Dopamine Signaling in the Dorsomedial Striatum Promotes Compulsive Behavior

Jillian L. Seiler[1,2], Caitlin V. Cosme[1,3], Venus N. Sherathiya[1], Joseph M. Bianco[1], and Talia N. Lerner[1,4*]

[1] Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago IL 60611 USA
[2] Department of Psychology, University of Illinois at Chicago, Chicago IL 60607 USA
[3] Present address: Teva Pharmaceutical Industries Ltd., USA
[4] Lead Contact
*Correspondence: talia.lerner@northwestern.edu

SUMMARY

Habits and compulsions are two aspects of behavior that often develop in parallel and together lead to inflexible responding. Both habits and compulsions are hypothesized to be involved in psychiatric disorders such as drug addiction and obsessive-compulsive disorder (OCD), but they are distinct behaviors that may rely on different brain circuitries. We developed an experimental paradigm to track the development of both habits and compulsions in individual animals while recording neural activity. We performed fiber photometry measurements of dopamine axon activity while mice engaged in reinforcement learning on a random interval (RI60) schedule and found that the emergence of compulsion was predicted by the dopamine signal in the dorsomedial striatum (DMS). By amplifying this DMS dopamine signal throughout training using optogenetics, we accelerated animals' transitions to compulsion, irrespective of habit formation. These results establish DMS dopamine signaling as a key controller of compulsions.

Keywords: dorsal striatum, dopamine, substantia nigra, instrumental learning, reinforcement learning, habit formation, compulsive behavior, punishment-resistant reward seeking, fiber photometry, optogenetics

INTRODUCTION

Animals learn about the consequences of their actions through reinforcement. As actions are reinforced by positive or negative outcomes, action-outcome associations are formed in the brain which allow an animal to predict action consequences and direct their behavior accordingly. This action-outcome learning relies on the dorsomedial striatum (DMS; Yin *et al.*, 2005; Yin, Knowlton and Balleine, 2005). However, as actions are repeated, animals transition away from goal-directed behavior relying on the DMS and towards a habitual strategy relying on the dorsolateral striatum (DLS; Yin, Knowlton and Balleine, 2004). Habits decouple actions from outcomes and instead promote the control of behavior through stimulus-response associations. The transition to habit can be measured as an insensitivity to outcome devaluation or as an insensitivity to changes in action-outcome contingency.

Habits are hypothesized to contribute to addiction by promoting compulsive reward seeking, defined as reward seeking in the face of negative consequences. As actions begin to produce bad as well as good outcomes, a reliance on habit may make it harder for an organism to adjust its reward-seeking behaviors. However, habits and compulsions are not the same. While habits produce inflexible behavior due to strong stimulus-response associations, they can be overridden by punishment. Compulsions, on the contrary, continue in the face of punishment. The failure of an animal to respond to the introduction of a new, negatively-valenced action-outcome contingency might depend on associative circuits distinct from those promoting habit.

The dorsolateral striatum (DLS) and its dopaminergic inputs from the substantia nigra pars compacta (SNc) are a proposed point of convergence between the mechanisms of habit and compulsion. Dopamine signaling in the DLS is required for habit (Faure et al., 2005), and blocking dopamine signaling in the DLS can inhibit drug seeking for cocaine or alcohol (Corbit et al., 2012; Hodebourg et al., 2018; Murray et al., 2012; Pacchioni et al., 2011; Vanderschuren et al., 2005). The inhibition of drug seeking by DLS dopamine signaling blockade precedes the development of compulsive drug-seeking, but can predict which animals will develop it (Giuliano et al., 2019; Willuhn et al., 2012).

Although the emergence of habitual, DLS-dependent control over reward seeking precedes and predicts the development of compulsions when animals are trained on repeated actions, the association between habits and compulsions is not absolutely required. When rats were trained to perform new action sequences each day to get cocaine, compulsive drug seeking developed, as measured by extinction and shock probes, that was independent of DLS dopamine signaling (Singer et al., 2018). Thus, constant learning of new action-outcome associations can prevent habit, but still lead to compulsion. This observation was made in the context of reward seeking for an addictive drug, cocaine. Whether similar dissociations between habit and compulsion might occur when animals are trained to work for natural rewards is unknown.

We set out to specifically examine whether training that is designed to promote habit formation during natural reward seeking also promotes compulsion and, if so, through what mechanisms. We trained mice on a random interval schedule (RI60) that promotes habit and probed for both compulsive and habitual responding. We found that compulsions and habits track together, as expected, in unmanipulated mice. However,

recordings of dopamine axon activity in the DMS and DLS during training revealed that it is in fact the DMS and not the DLS dopamine signal which effectively predicts which mice will become compulsive reward seekers. We then artificially manipulated the DMS dopamine signal and were able to drive the formation of compulsions. Notably, the observed increase in compulsive behavior was not accompanied by an increase in habit formation, suggesting that the formation of habits and compulsions rely on distinct dopaminergic reinforcement signals delivered to different subregions of the striatum.

RESULTS

A random interval, but not random ratio, schedule of reinforcement promotes punishment-resistant reward seeking

We first determined whether training paradigms used to elicit habitual responding also elicit punishment-resistant reward seeking. Previous literature has demonstrated that a random interval (RI60) schedule of reinforcement, but not a random ratio (RR20) schedule, promotes habit (Derusso et al., 2010; Gremel and Costa, 2013; Wiltgen et al., 2012; Yin et al., 2005a). We therefore compared training on RI60 and RR20 reinforcement schedules to assess whether these schedules have differential effects on punishment-resistant reward seeking. After initial magazine training and pre-training on a fixed ratio (FR1) schedule, mice were transitioned to either RI30 or RR10 schedules, and then finally to RI60 or RR20 schedules (Fig. 1A). We performed probes for punishment-resistant reward seeking at two time points: the first after 1-2 days of RI60/RR20 training and the second after 13-14 days of RI60/RR20 training. During the shock probe sessions, nosepokes were accompanied by a ⅓ risk of mild shock (0.2mA, 1s; Fig. 1B). The shock intensity was chosen based on previous studies of compulsive reward seeking (Harada et al., 2019). We verified that this shock intensity is aversive to the mice in a fear conditioning paradigm in which 12 tone-shock pairings were delivered. The next day, mice that had received tone-shock pairings showed increased freezing to the tone compared with mice that had only been exposed to the tone on the previous day (Fig. S1A; unpaired t test, $p<0.01$). To test whether punishment-resistant reward seeking developed in tandem with habit, a subset of mice were also tested at the end of training on an omission probe (Fig. 1A;Yu *et al.*, 2009; Derusso *et al.*, 2010; Rossi and Yin, 2012). In the omission probe, mice were required to withhold nosepokes to receive rewards, reversing the previously learned contingency (Fig. 1C).

We found a main effect of training (RI60 vs RR20) on the number of shocks mice were willing to receive in the shock probes, with RI60-trained mice willing to receive higher numbers of shocks than RR20-trained mice (Bonferroni, $p <0.05$, Fig. 1D). There was also a significant main effect of subject due to high inter-animal variability ($F_{30,30}=6.26$, $p<0.0001$). Nevertheless, by the end of training, RI60-trained mice, as a group, withstood significantly more footshocks than mice trained on an RR20 schedule for an equivalent period of time (Two-way ANOVA, $F_{1,31}=4.46$, $p<0.05$, Fig1D). These results indicate that it is not simply task experience but the specific reinforcement schedule that leads to an increase in punishment-resistant reward seeking.

Animals that were trained on RI60 showed some trend towards longer times to complete the omission probe than animals trained on RR20 (Fig. 1E; $p=0.24$). In both the shock probes and the omission probes, however, RI60-trained animals showed significant individual variability (omission F-test to compare variances, $F_{7,5}=241.6$, $p<0.0001$; shock 1 F test to compare variances $F_{17,7}=9.89$, $p<0.01$; shock 2 F test to compare variances $F_{24,7}= 8.07$, $p<0.01$). When looking at data from individual animals tested on both probes, we found that those who took longer to complete the omission (earn 50 rewards) also withstood a higher number of shocks on the second shock probe ($r=0.914$, $p<0.01$; Fig. 1F). Thus, RI60 schedules not only promote habit formation as

previously reported, they elicit compulsive responding. It is difficult to disentangle these two phenomena as they tend to occur together in the same individuals.

It is not clear what differences between RI60 and RR20 training might elicit differences in the development of habit or compulsion. An analysis of behavior during training showed that RI60-trained mice earned significantly more rewards per training session, on average, than RR20-trained mice despite, as previously reported, showing no significant differences in the rates of nosepoking or the rates of port entries (Fig. S1B-E; Wiltgen *et al.*, 2012; Gremel and Costa, 2013). RR20-trained animals made 24.19±0.8 nosepokes per reward, consistent with the structure of the task, which requires an average of 20 nosepokes per reward. RI60-trained animals, however, made significantly fewer nosepokes per reward - just 12.6±0.46 on average ($p<0.0001$; Fig. 1G). Therefore, these different task structures have different effort demands, which could influence learning trajectories.

Three behavioral phenotypes emerge with extended RI60 training

The large variation in behavior induced by RI60 training caused us to wonder whether individual mice were taking different strategies to "solve" the RI60 task, which could lead to differences in the development of punishment-resistant reward seeking. To analyze whether differences in performance in the shock probes were associated with different behavior in the RI60 task, we divided the RI60-trained mice into three analysis groups based on shock probe performance: "punishment resistant" (PR) mice that tolerated a high level of shocks in both the first and second probes, "delayed punishment resistant" (DPR) mice that increased the number of shocks they would tolerate from the first to the second probe, and "punishment sensitive" (PS) mice that would not tolerate many shocks on either the first or second probes (Fig. 2A; see methods for exact sorting criteria). As expected based on this sorting, a mixed effects analysis showed a significant interaction of phenotype and training time ($F_{11, 78} = 4.325$, $p<0.0001$), with a Bonferroni test revealing that only DPR mice showed a significant difference in shocks received between the early and late shock probes ($p<0.0001$; Fig. 2B).

When mice were divided into these groups, an analysis of behavior in RI60 sessions showed interesting differences (Figs. 2, S2). Both PR and DPR mice had higher rates of nosepoking than PS mice (Fig. 2C; mixed effects analysis, main effect of training stage $F_{4.44,85.64}=4.38$, $p<0.01$, main effect of phenotype $F_{2,21}=18.57$, $p <0.0001$, interaction $F_{26,251}=2.31$, $p<0.001$). However, port entry rates and rates of rewards earned per minute did not differ (Fig. 2D-E). As a result, PR and DPR mice made more nosepokes per reward than PS mice (Fig. 2F; main effect of training stage $F_{4.65,87.86}=2.42$, $p<0.05$, main effect of phenotype $F_{2,21}=22.77$, $p<0.0001$, interaction $F_{24,227}=1.76$, $p<0.05$).

PS mice are more "efficient," minimizing the effort exerted to earn rewards in the RI60 task. In order to achieve such efficiency, a mouse needs to understand when a nosepoke is likely to lead to reward. In the RI60 task, a mouse is unlikely to receive a reward if it has recently received one (on average, it will have to wait 60 seconds before its efforts become fruitful again). PS mice waited an average of 39 ± 3 seconds to resume nosepoking after retrieving a reward, significantly longer than the other groups

(Fig. 2G; Tukey's multiple comparison, PR=21s ± 2, $p<0.01$; DPR=21s ± 2, $p<0.01$). The behavior of PS mice is thus consistent with a model-based approach, in which the relationship between nosepoking behavior and reward delivery is understood, whereas the behavior of PR and DPR mice may not be model-based. As another test of this theory, we looked at whether mice would make additional nosepokes after a rewarded nosepoke, before going to the port to collect their reward. PS mice made significantly fewer "extra" nosepokes before reward retrieval than PR or DPR mice (Fig. 2H; Tukey's multiple comparison, PS=1.56 ± 0.23, PR= 4 ± 0.25, $p<0.01$, DPR= 3.9 ± 1.61, $p<0.05$), which helped maximize the efficiency of their behavior.

Why do only PS mice take this apparent model-based approach? Why would PR and DPR mice expend more effort than necessary to earn rewards? On an RI60 reinforcement schedule, the average interval between available rewards is 60 seconds, but there is variability. To ensure a nosepoke is likely to be rewarded, a mouse should wait to nosepoke, but to "catch" all instances when reward becomes available in the shortest possible amount of time, mice must nosepoke constantly. In other words, if mice are willing to expend the effort, high rates of nosepoking yield rewards at shorter intervals. Indeed, when we looked at the distribution of inter-reward intervals for PR, DPR and PS mice, we found that the inter-reward intervals for PS mice skew longer (Fig. 2I). The average inter-reward interval for PR is 72.55s ± 1.06 and for DPR mice is 74.61s ±1.93, while the average for PS mice is 82.61s ± 0.92 (K-S Test, PR vs PS, $p<0.0001$; DPR vs PS, $p<0.0001$). PR and DPR mice thus are better able to track the probability of reward availability imposed in the task, whereas PS mice have longer lags before they earn available rewards. Additionally, although we did not observe differences in rewards earned per minute on any particular day of RI60 training (Fig. 2E), when the number of rewards earned per session was averaged over all days, we found that PS received slightly fewer rewards (Fig. 2J; Tukey's multiple comparison, PS= 36.89 ± 1.76, PR= 43.99 ±1.49, $p<0.05$). The effect may be subtle since each session was capped at 50 rewards and PR and DPR mice are often at ceiling, nevertheless this difference is indicative of the advantage of the high-nosepoking approach: maximize rewards at any cost. As a whole, these behavioral results demonstrate that even on identical schedules of reinforcement, individual animals use different strategies in reward seeking.

What are the differences between PR and DPR mice? These two groups of mice appear to be taking similar reward seeking strategies in the RI60 task, so to examine differences between them, we looked at even earlier training data from FR1 and RI30 sessions. We observed that PR mice escalate their nosepoking more rapidly than DPR mice, as soon as they enter RI30 training (the first RI schedule they encounter; Fig. 2K; t-test of slopes, PR vs PS in RI30, $p<0.001$; PR vs DPR in RI60, $p<0.01$; DPR vs PS in RI60, $p<0.0001$). Thus, the timing of nosepoke escalation is a key behavioral predictor of mice who will show punishment-resistant reward seeking when probed.

Dopamine signals in the dorsal striatum at the time of a rewarded nosepoke differ across training and by phenotype

To better understand the neural circuits mediating differences in reward seeking behavior and the development of habitual or compulsive responding, we recorded the

activity of dopamine axons in the dorsal striatum during RI60 training. RI60 training is instrumental learning, which has been previously associated with the function of dorsal striatal subregions. In particular, the dorsomedial striatum (DMS) is required for goal-directed learning of action-outcome associations, whereas the dorsolateral striatum (DLS) is required for habit formation (Yin et al., 2004, 2005b, 2005a). Dopaminergic projections to the DMS and DLS are distinct, meaning that dopamine-mediated reinforcement learning can be separately effectuated in these two areas (Ikemoto, 2007; Lerner et al., 2015). We reasoned that by examining the activity of dopaminergic projections to the DMS and DLS during RI60 training, we could assess whether different dopamine signals were contributing to different aspects of behavior (perhaps allowing us to observe which parts of the signal might predict the development of compulsive or habitual responding). To record the activity of dopamine axons in the DMS and DLS in freely moving mice, we injected an adeno-associated virus (AAV) expressing cre-dependent GCaMP7b (AAV5-CAG-FLEX-jGCaMP7b-WPRE) into the substantia nigra pars compacta (SNc) of DAT-IRES-cre mice. We then implanted fiber optic probes above the DMS and DLS to record the activity of dopaminergic axons (Fig. 3A). Dopaminergic axon activity was recorded simultaneously in the DMS and DLS in all mice. Thus, we could be certain that any differences we observed in activity in the two regions were not due to differences in behavior between groups of mice.

To analyze this data, we first examined DMS dopamine activity occurring during rewarded and unrewarded nosepokes. We histologically verified that GCaMP was expressed in dopaminergic (TH+) neurons in the SNc and found that on average 81.63% of neurons expressing GCaMP also expressed TH (Fig. 3B). We also confirmed that probe locations in the DMS were correct (Fig. 3C, S3A), and histology verified robust GCaMP expression in dopaminergic axons in the DMS (Fig. 3C, inset). We compared DMS dopamine axon activity in RI60-trained mice (PR, DPR, PS) and in RR20-trained mice (RR20; Fig 3D-K). Peaks in DMS dopamine axon activity at the time of a rewarded nosepoke were much clearer in PR and DPR mice than in PS or RR mice, an observation that was not true simply due to poor signal in PS or RR20 mice, as all mice included in the analysis had similar frequencies and amplitudes of GCaMP events (Fig. S3B-C). We quantified the amplitude of DMS signal peak for rewarded nosepokes and found a significant main effect of phenotype/schedule of reinforcement ($F_{3, 18} = 4.742$, $p<0.05$). A Tukey's multiple comparison test showed that late in training, PR mice had significantly larger peaks in DMS at the time of a rewarded nosepoke than PS mice, and that DPR mice had significantly larger peaks than PS and RR mice (Fig. 3L, $p<0.05$ for each comparison). DMS dopamine signals in response to rewarded nosepokes were clearly different from unrewarded nosepokes. Mice were able to detect the rewarded nosepokes because of an audible "click" of the pellet dispenser.

The DMS dopamine signal at the time of an unrewarded nosepoke showed a clear time-locked negative peak, or "dip" under certain conditions. More compulsive phenotypes had a prominent dip (Fig. 3M). This dip in particular was only evident in PR mice early in training, but developed later in training for other groups (DPR and PS, but not RR). The combination of a positive peak in DMS dopamine activity for rewarded nosepokes and a negative peak in DMS dopamine activity for unrewarded nosepokes creates a notable difference in the dopamine levels experienced by the mouse in response to the same motor action depending on the outcome, which is information that

could be used to influence future behavior. We found a significant correlation between the amplitude of peaks in DMS at the time of a rewarded nosepoke and the amplitude of peaks in DMS at the time of an unrewarded nosepoke. In effect, if a mouse had a strong positive peak in DMS dopamine activity for rewarded nosepokes it also had a negative peak in activity for unrewarded nosepokes (Fig. 3N, $r=-0.45$, $p<0.01$).

In recordings from the same mice, we next examined DLS dopamine axon activity occurring during nosepokes to see if it differed from activity in the DMS (Fig. 4A). Similarly, we verified the expression of virus in SNc (Fig. 4B; 86.71% of GcaMP neurons also expressed TH) and the probe placements in DLS (Fig. 4C, S4A). We also noted that the DLS signals in all groups of mice had similar frequencies and amplitudes of GCaMP events (Fig. S4B-C). There appeared to be several salient features of the DLS signal at the time of a nosepoke that varied across training and by phenotype (Fig. 4D-K). First, we noted immediate peaks in the signal following a rewarded nosepoke (Fig. 4L). A mixed-effect analysis revealed a significant interaction of training and phenotype on the amplitude of the peak at the time of a rewarded nosepoke ($F_{3, 8}=4.229$, $p<0.05$). Second, we noted that, as in the DMS signal, in some cases there was a negative peak in the DLS signal following an unrewarded nosepoke (Fig. 4M). These dips occurred early in training in PR mice and late in training in both PR and DPR mice, suggesting that they have some association with the compulsive phenotype. However, RR20-trained mice also showed these dips but were not compulsive. In addition to peaks, we noted that many mice showed a prolonged elevation of DLS dopamine activity above baseline, often for 10 seconds or more. We quantified this prolonged activity as the area under the curve (AUC; see Methods). While AUC did not differ statistically across groups or training (Fig. 4N), there was a significant correlation between the DLS dopamine signal peak at the time of a rewarded nosepoke and the following AUC of the signal (Fig. 4O; $r=0.64$, $p<0.0001$).

Ramping of nigrostriatal dopamine signals at the time of reward retrieval is altered with increasing punishment resistance

Next, we examined the DMS and DLS dopamine axon signals surrounding the time of a rewarded port entry and noticed ramping activity beginning before the port entry, with peaks occurring either before or after port entry in different cases. Differences in this ramping between the DMS and DLS signals appeared to vary with stage of training and by phenotype (Fig. 5A-H). In order to quantify differences in peak timing, we found the time at which the dopamine signal peaked relative to the time of the port entry. Since only signals from PR and DPR mice had clear port entry-related peaks, only these signals were analyzed (Fig. 5I-J). In PR mice, DLS peaks occurred later than DMS peaks during early RI60 training (Bonferroni, $p<0.05$), but the difference was not significant after extended RI60 training. In DPR mice, we found a significant interaction between subregion and training stage ($F_{1,4}=8.33$, $p<0.05$), again with DLS peaks occurring significantly later than DMS peaks in early training ($p<0.01$) and DLS peaks becoming significantly closer to the time of the rewarded port entry between early and late training ($p<0.05$).

To analyze differences in the kinetics of the DMS and DLS signals as they ramped towards rewarded port entries, we also calculated the slope of the signals in

each subregion (Fig. 5K-L; see methods) and found that PR mice showed a main effect of subregion on slope ($F_{1,10}$=8.82, $p<0.05$) with slopes being significantly steeper in DMS than DLS early in training ($p<0.01$). DPR mice showed a similar trend, but no significant differences in slope across subregion or by training stage. This analysis suggests that the dopamine signal in DMS ramps more quickly than DLS, particularly early in training, for those animals which show, or go on to show, compulsive behavior. On the other hand, DLS signals seem to be more reactive to the reward retrieval following a port entry. In contrast to rewarded port entries, relatively little dopamine axon activity was observed in either striatal subregion following unrewarded port entries. Pre-port entry ramping was not apparent and in some cases dips in activity were observed, although these were inconsistent (Fig. S5).

DMS but not DLS dopamine signaling predicts punishment-resistant reward seeking

While there were numerous distinct features of activity in both DMS and DLS dopamine axons that could be interesting, we wanted to see which of these, if any, could predict an animal's resistance to punishment. We examined the correlation of these different features with the number of shocks an animal withstood on the shock probe that occurred the day after the signals were recorded. Both the DMS peak signal at the time of a rewarded nosepoke (Fig. 6A; $r= 0.5$, $p<0.001$) and DMS peak (positive or negative) signal at the time of an unrewarded nosepoke (Fig. 6B; $r=-0.4$, $p<0.05$) showed significant correlations with the number of shocks received. Conversely, neither the DLS peak signal at the time of a rewarded nosepoke (Fig. 6C) nor the AUC of this signal (Fig. 6D) were significantly correlated with the number of shocks received. This result was surprising since it is DLS dopamine which is classically linked to habit formation (Faure et al., 2005; Willuhn et al., 2012), and we had observed that punishment-resistant reward seeking in the shock probe and habit were strongly correlated (Fig. 1F). However, habit and compulsion are not the same, and we did not observe significant correlations between omission time and photometry signals (Fig. S6). We were intrigued by the idea that DMS dopamine might specifically relate to compulsion rather than to habit. Therefore, we designed an optogenetic experiment to test a causal relationship between DMS dopamine signaling and punishment-resistant rewarding seeking in our task.

Optogenetic excitation of dopamine terminals in DMS at the time of a rewarded nosepoke causes an increase in 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 decided to test if stimulation of DMS dopamine axons at the time of a rewarded nosepoke could cause punishment resistant behavior to emerge. We used the excitatory opsin ChR2 to stimulate DMS dopamine axons. 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 the DMS to allow light stimulation of dopamine terminals only within the DMS (Fig. 7A).

Previous work has shown that DMS-projecting dopamine neurons have minimal collateralization outside the DMS (Lerner et al., 2015), so this stimulation should be specific even if back-propagating action potentials are generated. We verified that this strategy led to the expression of ChR2 and EYFP in dopamine (TH+) neurons (Fig. 7B and C) and that fiber optics were appropriately placed in the DMS, matching the coordinates used for fiber photometry recordings (Fig. S7A).

Beginning with FR1 training, ChR2 animals received 1s, 20Hz trains of light stimulation on every rewarded nosepoke (Fig. 7D-E). Importantly, stimulation of DMS dopamine terminals was given during FR1/RI30/RI60 training but not during shock probes when punishment-resistant reward seeking was assessed. Therefore, any effects of the stimulation on probe performance are not due to acute effects of dopamine terminal stimulation, but are caused by differences in learning during the training sessions. In addition to animals expressing ChR2 and receiving stimulation on rewarded nosepokes, there were two control groups: EYFP controls that received a fluorophore-only control virus injected into the SNc (AAV5-EF1α-DIO-EYFP) and the same pattern of light stimulation, and "scrambled" controls that received the ChR2 virus, but had their DMS dopamine terminals stimulated on a random subset of nosepokes rather than on rewarded nosepokes. The scrambled control was important as these mice received the same amount of dopamine terminal stimulation as the ChR2 group and this stimulation reinforced the same action (a nosepoke). Therefore, the only difference between the ChR2 and ChR2 Scrambled groups was whether or not the dopamine terminal stimulation they received boosted vs degraded the ability of the natural DMS dopamine signal to differentiate between externally rewarded and unrewarded actions.

All three groups (ChR2, EYFP, and ChR2 Scrambled) learned and performed the RI60 task and received approximately the same number of rewards during RI60 sessions (Fig. S7B). However, ChR2 animals escalated their nosepoking much faster, from very early in training (Fig 7F; t-test of slopes, ChR2 vs EYFP in RI30, $p<0.01$). When tested on the first shock probe after early RI60 training, ChR2 mice were significantly more resistant to punishment than either EYFP or ChR2 Scrambled mice (Fig. 7G; Tukey's multiple comparison, $p<0.05$). Significant differences faded with more extended training as compulsive behavior also began to emerge naturally in the control groups. We categorized mice from this experiment as PR, DPR and PS based on our previously developed definitions and compared the fraction of mice in each category to the distribution expected based on results from our GCaMP-expressing cohort using chi-squared tests. We found that the ChR2 group showed higher than expected numbers of PR mice, with 86% of the group meeting the criteria for PR ($X^2_{(2,12)}=6.9$, $p<0.01$; Fig. S7C).

The finding that DMS dopamine terminal stimulation on rewarded nosepokes can promote the early development of punishment-resistant reward seeking was exciting, but confusing. We had observed in our previous GCaMP-expressing (but not causally manipulated) cohorts that compulsion and habit were well correlated. Yet DMS dopamine terminal stimulation is not expected to promote habit. Therefore, we also looked to see whether DMS terminal stimulation influenced performance on a final omission test after extended RI60 training. It did not. ChR2, EYFP, and ChR2 Scrambled mice all took the same amount of time to complete the omission probe (Fig.

7H). Although the correlation between shocks received in the shock probe and time to complete the omission was preserved in EYFP mice (which were essentially unmanipulated), the correlation was broken in ChR2 and ChR2 Scrambled mice (Fig. 7I-K). These final data emphasize that while compulsion and habit may often track together (e.g. during unmanipulated RI60 training), they are not inextricably linked. DMS dopamine plays a role in controlling the development of compulsions, but does not influence habit formation, and effectively unlinks these two processes.

DISCUSSION

Both habits and compulsions are hypothesized to contribute to disorders such as addiction and obsessive-compulsive disorder (OCD), however it has been unclear how these two distinct aspects of inflexible behavior are related to each other, if at all. Previous studies have shown a role for nigrostriatal dopamine signaling in habit formation (Faure et al., 2005; Murray et al., 2012; Willuhn et al., 2012), but less is understood about whether nigrostriatal dopamine is also involved in compulsive behavior (for review see Lerner, 2020; Lipton et al., 2019). Here, we addressed this question by using dual-site fiber photometry to record from dopamine terminals in the dorsomedial striatum (DMS) and dorsolateral striatum (DLS) during a task (RI60) that promotes both compulsions and habits. We identified task-relevant increases in DMS dopamine terminal activity as a key feature of neural circuit activity predicting which individual animals would display compulsive behavior. Both DMS dopamine peaks for rewarded nosepokes and DMS dopamine dips for unrewarded nosepokes were correlated with the development of compulsion. Therefore, we hypothesized that it is the discrimination of rewarded vs unrewarded actions by DMS dopamine -- not the general dopamine-mediated reinforcement of the nosepoking action -- that drives compulsion. To test this hypothesis, we optogenetically stimulated dopamine terminals in the DMS during learning. By creating peaks of DMS dopamine axon activity on rewarded nosepokes, we increased compulsive, but not habitual behavior. This result supports a model in which the DMS dopamine signal is specific to compulsion. It also suggests that the normal association of compulsive and habitual responding after extended RI60 training is due to a common upstream driver of DMS and DLS function and not to a promotion of habitual responding through DMS circuits. While we note that this finding is in conflict with the commonly cited “Ascending Spiral” hypothesis (Haber et al., 2000; Yin and Knowlton, 2006), we also recognize that it was not a direct test. Further investigations of the Ascending Spiral Hypothesis as it relates to the association of DMS and DLS function are necessary.

As a control for our optogenetics experiment, we included a group of mice that received DMS dopamine terminal stimulation on random nosepokes. Creating peaks in DMS dopamine on random nosepokes did not have the same effect as creating these peaks on rewarded nosepokes. If anything, this pattern of stimulation, which *disrupts* the discrimination between rewarded and unrewarded actions, inhibits the emergence of compulsion. The random stimulation data indicate that the pattern of dopamine activity in the DMS matters, but it remains to be determined *why* the pattern is important. For example, future studies could be designed to examine whether the pattern of cortical inputs to the DMS is different during rewarded vs unrewarded nosepokes. If different cortical inputs to the DMS are active during rewarded vs unrewarded nosepokes, dopamine release at these distinct times would reinforce the strength of different corticostriatal synapses.

We were surprised to see that DMS dopamine signaling promotes compulsive responding. Nevertheless, this observation fits with certain aspects of the previous literature. First, optogenetic stimulation of direct-pathway DMS neurons is reinforcing (Kravitz et al., 2012). The reinforcing effect persists even on a day of extinction training, indicating that the activation of DMS circuits is linked to inflexible behavior. Second,

stimulation of SNc dopamine neurons, which project to DMS and DLS, supports reinforcement (Corbett and Wise, 1980; Ilango et al., 2014; Keiflin et al., 2017; Reynolds et al., 2001; Rossi et al., 2013; Saunders et al., 2018; Wise, 1981). However, unlike ventral tegmental area (VTA) dopamine neurons, which support reinforcement by adding incentive salience to a reward-predicting cue (Robinson and Berridge, 1993), SNc dopamine neurons support reinforcement via the invigoration of movement (Crow, 1972; Saunders et al., 2018). Thus, DMS dopamine signaling is perfectly situated to translate histories of rewarded actions into the compulsive performance of those actions. Third, changes in DMS circuit function have been implicated in stereotyped, perhaps compulsive, grooming behaviors (Cromwell and Berridge, 1996; Kalueff et al., 2016). Sapap3$^{-/-}$ mice, a model of OCD that displays overgrooming, are more likely to form habits after training on RI60 (Burguiere et al., 2013; Hadjas et al., 2019), although it is unknown how Sapap3$^{-/-}$ mice would perform on a test for compulsion as in our study design. The answer is one worthy of future investigation as it could shed light on the neural circuit mechanisms of OCD-like compulsions.

In this study, we examined the behavior of mice during learning to pursue sucrose rewards. One previous study also examined the circuits underlying compulsive sucrose-seeking (Nieh et al., 2015). This study found a projection from the lateral hypothalamus (LH) to the VTA could bidirectionally control compulsive sucrose seeking. We do not know how this LH-VTA circuit might interact with DMS dopamine signaling, however, there are several possibilities. Most simply, some VTA dopamine neurons project to the DMS (Beier et al., 2015; Howe and Dombeck, 2016). Additionally, DMS-projecting dopamine neurons receive inputs from LH, VTA and NAc, any of which could easily interconnect the circuitry (Lerner et al., 2015).

Compulsive sucrose-seeking is relatively understudied, while many more previous studies have examined compulsive drug-seeking. Large individual variability in compulsivity has been observed in animals working for drugs such as cocaine and alcohol (Giuliano et al., 2019; Siciliano et al., 2019). Consistent with those findings, we also found large individual differences in compulsivity. We identified one reason for this variability: the different strategies used by individual animals to deal with uncertainty in the RI60 schedule. In our study, some mice maximized their rate of reward retrieval by nosepoking at high rates, while other mice conserved effort. Another interesting question for future research is whether mice working for cocaine or alcohol would show similar divisions of strategy. Not all mice (or people) who try drugs become compulsive. While the circuit hallmark of compulsive sucrose-seeking we have identified is exciting, it is far from clear whether it would also be a hallmark of compulsive drug-seeking. Conversely, it is unclear whether previously identified circuits for compulsive drug-seeking impact compulsive sucrose-seeking. In fact, one behavioral study directly examined whether the development of compulsive drug-seeking and compulsive sucrose-seeking were correlated and found that they were NOT: animals that became compulsive for one type of reward did not necessarily become compulsive for the other type (Datta et al., 2018). Thus, it is imperative to examine in more detail whether all rewards do or do not activate the same circuits for compulsivity.

Finally, further studies are needed to more explicitly examine the relationship between DLS dopamine signals and habit. Previous approaches using lesions and pharmacology are suggestive of a relationship, but do not provide temporal specificity.

Here, we observed some novel temporal dynamics in the DLS dopamine signal, which differentiate it from the signal in the DMS. The DLS dopamine signal has an immediate peak following a rewarded nosepoke as well as prolonged activity above baseline. We did not find strong correlations of this signal with individual behavior, although there was a significant overall training effect. Alternative tests for habit, such as outcome devaluation tests, may reveal a stronger association.

In summary, we have identified DMS dopamine signaling as a key part of the circuitry that drives compulsion. The data presented here set the stage for interesting new studies addressing questions about how DMS dopamine participates in the development of compulsions, both for natural reward-seeking and in the context of drug addiction and OCD.

ACKNOWLEDGEMENTS

We thank members of the Lerner laboratory for helpful discussions and critical feedback on the manuscript. We thank G. Palissery, A. Bridgemohan, and M. Holla for assistance with surgeries, histology, data collection and entry. pGP-AAV-CAG-FLEX-jGCaMP7b-WPRE was a gift from Douglas Kim & GENIE Project. AAV5-EF1α-DIO-hChR2(H134R)-EYFP was a gift from Karl Deisseroth. This work was supported by an NIH K99/R00 Award (R00MH109569) and a NARSAD Young Investigator Award from the Brain & Behavior Research Foundation to T.N.L., an NIH Diversity Supplement (R00MH109569-04S1) to support C.V.C.

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, Software, Validation, and Project Administration, J.L.S., C.V.C., T.N.L.; Investigation, J.L.S. and C.V.C.; Data Curation and Visualization, J.L.S., J.M.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.

Figure 1

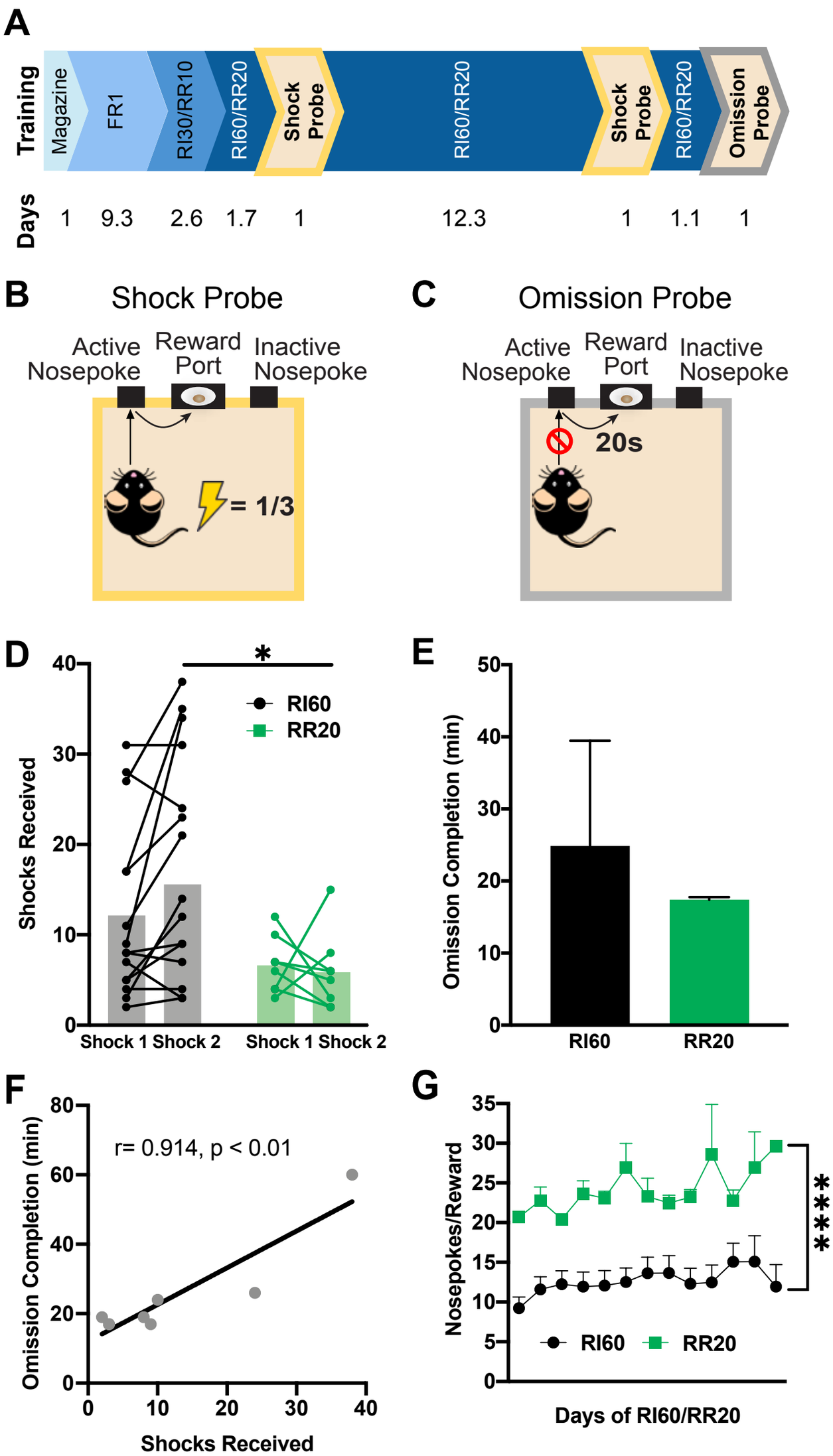

A
Training
Magazine
FR1
RI30/RR10
RI60/RR20
Shock Probe
RI60/RR20
Shock Probe
RI60/RR20
Omission Probe
Days
1
9.3
2.6
1.7
1
12.3
1
1.1
1
B
Shock Probe
Active Nosepoke
Reward Port
Inactive Nosepoke
= 1/3
C
Omission Probe
Active Nosepoke
Reward Port
Inactive Nosepoke
20s
D
Shocks Received
RI60
RR20
Shock 1 Shock 2
Shock 1 Shock 2
E
Omission Completion (min)
RI60
RR20
F
Omission Completion (min)
r= 0.914, p < 0.01
Shocks Received
G
Nosepokes/Reward
RI60
RR20
Days of RI60/RR20

FIGURE LEGENDS

Figure 1. A random interval, but not random ratio, schedule of reinforcement promotes both compulsive and habitual responding.
A. Timeline of operant training and probes. Animals progress from a fixed ratio to a random ratio or random interval schedule of reinforcement. A shock probe occurred both early and late in RI60/RR20 training and an omission probe occurred at the end of training. Average number of days in each stage of training is given below. **B.** Schematic of shock probe: active nosepokes had a ⅓ probability of incurring a shock consequence. **C.** Schematic of omission probe: rewards were delivered when mice refrained from nosepoking at the previously active port for 20s. **D.** Shocks received on early and late shock probes for RI60-trained (black; n=24) and RR20-trained (green; n=8) mice. Bars represent mean; points represent individual scores (connected with lines). *$p<0.05$. **E.** Average time to complete the omission probe (earn 50 rewards; max 60 minutes) for RI60-trained (black; n=7) and RR20-trained (green; n=7) mice. Error bars represent SEM. **F.** Correlation between shocks received on the late shock probe and omission completion time for all RI60-trained mice tested in both probes ($r=0.914$, $p<0.005$). **G.** Average nosepokes per reward across days of RI60/RR20 training. Error bars represent SEM. ****$p<0.0001$.

Figure 2

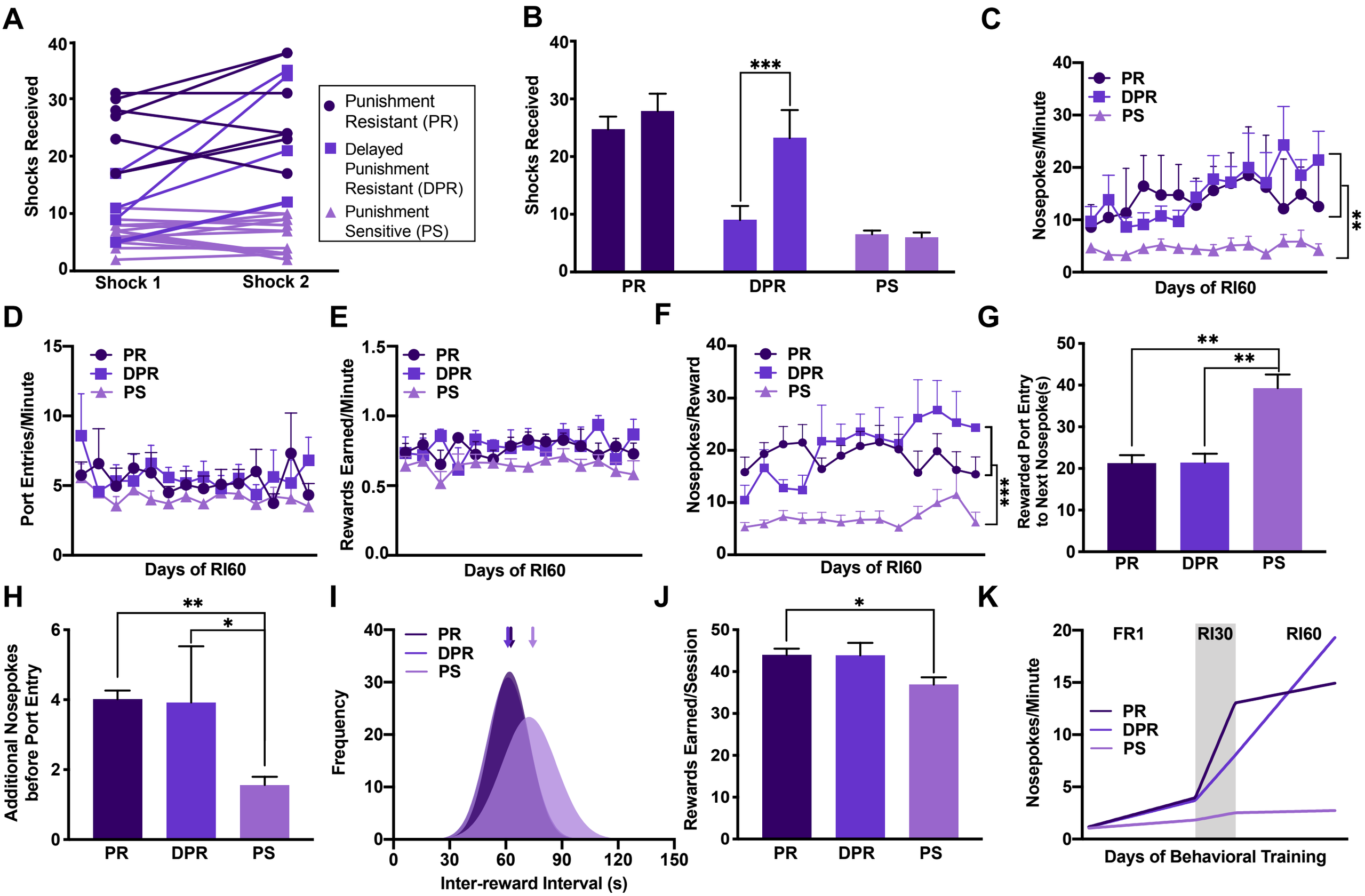

A
Shocks Received
Shock 1
Shock 2
Punishment Resistant (PR)
Delayed Punishment Resistant (DPR)
Punishment Sensitive (PS)
B
Shocks Received
***
PR
DPR
PS
C
Nosepokes/Minute
PR
DPR
PS
Days of RI60
**
D
Port Entries/Minute
Days of RI60
E
Rewards Earned/Minute
Days of RI60
F
Nosepokes/Reward
Days of RI60
***
G
Rewarded Port Entry to Next Nosepoke(s)
**
**
PR
DPR
PS
H
Additional Nosepokes before Port Entry
**
*
PR
DPR
PS
I
Frequency
Inter-reward Interval (s)
J
Rewards Earned/Session
*
PR
DPR
PS
K
Nosepokes/Minute
FR1
RI30
RI60
Days of Behavioral Training

Figure 2. Three behavioral phenotypes emerge with extended RI60 training.
A. Shocks received by mice trained under RI60 only (same data as shown in Fig. 1D). Mice were classified as punishment resistant (PR; dark purple), delayed punishment resistant (DPR; medium purple), or punishment sensitive (PS; light purple) based on the number of shocks they received during the early and late shock probes (see Methods). **B.** Average shocks received on early and late shock probes for each phenotype. Error bars represent SEM. ***$p<0.001$. **C.** Average nosepokes per minute across days of RI60 training. Error bars represent SEM **$p<0.01$. **D.** Average number of port entries per minute across days of RI60 training. Error bars represent SEM. **E.** Average number of rewards earned per minute across days of RI60 training. Error bars represent SEM. **F.** Average nosepokes made per reward across days of RI60 training. Error bars represent SEM ***$p<0.001$. **G.** Average time from a rewarded port entry to the next nosepoke. Error bars represent SEM **$p<0.01$. **H.** Average unrewarded nosepokes made following a rewarded nosepoke, prior to a rewarded port entry. Error bars represent SEM *$p<0.05$, **$p<0.01$. **I.** Distribution of inter-reward interval times in seconds for each group. **J.** Average rewards earned per day of RI60 training. Error bars represent SEM *$p<0.05$. **K.** Slope of nosepokes made per minute in FR1, RI30, and RI60 schedules. See also Supplemental Figure 2.

Figure 3

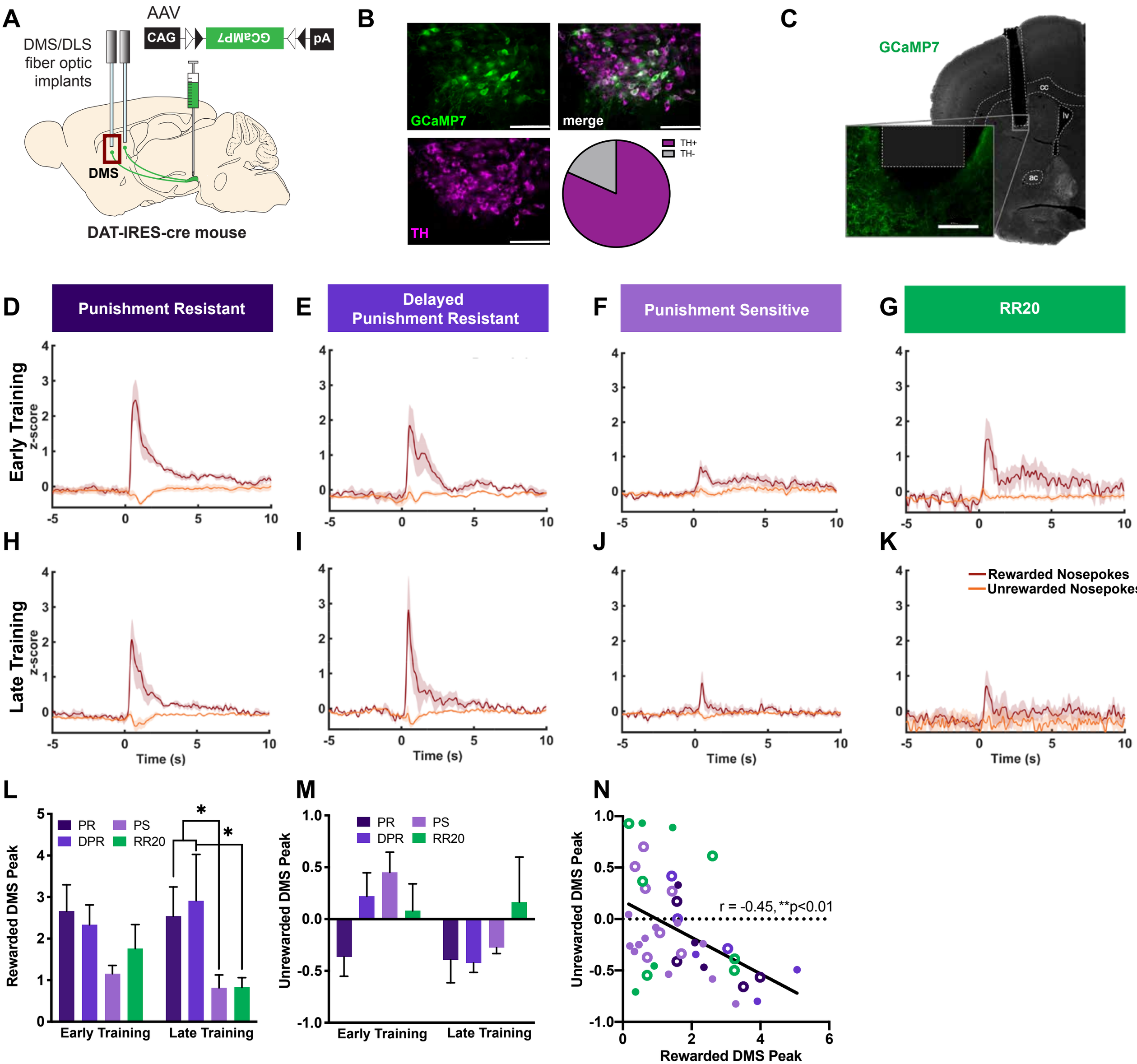

A
AAV
CAG
GCaMP7
pA
DMS/DLS
fiber optic
implants
DMS
DAT-IRES-cre mouse
B
GCaMP7
merge
TH
TH+
TH-
C
GCaMP7
cc
lv
ac
D
Punishment Resistant
E
Delayed
Punishment Resistant
F
Punishment Sensitive
G
RR20
Early Training
Late Training
z-score
Time (s)
Rewarded Nosepokes
Unrewarded Nosepokes
L
M
N
PR
DPR
PS
RR20
Rewarded DMS Peak
Unrewarded DMS Peak
Early Training
Late Training
r = -0.45, **p<0.01

Figure 3. Dopamine signals in the dorsomedial striatum at the time of a rewarded nosepoke differ across training and by phenotype. **A.** Viral injection and probe placement strategy for recording from dopamine terminals in the dorsomedial striatum. **B.** High magnification (40x) images of SNc showing GCaMP7b expression in green, TH positive cells in magenta, and the merged image. Scale bars are 100µm. Quantification of GCaMP7b-expressing cells that are TH+ is shown; n=492 cells. **C.** Representative image showing probe placement in DMS. Area of magnification shows GCaMP7b expression in dopaminergic axons near the probe site in green. Scale bar is 100µm. **D-G.** Peri-stimulus time histograms (PSTHs) showing the average signal from DMS at the time of rewarded (red) and unrewarded (orange) nosepokes early in training (day one or two). Shaded region represents SEM. **H-K.** PSTHs of the average signal from DMS at the time of rewarded (red) and unrewarded (orange) nosepokes late in training (day eleven or twelve). Shaded region represents SEM. **L.** Quantification of average peak signal in DMS at the time of a rewarded nosepoke. Error bars represent SEM. *$p<0.05$. **M.** Quantification of average peak signal (positive or negative) in DMS at the time of an unrewarded nosepoke. Error bars represent SEM. **N.** Correlation of the peak DMS signal at the time of an unrewarded vs rewarded nosepoke in individual mice. Open circles represent correlations from early training recordings, closed circles represent correlations from late training recordings. See also Supplemental Figure 3.

Figure 4

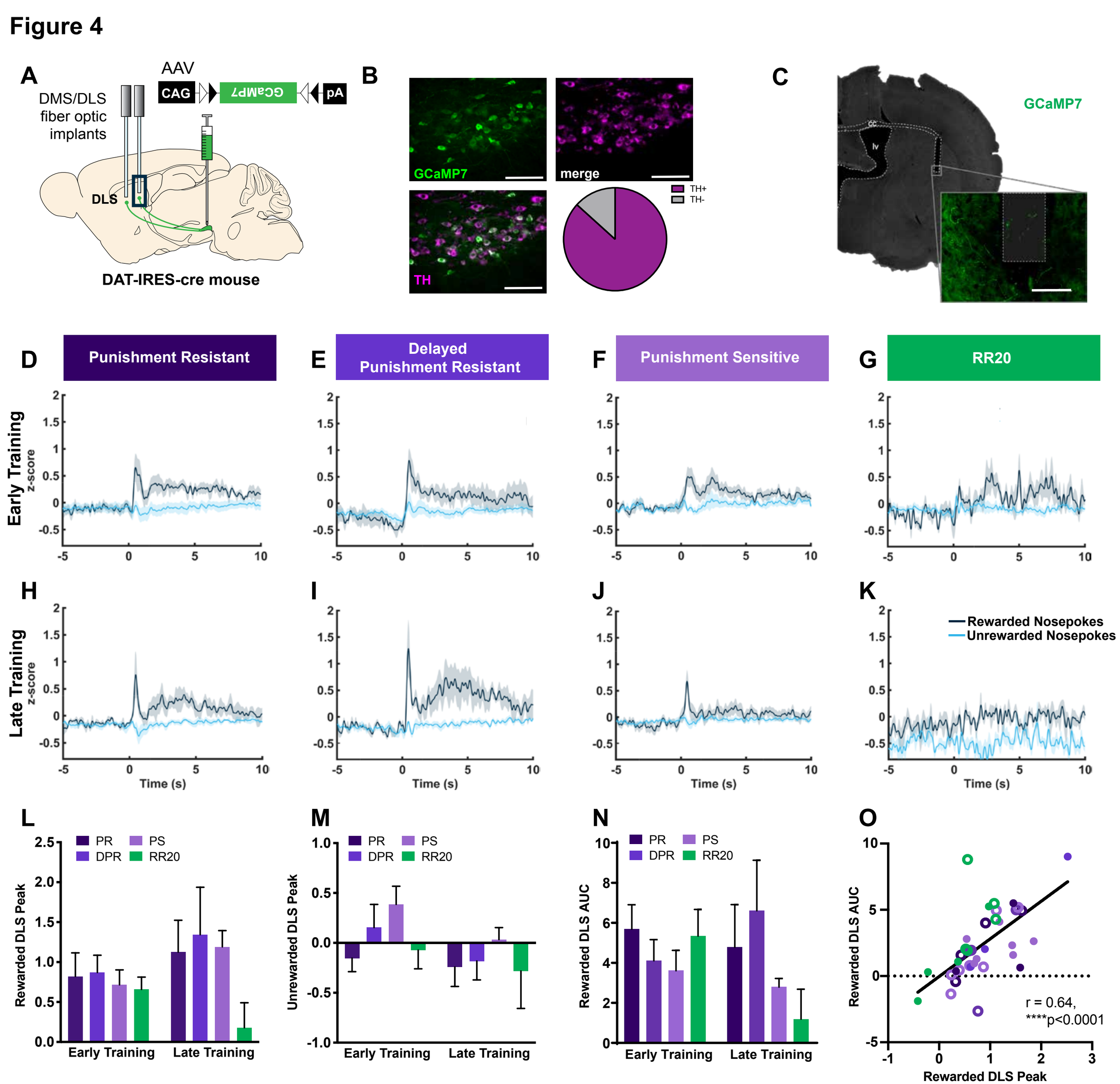

A
AAV
CAG
GCaMP7
pA
DMS/DLS
fiber optic
implants
DLS
DAT-IRES-cre mouse
B
GCaMP7
merge
TH
TH+
TH-
C
GCaMP7
cc
lv
D
Punishment Resistant
E
Delayed
Punishment Resistant
F
Punishment Sensitive
G
RR20
Early Training
z-score
Late Training
z-score
Time (s)
Rewarded Nosepokes
Unrewarded Nosepokes
L
PR
PS
DPR
RR20
Rewarded DLS Peak
Early Training
Late Training
M
Unrewarded DLS Peak
N
Rewarded DLS AUC
O
Rewarded DLS AUC
Rewarded DLS Peak
r = 0.64,
****p<0.0001

Figure 4. Dopamine signals in the dorsolateral striatum at the time of a rewarded nosepoke differ across training and by phenotype. **A.** Viral injection and probe placement strategy for recording from dopamine terminals in the dorsolateral striatum. **B.** High magnification (40x) images of SNc showing GCaMP7b expression in green, TH positive cells in magenta, and the merged image. Scale bars are 100µm. Quantification of GCaMP7b-expressing cells that are TH+ is shown; n=132 cells. **C.** Representative image showing probe placement in DLS. Area of magnification shows GCaMP7b expression in dopaminergic axons near the probe site in green. Scale bar is 100µm. **D-G.** Peri-stimulus time histograms (PSTHs) showing the average signal from DLS at the time of rewarded (dark blue) and unrewarded (light blue) nosepokes early in training (day one or two). Shaded region represents SEM. **H-K.** PSTHs of the average signal from DLS at the time of rewarded (dark blue) and unrewarded (light blue) nosepokes late in training (day eleven or twelve). Shaded region represents SEM. **L.** Quantification of average peak signal in DLS at the time of a rewarded nosepoke. Error bars represent SEM. **M.** Quantification of the average peak signal (positive or negative) in DLS at the time of an unrewarded nosepoke. Error bars represent SEM. **N.** Quantification of average area under the curve (AUC) of the DLS signal following a rewarded nosepoke divided by 1000. Error bars represent SEM. **O.** Correlation of AUC of the DLS signal divided by 1000 following a rewarded nosepoke and the peak DLS signal at the time of a rewarded nosepoke in individual mice. Open circles represent correlations from early training recordings, closed circles represent correlations from late training recordings. ****$p<0.0001$. See also Supplemental Figure 4.

Figure 5

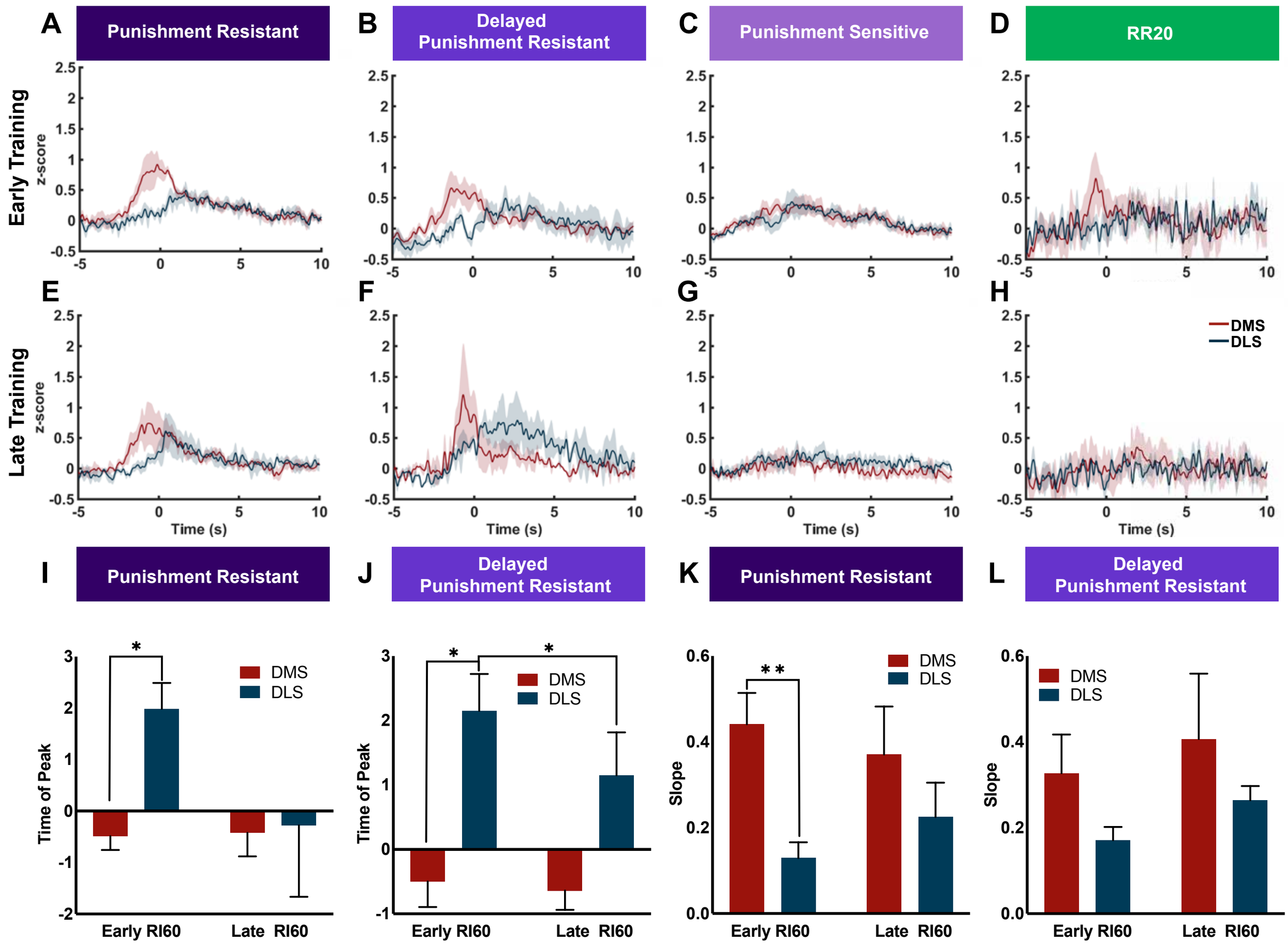
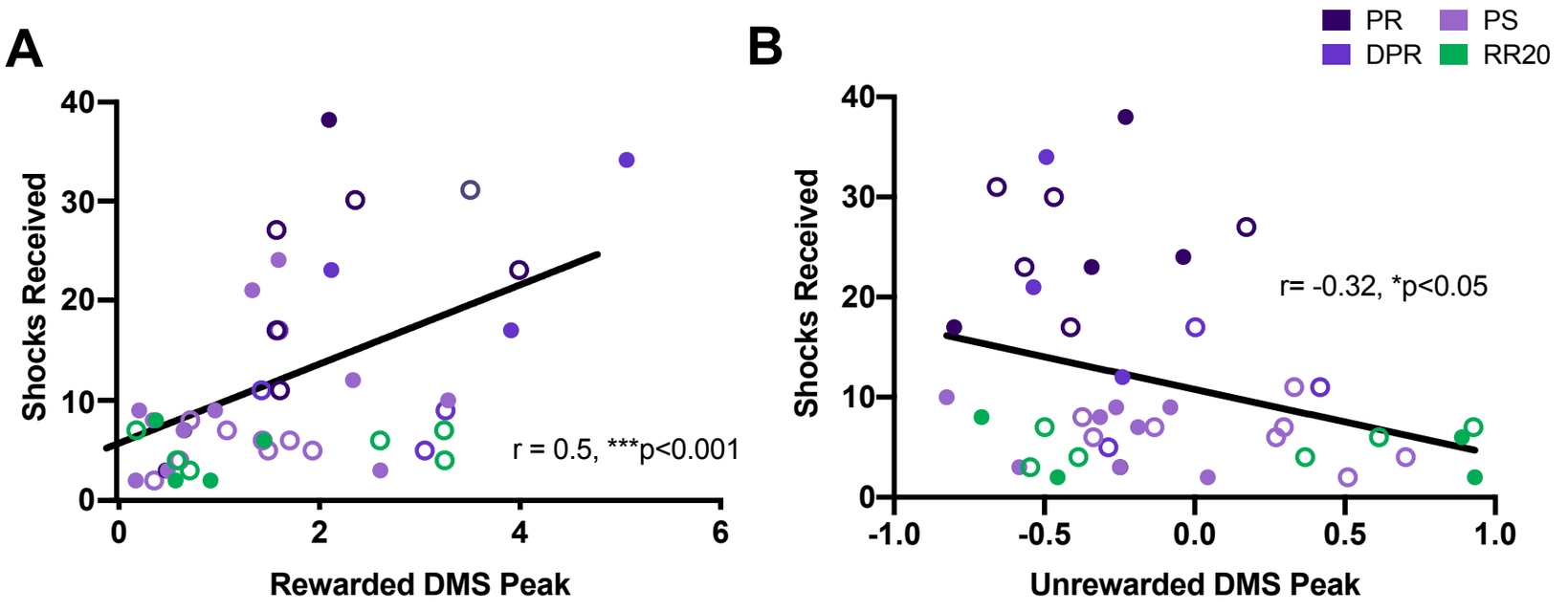

A
Punishment Resistant
B
Delayed Punishment Resistant
C
Punishment Sensitive
D
RR20
Early Training
Late Training
z-score
Time (s)
E
F
G
H
DMS
DLS
I
Punishment Resistant
J
Delayed Punishment Resistant
K
Punishment Resistant
L
Delayed Punishment Resistant
Time of Peak
Slope
Early RI60
Late RI60

Figure 5. Ramping of nigrostriatal dopamine signals at the time of reward retrieval is altered with increasing punishment resistance. A-D. Peri-stimulus time histograms (PSTHs) showing the average signal from DMS (red) and DLS (blue) at the time of a rewarded port entry early in training (day one or two). Shaded region represents SEM. **E-H.** PSTHs showing the average signal from DMS (red) and DLS (blue) at the time of a rewarded port entry late in training (day eleven or twelve). Shaded region represents SEM. **I.** Average time of the peak signal in DMS or DLS relative to the time of rewarded port entries early and late in RI60 training for punishment resistant animals. Error bars represent SEM. *$p<0.05$. **J.** Average time of the peak signal in DMS or DLS relative to the time of rewarded port entries early and late in RI60 training for delayed punishment resistant animals. Error bars represent SEM. *$p<0.05$. **K.** Quantification of the slope of the rise to the peak signal in DMS and DLS early and late in training for punishment resistant animals. Error bars represent SEM. **$p<0.01$. **L.** Quantification of the slope of the rise to the peak signal in DMS and DLS early and late in training for delayed punishment resistant animals. Error bars represent SEM. See also Supplemental Figure 5.

Figure 6

DMS Signal Correlations



DLS Signal Correlations

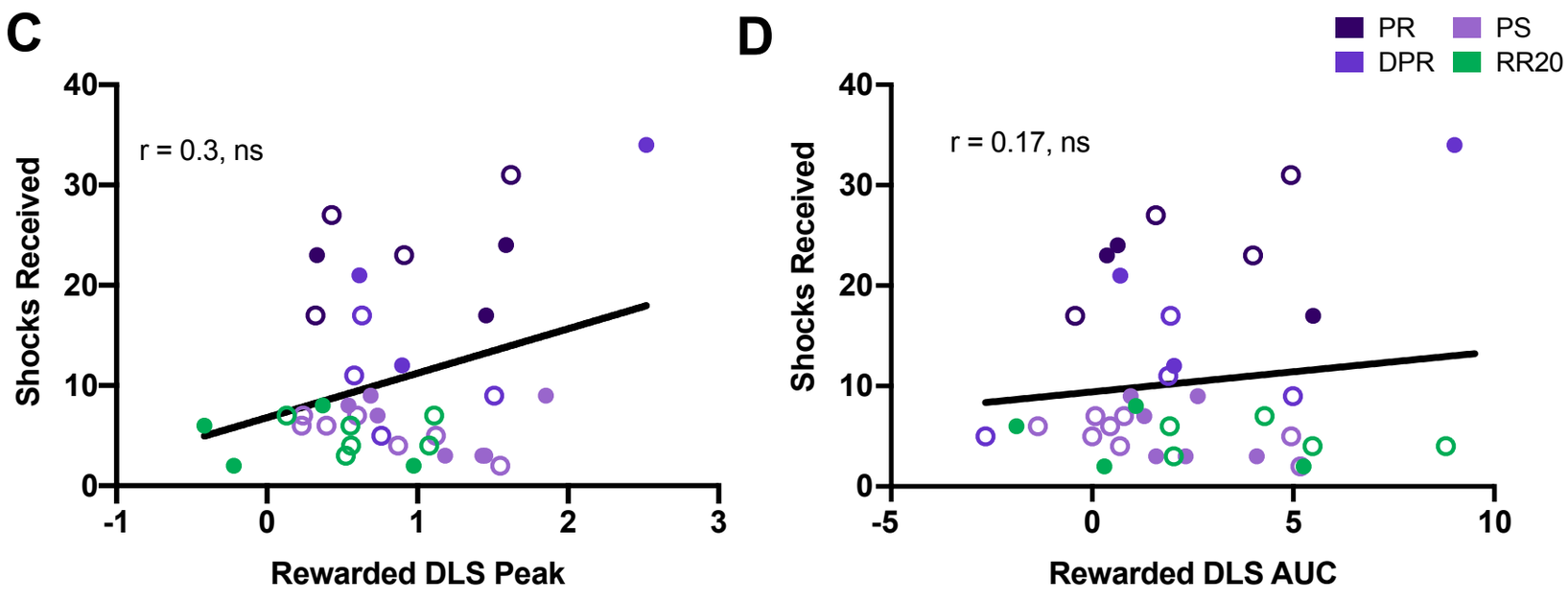


Figure 6. DMS but not DLS dopamine signaling predicts punishment-resistant reward seeking. **A.** Correlation of shocks received and peak signal in DMS at the time of a rewarded nosepoke. ***$p<0.001$. **B.** Correlation of shocks received and peak signal in DMS at the time of an unrewarded nosepoke. *$p<0.05$. **C.** Correlation of shocks received and peak signal in DLS at the time of a rewarded nosepoke. **D.** Correlation of shocks received and the area under the curve for the signal in DLS divided by 1000 following a rewarded nosepoke. Open circles represent correlations between shocks received on the early shock probe and recordings from early training (the day preceding the early shock probe). Closed circles represent correlations between shocks received on the late shock probe and recordings from late training (the day preceding the late shock probe). See also Supplemental Figure 6.

Figure 7

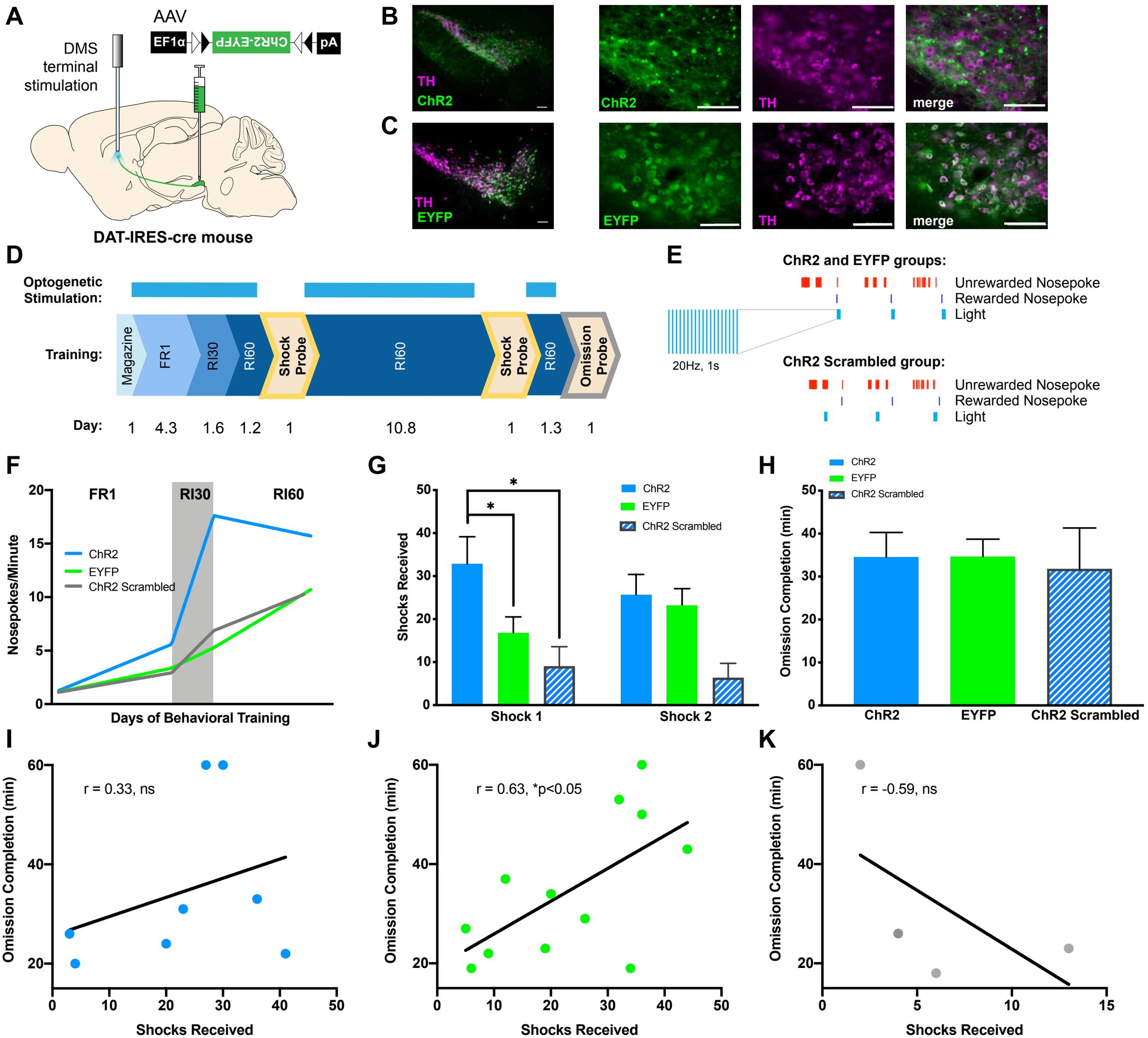

A
AAV
EF1α
ChR2-EYFP
pA
DMS terminal stimulation
DAT-IRES-cre mouse
B
TH
ChR2
ChR2
TH
merge
C
TH
EYFP
EYFP
TH
merge
D
Optogenetic Stimulation:
Training:
Magazine
FR1
RI30
RI60
Shock Probe
RI60
Shock Probe
RI60
Omission Probe
Day: 1 4.3 1.6 1.2 1 10.8 1 1.3 1
E
ChR2 and EYFP groups:
Unrewarded Nosepoke
Rewarded Nosepoke
Light
20Hz, 1s
ChR2 Scrambled group:
Unrewarded Nosepoke
Rewarded Nosepoke
Light
F
FR1
RI30
RI60
Nosepokes/Minute
ChR2
EYFP
ChR2 Scrambled
Days of Behavioral Training
G
Shocks Received
ChR2
EYFP
ChR2 Scrambled
Shock 1
Shock 2
H
ChR2
EYFP
ChR2 Scrambled
Omission Completion (min)
ChR2
EYFP
ChR2 Scrambled
I
r = 0.33, ns
Omission Completion (min)
Shocks Received
J
r = 0.63, *p<0.05
Omission Completion (min)
Shocks Received
K
r = -0.59, ns
Omission Completion (min)
Shocks Received

Figure 7. Optogenetic excitation of DA terminals in DMS at the time of a rewarded nosepoke causes an increase in punishment resistance. A. Viral injection and probe placement strategy for stimulation of dopamine terminals in the DMS. **B.** Low (10x) and high magnification (40x) images of SNc showing ChR2-EYFP expression in green, TH positive cells in magenta, and the merged image. Scale bars are 100µm. **C.** Low (10x) and high magnification (40x) images of SNc showing EYFP expression in green, TH positive cells in magenta, and the merged image. Scale bars are 100µm. **D.** Training timeline showing sessions during which optogenetic stimulation was delivered (FR1, RI30, RI60). Average number of days in each stage of training is given below. **E.** Schematic of stimulation parameters. A 1s 20Hz 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.** Average shocks received on early and late shock probes for each manipulation. Error bars represent SEM; *$p<0.05$. **F.** Slope of nosepokes made per minute in FR1, RI30, and RI60 schedules. G. Average time to complete the omission probe (earn 50 rewards; max 60 minutes) for ChR2 (blue; n=7), EYFP (green; n=12) and ChR2 scrambled (blue stripes; n=4) mice. Error bars represent SEM. **I-K.** Correlation of time to earn 50 rewards during the omission probe and shocks received on the late shock probe for each manipulation. See also Supplemental Figure 7.

METHODS

Lead Contact and Materials Availability

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). This study did not generate new unique reagents.

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-10 males, 14 females; optogenetics- 12 males, 12 females). Adult mice at least 10 weeks of age were used in all experiments. Mice were group housed under a conventional 12h 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 hour. They were then trained to retrieve food rewards (45 mg sucrose 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 hour. Next, operant training began, with all training sessions lasting one hour or until 50 rewards had been earned. Animals were trained to associate nosepoking with reward on a fixed ratio (FR1) schedule where both nosepokes delivered a reward. Animals had to retrieve the reward (as measured by making a port entry following a rewarded nosepoke) before they could earn the next reward. After animals showed a preference for one nosepoke (>25 rewards on that side), 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. Animals were then moved to either a random interval (n=24) or ratio (n=7) schedule of reinforcement. Animals on the random interval schedule were trained on RI30 until they earned >30 rewards in one hour, and then trained on RI60. Animals on a random ratio schedule of reinforcement were trained on RR10 until they earned >30 rewards in one hour, and then trained on RR20 (Fig. 1A).

Shock Probe

Animals were subjected to a footshock probe early and late in training (Fig. 1B) to evaluate their levels of resistance to negative punishments. These probes were under a FR1 schedule of reinforcement where a mild footshock (0.2mA, 1s) 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. During shock probes, the session

ended after 60 minutes or an animal was inactive (no nosepokes on the rewarded side) for >10 minutes. There was no maximum number of rewards.

Omission Probe
Animals 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 seconds in order to receive a single sucrose pellet. A nosepoke reset the 20 second timer. The session ended after animals received 50 rewards or 60 minutes had elapsed.

Fear Conditioning
A trace fear conditioning paradigm, adapted from Lugo, Smith, and Holly (2014) was used in a naïve cohort of wild-type mice (n=13) to verify that our shock intensity (0.2mA) is aversive to the mice (Lugo et al., 2014). 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.

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. Animals 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. Animals then recovered for three weeks before behavioral experiments began.

Fiber Photometry
Animals 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. 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 on the first and last days of RI60 training as

well as on both footshock probes for each animal. 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.

Optogenetic Stimulation

Animals for 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 SNc (AP -3.1, ML 0.8, DV -4.7) and a single fiber optic implant (Prizmatix; 0.66 NA) over ipsilateral DMS (AP 0.8, ML 1.5, DV -2.8). During operant training (beginning with FR1), each rewarded nosepoke was paired with a train of light (1s, 20 Hz, 15 mW; Prizmatix). A subset of animals ("ChR2 Scrambled") received the same train 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 hour 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 hour 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 (Franklin and Paxinos, 2008). GCaMP and ChR2 neurons expressing YFP were counted and colocalized with TH+ neurons using ImageJ software.

Quantification and Statistical Analysis

Behavioral Analysis

Shock-evoked freezing during fear conditioning was scored manually with EthoVision software (Noldus) from a recording of the fear conditioning test session by two blind observers and then 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. To sort mice into PR, DPR, and PS groups, we calculated the percent change in shocks received from the early to late shock probe for each animal--animals that increased the number of footshocks withstood by greater than 89% were classified as delayed punishment resistant (n=5). The remaining animals were sorted by a median split, with animals earning more than 13 footshocks on the first probe classified as punishment resistant (n=7) and those earning fewer as punishment sensitive (n=13, total n=24, Fig. 2A).

Plots in Fig. 2K and 7H were generated by plotting the average slope of nosepokes/minute across FR1, RI30, and RI60 training to reveal escalation of nosepoke behavior. Inter-reward intervals were calculated as the time from a rewarded nosepoke to the subsequent rewarded nosepoke (Fig. 2J) on each day of RI60 training. A frequency distribution was created and plotted using GraphPad (Prism) software.

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. $\Delta F/F$ 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 $\Delta F/F$ 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 seconds following behavioral event (i.e. nosepoke, port entry). AUC was calculated using trap function in MATLAB from the time of behavioral event to the time at which the signal returned to baseline. In order to get the slope of the rise the signal, we first applied a smoothing spline function to filter the data. We then characterized the inflection point by creating a threshold slope and finding the first point which reaches that slope, which was used to generate a line from the inflection point to the absolute maximum and find the approximate slope of each signal. We used a customized logic for peak detection in Supplemental Figures 3 and 4 adapted from Holly et al (2019) and Muir et al (2018) (Holly et al., 2019; Muir et al., 2018). Events having amplitudes greater than the summation of a median of 30 seconds 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 8 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. Four animals were excluded from the fiber photometry study due to probe placement or loss of probes. Two animals were excluded from the optogenetics study—one due to improper probe placement and one because of illness.

Data and Code Availability

The code generated during this study is available at [will make public GitHub repository on publication].

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures.

REFERENCES

1. Beier, K.T., Steinberg, E.E., DeLoach, K.E., Xie, S., Miyamichi, K., Schwarz, L., Gao, X.J., Kremer, E.J., Malenka, R.C., and Luo, L. (2015). Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell *162*, 622–634.
2. Burguiere, E., Monteiro, P., Feng, G., and Graybiel, A.M. (2013). Optogenetic Stimulation of Lateral Orbitofronto-Striatal Pathway Suppresses Compulsive Behaviors. Science (80-.). *340*, 1243–1246.
3. Corbett, D., and Wise, R.A. (1980). Intracranial Self-Stimulation in Relation to the Ascending Dopaminergic Systems of the Midbrain: A Move-able electrode mapping study. Brain Res. *185*, 1–15.
4. Corbit, L.H., Nie, H., and Janak, P.H. (2012). Habitual Alcohol Seeking: Time Course and the Contribution of Subregions of the Dorsal Striatum. Biol. Psychiatry *72*, 389–395.
5. Cromwell, H.C., and Berridge, K.C. (1996). Implementation of action sequences by a neostriatal site: a lesion mapping study of grooming syntax. J. Neurosci. *16*, 3444–3458.
6. Crow, T.J. (1972). A map of the rat mesencephalon for electrical self-stimulation. Brain Res. *36*, 265–273.
7. Datta, U., Martini, M., and Sun, W. (2018). Different functional domains measured by cocaine self-administration under the progressive-ratio and punishment schedules in male Wistar rats. Psychopharmacology (Berl). *235*, 897–907.
8. 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. Neurosci. *4*.
9. 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.
10. Franklin, K.B.J., and Paxinos, G. (2008). The mouse brain in stereotaxic coordinates (Boston).
11. 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*, 2615–2618.
12. 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.
13. Haber, S.N., Fudge, J.L., and McFarland, N.R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. J. Neurosci. *20*, 2369–2382.
14. Hadjas, L.C., Lüscher, C., and Simmler, L.D. (2019). Aberrant habit formation in the Sapap3-knockout mouse model of obsessive-compulsive disorder. Sci. Rep. *9*, 12061.
15. Harada, M., Hiver, A., Pascoli, V., and Lüscher, C. (2019). Cortico-striatal synaptic plasticity underlying compulsive reward seeking. BioRxiv 789495.
16. Hodebourg, R., Murray, J.E., Fouyssac, M., Puaud, M., Everitt, B.J., and Belin, D. (2018). Heroin seeking becomes dependent on dorsal striatal dopaminergic mechanisms and can be decreased by N-acetylcysteine. Eur. J. Neurosci.
17. 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.
18. Howe, M.W., and Dombeck, D.A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature.
19. Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. Brain Res. Rev. *56*, 27–78.

20. Ilango, A., Kesner, A.J., Keller, K.L., Stuber, G.D., Bonci, A., and Ikemoto, S. (2014). Similar Roles of Substantia Nigra and Ventral Tegmental Dopamine Neurons in Reward and Aversion. J. Neurosci. *34*, 817–822.
21. Kalueff, A. V., Stewart, A.M., Song, C., Berridge, K.C., Graybiel, A.M., and Fentress, J.C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. Nat. Rev. Neurosci. *17*, 45–59.
22. Keiflin, R., Pribut, H.J., Shah, N.B., and Janak, P.H. (2017). Phasic Activation of Ventral Tegmental, but not Substantia Nigra, Dopamine Neurons Promotes Model-Based Pavlovian Reward Learning. BioRxiv 232678.
23. Kravitz, A. V, Tye, L.D., and Kreitzer, A.C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. *15*, 816–818.
24. Lerner, T.N. (2020). Interfacing behavioral and neural circuit models for habit formation. J. Neurosci. Res. jnr.24581.
25. 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., et al. (2015). Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell *162*, 635–647.
26. Lipton, D.M., Gonzales, B.J., and Citri, A. (2019). Dorsal Striatal Circuits for Habits, Compulsions and Addictions. Front. Syst. Neurosci. *13*, 28.
27. Lugo, J.N., Smith, G.D., and Holley, A.J. (2014). Trace fear conditioning in mice. J. Vis. Exp.
28. 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.
29. Murray, J.E., Belin, D., and Everitt, B.J. (2012). Double dissociation of the dorsomedial and dorsolateral striatal control over the acquisition and performance of cocaine seeking. Neuropsychopharmacology *37*, 2456–2466.
30. Nieh, E.H., Matthews, G.A., Allsop, S.A., Presbrey, K.N., Leppla, C.A., Wichmann, R., Neve, R., Wildes, C.P., and Tye, K.M. (2015). Decoding Neural Circuits that Control Compulsive Sucrose Seeking. Cell *160*, 528–541.
31. Pacchioni, A.M., Gabriele, A., and See, R.E. (2011). Dorsal striatum mediation of cocaine-seeking after withdrawal from short or long daily access cocaine self-administration in rats. Behav. Brain Res. *218*, 296–300.
32. Reynolds, J.N.J., Hyland, B.I., and Wickens, J.R. (2001). A cellular mechanism of reward-related learning. Nature *413*, 67–70.
33. Robinson, T.E., and Berridge, K.C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res. Brain Res. Rev. *18*, 247–291.
34. Rossi, M.A., and Yin, H.H. (2012). Methods for Studying Habitual Behavior in Mice. In Current Protocols in Neuroscience, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. Unit 8.29.
35. Rossi, M.A., Sukharnikova, T., Hayrapetyan, V.Y., Yang, L., and Yin, H.H. (2013). Operant Self-Stimulation of Dopamine Neurons in the Substantia Nigra. PLoS One *8*, e65799.
36. Saunders, B.T., Richard, J.M., Margolis, E.B., and Janak, P.H. (2018). Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. Nat. Neurosci. *21*, 1072–1083.
37. 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.
38. 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.

39. Vanderschuren, L.J.M.J., Di Ciano, P., and Everitt, B.J. (2005). Involvement of the Dorsal Striatum in Cue-Controlled Cocaine Seeking. J. Neurosci. *25*, 8665–8670.
40. 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. U. S. A. *109*, 20703–20708.
41. 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.
42. Wise, R.A. (1981). Intracranial self-stimulation: mapping against the lateral boundaries of the dopaminergic cells of the substantia nigra. Brain Res. *213*, 190–194.
43. Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476.
44. 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.
45. Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005a). The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.
46. Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2005b). Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur. J. Neurosci. *22*, 505–512.
47. 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.