.56%

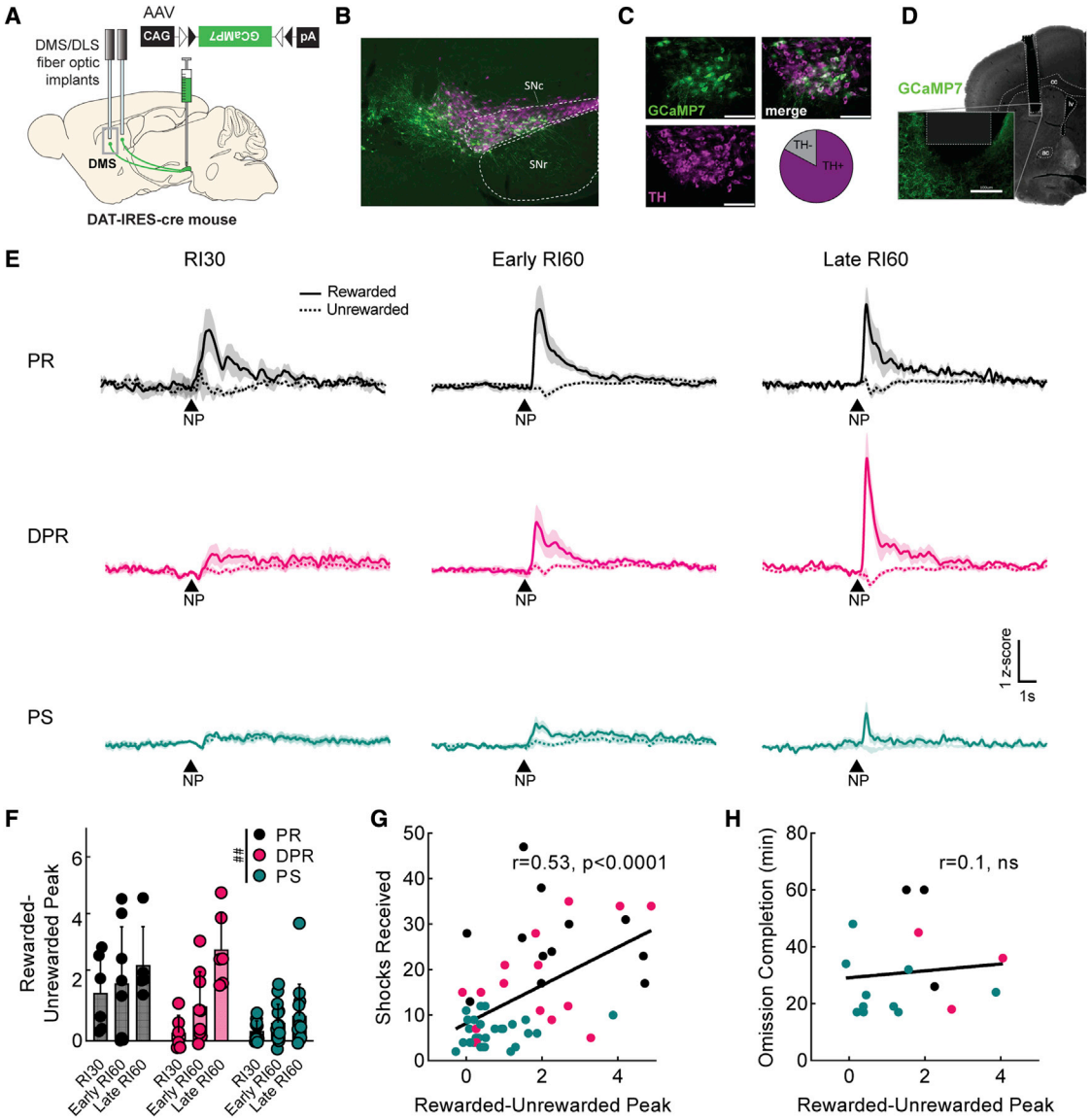


Figure 3. Dopamine axon signals in DMS predict punishment-resistant reward seeking

(A) Viral injection and probe placement strategy.

(B) Representative image (4×) showing viral spread of GCaMP7b (green, all images) and tyrosine hydroxylase (TH) positive cells (magenta, all images) in midbrain. Scale bar is 100 μm (all images). SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata.

(C) 40× images of SNc showing GCaMP7b, TH positive cells, and merged image. Quantification of GCaMP7b-expressing cells that are TH+ is shown; n = 572 cells.

(D) Representative image showing probe placement in DMS. Area of magnification shows GCaMP7b expression in dopaminergic axons near probe. cc, corpus callosum; lv, lateral ventricle; ac, anterior commissure.

(E) Peri-stimulus time histograms (PSTHs) showing average signal from DMS dopamine terminals at rewarded (solid) and unrewarded (dashed) nosepokes (NP) for each phenotype during RI30 training, early, and late RI60 training. Shaded region represents SEM. Punishment resistant (PR; black, RI30 n = 6, early n = 7, late n = 5 for all panels), delayed punishment resistant (DPR; pink, RI30 n = 6, early n = 9, late n = 6 for all panels), or punishment sensitive (PS; teal, RI30 n = 11, early n = 15, late n = 13 for all panels).

(F) Quantification of average rewarded-unrewarded peak for DMS dopamine terminal signals in response to nose pokes. Error bars represent SD, main effect $^{\#\#}p < 0.01$

(G) Correlation of shocks received in shock probes and rewarded-unrewarded peaks in DMS dopamine terminals (r = 0.53, p < 0.0001).

(H) Correlation of omission completion time and rewarded-unrewarded peaks in DMS dopamine terminals (r = 0.1, ns).

See also Figure S3.

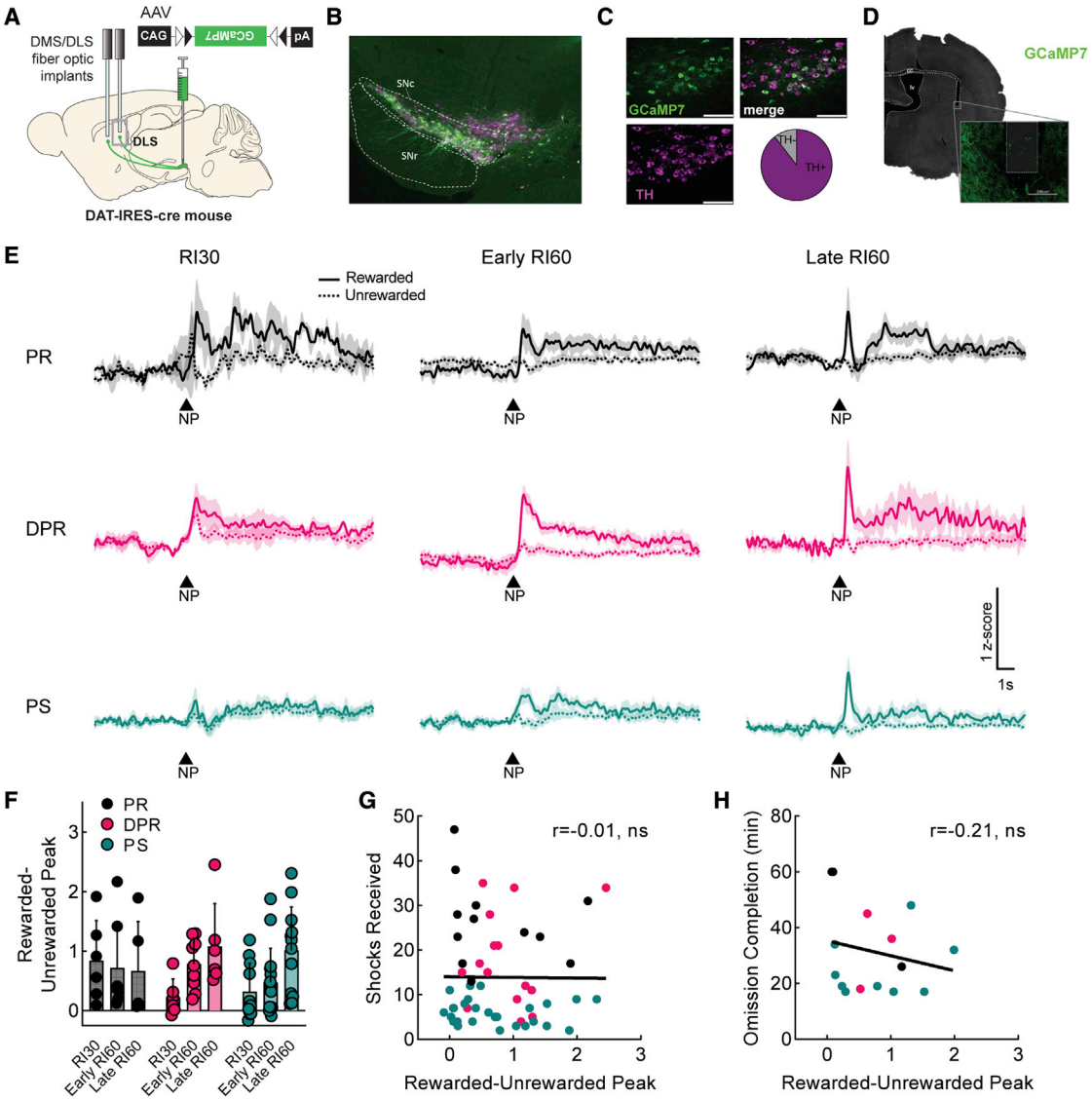


Figure 4. Dopamine signals in DLS do not predict punishment-resistant reward seeking

(A) Viral injection and probe placement strategy.

(B) Representative image (4×) showing viral spread of GCaMP7b (green, all images) and TH positive cells (magenta, all images) in midbrain. Scale bar is 100 μm (all images). SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata.

(C) 40× images of SNc showing GCaMP7b expression, TH positive cells, and merged image. Quantification of GCaMP7b-expressing cells that are TH+ is shown; n = 193 cells.

(D) Representative image showing probe placement in DLS. Area of magnification shows GCaMP7b expression in dopaminergic axons near probe. cc, corpus callosum; lv, lateral ventricle.

(E) Peri-stimulus time histograms (PSTHs) showing average signal from DLS dopamine terminals at rewarded (solid) and unrewarded (dashed) nosepokes (NP) for each phenotype during RI30 training, early, and late RI60 training. Shaded region represents SEM. Punishment resistant (PR; black, RI30 n = 6, early n = 7, late n = 5 for all panels), delayed punishment resistant (DPR; pink, RI30 n = 6, early n = 9, late n = 6 for all panels), or punishment sensitive (PS; teal, RI30 n = 11, early n = 15, late n = 13 for all panels).

(F) Quantification of average rewarded-unrewarded peak for DLS dopamine terminal signals in response to nosepokes. Error bars represent SD.

(G) Correlation of shocks received in shock probes and rewarded-unrewarded peak in DLS dopamine terminals (r = −0.01, ns).

(H) Correlation of omission completion time and rewarded-unrewarded peak in DLS dopamine terminals (r = −0.21, ns).

See also Figure S4.

of GCaMP neurons also expressed TH; Figures 4B and 4C), [29] and the probe placements in DLS (Figures 4D and S4A). DLS signals in all groups had similar frequencies and amplitudes of GCaMP events (Figures S4B and S4C). At the time of a rewarded nosepoke, DLS dopamine axon signals differed from DMS in that they had both an immediate component and a prolonged component (Figures 4E and S4D). To test whether these DLS dopamine axon signals bore any relationship to punishment

resistance, we calculated a rewarded-unrewarded peak score as above. There was a significant effect of training stage (mixed-effects analysis; $F_{2,35} = 3.53$, $p < 0.05$), indicating the signals do change across training, but no significant effect of group (Figure 4F). We also looked at whether these signals correlated with shock or omission probe performance on an individual basis and found no correlations ($r = -0.01$ and $r = -0.21$, respectively; Figures 4G and 4H).

We next assessed whether the prolonged elevation of DLS dopamine axon activity after a rewarded nosepoke was related to the development of punishment or omission resistance. We quantified the prolonged activity as the AUC 2–10 s after the nosepoke. AUC did not correlate with the performance of individuals on the shock or omission probes (Figures S4E and S4F) nor did it differ statistically across groups or training time (Figure S4G).

We were surprised we could not observe a correlation of DLS dopamine axon signals with behavior, even though they changed over time. To get another view of these signals, we grouped mice by their performance on the omission probe rather than the shock probe. However, we saw no differences in DMS or DLS dopamine axon signals depending on omission probe performance (Figure S4H).

Finally, we looked at DLS dopamine axon signals aligned to port entries (Figure S4I). We observed weak ramping before a rewarded port entry, with the peak of the signal occurring after port entry. Quantification of the AUC −5–0 s relative to the rewarded port entry showed no differences by behavioral phenotype (Figure S4J). The lack of ramping in DLS dopamine axons is similar to observations in another recent study.[40]

Optogenetic excitation of DMS dopamine terminals at the time of a rewarded nosepoke accelerates the development of punishment-resistant reward seeking

Since peaks in DMS dopamine axon activity in response to rewarded nosepokes predicted the development of punishment-resistant reward seeking, we tested if stimulation of DMS dopamine axons using the excitatory opsin ChR2 caused punishment-resistant behavior to emerge. An AAV expressing cre-dependent ChR2 (AAV5-EF1α-DIO-hChR2(H134R)-EYFP) was injected into the SNc of DAT-IRES-cre mice to express ChR2 specifically in dopamine neurons. A fiber optic probe was placed above DMS (Figure 5A). We verified this strategy led to the expression of ChR2 in dopamine (TH+) neurons (Figures 5B and 5C), and probes were correctly placed in DMS, matching the coordinates used for fiber photometry (Figure S5A).

Beginning with FR1, ChR2 mice received 1-s, 20-Hz trains of light stimulation on every rewarded nosepoke (Figures 5D and 5E). Importantly, stimulation was delivered during FR1/RI30/RI60 training, but not during shock probes where punishment-resistant reward seeking was assessed. Therefore, the effects of stimulation on probe performance are not due to acute effects of stimulation but are caused by differences in learning during the training sessions. In addition to ChR2 mice, there were two control groups: EYFP controls received a fluorophore-only virus (AAV5-EF1α-DIO-EYFP) and the same pattern of light stimulation, and "scrambled" controls received the ChR2 virus but received stimulation on random nosepokes. The scrambled control was important as these mice received at least as many dopamine terminal stimulations as the ChR2 group, reinforcing the same action (a nosepoke). Therefore, the main difference between the ChR2 and ChR2 scrambled groups was whether the dopamine terminal stimulation they received boosted or degraded the ability of the natural DMS dopamine signal to differentiate between externally rewarded and unrewarded actions.

All 3 groups (ChR2, EYFP, and ChR2 scrambled) learned FR1 and were advanced to RI30 and RI60 (Figure S5B). However, ChR2 scrambled mice took significantly longer to reach FR1 criterion, indicating scrambled stimulation caused an initial learning impairment (one-way ANOVA, main effect of manipulation $F_{2,36} = 3.86$, $p < 0.05$; Tukey's multiple comparison, EYFP versus scrambled, $p < 0.05$; Figure S5C). In RI60, there was a small difference in the rewards per minute on a day-by-day basis (mixed-effects analysis; main effect of manipulation $F_{2,34} = 3.85$, $p < 0.05$; Figure S5D); however, all mice received approximately the same number of rewards per session pooled across days (Figure S5E). ChR2 mice escalated their nosepoking much faster than the control groups during FR1 and RI30 then leveled off during RI60 (Figure 5F). On the first shock probe, ChR2 mice were significantly more resistant to punishment than EYFP mice (two-way ANOVA; interaction of time and manipulation $F_{2,35} = 3.65$, $p < 0.05$; Tukey's multiple comparison, $p < 0.05$; Figure 5G). This difference faded by the second probe as punishment resistance emerged naturally in the controls. We categorized mice from this experiment as PR, DPR, and PS using our previously defined post hoc criteria (Methods S1B). ChR2 mice were extremely likely to be categorized as PR, whereas EYFP and ChR2 scrambled mice were distributed as expected across groups (Figure 5H). Notably, under ChR2 stimulation, 100% of male mice were PR mice (Figure S5F). A large majority of female mice (~71%) were also PR—despite the fact they were unlikely to be PR otherwise—indicating DMS dopamine terminal stimulation drives both sexes to develop punishment-resistant reward seeking (Figure S5F; cf. Figures 2A and S2A). We also looked to see whether DMS dopamine terminal stimulation influenced performance on the omission probe. It did not (Figure 5I).

Optogenetic inhibition of dopamine terminals in DMS delays the development of punishment-resistant reward seeking

Promoting DMS dopamine in response to rewarded nosepokes accelerated the development of punishment-resistant reward seeking. Would inhibiting DMS dopamine delay its development? We performed bilateral inhibition of DMS dopamine axons using the inhibitory opsin eNpHR3.0.[41] An AAV expressing cre-dependent NpHR (AAV5-EF1α-DIO-eNpHR3.0-EYFP) or a fluorophore-only control virus (AAV5-EF1α-DIO-EYFP) was injected into the SNc of DAT-IRES-cre mice to express NpHR (or EYFP) specifically in dopamine neurons (Figures 6A–6C). Fiber optic probes were placed above DMS (Figures 6A and S6A).

Mice were divided into 3 groups. NpHR mice received a 1-s continuous pulse of light on every rewarded nosepoke (Figures 6D and 6E). EYFP mice received the same light stimulation but lacked NpHR. NpHR scrambled mice received a 1-s continuous pulse of light on random nosepokes. In an initial experiment, light delivery began during FR1 training (Figure 6D [group 1]), paralleling the design of the stimulation experiment (Figure 5). However, DMS dopamine terminal inhibition resulted in a learning deficit. 25% of NpHR mice and 22% of NpHR

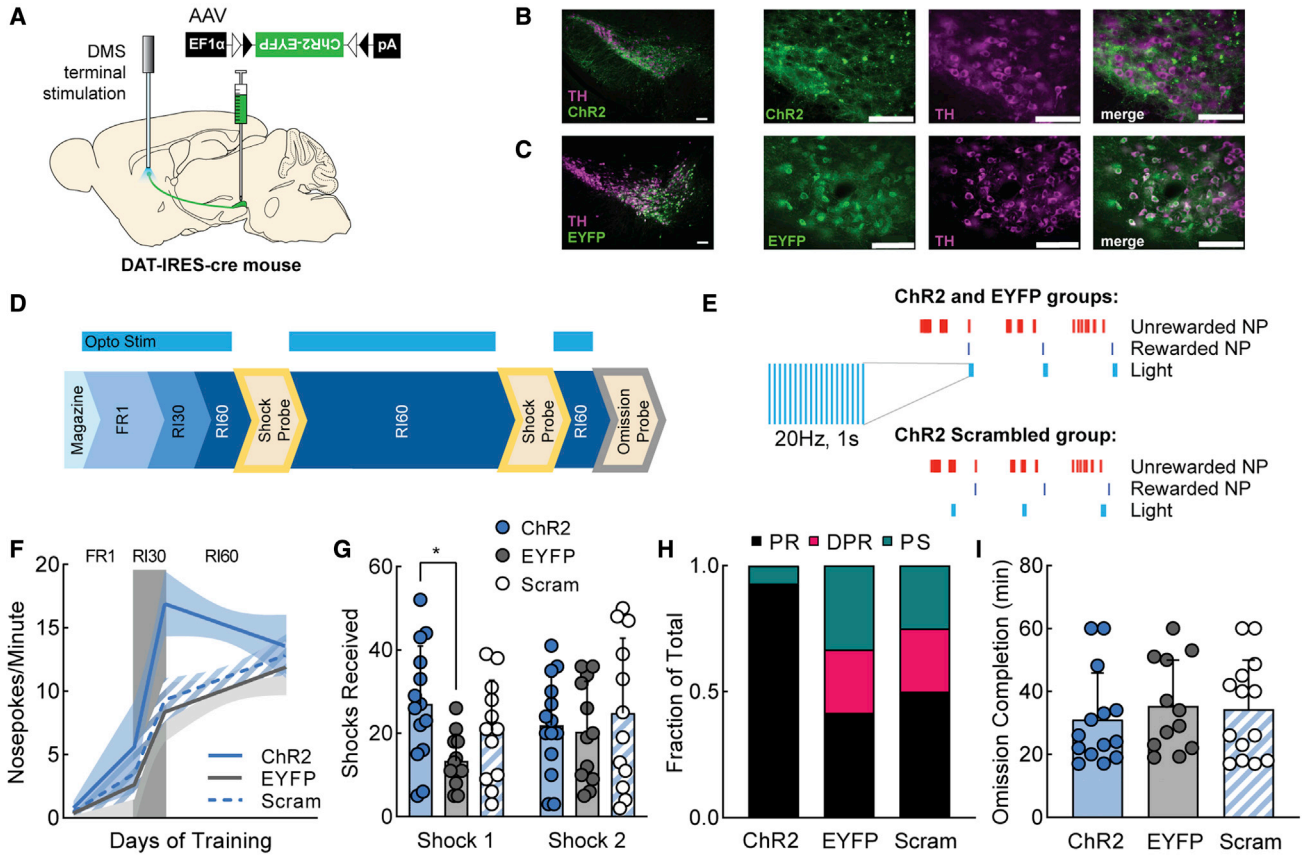


Figure 5. Optogenetic excitation of dopamine terminals in DMS at the time of a rewarded nosepoke accelerates development of punishment-resistant reward seeking

(A) Viral injection and probe placement strategy.

(B and C) 10× and 40× images of SNc showing ChR2-EYFP (top) or EYFP (bottom) expression in green, TH positive cells in magenta, and merged image. Scale bars are 100 μm (all images).

(D) Training timeline showing when optogenetic stimulation was delivered (FR1, RI30, or RI60).

(E) Schematic of stimulation parameters. A 1-s, 20-Hz burst of stimulation was paired with rewarded nosepokes for ChR2 and EYFP groups, and the same stimulation was paired with a random subset of nosepokes for ChR2 scrambled animals.

(F) Segmental linear regression showing slope of nosepokes made per minute in FR1, RI30, and RI60 schedules. Shaded region represents 95% confidence bands.

(G) Average shocks received on early and late shock probes for each manipulation.

(H) Fraction of each behavioral phenotype (PR, black; DPR, pink; PS, teal) per manipulation.

(I) Average omission completion time per manipulation. All error bars represent SD. *$p < 0.05$.

See also Figure S5.

scrambled mice were dropped from the study after >14 days (mean + 2 SD) of unsuccessful FR1 training. Mice that completed FR1 were able to reach RI30 criterion and continue (Figure 6F). In a second experiment, light delivery began during RI30 (Figure 6D [group 2]). In this case, 29% of NpHR mice were dropped from the study because they could not learn RI30. NpHR scrambled mice in this second experiment were all able to learn RI30 (during RI30, the NpHR scrambled condition is much more distinct from the NpHR condition since there are more unrewarded nosepokes on which inhibition can occur randomly). In both inhibition experiments, NpHR and NpHR scrambled mice had reduced nosepoke escalation compared to EYFP controls (Figures S6B and S6C).

To assess effects on punishment-resistant reward seeking in NpHR mice, we pooled mice from both experiments that passed FR1 and RI30 criteria and were able to continue to RI60. NpHR, EYFP, and NpHR scrambled mice did not differ in the number of rewards earned during RI60 training (Figures S6D and S6E). NpHR inhibition of DMS dopamine terminals during RI60 training delayed the development of punishment-resistant reward seeking with a significant effect on the second shock probe (two-way ANOVA, main effect of training $F_{2,36} = 4.72$, $p < 0.05$; Tukey's multiple comparisons, shock 2 NpHR versus EYFP, $p < 0.01$; Figure 6G). We sorted these mice into PR, DPR, and PS groups and noted that the NpHR group had an increased incidence of PS mice, while the NpHR scrambled group had an increased incidence of PR mice (Figure 6H). The effects of NpHR inhibition on PR/DPR/PS phenotype are driven by stark effects in male mice (Figure S6F). No significant differences in omission time were observed (Figure 6I). These data suggest

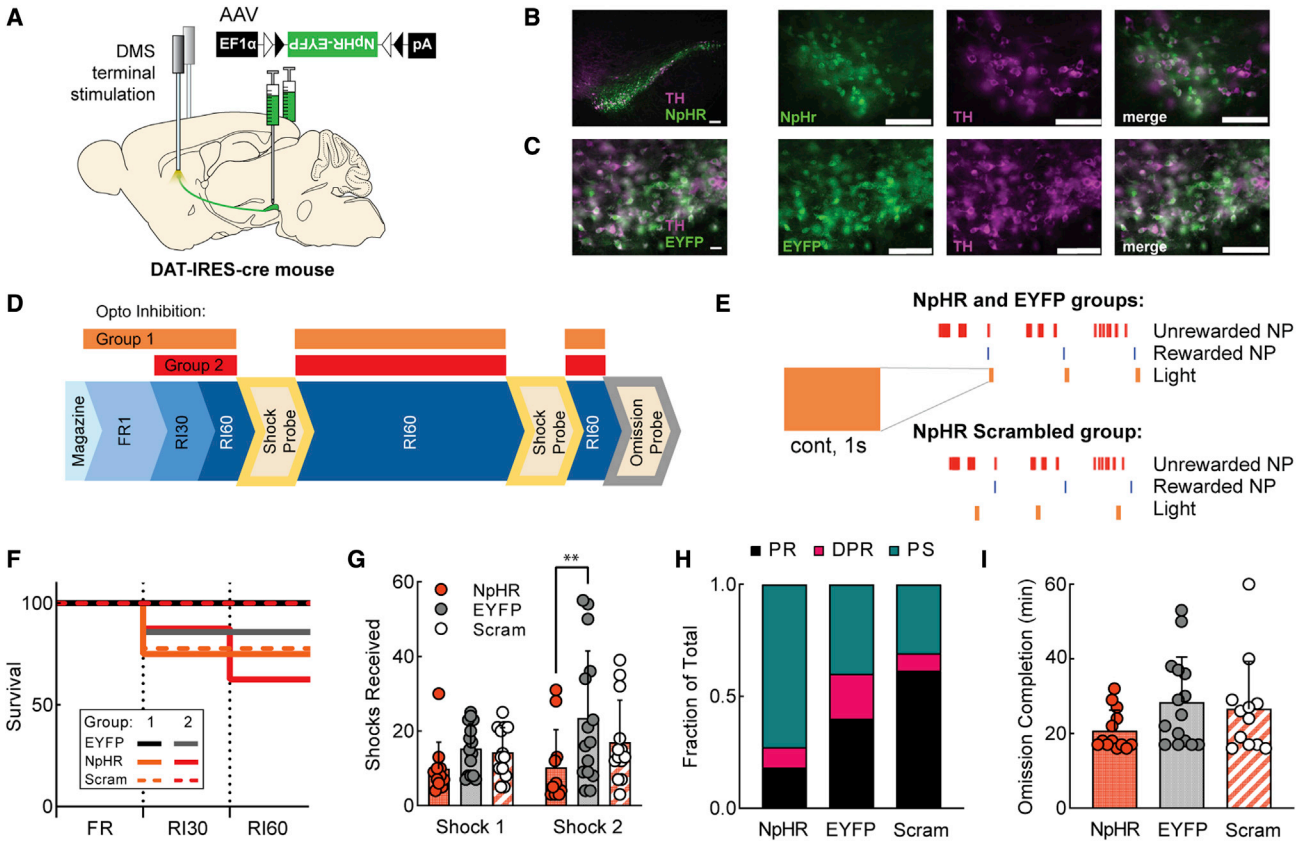


Figure 6. Optogenetic inhibition of dopamine terminals in DMS delays development of punishment-resistant reward seeking

(A) Viral injection and probe placement strategy.

(B and C) 10× and 40× images of SNc showing NpHR-EYFP (top) or EYFP (bottom) expression in green, TH positive cells in magenta, and merged image. Scale bars are 100 μm (all images).

(D) Training timeline showing when optogenetic inhibition was delivered (beginning with FR1 for group 1, RI30 for group 2).

(E) Schematic of light parameters: 1-s continuous light delivery was paired with rewarded nosepokes for NpHR and EYFP groups, and the same light was paired with a random subset of nosepokes for NpHR scrambled animals.

(F) Survival plot showing percentage of animals reaching criterion for each stage of training (see STAR Methods). NpHR (group 1, orange, n = 8; group 2, red, n = 8), EYFP (group 1, black, n = 9; group 2, gray, n = 7), NpHR scrambled (group 1, orange dash, n = 9; group 2, red dash, n = 6).

(G) Average shocks received on early and late shock probes for each manipulation, groups 1 and 2 combined

(H) Fraction of each behavioral phenotype (PR, black; DPR, pink; PS, teal) per manipulation.

(I) Average omission completion time per manipulation. All error bars represent SD. **$p < 0.01$.

See also Figure S6.

that after initial learning delays due to DMS dopamine terminal inhibition are overcome, inhibition on rewarded nosepokes delays the development of punishment resistance.

Optogenetic excitation of dopamine terminals in DLS at the time of a rewarded nosepoke does not influence instrumental learning or behavioral flexibility

Peaks in DLS dopamine axon activity in response to rewarded nosepokes were not correlated with the development of punishment-resistant reward seeking (Figure 4G). Therefore, we hypothesized stimulation of DLS dopamine terminals following rewarded nosepokes would not affect the development of this behavior. Nevertheless, we tested the effects of DLS dopamine terminal stimulation as a counterpoint to the effects of DMS dopamine terminal stimulation (Figures 5 and S5). In other words, can stimulating *any* dopamine signal boost the development of punishment resistance, or is this effect specific to DMS?

We performed the same experiment as in Figure 5 but targeting DLS instead of DMS. An AAV expressing cre-dependent ChR2 (AAV5-EF1α-DIO-hChR2(H134R)-EYFP) or a fluorophore-only control virus (AAV5-EF1α-DIO-EYFP) was injected into the SNc of DAT-IRES-cre mice to express ChR2 or EYFP specifically in dopamine neurons (Figures 7A–7C). A fiber optic probe was placed above DLS (Figures 7A and S7A). Expression levels of ChR2 in this experiment were closely matched with the previous DMS ChR2 experiment (Figure S7D). Light stimulation (1 s, 20 Hz) was delivered during training sessions, beginning with FR1 (Figures 7D and 7E). All mice quickly reached FR1 criterion (Figures S7B and S7C). All groups of mice (ChR2, EYFP, and ChR2 scrambled) behaved similarly, escalating their nosepoking at the same rates (Figure 7F), and receiving similar numbers of

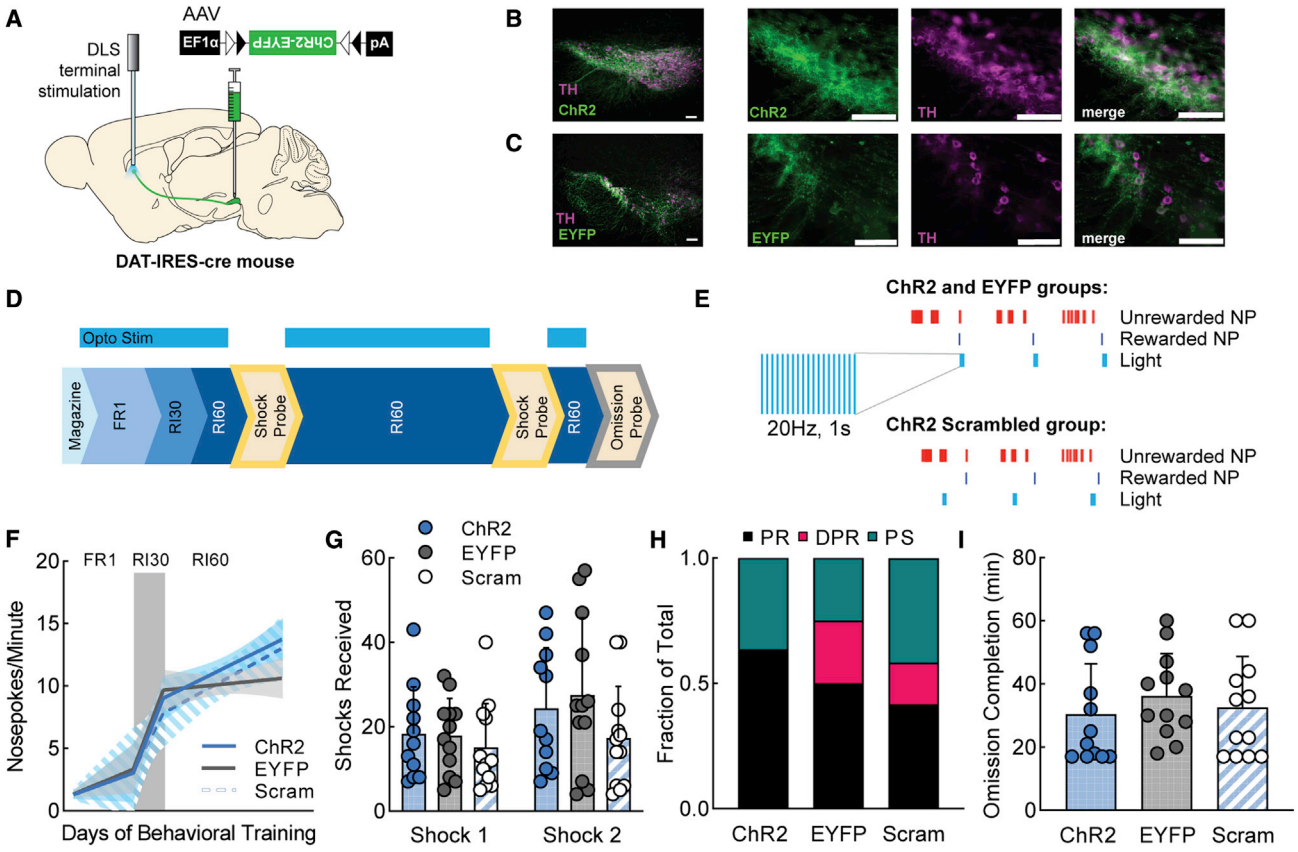


Figure 7. Optogenetic excitation of dopamine terminals in DLS at the time of a rewarded nosepoke does not influence instrumental learning or behavioral flexibility

(A) Viral injection and probe placement strategy.

(B and C) 10× and 40× images of SNc showing ChR2-EYFP (top) or EYFP (bottom) expression in green, TH positive cells in magenta, and merged image. Scale bars are 100 μm (all images).

(D) Training timeline showing when optogenetic stimulation was delivered (FR1, RI30, or RI60).

(E) Schematic of stimulation parameters.

(F) Segmental linear regression showing slope of nosepokes made per minute in FR1, RI30, and RI60 schedules. Shaded region represents 95% confidence bands.

(G) Average shocks received on early and late shock probes for each manipulation.

(H) Fraction of each behavioral phenotype (PR, black; DPR, pink; PS, teal) per manipulation.

(I) Average omission completion time per manipulation. All error bars represent SD.

See also Figure S7.

shocks (Figure 7G). All groups had similar distributions of PR, DPR, and PS mice (Figure 7H). No significant differences were observed in omission completion time (Figure 7I). We conclude DLS dopamine terminal stimulation immediately following a rewarded nosepoke does not influence the development of punishment resistance.

DISCUSSION

Compulsive behavior is a defining feature of disorders such as substance use disorder and is often modeled in rodents as punishment-resistant reward seeking. There is some evidence suggesting that corticostriatal circuits control the expression of established compulsions, but little is known about the mechanisms regulating the development of compulsions.[5] We hypothesized that dopamine—a key neuromodulator regulating corticostriatal synaptic plasticity—could play a role in sculpting the emergence of punishment-resistant reward seeking. Using dual-site fiber photometry to record the activity of dopamine axons in DMS and DLS during a task (RI60) that promotes punishment-resistant reward seeking, we found that DMS dopamine responses on rewarded actions predicted the development of punishment resistance. We confirmed a causal relationship between DMS dopamine signaling and the development of punishment resistance using excitatory and inhibitory optogenetics. Although DMS dopamine signaling was related to punishment-resistant reward seeking, it was not related to another form of inflexible behavior involving a contingency reversal (omission). The omission probe differs from the shock probe because it requires response inhibition, while the shock probe requires weighing the cost of the shock versus the benefit of the reward. We speculate that DMS dopamine is related only to the latter evaluation because it involves a goal-directed cost-benefit analysis rather than being related to impulsivity or habit.

Although our data are not conclusive, given the well-known role of DMS in action-outcome learning, we favor the hypothesis that DMS dopamine signaling promotes punishment-resistant reward seeking by boosting the predicted value of reward-seeking actions, strengthening action-outcome associations to make them robust to occasional punishment. This role for goal-directed behavior in punishment resistance could be a mechanism for adaptive resilience in challenging natural environments, where the benefits of seeking reliably available rewards might outweigh potential dangers. A goal-directed account of punishment resistance would be consistent with data showing that people with SUDs can still respond to incentives, for example in contingency management therapy.[42]

Some previous studies have observed a progression from habit to punishment-resistant reward seeking after extended training.[7,43,44] Such observations could be due to a common upstream driver of DMS and DLS function rather than a direct and necessary link between habit formation and punishment resistance. In our experiments, extended RI60 training led to habit-like omission-resistant reward seeking (as previously documented) [23,45] in addition to punishment resistance. However, by analyzing individual differences in behavior, we determined that the habits and punishment resistance do not inevitably develop together, consistent with the findings of Singer et al. and others.[11,46] Our results do not rule out the possibility that there are both DMS- and DLS-dependent routes to developing punishment resistance, which could be invoked under different circumstances. Interesting, in this issue, van Elzelingen et al. also identify DMS dopamine signaling as a key component of the shift to habit, questioning the previously hypothesized role of DLS in both habits and punishment resistance [47].

Further studies are needed to examine the relationship between DLS dopamine signals and behavior. Here, we observed novel temporal dynamics in the DLS dopamine axon signal, but the importance of these signals is mysterious. Alternative tasks or outcome measures might be used in future experiments. For example, 1 recent study linked high levels of extracellular dopamine in DLS with high impulsivity in a delay-discounting task,[48] while another study linked a molecularly defined population of dopamine neurons that primarily projects to DLS (Aldh1a1+ dopamine neurons) to motor learning on the accelerating rotarod.[49]

Future work should also examine temporal patterns. Creating peaks in DMS dopamine on random nosepokes did not have the same effect as creating these peaks on rewarded nosepokes in our experiments. It remains to be determined *why* these conditions differ. For example, if different cortical inputs to DMS are active during rewarded versus unrewarded nosepokes, dopamine release at these distinct times would reinforce the strength of different corticostriatal synapses.

Our findings emphasize the importance of grappling with individual differences in behavior. Not all people who try drugs become compulsive users. Large individual variability in compulsivity has been observed in animals working for drugs such as cocaine and alcohol.[7,50] We identified one reason for this variability: the different strategies used by individual animals to deal with uncertainty in reward availability. Our findings suggest there is a predisposition to punishment resistance present in some individuals before they confront punishments (analogous to our PR mice), rather than a stochastic process occurring during the experience of punishment.

One source of individual variability was sex: male mice were more likely to be punishment resistant than females. Nevertheless, it is important to note that sex is not deterministic. The correlation between DMS dopamine axon signaling and the development of punishment resistance was not sex-dependent, and DMS dopamine terminal stimulation induces punishment resistance in both sexes. We therefore suspect that sex differences in the propensity to develop punishment resistance occur upstream of dopamine neurons.

Finally, it is important to understand compulsive behavior for natural rewards to better understand the evolutionary context under which this behavior developed. Understanding how compulsive drug seeking and compulsive sucrose seeking relate to each other could also elucidate how concepts from SUD should be applied to understand behavioral addictions like gambling.

In summary, we have identified DMS dopamine signaling as a key part of the circuitry that drives the emergence of compulsive behavior in the context of natural reward seeking. The data presented here set the stage for interesting new studies in a variety of areas. Examining how the mechanisms we have identified contribute to the etiology of disorders such as SUD is of particular importance for translational impact.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Operant behavior
  - Shock probe
  - Omission probe
  - Fear conditioning
  - Open field locomotion
  - Stereotaxic surgery
  - Fiber photometry
  - Quantitative immunoblotting
  - Excitatory optogenetic stimulation
  - Inhibitory optogenetic stimulation
  - Transcardial perfusions
  - Histology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral analysis
  - Fiber photometry analysis
  - Statistical methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.01.055.

ACKNOWLEDGMENTS

We thank members of the Lerner Laboratory for helpful discussions and critical feedback on the manuscript. We thank Gates Palissery, Louis Van Camp, Hayden Sikora, and Meghana Holla for assistance with surgeries, histology, and data collection and entry. This work was supported by an NIH K99/R00 Award (R00MH109569), a NARSAD Young Investigator Award from the Brain & Behavior Research Foundation to T.N.L., and an NIH Diversity Supplement (R00MH109569-04S1) to support C.V.C. Graphical abstract created with https://biorender.com/.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, software, validation, and project administration, J.L.S., C.V.C., and T.N.L.; investigation, J.L.S., C.V.C., M.D.S., and A.S.B.; data curation and visualization, J.L.S., M.D.S., J.M.B., A.S.B., and T.N.L.; formal analysis, J.L.S., V.N.S., and J.M.B.; writing – original draft, J.L.S.; resources, writing – review & editing, supervision, and funding acquisition, T.N.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: August 19, 2021
Revised: December 2, 2021
Accepted: January 20, 2022
Published: February 7, 2022

REFERENCES

1. Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005). The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.
2. Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur. J. Neurosci. *22*, 505–512.
3. Lipton, D.M., Gonzales, B.J., and Citri, A. (2019). Dorsal Striatal Circuits for Habits, Compulsions and Addictions. Front. Syst. Neurosci. *13*, 28.
4. Lüscher, C., Robbins, T.W., and Everitt, B.J. (2020). The transition to compulsion in addiction. Nat. Rev. Neurosci. *21*, 247–263.
5. Lüscher, C., and Janak, P.H. (2021). Consolidating the Circuit Model for Addiction. Annu. Rev. Neurosci. *44*, 173–195.
6. Gillan, C.M., Robbins, T.W., Sahakian, B.J., van den Heuvel, O.A., and van Wingen, G. (2016). The role of habit in compulsivity. Eur. Neuropsychopharmacol. *26*, 828–840.
7. Giuliano, C., Belin, D., and Everitt, B.J. (2019). Compulsive Alcohol Seeking Results from a Failure to Disengage Dorsolateral Striatal Control over Behavior. J. Neurosci. *39*, 1744–1754.
8. Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur. J. Neurosci. *19*, 181–189.
9. Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476.
10. Willuhn, I., Burgeno, L.M., Everitt, B.J., and Phillips, P.E.M. (2012). Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. Proc. Natl. Acad. Sci. USA *109*, 20703–20708.
11. Singer, B.F., Fadanelli, M., Kawa, A.B., and Robinson, T.E. (2018). Are cocaine-seeking "habits" necessary for the development of addiction-like behavior in rats? J. Neurosci. *38*, 60–73.
12. Harada, M., Pascoli, V., Hiver, A., Flakowski, J., and Lüscher, C. (2021). Corticostriatal Activity Driving Compulsive Reward Seeking. Biol. Psychiatry *90*, 808–818.
13. Hu, Y., Salmeron, B.J., Krasnova, I.N., Gu, H., Lu, H., Bonci, A., Cadet, J.L., Stein, E.A., and Yang, Y. (2019). Compulsive drug use is associated with imbalance of orbitofrontal- and prelimbic-striatal circuits in punishment-resistant individuals. Proc. Natl. Acad. Sci. USA *116*, 9066–9071.
14. Pascoli, V., Terrier, J., Hiver, A., and Lüscher, C. (2015). Sufficiency of Mesolimbic Dopamine Neuron Stimulation for the Progression to Addiction. Neuron *88*, 1054–1066.
15. Pascoli, V., Hiver, A., Van Zessen, R., Loureiro, M., Achargui, R., Harada, M., Flakowski, J., and Lüscher, C. (2018). Stochastic synaptic plasticity underlying compulsion in a model of addiction. Nature *564*, 366–371.
16. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.
17. Lerner, T.N., Holloway, A.L., and Seiler, J.L. (2021). Dopamine, Updated: Reward Prediction Error and Beyond. Curr. Opin. Neurobiol. *67*, 123–130.
18. Nicola, S.M., Surmeier, J., and Malenka, R.C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. Annu. Rev. Neurosci. *23*, 185–215.
19. Kreitzer, A.C., and Malenka, R.C. (2008). Striatal plasticity and basal ganglia circuit function. Neuron *60*, 543–554.
20. Lovinger, D.M. (2010). Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. Neuropharmacology *58*, 951–961.
21. Yin, H.H., Mulcare, S.P., Hilário, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., and Costa, R.M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat. Neurosci. *12*, 333–341.
22. Faure, A., Haberland, U., Condé, F., and El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. J. Neurosci. *25*, 2771–2780.
23. Derusso, A.L., Fan, D., Gupta, J., Shelest, O., Costa, R.M., and Yin, H.H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. Front. Integr. Nuerosci. *4*, 17.
24. Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat. Commun. *4*, 2264.
25. Wiltgen, B.J., Sinclair, C., Lane, C., Barrows, F., Molina, M., and Chabanon-Hicks, C. (2012). The effect of ratio and interval training on Pavlovian-instrumental transfer in mice. PLoS ONE *7*, e48227.
26. Lerner, T.N. (2020). Interfacing behavioral and neural circuit models for habit formation. J. Neurosci. Res. *98*, 1031–1045.
27. Yu, C., Gupta, J., Chen, J.-F., and Yin, H.H. (2009). Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. J. Neurosci. *29*, 15100–15103.
28. Rossi, M.A., and Yin, H.H. (2012). Methods for Studying Habitual Behavior in Mice. Curr. Protoc. Neurosci. *8*, Unit 8.29.
29. Farassat, N., Costa, K.M., Stojanovic, S., Albert, S., Kovacheva, L., Shin, J., Egger, R., Somayaji, M., Duvarci, S., Schneider, G., and Roeper, J. (2019). In vivo functional diversity of midbrain dopamine neurons within identified axonal projections. eLife *8*, e48408.
30. Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. Brain Res. Brain Res. Rev. *56*, 27–78.
31. Lerner, T.N., Shilyansky, C., Davidson, T.J., Evans, K.E., Beier, K.T., Zalocusky, K.A., Crow, A.K., Malenka, R.C., Luo, L., Tomer, R., and

Deisseroth, K. (2015). Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell *162*, 635–647.

32. Dana, H., Sun, Y., Mohar, B., Hulse, B.K., Kerlin, A.M., Hasseman, J.P., et al. (2019). High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nature Methods *16* (7), 649–657.
33. Bäckman, C.M., Malik, N., Zhang, Y., Shan, L., Grinberg, A., Hoffer, B.J., Westphal, H., and Tomac, A.C. (2006). Characterization of a mouse strain expressing Cre recombinase from the 3¢ untranslated region of the dopamine transporter locus. Genesis *44*, 383–390.
34. Chohan, M.O., Esses, S., Haft, J., Ahmari, S.E., and Veenstra-VanderWeele, J. (2020). Altered baseline and amphetamine-mediated behavioral profiles in dopamine transporter Cre (DAT-Ires-Cre) mice compared to tyrosine hydroxylase Cre (TH-Cre) mice. Psychopharmacology (Berl.) *237*, 3553–3568.
35. Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. Nat. Neurosci. *19*, 117–126.
36. Mohebi, A., Pettibone, J.R., Hamid, A.A., Wong, J.-M.T., Vinson, L.T., Patriarchi, T., Tian, L., Kennedy, R.T., and Berke, J.D. (2019). Dissociable dopamine dynamics for learning and motivation. Nature *570*, 65–70.
37. Kim, H.R., Malik, A.N., Mikhael, J.G., Bech, P., Tsutsui-Kimura, I., Sun, F., Zhang, Y., Li, Y., Watabe-Uchida, M., Gershman, S.J., and Uchida, N. (2020). A Unified Framework for Dopamine Signals across Timescales. Cell *183*, 1600–1616.e25.
38. Guru, A., Seo, C., Post, R., Kullakanda, D., Schaffer, J., and Warden, M. (2020). Ramping activity in midbrain dopamine neurons signifies the use of a cognitive map. bioRxiv. https://doi.org/10.1101/2020.05.21.108886.
39. Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M., and Graybiel, A.M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature *500*, 575–579.
40. Hamid, A.A., Frank, M.J., and Moore, C.I. (2021). Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. Cell *184*, 2733–2749.e16.
41. Gradinaru, V., Zhang, F., Ramakrishnan, C., Mattis, J., Prakash, R., Diester, I., Goshen, I., Thompson, K.R., and Deisseroth, K. (2010). Molecular and cellular approaches for diversifying and extending optogenetics. Cell *141*, 154–165.
42. Pickard, H. (2021). Is addiction a brain disease? A plea for agnosticism and heterogeneity. Psychopharmacology (Berl.). Published online November 26, 2021. https://doi.org/10.1007/s00213-021-06013-4.
43. Everitt, B.J., and Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat. Neurosci. *8*, 1481–1489.
44. Everitt, B.J., and Robbins, T.W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. Annu. Rev. Psychol. *67*, 23–50.
45. Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav. Brain Res. *166*, 189–196.
46. Olmstead, M.C., Lafond, M.V., Everitt, B.J., and Dickinson, A. (2001). Cocaine seeking by rats is a goal-directed action. Behav. Neurosci. *115*, 394–402.
47. van Elzelingen, W., Warnaar, P., Matos, J., Bastet, W., Jonkman, R., Smulders, D., Goedhoop, J., Denys, D., Arbab, T., and Willuhn, I. (2022). Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. Curr. Biol. Published February 7, 2022. https://doi.org/10.1016/j.cub.2021.12.027.
48. Moreno, M., Azocar, V., Vergés, A., and Fuentealba, J.A. (2021). High impulsive choice is accompanied by an increase in dopamine release in rat dorsolateral striatum. Behav. Brain Res. *405*, 113199.
49. Wu, J., Kung, J., Dong, J., Chang, L., Xie, C., Habib, A., Hawes, S., Yang, N., Chen, V., Liu, Z., et al. (2019). Distinct Connectivity and Functionality of Aldehyde Dehydrogenase 1a1-Positive Nigrostriatal Dopaminergic Neurons in Motor Learning. Cell Rep. *28*, 1167–1181.e7.
50. Siciliano, C.A., Noamany, H., Chang, C.-J., Brown, A.R., Chen, X., Leible, D., Lee, J.J., Wang, J., Vernon, A.N., Vander Weele, C.M., et al. (2019). A cortical-brainstem circuit predicts and governs compulsive alcohol drinking. Science *366*, 1008–1012.
51. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.
52. Lugo, J.N., Smith, G.D., and Holley, A.J. (2014). Trace Fear Conditioning in Mice. J. Vis. Exp. *80*, 51180.
53. Seibenhener, M.L., and Wooten, M.C. (2015). Use of the Open Field Maze to measure locomotor and anxiety-like behavior in mice. J. Vis. Exp. *96*, e52434.
54. Franklin, K.B.J., and Paxinos, G. (2013). The mouse brain in stereotaxic coordinates (Academic Press).
55. Muir, J., Lorsch, Z.S., Ramakrishnan, C., Deisseroth, K., Nestler, E.J., Calipari, E.S., and Bagot, R.C. (2018). In Vivo Fiber Photometry Reveals Signature of Future Stress Susceptibility in Nucleus Accumbens. Neuropsychopharmacology *43*, 255–263.
56. Holly, E.N., Davatolhagh, M.F., Choi, K., Alabi, O.O., Vargas Cifuentes, L., and Fuccillo, M.V. (2019). Striatal Low-Threshold Spiking Interneurons Regulate Goal-Directed Learning. Neuron *103*, 92–101.e6.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
goat anti-chicken Alexa Fluor 647	Life Technologies	Cat#A-21449: RRID:AB_2535866
Tyrosine Hydroxylase Antibody cocktail	Aves Labs	Cat#TYH; RRID:AB_10013440
Anti-Dopamine Transporter (rabbit polyclonal)	MilliporeSigma	Cat#AB2231; RRID:AB_1586991
Monoclonal Beta-actin antibody	MilliporeSigma	Cat#A2228; RRID:AB_476697
Bacterial and virus strains		
AAV5-EF1α-DIO-EYFP	UNC Vector Core	Lot#AV4310K
AAV5-CAG-FLEX-jGCaMP7b-WPRE	Addgene	Lot#18-429; RRID:Addgene_104497
AAV5-EF1α-DIO-hChR2(H134R)-EYFP	Addgene	Lot#v17652; RRID:Addgene_55639
AAV5-EF1α-DIO-eNpHR3.0-EYFP	Addgene	Lot#v32533; RRID:Addgene_26966
Chemicals, peptides, and recombinant proteins		
Isoflurane	Henry Schein	N/A
Buprenorphine SR	Zoopharm	Lot#1-212403
Carpofen	Zoetis	N/A
Euthasol	Virbac	N/A
Normal Goat Serum	Jackson ImmunoResearch Laboratories	Lot#153636; RRID:AB_2336990
Fluoromont-G	Southern Biotech	Cat#0100-01
Triton X	Sigma	Cat#X100-1L
Critical commercial assays		
Pierce BCA protein assay kit	Thermo Fisher	Lot#WG332025
Deposited data		
Photometry Analysis Code (MATLAB)	Generated by study	GitHub: https://doi.org/10.5281/zenodo.5828906
Experimental models: Organisms/strains		
Mouse: DAT-IRES-Cre: B6.SJL-*Slc6a3tm1.1(cre)Bkmn*/J	Bred in house	N/A
Mouse: WT: C57BL/6J	Jackson Laboratories	Strain #000664, RRID:IMSR_JAX:000664
Software and algorithms		
Synapse	Tucker Davis Technologies	https://www.tdt.com/component/synapse-software/
ImageJ	[51]	https://imagej.nih.gov/ij/; RRID:SCR_003070
MED-PC V	Med Associates	https://www.med-associates.com/med-pc-v/; RRID:SCR_014721
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html; RRID:SCR_001622
Ethovision XT	Noldus	https://www.noldus.com/ethovision-xt; RRID:SCR_000441

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Talia Lerner (talia.lerner@northwestern.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- All original code for fiber photometry analysis has been deposited to Github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Male and female *WT* (C57BL/6J) and (DAT)::IRES-Cre knockin mice (JAX006660) were obtained from The Jackson Laboratory and crossed in house. Only heterozygote transgenic mice, obtained by backcrossing to C57BL/6J wildtypes, were used for experiments. Littermates of the same sex were randomly assigned to experimental groups (fiber photometry-14 males, 22 females; DMS excitatory optogenetics- 20 males, 19 females; DMS inhibitory optogenetics- 13 males, 13 females; DLS excitatory optogenetics- 18 males, 18 females). Adult mice at least 10 weeks of age were used in all experiments. Mice were group housed under a conventional 12 h light cycle (dark from 7:00pm to 7:00am) with *ad libitum* access to food and water prior to operant training. All experiments were approved by the Northwestern University Institutional Animal Care and Use Committee.

METHOD DETAILS

Operant behavior

Mice were food restricted to 85% of *ad libitum* body weight for the duration of operant training. Mice were given one day of habituation to operant chambers (Med Associates) and tethering with patch cords (Doric Lenses) for one h. They were then trained to retrieve food rewards (45 mg purified pellet, Bio-Serv) from a magazine port. For this magazine training, pellets were delivered to the port on a random interval (RI60) schedule non-contingently for one h. Next, operant training began, with all training sessions lasting one h or until 50 rewards had been earned. Mice were trained to associate nosepoking with reward on a fixed ratio (FR1) schedule where both nosepokes delivered a reward. They had to retrieve the reward (as measured by making a port entry following a rewarded nosepoke) before they could earn the next reward. After a mouse showed a preference for one nosepoke (> 25 rewards on that side; average of 3 days), they were trained on FR1 on their preferred side only, with nosepokes on the other side having no consequence, until they received > 30 rewards for a minimum of two consecutive days (average of 6 days). Mice that did not reached this criterion after 14 days of FR1 training (mean+2 SD), were removed from the study. Mice passing the FR1 criterion were then moved to either a random interval (n = 36) or random ratio (n = 7) schedule of reinforcement. Mice on the random interval schedule were trained on RI30 until they earned > 30 rewards in one h (average of 2 days). Mice that did not reach this criterion after 5 days of RI30 training were removed from the study. Mice passing RI30 criterion were then trained on RI60. Mice on a random ratio schedule of reinforcement were trained on RR10 until they earned > 30 rewards in one h (average of 3 days), and then trained on RR20 (Figure 1A). For random interval and random ratio schedules, a normal distribution centered around the number indicated in the name of the schedule was used to create the schedule. The range for RI30 was from 15-45 s, RI60 from 30-90 s, RR10 from 6-14 nosepokes, and RR20 from 14-28 nosepokes.

Shock probe

Mice were subjected to a footshock probe early and late in training (Figure 1B) to evaluate their levels of punishment-resistance reward-seeking. These probes were performed under an FR1 schedule of reinforcement where a mild footshock (0.2mA, 1 s) was paired with a subset of rewarded nosepokes on a RR3 schedule, so that, on average, every third rewarded nosepoke was paired with a footshock. The first five rewarded nosepokes were never paired with shock. During shock probes, the session ended after 60 min or a mouse was inactive (no nosepokes on the rewarded side) for > 10 min. There was no maximum number of rewards.

Omission probe

A subset of mice (n = 20) were returned to RI60/RR20 training after the late footshock probe until their nosepoke rates returned to pre-shock levels. They then received a single omission probe session where they had to withhold nosepoking for 20 s in order to receive a single reward pellet. A nosepoke reset the 20 s timer. Each session ended after a mouse received 50 rewards or 60 min had elapsed.

Fear conditioning

A trace fear conditioning paradigm, adapted from Lugo, Smith, and Holley was used in a naive cohort of wild-type mice (n = 13) to verify that our shock intensity (0.2mA) is aversive to the mice.[52] Mice were randomly assigned to cued or non-cued groups. On the first day, mice received 12 tone only or tone-shock pairings (2900 Hz tone) in a standard operant chamber (Med Associates). The next

day, mice were placed in a different context (using white walls, white plastic flooring, and vanilla scent) and 12 tones were presented. All sessions were recorded using Med Associates Video Monitor software.

Open field locomotion

Open-field test was performed as previously described.[53] Mice were placed into the center of a 28x28x28 cm arena exposed to white fluorescent light and video recorded for 5 min using a GigE camera (Basler AG) and the Noldus XT video tracking software (Noldus Information Tech Inc). The arena was separated into 3 zones: center (12 cm square in center), border (5 cm area surrounding center), and wall (3 cm area adjacent to walls). Cumulative time spent in each zone and movement parameters were calculated by the tracking software based off of the mouse's center body point.

Stereotaxic surgery

Viral infusions and optic fiber implant surgeries took place under isoflurane anesthesia (Henry Schein). Mice were anesthetized in an isoflurane induction chamber at 3%–4% isoflurane, and then injected with buprenorphine SR (Zoopharm, 0.5 mg/kg s.q.) and carpofen (Zoetis, 5 mg/kg s.q.) prior to the start of surgery. Mice were placed on a stereotaxic frame (Stoetling) and hair was removed from the scalp using Nair. The skin was cleaned with alcohol and a povidone-iodine solution prior to incision. The scalp was opened using a sterile scalpel and holes were drilled in the skull at the appropriate stereotaxic coordinates. Viruses were infused at 100 nL/min through a blunt 33-gauge injection needle using a syringe pump (World Precision Instruments). The needle was left in place for 5 min following the end of the injection, then slowly retracted to avoid leakage up the injection tract. Implants were secured to the skull with Metabond (Parkell) and Flow-it ALC blue light-curing dental epoxy (Pentron). After surgery, mice were allowed to recover until ambulatory on a heated pad, then returned to their homecage with moistened chow or DietGel available. Mice then recovered for three weeks before behavioral experiments began.

Fiber photometry

Mice for fiber photometry experiments received infusions of 1μl of AAV5-CAG-FLEX-jGCaMP7b-WPRE (1.02e13 vg/mL, Addgene, lot 18-429) into lateral SNc (AP −3.1, ML 1.3, DV −4.2) in one hemisphere and medial SNc (AP −3.1, ML 0.8, DV −4.7) in the other. Hemispheres were counterbalanced between mice. Fiber optic implants (Doric Lenses; 400 μm, 0.48 NA) were placed above DMS (AP 0.8, ML 1.5, DV −2.8) and DLS (AP −0.1, ML 2.8, DV −3.5). The DMS implant was placed in the hemisphere receiving a medial SNc viral injection, while the DLS implant was placed in the hemisphere receiving a lateral SNc viral injection. Calcium signals from dopamine terminals in DMS and DLS were recorded during RI30, on the first and last days of RI60/RR20 training as well as on both footshock probes for each mouse. All recordings were done using a fiber photometry rig with optical components from Doric lenses controlled by a real-time processor from Tucker Davis Technologies (TDT; RZ5P). TDT Synapse software was used for data acquisition. 465nm and 405nm LEDs were modulated at 211 Hz and 330 Hz, respectively, for DMS probes. 465nm and 405nm LEDs were modulated at 450 Hz and 270 Hz, respectively for DLS probes. LED currents were adjusted in order to return a voltage between 150-200mV for each signal, were offset by 5 mA, were demodulated using a 4 Hz lowpass frequency filter. Behavioral timestamps, e.g., for nosepokes and port entries, were fed into the real-time processor as TTL signals from the operant chambers (MED Associates) for alignment with the neural data.

Quantitative immunoblotting

Brains were harvested from mice following cervical dislocation, immediately snap frozen in liquid nitrogen and stored at −80°C. In order to collect tissue punches, brains were kept frozen and cut into 1mm coronal slices using a stainless steel adult brain matrix (Zinc Instruments). A one millimeter frozen tissue punch was collected from both DMS and DLS. Tissue samples were lysed in cold RIPA buffer supplemented with phosphatase and protease inhibitor tablets. Tissues were lysed in microfuge tubes using a motorized pestle grinder (Cole Palmer), centrifuged at 15,000 g for 10 min at 4°C and supernatant was collected. Lysate protein concentration was determined using a Pierce BCA protein assay kit (Thermo Fisher Scientific). 50μg of protein was separated by SDS-page electrophoresis on 12% resolving gels (Bio-Rad) and transferred to a PVDF membrane (Thermo Fisher Scientific). Dopamine transporter antibody (AB2231) and monoclonal β-actin (A2228) antibody were purchased from MilliporeSigma. Secondary IRDye 680RD (926-6807) and IRDye 800CW (926-32210) were purchased from Li-Cor. Imaging was performed using a Li-Cor Odyssey FC imaging station. Quantification was performed by densitometry using NIH ImageJ software.[51]

Excitatory optogenetic stimulation

Mice for DMS (Figure 5) and DLS (Figure 7) excitatory optogenetics experiments received 1 μl of AAV5-EF1α-DIO-hChR2(H134R)-EYFP (3.3e13 GC/mL, Addgene, lot v17652) or the control fluorophore-only virus AAV5-EF1α-DIO-EYFP (3.5e12 virus molecules/mL, UNC Vector Core, lot AV4310K) in medial (AP −3.1, ML 0.8, DV −4.7) or lateral SNc (AP −3.1, ML 1.3, DV −4.2) and a single fiber optic implant (Prizmatix; 250μm core, 0.66 NA) over ipsilateral DMS (AP 0.8, ML 1.5, DV −2.8) or DLS (AP −0.1, ML 2.8, DV −3.5). Hemispheres were counterbalanced between mice. During operant training (beginning with FR1), each rewarded nosepoke was paired with a train of blue light (460nm, 1 s, 20 Hz, 15 mW) generated by an LED light source and pulse generator (Prizmatix). A subset of mice ("ChR2 Scrambled") received the same train of light but paired with random nosepokes on a separate RI60 schedule.

Inhibitory optogenetic stimulation

Mice for DMS inhibitory optogenetics experiments received 1 μl per side of AAV5-EF1α-DIO-eNpHR3.0-EYFP (1.1e13 GC/mL, Addgene, lot v32533) or the control fluorophore-only virus AAV5-EF1α-DIO-EYFP (3.5e12 virus molecules/mL, UNC Vector Core, lot AV4310K) in bilateral medial SNc (AP −3.1, ML 0.8, DV −4.7) and bilateral fiber optic implants (Prizmatix; 500μm core, 0.66 NA) in DMS (AP 0.8, ML ± 1.5, DV −2.8). There were two groups of inhibitory optogenetics animals. Group 1 received inhibitory stimulation during operant training beginning with FR1, Since a subset of animals in this group were unable to learn the operant task, we also ran another group (Group 2) that received inhibitory stimulation during operant training beginning with RI30. These groups are combined for analysis of behaviors occurring after RI training has begun. For both groups, each rewarded nosepoke was paired with a continuous pulse of orange/red light (625nm, 1 s, 15 mW) generated by an LED light source and pulse generator (Prizmatix). A subset of mice ("NpHR Scrambled") received the same continuous pulse of light but paired with random nosepokes on a separate RI60 schedule.

Transcardial perfusions

Mice received lethal i.p. injections of Euthasol (Virbac, 1mg/kg) a combination of sodium pentobarbital (390 mg/mL) and sodium phenytoin (50 mg/mL), to induce a smooth and rapid onset of unconsciousness and death. Once unresponsive to a firm toe pinch, an incision was made up the middle of the body cavity. An injection needle was inserted into the left ventricle of the heart, the right atrium was punctured and solution (PBS followed by 4% PFA) was infused as the mouse was exsanguinated. The mouse was then decapitated and its brain was removed and fixed overnight at 4°C in 4% PFA.

Histology

After perfusion and fixation, brains were transferred to a solution of 30% sucrose in PBS, where they were stored for at least two overnights at 4°C before sectioning. Tissue was sectioned on a freezing microtome (Leica) at 30 μm, stored in cryoprotectant (30% sucrose, 30% ethylene glycol, 1% polyvinyl pyrrolidone in PB) at 4°C until immunostaining. Tyrosine hydroxlase (TH) staining was performed on free floating sections, which were blocked with 3% normal goat serum in PBS-T for 1 h at room temperature, then stained with 1:500 primary antibody (Aves Labs, Cat No. TYH) in blocking solution at 4°C overnight. Secondary staining was performed using 1:500 goat anti-chicken Alexa Fluor 647 secondary antibody (Life Technologies, Cat. No. A-21449). Anti-GFP staining was performed on free floating sections to amplify signals from GCaMP7b. This staining was performed by blocking in 3% normal goat serum in PBS-T for 1 h at room temperature, then using 1:500 primary antibody conjugated directly to Alexa Fluor 488 (Life Technologies, Cat. No. A-21311) in blocking solution at 4°C overnight. Tissue was mounted on slides in PBS and coverslips were secured with Fluoromont-G (Southern Biotech). Slides were imaged using a fluorescent microscope (Keyence BZ-X800) with 5x and 40x air immersion objectives. Probe placements were determined by comparing to the Mouse Brain Atlas.[54] GCaMP neurons expressing YFP were counted and colocalized with TH+ neurons using ImageJ software.[51] Mean fluorescence (Figure S7E) was calculated as mean gray value in a fixed region of interest surrounding the end of the probe using ImageJ software.[51]

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

Cue-evoked freezing during fear conditioning was scored manually by two blind observers from a recording of the fear conditioning test session using EthoVision software (Noldus). Scores from the two observers were averaged. Freezing was measured throughout the session as a mouse remaining still for more than two seconds.

For all other studies, behavioral data was collected automatically by MED-PC software (Med Associates). For behavioral and fiber photometry experiments (Figures 1, 2, 3, and4) mice were trained on RI60 or RR20 reward schedules, then sorted into PR, DPR and PS groups based on post hoc analysis of their performance in the shock probe sessions. For optogenetics experiments (Figures 5, 6, and 7), mice were trained on an RI60 schedule and sorted into *a priori* stimulation groups. *Post hoc* behavioral classifications were determined after the experiment as a means of analyzing whether the optogenetic manipulations influenced the punishment-resistant phenotype (see Methods S1).

In the initial behavioral experiments, we sorted mice into PR, DPR, and PS groups by calculating the percent change in shocks received from the early to late shock probe for each mouse. Mice in the top quartile of changers (who increased the number of shocks received by greater than 85%) were classified as delayed punishment resistant (DPR; n = 9). The remaining mice were sorted by a median split, with mice receiving more than 13 shocks on the first probe classified as punishment resistant (PR; n = 9) and those earning fewer as punishment sensitive (PS; n = 18, total n = 36, Figure 2A). In subsequent optogenetics experiments, PR, DPR and PS groups were determined based on the absolute median of the RI60-trained animals in the fiber photometry experiment, so that criteria for the phenotypes remains consistent across experiments.

The subset of mice that received the omission probe were also sorted by a median split of omission completion time (time to 50 rewards), with mice taking more than 29 min classified as long omission (n = 10) and those taking less time as short omission (n = 10, Figure S4H).

Plots in Figures 2O, 5F, 7F, S6B, and S6C were generated by plotting a segmental linear regression with lines for the average slope of nosepokes/minute across FR1, RI30, and RI60 training to reveal escalation of nosepoke behavior. The shaded area shows the 95% confidence band surrounding each slope. This analysis was done using GraphPad (Prism) software. Inter-reward intervals were

calculated as the time from a rewarded nosepoke to the subsequent rewarded nosepoke (Figure 2L) on each day of RI60 training. A frequency distribution was created and plotted using GraphPad (Prism) software. Plots in Figures S1B and S1C were generated by binning the number of nosepokes per five minutes during probe sessions and dividing by nosepokes in the same five-minute bin on the most recent day of RI60/RR20 training.

Fiber photometry analysis

All analysis was done using custom MATLAB (Mathworks) and Python code. Raw data from 465nm and 405nm channels were passed through a zero-phase digital filter (filtfilt function in MATLAB) and a least-squares linear fit (parameters derived with polyfit function) was applied to the 405nm control signal to align it to the 465nm signal. Recordings with excessive artifacts (due to patch cord movement, mouse pulling cords off, etc) after fitting signal that were therefore too noisy were excluded from group analysis. All n's for each group are shown in figure legends. ΔF/F was calculated with the following formula: (465nm signal - fitted 405nm signal) / (fitted 405nm signal). To facilitate comparisons across animals, z-scores were calculated by subtracting the mean ΔF/F calculated across the entire session and dividing by the standard deviation. Peri-stimulus time histograms (PSTHs) were created using the TTL timestamps corresponding to behavioral events. Maximum and minimum peak values and locations from PSTHs in main figures were generated using max and min functions in MATLAB for the 1.5 s following behavioral event (ie nosepoke, port entry). Rewarded-unrewarded peak were calculated by subtracting the minimum peak in the average unrewarded nosepoke PSTH from the maximum in the average rewarded nosepoke PSTH. AUC was calculated using trap function in MATLAB. We used a customized logic for peak detection in Figures S3 and S4 adapted from Holly et al. and Muir et al.[55,56] Events having amplitudes greater than the summation of a median of 30 s moving window and two times median absolute deviation (MADs), were filtered out and the median of the resultant trace was calculated. Peaks having local maxima greater than three times MADs of the resultant trace above the median were considered as events.

Statistical methods

Statistical analysis was done using Prism 9 software (GraphPad). One and two-way ANOVAs, or mixed effects analyses were performed with Tukey's multiple comparisons and Bonferroni post hoc analyses when statistically significant main effects or interactions were found. A Kolmogorov-Smirnov test was used to compare the distributions of inter-reward intervals. One RI60 mouse was excluded from the fiber photometry study due to improper fiber placement and two due to poor photometry signal. A total of six mice were excluded from the optogenetics studies—five due to improper probe placement and one because of illness. All n values listed above do not include these mice.