Striatal cell-type specific stability and reorganization underlying agency and habit

Melissa Malvaez[1], Alvina Liang[1], Baila S. Hall[1], Jacqueline R. Giovanniello[1], Natalie Paredes[1], Julia Y. Gonzalez[1], Garrett J. Blair[1], Ana C. Sias[1], Michael D. Murphy[1], Wanyi Guo[1], Alicia Wang[1], Malika Singh[1], Nicholas K. Griffin[1], Samuel P. Bridges[1], Anna Wiener[1], Jenna S. Pimenta[1] Sandra M. Holley[3], Carlos Cepeda[2, 3], Michael S. Levine[2, 3], H. Tad Blair[1, 2], Andrew M. Wikenheiser[1, 2], Kate M. Wassum[1, 2]

[1]Dept. of Psychology, UCLA, Los Angeles, CA 90095. [2]Brain Research Institute, UCLA, Los Angeles, CA 90095, USA.[3]Intellectual and Developmental Disabilities Research Center, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, United States

Correspondence:
Kate Wassum: kwassum@ucla.edu
Dept. of Psychology, UCLA
1285 Franz Hall
Box 951563
Los Angeles, CA 90095-1563

Key words: learning, decision making, instrumental conditioning, reward, striatum, habit, devaluation,

Figures: 6
Tables: 0
Extended Data Figures: 27
Supplemental Tables: 4

ABSTRACT
Adaptive decision making requires agency, knowledge that actions produce particular outcomes. For well-practiced routines, agency is relinquished in favor of habit. Here, we asked how dorsomedial striatum $D1^+$ and $D2/A2A^+$ neurons contribute to agency and habit. We imaged calcium activity of these neurons as mice learned to lever press with agency and formed habits with overtraining. Whereas many $D1^+$ neurons stably encoded actions throughout learning and developed encoding of reward outcomes, $A2A^+$ neurons reorganized their encoding of actions from initial action-outcome learning to habit formation. Chemogenetic manipulations indicated that both $D1^+$ and $A2A^+$ neurons support action-outcome learning, but only $D1^+$ neurons enable the use of such agency for adaptive, goal-directed decision making. These data reveal coordinated dorsomedial striatum $D1^+$ and $A2A^+$ function for the development of agency, cell-type specific stability and reorganization underlying agency and habit, and important insights into the neuronal circuits of how we learn and decide.

When making a decision, we often engage our agency. We use our understanding of our actions and their consequences to prospectively evaluate our options and choose actions that will cause desirable consequences and avoid those that will lead to outcomes that are not currently beneficial[1-3]. This goal-directed strategy is model-based. An internal model of action-outcome associations supports the predictions and inferences needed for adaptive decision making[2-8]. This strategy is highly adaptive. Should an outcome become unbeneficial, we will reduce the actions that cause it. But it is also cognitively taxing. So, we have a model-free strategy for well-practiced routines: habits. Instead of relying on learned action-outcome associations and forethought of consequences, habits are executed automatically based on past success[2, 9-12]. This strategy is, thus, resource efficient, but inflexible. We might continue a habit, even when its outcome has become disadvantageous. Balance between goal-directed and habitual control allows behavior to be adaptive when needed, yet efficient when appropriate[13, 14]. Disrupted agency and overreliance on habit can, however, cause inadequate consideration of consequences, inflexible behavior, a lower threshold for compulsivity, and disrupted decision making[15-18]. This can contribute to cognitive deficits in addiction[19-30] and myriad other mental health conditions[17, 26, 31-40]. Yet, despite importance for understanding adaptive and maladaptive behaviors, major questions remain about how the brain forms action-outcome associations to support agency and what changes as habits form.

The dorsomedial striatum (DMS) is an evolutionally conserved hub for action-outcome learning and goal-directed decision making[9, 41-51]. Suppression of DMS activity prevents goal-directed decisions and promotes inflexible habits[44, 52]. The striatum has two major projection neuron subtypes, one expressing dopamine D1 receptors and another expressing dopamine D2 and adenosine Adora2A (A2A) receptors. Whereas, D1$^+$ neurons are more dominant in the direct pathway to basal ganglia output nuclei, D2/A2A$^+$ neurons are associated with the multisynaptic, indirect output pathway[53-55], though this organization is likely more complex than originally thought[56]. The canonical view is that these striatal neuron subtypes have opposing function[55, 57-61], but this has been challenged by evidence of more coordinated function[62, 63]. Very little is known of how these two important DMS neuron subtypes organize their activity to support action-outcome learning and agency, and even less of whether or how this changes as habits form. Here we fill these gaps in knowledge by characterizing the activity and function of DMS D1$^+$ and A2A$^+$ neurons during action-outcome learning and habit formation.

RESULTS

DMS D1$^+$ neurons encode actions and outcomes at each phase of instrumental learning

We first characterized the activity of DMS D1$^+$ neurons during action-outcome learning and habit formation. We used UCLA miniscopes[64] with a gradient refractive index (GRIN) lens for cellular resolution, microendoscopic imaging of the genetically encoded calcium indicator jGCaMP7s[65] selectively expressed in D1$^+$ neurons in the DMS of Drd1a-cre mice (Figure 1a-c). We recorded calcium activity as mice learned to press a single lever to earn food-pellet rewards in a self-initiated instrumental task (Figure 1d). We used a random-interval schedule of reinforcement that enables action-outcome learning and flexible goal-directed decision making early in training and promotes the formation of inflexible habits with overtraining (Extended Data Figure 1-1). All mice acquired the instrumental behavior (Figure 1e; see Extended Data Figure 1-2 for food-port entry data). To evaluate behavioral control

strategy, we used the outcome-specific devaluation test[2, 66]. Mice were given 90-min, non-contingent access to the food pellet earned during training to induce a sensory-specific satiety rendering that specific food pellet temporarily devalued. Lever pressing was then assessed in an immediately following 5-min, non-reinforced probe test. Performance was compared to that following satiation on an alternate food pellet to control for general satiety (Valued state; test order counterbalanced). As expected, following overtraining, mice were insensitive to devaluation (Figure 1f-g), indicating the lack of consideration of action consequences that marks inflexible habits[9, 13]. To capture activity related to initial action-outcome learning and the transition to habit, we analyzed calcium activity at the beginning (random-interval training session 1), middle (session 4), and end (overtraining, session 8) of instrumental training. Individual neuron footprints and activity were extracted using Constrained Non-negative Matrix Factorization for Endoscopic data (CNMF-E)[67, 68]. We recorded 870 – 999 D1$^+$ neurons/session from all mice (see Supplemental Table 1 for neurons/mouse/session). Individual fluorescent signals were deconvolved to estimate temporally constrained neural activity for each neuron[69].

DMS D1$^+$ neurons are active around actions and earned rewards. To ask how DMS D1$^+$ neurons encode task events during action-outcome learning and the transition to habit, we computed area under the receiver operating characteristic (auROC) values and classified neurons as significantly modulated by critical behavioral events: lever-press action initiation (first press in the session, after earned reward collection, or after a non-reinforced food-port check), action termination (last press before reward collection or non-reinforced food-port check), intervening lever presses (all other presses), non-reinforced food-port checks, and reward collection (Figure 1h-p). 76 - 85% of the neurons significantly increased or decreased their activity within ±2.9 s around one or more of these events. 11 - 14% of DMS D1$^+$ neurons were activated around, largely immediately prior to, action initiation (Figure 1h-p; Extended Data Figure 1-3). Approximately 10% of neurons were activated around, mostly prior to, action termination. 13 – 15% of neurons responded to the earned reward. Similar proportions of neurons were activated by each task event across training, though there was a slight trend towards more neurons being inhibited around actions with overtraining (Extended Data Figure 1-3). These data accord with prior electrophysiological evidence of DMS neuronal encoding of actions and rewards during instrumental behavior[50, 62, 70, 71]. We find that DMS D1$^+$ neurons encode actions and outcomes during learning and as habits form with overtraining.

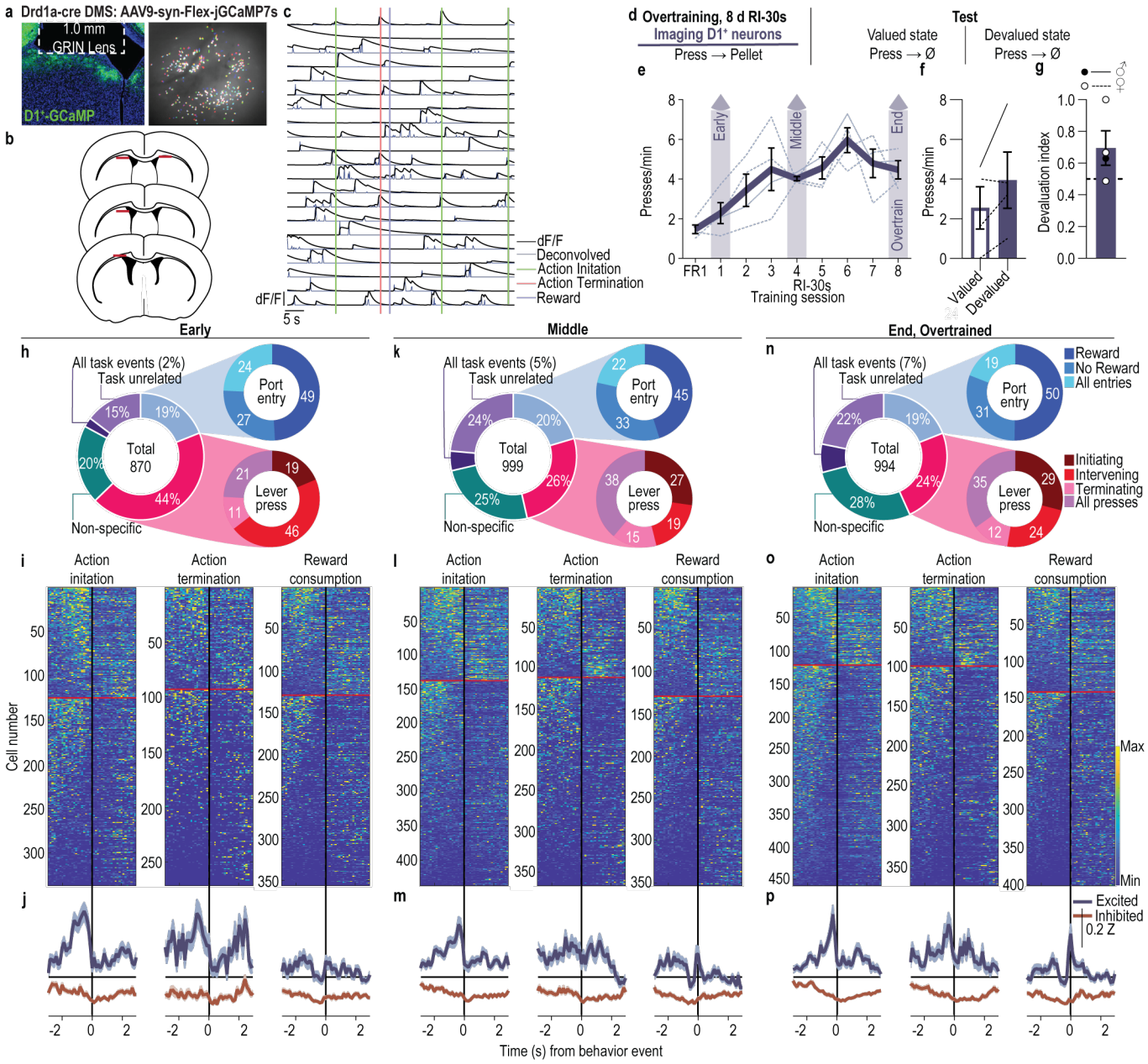


Figure 1: DMS D1⁺ neurons are modulated by actions and rewards during instrumental learning and overtraining. (a) Representative cre-dependent jGCaMP7s expression in DMS D1⁺ neurons (right) and maximum intensity projection of the field-of-view for one session (left). **(b)** Map of DMS GRIN lens placements. **(c)** Representative dF/F and deconvolved calcium signal. **(d)** Procedure. RI, random-interval reinforcement schedule. **(e)** Training press rate. 1-way ANOVA, Training: $F_{1.81, 5.43} = 4.84$, $P = 0.06$. **(f)** Test press rate. 2-tailed t-test, $t_3 = 1.98$, $P = 0.14$, 95% confidence interval (CI) -0.85 - 3.65. **(g)** Devaluation index [(Devalued presses)/(Valued presses + Devalued presses)]. One-tailed Bayes factor, $BF_{10} = 0.18$. **(h-j)** Activity of DMS D1⁺ neurons on the 1st (Early) session of RI training. **(h)** Percent of all recorded neurons ($N = 870$) significantly modulated by lever presses, food-delivery port checks, and reward. **(i)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each D1⁺ neuron significantly modulated around lever-press action initiation (left), action termination (middle), or reward consumption (right). Above red line = excited above cutoff criterion, below = inhibited. **(j)** Z-scored activity of each population of modulated neurons. **(k-m)** Activity of DMS D1⁺ neurons on the 4th (Middle) training session. **(k)** Percent of all recorded neurons ($N = 999$) significantly modulated by lever presses, food-delivery port checks, and reward. **(l)** Heat map of each D1⁺ neuron significantly modulated around lever-press action initiation, action termination, or reward consumption. **(m)** Z-scored activity of each population of modulated neurons. **(n-p)** Activity of DMS D1⁺ neurons on the 8th (End, Overtrain) training session. **(k)** Percent of all recorded neurons ($N = 994$) significantly modulated by lever presses, food-delivery port checks, and reward. **(l)** Heat map of each DMS D1⁺ neuron significantly modulated around lever-press action initiation, action termination, or reward consumption. **(m)** Z-scored activity of each population of modulated neurons. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines.

DMS D1$^+$ neurons stably encode action initiation throughout action-outcome learning and habit formation

We next leveraged the ability to track neurons longitudinally to ask how DMS D1$^+$ neurons change their activity and encoding with action-outcome learning and habit formation. We were able to coregister on average 56% (s.e.m. = 3.90%) of the D1$^+$ neurons across all three (early, middle, overtrain) phases of training (Figure 2a-c; Extended Data Figure 2-1; Supplemental Table 1).

Instrumental actions and reward outcomes can be decoded from DMS D1$^+$ neuron population activity. We first probed the information content available in the activity of the tracked DMS D1$^+$ neuronal population using two within-session decoding approaches. We trained a multi-class support vector machine (SVM) algorithm to classify behavioral events (lever-press action initiation, action termination, non-reinforced food-port checks, and earned reward consumption) from the activity of all coregistered D1$^+$ neurons. Across training, the decoder classified each behavioral event more accurately than one trained on shuffled data (Figure 2d). Decoding of action initiation improved during the first half of training, but did not improve further with overtraining. Next, we trained a linear regression model to decode lever-press action rate from D1$^+$ neuron ensemble activity. This, too, could be decoded with high accuracy across training (Figure 2e). We found similar results if we used only those neurons that were significantly excited or inhibited by action initiation (Extended Data Figure 2-2). Thus, the DMS D1$^+$ neuron population can convey information about actions, action rate, and outcomes.

DMS D1$^+$ neurons stably encode action initiation throughout action-outcome learning and habit formation. We next asked how the activity and encoding of individual DMS D1$^+$ neurons changes with action-outcome learning and habit formation. As with the entire population, a substantial proportion of tracked D1$^+$ neurons were significantly modulated (Figure 2f), excited (Figure 2g) or inhibited (Figure 2h), by action initiation across training (see Extended Data Figure 2-3 for percentage of neurons excited and inhibited by action termination and reward). We classified neurons significantly excited by action initiation during initial training. These *'early action-initiation excited'* D1$^+$ neurons were more active (Figure 2i-j) and modulated (Figure 2k; Extended Data Figure 2-4) prior to action initiation, indicating these cells encode the initiation of action. We looked prospectively to ask how this ensemble encodes action initiation as training progresses and found that these neurons continued to be active prior to action initiation throughout training (Figure 2i-j). This refined, occurring tighter prior to action initiation, with learning. Accordingly, the auROC modulation index immediately prior to action initiation of these activated neurons was stable throughout training (Figure 2k; Extended Data Figure 2-4). Further demonstrating the stability of the ensemble, the activity around action initiation of both individual (Figure 2l) and the population (Figure 2m) of early action-initiation excited D1$^+$ neurons was significantly correlated across training phases. We also found significant cross-session decoding of action rate. We trained a linear regression model to decode action rate from the activity of the early action-initiation excited D1$^+$ ensemble on the 1st training session. We could not only reliably decode action rate for this early session, but could also reliably decode action rate from the middle and overtraining training sessions (Figure 2n). Lastly, nearly half of the early action-initiation excited D1$^+$ neuron ensemble continued to significantly encode action initiation across training (Figure 2o). Thus, early action-

initiation excited D1⁺ neurons stably encode action initiation throughout action-outcome learning and habit formation with overtraining.

To provide converging evidence for the stability of DMS D1⁺ neuronal encoding of action initiation, we next took a retrospective approach. We classified D1⁺ neurons significantly excited by action initiation during overtraining. These *'overtrain action-initiation excited'* neurons, too, were more active (Figure 2p-q) and modulated (Figure 2r; Extended Data Figure 2-4) prior to action initiation. We looked retrospectively to ask whether this ensemble encoded action initiation at earlier training phases. It did. Overtrain action-initiation excited D1⁺ neurons also showed elevated activity (Figure 2p-q) and modulation (Figure 2r) prior to action initiation during early and middle training. Encoding of action initiation by these neurons improved with training (Figure 2r; Extended Data Figure 2-4). Like the early action-initiation excited ensemble, the activity around action initiation of individual (Figure 2s) and the population (Figure 2t) of overtrain action-initiation excited D1⁺ neurons was significantly correlated across training phases. We could also decode action rate across sessions from the activity of this ensemble. When we trained a linear regression model to decode action rate from the activity of overtrain action-initiation excited D1⁺ neurons on the last training session, we could not only reliably decode action rate from this session, but also the preceding training sessions (Figure 2u). Again, about half the overtrain action-initiation excited D1⁺ ensemble significantly encoded action initiation across training (Figure 2v). Thus, regardless of whether we identified action-initiation neurons early in training and looked prospectively or during overtraining and looked retrospectively, we found a substantial subpopulation of DMS D1⁺ neurons that stably encoded action initiation throughout action-outcome learning and habit formation. These neurons encoded action initiation with high trial-by-trial fidelity (Extended Data Figure 2-5). D1⁺ action-initiation neurons were excited on more than half the action-initiation events across training. The activity of these neurons around action initiation was also highly correlated within a training session and became more correlated with training. Thus, there is an ensemble of DMS D1⁺ neurons that stably encode action initiation (Figure 2w) with high fidelity from initial action-outcome learning throughout training even as habits form. The D1⁺ neuron ensemble that was inhibited around action initiation was similarly stable (Extended Data Figure 2-6). D1⁺ ensembles also encoded action termination and earned reward across training with modest stability (Extended Data Figure 2-7).

A subensemble of stable DMS D1⁺ action-initiation neurons developed encoding of action termination and earned reward. If activity of the stable DMS D1⁺ action-initiation ensemble relates to action-outcome learning, we reasoned that it might not solely encode action initiation, but, with learning, might also develop encoding of action termination and earned reward outcomes. We found evidence of this (Figure 2w, Extended Data Figure 2-8). Early in training 41.3% (s.e.m. = 15.20%) of the stable action-initiation excited ensemble encoded not only action initiation, but also termination. With learning, almost all of these neurons (87.03%, s.e.m. = 7.65%) significantly encoded action termination (Figure 2x; Extended Data Figure 2-8). Thus, this stable action-initiation excited ensemble shows action bracketing, being excited before the start and end of an action bout. The proportion of the stable action-initiation neurons that also encoded earned reward grew from 18.7% (s.e.m. = 12.99) to 63.9% (s.e.m. = 12.95%). Early in training only 10.7% (s.e.m. = 10.71%) of the stable action-initiation excited ensemble encoded action initiation, termination, and earned reward. With learning over half (average =

58.9%, s.e.m. = 14.34%) of this ensemble significantly encoded all three variables (Figure 2x). This did not improve further with overtraining. Thus, a subensemble of DMS D1⁺ neurons not only stably encode actions, but also their reward outcomes. Together these data indicate that DMS D1⁺ neurons convey information about actions and outcomes during learning and overtraining, with an ensemble stably encoding action initiation and developing encoding of action termination and earned reward outcomes.

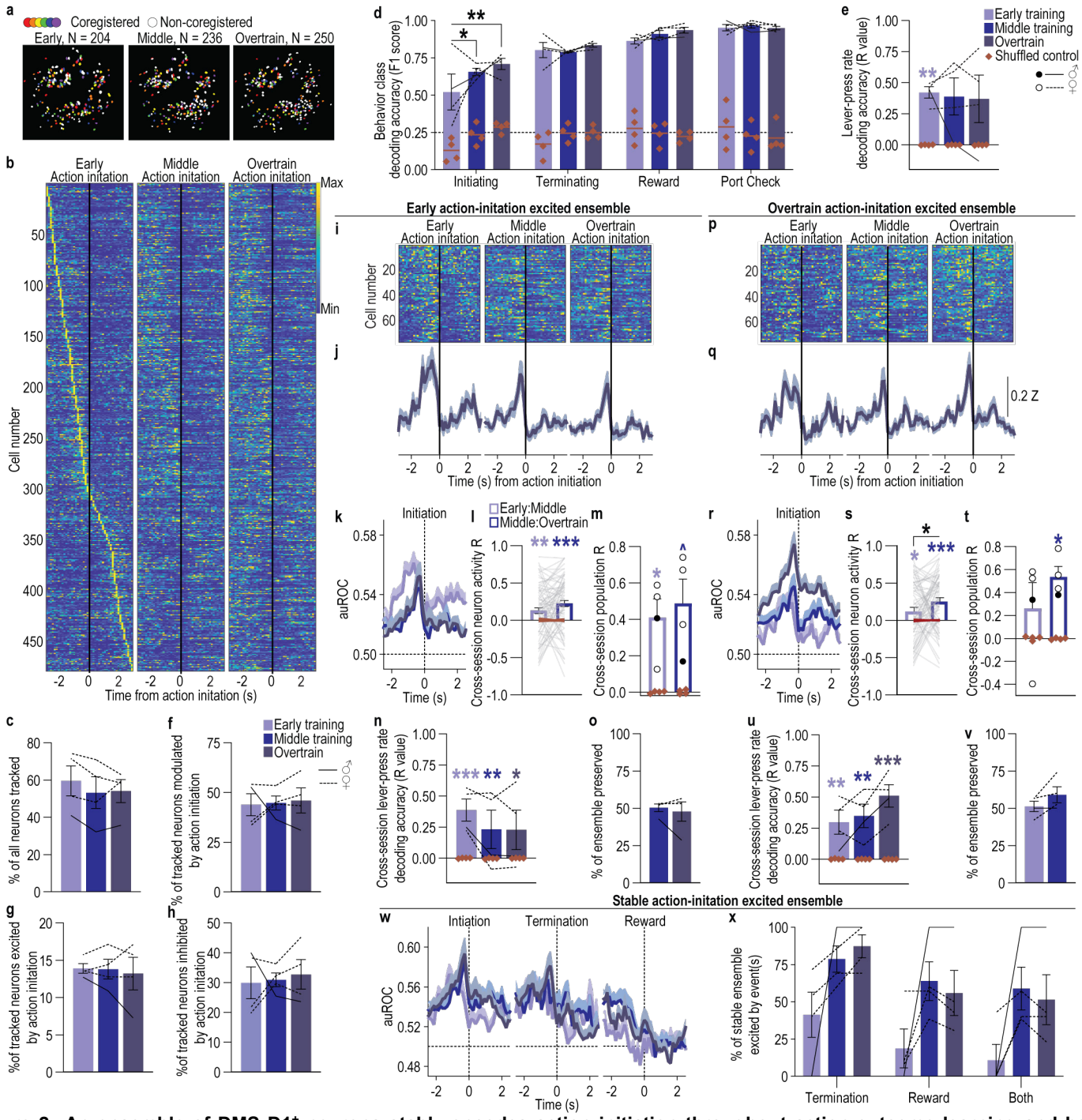


Figure 2: An ensemble of DMS D1⁺ neurons stably encodes action initiation throughout action-outcome learning and habit formation. (a) Representative recorded D1⁺ neuron spatial footprints during the first (early, left), 4th (middle), and 8th (overtrain) random-interval training sessions. Colored, co-registered neurons; white, non-co-registered neurons. **(b)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each coregistered DMS D1⁺ neuron around lever-press action initiation. **(c)** Percent of all neurons coregistered across training. 1-way ANOVA, $F_{1.19, 3.58} = 1.62$, $P = 0.29$. **(d)** Behavior class (initiating lever press, terminating press, reward collection, non-reinforced food-port check) decoding accuracy from D1⁺ coregistered neuron activity compared to shuffled

control. Line at 0.25 = chance. 3-way ANOVA, Neuron activity (v. shuffled): $F_{1, 3} = 1092.83$, $P < 0.001$; Training session: $F_{1.17, 3.51} = 2.40$, $P = 0.21$; Behavior class: $F_{1.76, 5.28} = 15.56$, $P = 0.007$; Neuron activity x Training: $F_{1.24, 3.72} = 4.60$, $P = 0.10$; Neuron activity x Behavior Class: $F_{1.81, 5.43} = 6.49$, $P = 0.04$; Training x Behavior Class: $F_{2.15, 6.46} = 3.25$, $P = 0.10$; Neuron activity x Training x Behavior class: $F_{1.95, 5.86} = 0.69$, $P = 0.54$. **(e)** Lever-press rate decoding accuracy from D1⁺ coregistered neuron activity. R = correlation coefficient between actual and decoded press rate. 2-way ANOVA, Neuron activity: $F_{1, 3} = 11.46$, $P = 0.04$; Training: $F_{1.22, 3.66} = 0.07$, $P = 0.85$; Neuron activity x Training: $F_{1.22, 3.66} = 0.07$, $P = 0.85$. **(f-h)** Percent of coregistered neurons (Average 137 coregistered neurons/mouse, s.e.m. 37.73) significantly modulated (f; 1-way ANOVA, $F_{1.12, 3.37} = 0.06$, $P = 0.85$), excited (g; 1-way ANOVA, $F_{1.13, 3.63} = 0.14$, $P = 0.76$), or inhibited (h; 1-way ANOVA, $F_{1.29, 3.86} = 0.14$, $P = 0.79$) around action initiation. **(i-k)** Activity and modulation across training of DMS D1⁺ early action-initiation excited neurons ($N = 77$ neurons/4 mice; average 19.25 neurons/mouse, s.e.m. = 5.36). Heat map (i), Z-scored activity (j), and area under the receiver operating characteristic curve (auROC) modulation index (k) of early action-initiation excited neurons around action initiation across training. **(l-m)** Cross-session correlation of the activity around action initiation of each early action-initiation excited neuron (l; 2-way ANCOVA, Neuron activity: $F_{1, 75} = 22.92$, $P < 0.001$; Training: $F_{1, 75} = 0.40$, $P = 0.53$; Activity x Time: $F_{1, 75} = 0.24$, $P = 0.63$) or the population activity of these neurons (m; 2-way ANOVA, Neuron activity: $F_{1, 3} = 39.10$, $P = 0.008$; Training: $F_{1, 3} = 0.70$, $P = 0.18$; Activity x Time: $F_{1, 3} = 0.16$, $P = 0.72$). **(n)** Cross-session decoding accuracy of lever-press rate from the activity of D1⁺ early action-initiation-excited neuron population activity on the 1st training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: $t_6 = 7.94$, $P = 0.0006$, 95% CI 0.23 - 0.55; Middle: $t_6 = 4.83$, $P = 0.009$, 95% CI 0.08 - 0.40; Overtrain: $t_6 = 4.68$, $P = 0.01$, 95% CI 0.07 - 0.39. **(o)** Percent of D1⁺ early action-initiation excited neurons that continued to be significantly excited by action initiation on the 4th and 8th training sessions. 2-tailed Wilcoxon signed rank test, W = -1.00, P > 0.99. **(p-r)** Activity and modulation across training of D1⁺ overtrain action-initiation excited neurons ($N = 76$ neurons/4 mice; average 19 neurons/mouse, s.e.m. 5.49). Heat map (p), Z-scored activity (q), and auROC modulation index (r) of overtrain action-initiation excited neurons around action initiation across training. **(s-t)** Cross-session correlation of the activity around action initiation of each overtrain action-initiation excited neuron (s; 2-way ANCOVA, Neuron activity: $F_{1, 74} = 21.06$, $P < 0.001$; Training: $F_{1, 74} = 0.23$, $P = 0.64$; Activity x Time: $F_{1, 74} = 0.26$, $P = 0.61$) or the population activity of these neurons (t; 2-way ANOVA, Neuron activity: $F_{1, 3} = 8.25$, $P = 0.06$; Training: $F_{1, 3} = 1.76$, $P = 0.28$; Activity x Time: $F_{1, 3} = 2.29$, $P = 0.23$). **(u)** Cross-session decoding accuracy of lever-press rate from the activity of D1⁺ overtrain action-initiation-excited neuron population activity on the 8th training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: $t_6 = 3.64$, $P = 0.005$, 95% CI 0.12 to 0.47; Middle: $t_6 = 6.48$, $P = 0.002$, 95% CI 0.17 - 0.53; Overtrain: $t_6 = 9.49$, $P = 0.002$, 95% CI 0.33 - 0.69. **(v)** Percent of D1⁺ overtrain action-initiation excited neurons that were also significantly excited by action initiation on 1st and 4th training sessions. 2-tailed t-test, $t_3 = 2.02$, $P = 0.14$, 95% CI -4.52 - 20.15. **(w)** Modulation around action initiation, termination, and earned reward across training of DMS D1⁺ stable action-initiation excited neurons ($N = 31$ neurons/4 mice; average 7.75 neurons/mouse, s.e.m. = 2.56). **(x)** Percent of coregistered D1⁺ neurons stably excited by action initiation across training that were also significantly modulated by action termination and reward collection. 2-way ANOVA, Event: $F_{1, 11} = 7.81$, $P = 0.06$; Training: $F_{1.02, 3.05} = 4.37$, $P = 0.13$; Event x Time: $F_{1.86, 5.59} = 1.76$, $P = 0.25$. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

DMS D1⁺ neurons drive action-outcome learning and agency

DMS D1⁺ neuronal activity is necessary for action-outcome learning. Given that DMS D1⁺ neurons convey and encode information about actions and outcomes during learning, we next reasoned they might mediate action-outcome learning and, therefore, agency. To test this, we chemogenetically inhibited DMS D1⁺ neurons during learning and then probed for action-outcome learning using the devaluation test. We chose chemogenetic manipulation because we found DMS D1⁺ neurons were endogenously active at multiple time points during learning. Such manipulations also avoid artificially synchronizing subpopulations neurons with different activity profiles and avoid reinforcing/punishing[72, 73] and locomotor[58, 63, 74] effects, which occur with optogenetic manipulations of these cells. We expressed the inhibitory designer receptor human M4 muscarinic receptor (hM4Di) or fluorophore control selectively in DMS D1⁺ neurons of DRD1a-cre mice (Figure 3a-b). Mice were trained to lever press to earn food-pellet rewards on a random-interval schedule of reinforcement (Figure 3c). We gave mice limited (4 sessions) training to preserve dominance of goal-directed behavior in control subjects (Extended Data Figure 1-1). Prior to each training session, mice received the hM4Di ligand clozapine-N-oxide (CNO; 2.0 mg/kg[75-77] i.p.) to, in hM4Di-expressing subjects, inactivate DMS D1⁺ neurons (see Extended Data Figure 3-1 for validation). Chemogenetic inactivation of DMS D1⁺ neurons did not alter acquisition of the instrumental behavior (Figure 3d; see Extended Data Figure 3-2 for food-port entry and presses/reward data). Thus, DMS D1⁺ neuronal activity is not necessary for general acquisition or execution of instrumental behavior. Training was followed by the outcome-specific devaluation test. No CNO was given on test. If subjects have learned the action-outcome relationship and are using this to support prospective consideration of action

consequences for flexible, goal-directed decision making, they will reduce lever pressing when the outcome is devalued. We saw such agency in control subjects (Figure 3e-f). DMS D1$^+$ neuron inhibition disrupted action-outcome learning as evidenced by subsequent insensitivity to outcome devaluation (Figure 3e-f). Thus, DMS D1$^+$ neurons are active during instrumental learning and this activity mediates the action-outcome learning that supports agency for flexible, goal-directed decision making.

DMS D1$^+$ neuronal activity is sufficient to drive action-outcome learning to promote agency. If DMS D1$^+$ neuronal activity drives action-outcome learning, we reasoned that augmenting this activity might promote the dominance of such learning leading to continued agency, even after the overtraining that typically promotes habits. To test this, we chemogenetically activated DMS D1$^+$ neurons during learning and then probed behavioral control strategy using the devaluation test. We expressed the excitatory designer receptor human M3 muscarinic receptor (hM3Dq) or fluorophore control selectively in DMS D1$^+$ neurons of DRD1a-cre mice (Figure 3g-h). Mice were trained to lever press to earn food-pellet rewards (Figure 3i). We overtrained (8 sessions) mice to promote habit formation in control subjects (Extended Data Figure 1-1). Prior to each training session, mice received the hM3Dq ligand CNO (0.2 mg/kg[78-81] i.p.) to, in hM3Dq-expressing subjects, activate DMS D1$^+$ neurons (Extended Data Figure 3-1). Chemogenetic activation of DMS D1$^+$ neurons did not alter acquisition of the instrumental behavior (Figure 3j), but did promote the dominance of goal-directed behavioral control. Whereas controls showed evidence of habit formation, insensitivity to devaluation at test, subjects for which we activated DMS D1$^+$ neurons during learning were sensitive to devaluation, indicating preserved agency for goal-directed decision making (Figure 3k-l). Thus, enhanced DMS D1$^+$ neuronal activity promotes action-outcome learning to encourage dominance of flexible, goal-directed decision making over inflexible habitual behavioral control. Together, these data indicate that DMS D1$^+$ neurons drive the action-outcome learning that supports agency.

DMS D1$^+$ neuronal activity mediates the expression of agency. Since DMS D1$^+$ neurons are fundamental for action-outcome learning, we next asked whether they also support the use of such agency for flexible goal-directed decision making. To test this, we again chemogenetically inhibited DMS D1$^+$ neurons, this time at test rather than during learning. We expressed hM4Di or fluorophore control selectively in DMS D1$^+$ neurons of DRD1a-cre mice (Figure 3m-n). Mice received limited training to lever press to earn food-pellet rewards (Figure 3o). All mice acquired the instrumental behavior (Figure 3p). Mice then received the outcome-specific devaluation test. After sensory-specific satiety, prior to the lever-pressing probe test mice received CNO (2.0 mg/kg i.p.) to, in hM4Di-expressing subjects, inactivate DMS D1$^+$ neurons. Controls showed flexible, goal-directed decision making, reducing action performance when the outcome was devalued (Figure 3q-r). Chemogenetic inhibition of DMS D1$^+$ neurons disrupted the expression of such agency, as evidenced by insensitivity to devaluation (Figure 3q-r). Thus, DMS D1$^+$ neurons mediate both action-outcome learning and the application of this learned agency for flexible, goal-directed decision making.

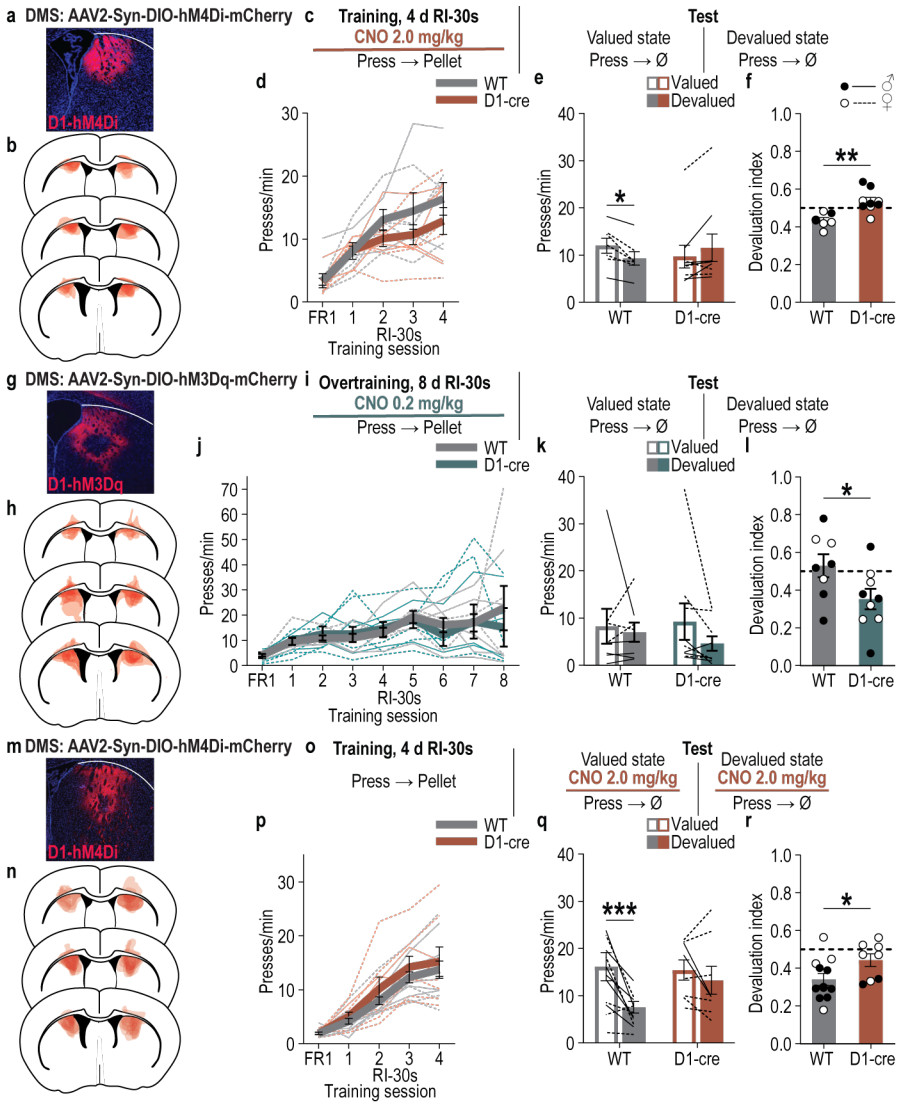


Figure 3: DMS D1⁺ neurons drive action-outcome learning and goal-directed decision making. (a-f) Chemogenetic inactivation of DMS D1⁺ neurons during learning. WT: $N = 7$ (2 males); D1-cre: $N = 9$ (5 males). **(a)** Representative immunofluorescent image of cre-dependent hM4Di expression in DMS. **(b)** Map of DMS cre-dependent hM4Di expression for all subjects. **(c)** Procedure. RI, random-interval reinforcement schedule; CNO, clozapine-N-oxide. **(d)** Training press rate. 2-way ANOVA, Training: $F_{2.22, 31.01} = 27.65$, $P < 0.0001$; Genotype: $F_{1, 14} = 1.27$, $P = 0.28$; Training x Genotype: $F_{4, 56} = 1.24$, $P = 0.30$. **(e)** Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 14} = 14.30$, $P = 0.002$; Value: $F_{1, 14} = 0.49$, $P = 0.49$; Genotype: $F_{1, 14} = 0.00005$, $P = 0.99$. **(f)** Devaluation index [(Devalued condition presses)/(Valued condition presses + Devalued presses)]. 2-tailed t-test, $t_{14} = 4.01$; $P = 0.001$, 95% CI -0.16 - -0.05. **(g-l)** Chemogenetic activation of DMS D1⁺ neurons during overtraining. WT: $N = 6$ (4 males); D1-cre: $N = 6$ (3 males). **(g)** Representative immunofluorescent image of cre-dependent hM3Dq expression in DMS. **(h)** Map of DMS cre-dependent hM3Dq expression. **(i)** Procedure. **(j)** Training press rate. 2-way ANOVA, Training: $F_{2.44, 24.43} = 3.23$, $P = 0.048$; Genotype: $F_{1, 10} = 0.31$, $P = 0.59$; Training x Genotype: $F_{8, 80} = 0.68$, $P = 0.71$. **(k)** Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 10} = 3.37$, $P = 0.10$; Value: $F_{1, 10} = 0.56$, $P = 0.47$; Genotype: $F_{1, 10} = 0.30$, $P = 0.59$. **(l)** Devaluation index. 2-tailed t-test, $t_{10} = 3.07$; $P = 0.01$, 95% CI -0.45 - -0.07. **(m-r)** Chemogenetic inactivation of DMS D1⁺ neurons during test of behavioral control strategy after learning. WT: $N = 12$ (7 males); D1-cre: $N = 12$ (5 males). **(m)** Representative immunofluorescent image of cre-dependent hM4Di expression in DMS. **(n)** Map of DMS cre-dependent hM4Di expression. **(o)** Procedure. **(p)** Training press rate. 2-way ANOVA, Training: $F_{1.59, 28.63} = 68.95$, $P < 0.0001$; Genotype: $F_{1, 18} = 0.74$, $P = 0.40$; Training x Genotype: $F_{4, 72} = 0.44$, $P = 0.78$. **(q)** Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 18} = 4.23$, $P = 0.05$; Value: $F_{1, 18} = 11.93$, $P = 0.003$; Genotype: $F_{1, 18} = 0.64$, $P = 0.44$. **(r)** Devaluation index. 2-tailed t-test, $t_{18} = 2.15$; $P = 0.045$, 95% CI 0.003 - 0.20. Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

DMS A2A⁺ neurons encode actions across phases of instrumental learning

We next characterized the activity of DMS D2/A2A⁺ neurons during action-outcome learning and habit formation. We used UCLA miniscopes with a GRIN lens for cellular resolution, microendoscopic imaging of jGCaMP7s selectively expressed in A2A⁺ neurons of A2A-cre mice (Figure 4a-c). We recorded calcium activity in DMS A2A⁺ neurons as mice learned to lever press to earn food-pellet rewards on a random-interval schedule of reinforcement

(Figure 4d). All mice acquired the instrumental behavior (Figure 4e; see Extended Data Figure 4-1 for food-port entry data). Following overtraining, half the mice were insensitive to devaluation (Figure 4f-g), indicating they formed routine habits. Half the mice did not form habits with overtraining (Extended Data Figure 4-2). We analyzed calcium activity at the beginning, middle, and end of instrumental training to characterize A2A$^+$ neuronal activity as behavioral strategy transitioned from goal-directed to habitual.

DMS A2A$^+$ neurons are active around actions. In the subjects that formed habits, we recorded the activity of 671 - 697 A2A$^+$ neurons/session and deconvolved the fluorescent signals to estimate temporally constrained neural activity for each neuron. 54-72% of the A2A$^+$ neurons significantly increased or decreased their activity around one or more behavioral events (Figure 4i-p). 8 - 12% of A2A$^+$ neurons were activated around action initiation (Extended Data Figure 4-3), largely prior to the action (Figure 4h-p). A2A$^+$ neurons were also excited around action termination (6 - 10%). Thus, like D1$^+$ neurons, DMS A2A$^+$ neurons encode actions during action-outcome learning and as habits form with overtraining. Only 4 – 6% of A2A$^+$ neurons were excited by the earned reward, significantly fewer than the proportion excited by action initiation (Extended Data Figure 4-3). Rather, A2A$^+$ neurons appeared to quiet their activity during consumption of the earned reward (Figure 4j, m, p). The proportion of A2A$^+$ neurons modulated by each behavior was not affected by training (Extended Data Figure 4-3). In subjects that did not form habits with overtraining, A2A$^+$ neurons were also excited prior to action initiation and termination and a subset were excited by the earned reward (Extended Data Figure 4-2).

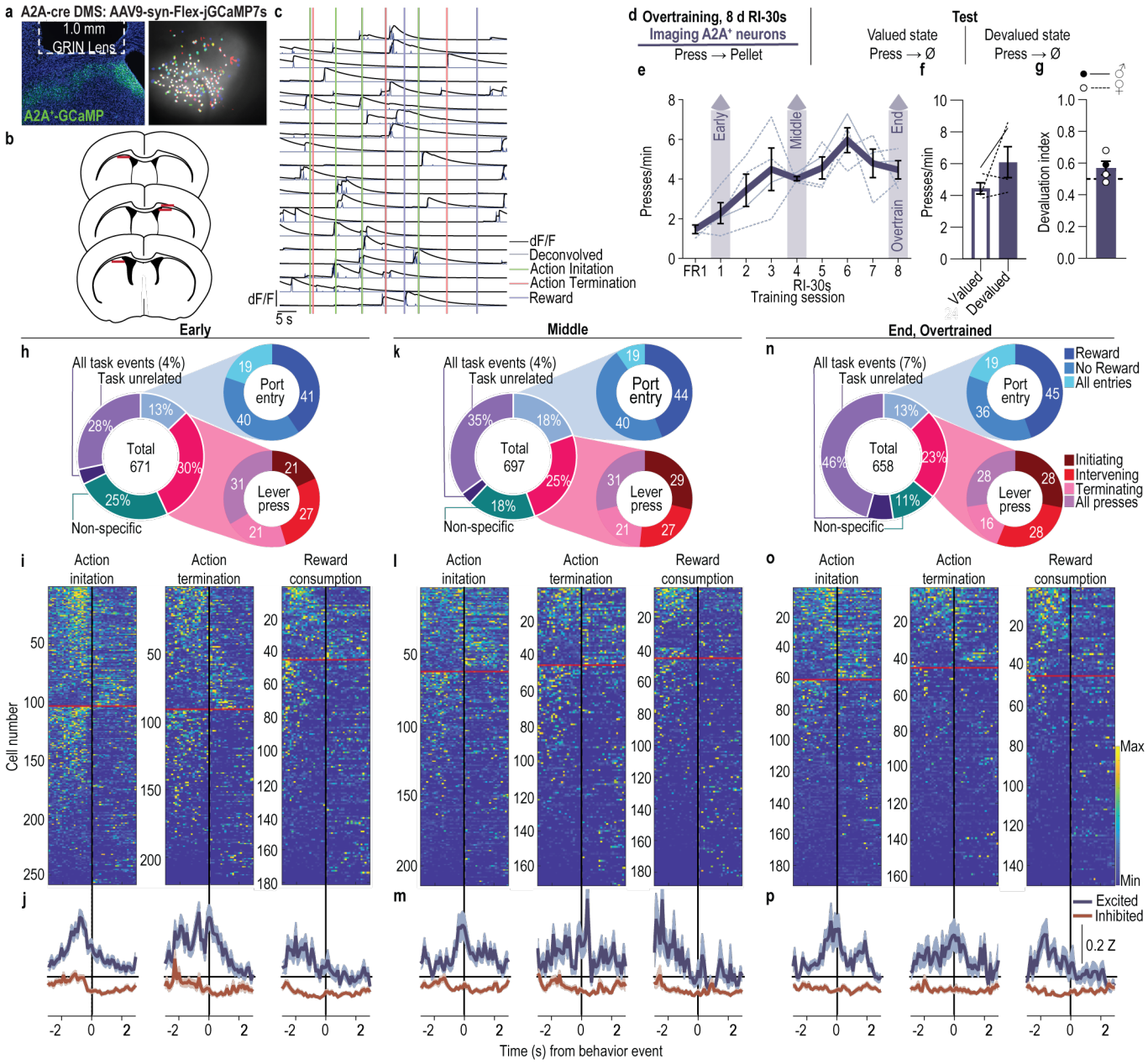


Figure 4: DMS A2A⁺ neurons are modulated by actions during instrumental learning and overtraining. (a) Representative image of cre-dependent jGCaMP7s expression in DMS A2A⁺ neurons (right) and maximum intensity projection of the field-of-view for one session (left). **(b)** Map of DMS GRIN lens placements. **(c)** Representative dF/F and deconvolved calcium signal. **(d)** Procedure. RI, random-interval reinforcement schedule. **(e)** Training press rate. 2-way ANOVA, Training: $F_{1.39, 4.18} = 3.56$, $P = 0.12$. **(f)** Test press rate. 2-tailed t-test, $t_3 = 1.50$, $P = 0.23$, 95% CI -1.86 - 5.16. **(g)** Devaluation index [(Devalued condition presses)/(Valued condition presses + Devalued presses)]. One-tailed Bayes factor, $BF_{10} = 0.19$. **(h-j)** Activity of DMS A2A⁺ neurons on the 1st (Early) session of RI training. **(h)** Percent of all recorded neurons ($N = 671$) significantly modulated by lever presses, food-delivery port checks, and reward. **(i)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each D4-3MS A2A⁺ neuron significantly modulated around lever-press action initiation (right), action termination (middle), or reward consumption (left). Above red line = excited, below = inhibited. **(j)** Z-scored activity of each population of modulated neurons. **(k-m)** Activity of DMS A2A⁺ neurons on the 4th (Middle) training session. **(k)** Percent of all recorded neurons ($N = 697$) significantly modulated by lever presses, food-delivery port checks, and reward. **(l)** Heat map of minimum to maximum deconvolved activity of each DMS A2A⁺ neuron significantly modulated around lever-press action initiation, action termination, or reward. **(m)** Z-scored activity of each population of modulated neurons. **(n-p)** Activity of DMS A2A⁺ neurons on the 8th (End, Overtrain) training session. **(k)** Percent of all recorded neurons ($N = 658$) significantly modulated by lever presses, food-delivery port checks, and reward. **(l)** Heat map of minimum to maximum deconvolved activity of each DMS A2A⁺ neuron significantly modulated around lever-press action initiation, action termination, or reward. **(m)** Z-scored activity of each population of modulated neurons. Data presented as mean ± s.e.m. A2A-cre: $N = 4$ (1 male). Males = closed circles/solid lines, Females = open circles/dashed lines.

The ensemble of DMS A2A⁺ neurons that encodes action initiation realigns with habit formation

We next used the 47% (s.e.m. 3.60%) of the A2A$^+$ neurons we were able to coregister across all three phases of training in subjects that formed habits (Figure 5a-c; Extended Data Figure 5-1; Supplemental Table 1) to ask how DMS A2A$^+$ neurons change their activity and encoding with action-outcome learning and habit formation.

Instrumental actions and reward outcomes can be decoded from DMS A2A$^+$ population activity. We again used decoding to probe the information content available in the activity of the tracked A2A$^+$ neuronal population. First, we found that a multi-class SVM algorithm trained to classify behavioral events from the activity of all coregistered A2A$^+$ neurons was able to classify each behavioral event more accurately than shuffled control data (Figure 5d). Decoding of action initiation, in particular, improved with continued action-outcome training, but decreased with overtraining. We found similar results if we used only those neurons that were significantly excited by action initiation (Extended Data Figure 5-2). We could also decode action (lever-press) rate from the A2A$^+$ neuron population activity using a linear regression model (Figure 5e). However, this same information was not decodable from the action-initiation excited neurons (Extended Data Figure 5-2). Thus, at the population level, DMS A2A$^+$ neurons can convey information about actions, action-rate, and outcomes.

DMS A2A$^+$ neurons realign their action encoding as habits form. We next asked how the activity and encoding of individual DMS A2A$^+$ neurons changes with action-outcome learning and habit formation. As with the entire A2A$^+$ population, a substantial proportion of tracked A2A$^+$ neurons were significantly modulated (Figure 5f), either excited (Figure 5g) or inhibited (Figure 5h), by action initiation across training. Fewer tracked A2A$^+$ neurons were excited by action termination or reward, though substantial proportions were inhibited by these events (Extended Data Figure 5-3). We classified neurons that were significantly excited by action initiation early in training. These early action-initiation excited A2A$^+$ neurons were more active (Figure 5i-j) and modulated (Figure 5k; Extended Data Figure 5-4) prior to action initiation than after, indicating they encode the initiation of an action. 43.12% (s.e.m. = 9.75%) of these neurons were also excited by action termination, indicating bracketing of the start and stop of action bouts (Extended Data Figure 5-5). Unlike D1$^+$ neurons, A2A$^+$ neurons lost their encoding of action initiation with training (Figure 5i-j). A2A$^+$ early action-initiation excited neurons became less modulated around action initiation with training (Figure 5k; Extended Data Figure 5-4). Moreover, neither the activity around action initiation of individual (Figure 5l), nor the population (Figure 5m) of early action-initiation excited A2A$^+$ neurons was significantly correlated across training phases. We were also unable to decode action rate across sessions (Figure 5n). Indeed, after overtraining only 15% (s.e.m. 8.68%) of the early action-initiation excited A2A$^+$ neuron ensemble continued to significantly encode action initiation (Figure 5o). Thus, early action-initiation excited DMS A2A$^+$ neurons transiently encode action initiation and this diminishes as habits form with overtraining. These neurons did not retune to become excited by other task events, though some became significantly inhibited by actions or rewards (Extended Data Figure 5-5).

Although the early action-initiation excited A2A$^+$ neurons stopped encoding actions, after overtraining there were neurons that encoded action initiation. We classified these overtrain action-initiation excited A2A$^+$ neurons and found that they, too, were more active (Figure 5p-q) and modulated (Figure 5r, Extended Data Figure 5-4) prior to action initiation. 38.06% (s.e.m. = 16.04%) of these neurons showed action-bracketing, also being excited

by action termination (Extended Data Figure 5-5). We looked retrospectively and found that activity around and modulation by action initiation of this A2A$^+$ ensemble grew with training (Figure 5p-r; Extended Data Figure 5-4). We did not find significant cross-session correlations between the activity of individual or the population of overtrain action-initiation excited A2A$^+$ neurons (Figure 5s-t) and were unable to decode action rate across sessions from these neurons (Figure 5u). Only about 19% (average 19.71-18.99%, s.e.m. 12.81-10.97%) of the overtrain action-initiation A2A$^+$ ensemble significantly encoded action initiation at earlier training phases (Figure 5v). Neurons that were incorporated into the overtrain action-initiation A2A$^+$ neurons ensemble were largely not excited by other events earlier in training, suggesting incorporation did not result from retuning (Extended Data Figure 5-5). Though, a small proportion of the incorporated neurons were inhibited by actions at earlier task phases. Both the early and overtrain populations of A2A$^+$ action-initiation excited neurons encoded action initiation with moderate trial-by-trial fidelity (Extended Data Figure 5-6).

Thus, whereas DMS D1$^+$ neurons stably encode action initiation with action-outcome learning and habit formation, one ensemble of A2A$^+$ neurons transiently encodes actions during early action-outcome learning and another slowly starts to encode actions as routine habits form. Indeed, significantly fewer A2A$^+$ than D1$^+$ neurons stably encoded action initiation (Figure 5w). A2A$^+$ ensembles also encoded action termination and this also reorganized with habit formation (Extended Data Figure 5-8).

DMS A2A$^+$ action-initiation inhibited neurons convey information about action rate. The DMS A2A$^+$ ensemble that was inhibited around action initiation was similarly unstable (Extended Data Figure 5-7). Both early in training (Figure 5x-y) and after overtraining (Figure 5z), a population of DMS A2A$^+$ neurons was inhibited around action initiation, but these were largely non-overlapping neurons (Extended Data Figure 5-7). Interestingly, however, and in contrast to the A2A$^+$ action-initiation excited ensemble (Extended Data Figure 5-3), action rate could be decoded from the activity of the overtrain action-initiation inhibited neurons (Figure 5aa). Thus, whereas DMS D1$^+$ action-initiation excited and inhibited neurons convey action rate, only the A2A$^+$ action-initiation inhibited neurons convey this information.

The ensemble of DMS A2A$^+$ neurons that encodes action initiation is more stable if habits do not form. We next asked whether habit formation contributes to the realignment of the DMS A2A$^+$ ensemble encoding action initiation by exploiting the serendipity that half of the A2A-cre subjects did not form habits with overtraining. We recorded and tracked similar numbers of A2A$^+$ neurons in these subjects (Extended Data Figure 5-9, Supplemental Table 1). 9 - 15% of DMS A2A$^+$ neurons were activated around action initiation, largely prior to action initiation in these subjects (Extended Data Figure 5-9). The activity and modulation of A2A$^+$ action-initiation excited neurons were stable across action-outcome learning and overtraining (Extended Data Figure 5-9). The activity around action initiation of the population of early action-initiation excited A2A$^+$ neurons was significantly correlated across training phases. In these subjects, 32% (s.e.m. = 17.32%) of the early and 28% (s.e.m. = 10.40%) of the overtrain action-initiation excited A2A$^+$ ensemble stably encoded action initiation. Thus, A2A$^+$ neurons more stably encode action with training if habits do not form, suggesting habit formation may contribute to the realignment of the A2A$^+$ ensemble encoding action initiation.

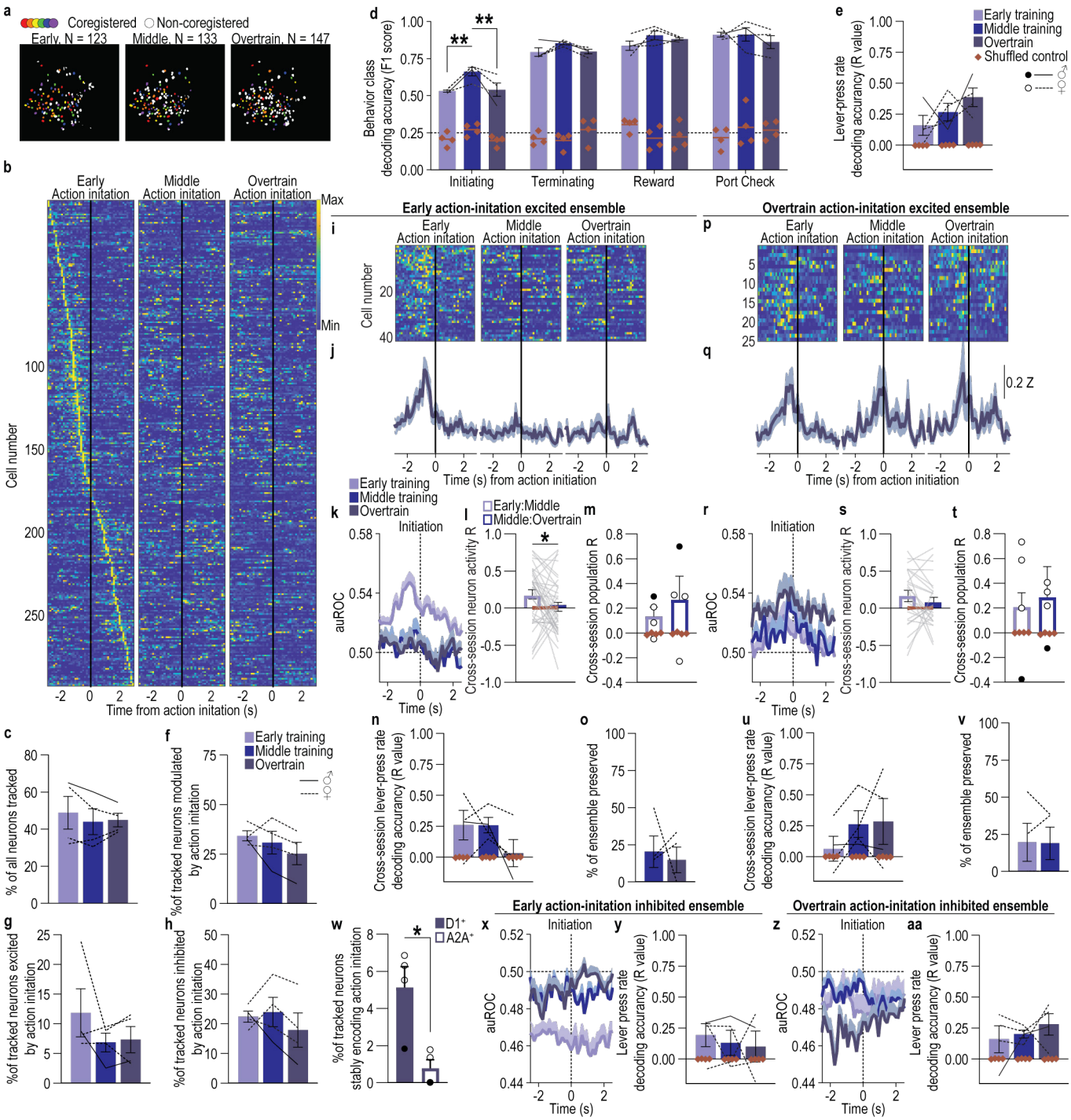


Figure 5: The ensemble of DMS A2A+ neurons that encodes action initiation shifts as habits form. (a) Representative recorded A2A$^+$ neuron spatial footprints during the first (early, left), 4th (middle), and 8th (overtrain) random-interval training session. Colored, co-registered neurons; white, non-co-registered neurons. **(b)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each coregistered DMS A2A$^+$ neuron around lever-press action initiation. **(c)** Percent of all neurons coregistered across training. 1-way ANOVA, $F_{1.18, 3.54} = 0.78$, $P = 0.45$. **(d)** Behavior class (initiating lever press, terminating press, reward collection, non-reinforced food-port check) decoding accuracy from A2A$^+$ coregistered neuron activity compared to shuffled control. Line at 0.25 = chance. 3-way ANOVA, Neuron activity (v. shuffled): $F_{1, 3} = 1570.57$, $P < 0.001$; Training session: $F_{2, 6} = 5.14$, $P = 0.05$; Behavior class: $F_{3, 9} = 23.33$, $P < 0.001$; Neuron activity x Training: $F_{2, 6} = 3.72$, $P = 0.09$; Neuron activity x Behavior Class: $F_{3, 9} = 17.65$, $P < 0.001$; Training x Behavior Class: $F_{6, 18} = 2.00$, $P = 0.12$; Neuron activity x Training x Behavior class: $F_{6, 18} = 1.66$, $P = 0.19$. **(e)** Lever-press rate decoding accuracy from A2A$^+$ coregistered neuron activity. R = correlation coefficient between actual and decoded press rate. 2-way ANOVA, Neuron activity: $F_{1, 3} = 72.91$, $P = 0.003$; Training: $F_{1.33, 4.00} = 1.91$, $P = 0.25$; Neuron activity x Training: $F_{1.35, 4.04} = 1.74$, $P = 0.27$. **(f-h)** Percent of coregistered neurons (Average 73.25 coregistered neurons/mouse, s.e.m. 17.09) significantly modulated (f; 1-way ANOVA, $F_{1.02, 3.05} = 2.09$, $P = 0.24$), excited (g; 1-way ANOVA, $F_{1.16, 3.48} = 1.79$, $P = 0.27$), or inhibited (h; 1-way ANOVA, $F_{1.05, 3.16} = 1.14$, $P = 0.37$) around action initiation. **(i-k)** Activity and modulation across training of DMS A2A$^+$ early action-initiation excited neurons (N = 42 neurons/4 mice; average 16.8 neurons/mouse, s.e.m. = 5.61). Heat map (i), Z-scored activity (j), and area under the receiver operating characteristic curve (auROC) modulation index (k) of early action-initiation excited neurons around action initiation across training. **(l-m)** Cross-session correlation of

the activity around action initiation of each early action-initiation excited neuron (l; 2-way ANCOVA, Neuron activity: $F_{1, 38} = 7.59$, $P = 0.009$; Training: $F_{1, 38} = 1.12$, $P = 0.30$; Activity x Time: $F_{1, 38} = 1.04$, $P = 0.32$) or the population activity of these neurons (m; 2-way ANOVA, Neuron activity: $F_{1, 3} = 3.07$, $P = 0.18$; Training: $F_{1, 3} = 0.68$, $P = 0.47$; Activity x Time: $F_{1, 3} = 0.64$, $P = 0.48$). **(n)** Cross-session decoding accuracy of lever-press rate from the activity of A2A$^+$ early action-initiation-excited neuron population activity on the 1st training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: $t_6 = 2.94$, $P = 0.08$, 95% CI -0.03 - 0.55; Middle: $t_6 = 2.92$, $P = 0.08$, 95% CI -0.03 - 0.55; Overtrain: $t_6 = 0.37$, $P > 0.9999$, 95% CI -0.26 - 0.33. **(o)** Percent of A2A$^+$ early action-initiation excited neurons that continued to be significantly excited by action initiation on the 4th and 8th training sessions. 2-tailed t-test, $t_3 = 0.37$, $P = 0.74$, 95% CI -53.97 - 42.86. **(p-r)** Activity and modulation across training of A2A$^+$ overtrain action-initiation excited neurons ($N = 25$ neurons/4 mice; average 6.25 neurons/mouse, s.e.m. 2.69). Heat map (p), Z-scored activity (q), and auROC modulation index (r) of overtrain action-initiation excited neurons around action initiation across training. **(s-t)** Cross-session correlation of the activity around action initiation of each overtrain action-initiation excited neuron (s; 2-way ANCOVA, Neuron activity: $F_{1, 23} = 7.13$, $P = 0.01$; Training: $F_{1, 23} = 0.18$, $P = 0.68$; Activity x Time: $F_{1, 23} = 0.14$, $P = 0.72$) or the population activity of these neurons (t; 2-way ANOVA, Neuron activity: $F_{1, 3} = 2.01$, $P = 0.25$; Training: $F_{1, 3} = 0.20$, $P = 0.69$; Activity x Time: $F_{1, 3} = 0.31$, $P = 0.62$). **(u)** Cross-session decoding accuracy of lever-press rate from the activity of A2A$^+$ overtrain action-initiation-excited neuron population activity on the 8th training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: $t_6 = 0.61$, $P > 0.999$, 95% CI -0.27 to 0.39; Middle: $t_6 = 2.63$, $P = 0.12$, 95% CI -0.07 - 0.59; Overtrain: $t_6 = 2.86$, $P = 0.09$, 95% CI -0.04 - 0.62. **(v)** Percent of A2A$^+$ overtrain action-initiation excited neurons that were also significantly excited by action initiation on 1st and 4th training sessions. 2-tailed t-test, $t_3 = 0.13$, $P = 0.31$, 95% CI -18.88 - 17.44. **(w)** Percent of all coregistered DMS D1$^+$ and A2A$^+$ neurons that are significantly modulated by action initiation across all phases of training ('stable action-initiation ensemble'). 2-tailed t-test, $t_3 = 3.55$, $P = 0.01$, 95% CI -7.34 - -1.35. **(x)** Modulation across training of A2A$^+$ early action-initiation inhibited neurons. **(y)** Accuracy with which lever-press rate can be can be decoded from the activity of A2A$^+$ early action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 7.44$, $P = 0.07$; Training: $F_{1.40, 4.21} = 0.17$, $P = 0.77$; Neuron activity x Training: $F_{1.41, 4.22} = 0.16$, $P = 0.79$. **(z)** Modulation across training of A2A$^+$ overtrain action-initiation inhibited neurons. **(aa)** Accuracy with which lever-press rate can be can be decoded from the activity of DMS A2A$^+$ overtrain action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 30.03$, $P = 0.01$; Training: $F_{1.52, 4.57} = 0.45$, $P = 0.61$; Neuron activity x Training: $F_{1.51, 4.54} = 0.48$, $P = 0.60$. A2A-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. *$P < 0.05$, **$P < 0.01$.

DMS A2A$^+$ neurons transiently support action-outcome learning, but do not mediate agency

DMS A2A$^+$ neuronal activity is necessary for action-outcome learning. Because a population of DMS A2A$^+$ neurons encoded actions during early action-outcome learning, we next asked whether this neuronal activity enables action-outcome learning to support agency. To test this, we chemogenetically inhibited DMS A2A$^+$ neurons during learning and then probed for action-outcome learning using the devaluation test. We expressed hM4Di or a fluorophore control selectively in DMS A2A$^+$ neurons of A2A-cre mice (Figure 6a-b). Mice were trained to lever press to earn food-pellet rewards (Figure 6c). Prior to each of the 4 random-interval training sessions, mice received CNO (2.0 mg/kg, i.p.) to, in hM4Di-expressing subjects, inactivate DMS A2A$^+$ neurons (Extended Data Figure 3-1). Chemogenetic inactivation of DMS A2A$^+$ neurons did not alter acquisition of the instrumental lever-press behavior (Figure 6d; see Extended Data Figure 6-1 for food-port entry and press/earned outcome data). Thus, DMS A2A$^+$ neurons are not necessary for general acquisition or execution of instrumental behavior. A2A$^+$ neuron inhibition did cause more entries into the food-delivery port, suggesting an inability to suppress this competing behavior (Extended Data Figure 6-1). It also disrupted action-outcome learning as evidenced by subsequent insensitivity to outcome devaluation (Figure 6e-f). Thus, like D1$^+$ neurons, DMS A2A$^+$ neurons are active during learning and this activity enables the action-outcome learning that supports agency for flexible, goal-directed decision making.

DMS A2A$^+$ neuronal activity is not sufficient to promote action-outcome learning. If, like D1$^+$ neurons, DMS A2A$^+$ neuronal activity drives action-outcome learning, then augmenting this activity should also promote the dominance of goal-directed behavioral control after overtraining. To test this, we chemogenetically activated DMS A2A$^+$ neurons during learning and then probed behavioral control strategy using the devaluation test. We expressed hM3Dq or a fluorophore control selectively in DMS A2A$^+$ neurons of A2A-cre mice (Figure 6g-h). Mice were trained to lever press to earn food-pellet rewards (Figure 6i) and were overtrained. Prior to each training session,

mice received CNO (0.2 mg/kg, i.p.) to, in hM3Dq-expressing subjects, activate DMS A2A$^+$ neurons (Extended Data Figure 3-1). Chemogenetic activation of DMS A2A$^+$ neurons did not alter acquisition of the instrumental lever-press behavior (Figure 6j). It also did not affect habit formation (Figure 6k-l). Both controls and subjects for which we activated DMS A2A$^+$ neurons during learning were insensitive to devaluation at test. Thus, unlike D1$^+$ neurons, DMS A2A$^+$ neuronal activity is not sufficient to drive action-outcome or disrupt habit formation.

DMS A2A$^+$ neuronal activity is not necessary to express agency. Since DMS A2A$^+$ neurons mediate action-outcome learning, we next asked whether they also support the use of such agency for flexible goal-directed decision making. To test this, we chemogenetically inhibited DMS A2A$^+$ neurons at test, rather than during learning. We expressed hM4Di or fluorophore control selectively in DMS A2A$^+$ neurons of A2A-cre mice (Figure 6m-n). Mice received limited training to lever press to earn food-pellet rewards (Figure 6o). All mice acquired the instrumental behavior (Figure 6p). Mice then received the outcome-specific devaluation test and were given CNO (2.0 mg/kg, i.p.) after sensory-specific satiety, prior to the lever-pressing probe test. Presses were expressed as a percent of baseline (training session prior to test) to normalize for the pre-existing differences in press rate between groups (see Extended Data Figure 6-2 for raw press rate data). Inactivation of DMS A2A$^+$ neurons did not affect flexible, goal-directed decision making, as evidenced by sensitivity to devaluation in both groups (Figure 6q-r). Thus, whereas DMS D1$^+$ neurons are fundamental for both action-outcome learning and flexible goal-directed decision making, A2A$^+$ neurons only transiently support action-outcome learning and become unnecessary for using such agency for flexible, goal-directed decision making.

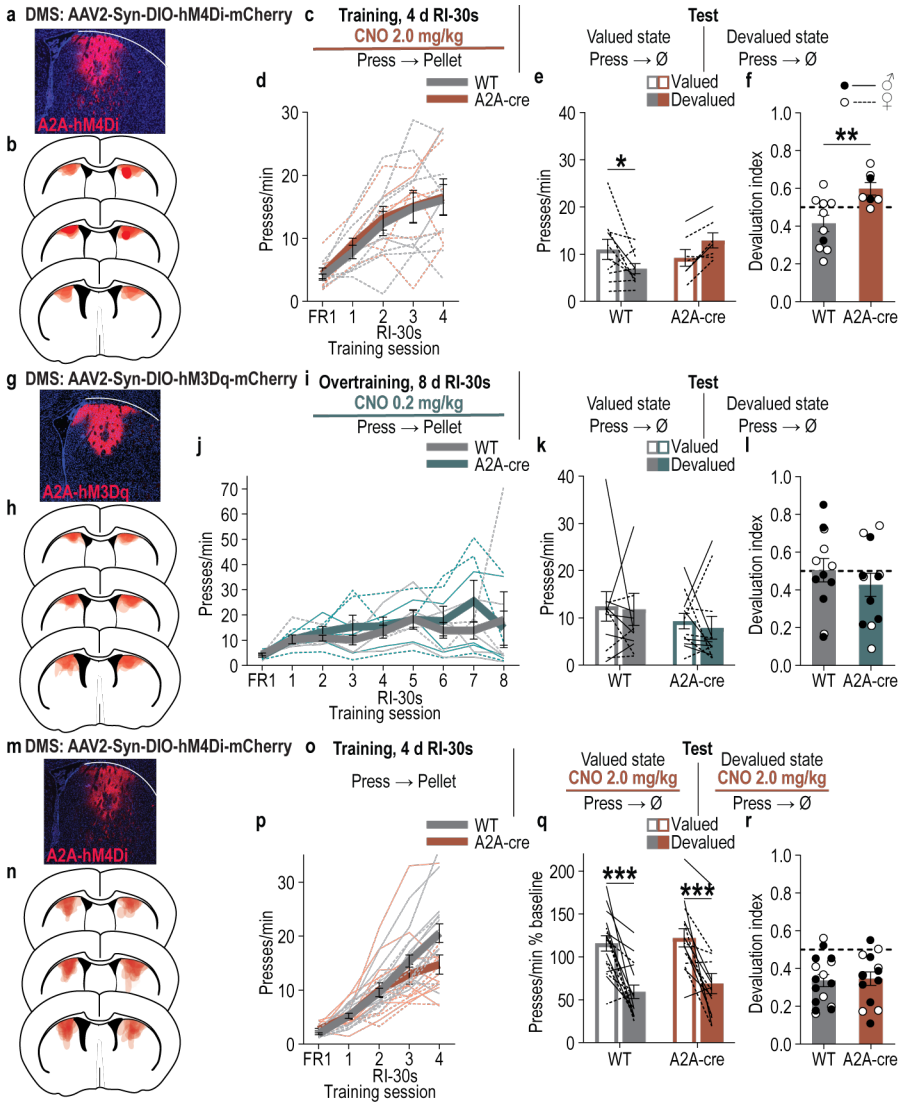


Figure 6: DMS A2A⁺ neurons are necessary for action-outcome learning, but not goal-directed decision making. (a-f) Chemogenetic inactivation of DMS A2A⁺ neurons during learning. WT: N = 10 (1 male); A2A-cre: N = 7 (2 males). **(a)** Representative immunofluorescent image of cre-dependent hM4Di expression in DMS. **(b)** Map of DMS cre-dependent hM4Di expression for all subjects. **(c)** Procedure. RI, random-interval reinforcement schedule; CNO, clozapine-N-oxide. **(d)** Training press rate. 2-way ANOVA, Training: $F_{2.39, 35.84} = 30.54$, $P < 0.0001$; Genotype: $F_{1, 15} = 0.07$, $P = 0.79$; Training x Genotype: $F_{4, 60} = 0.05$, $P = 0.99$. **(e)** Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 15} = 9.67$, $P = 0.007$; Value: $F_{1, 15} = 0.02$, $P = 0.88$; Genotype: $F_{1, 15} = 0.96$, $P = 0.34$. **(f)** Devaluation index [(Devalued condition presses)/(Valued condition presses ⁺ Devalued presses)]. 2-tailed t-test, $t_{15} = 3.10$; $P = 0.007$, 95% CI 0.057 - 0.31. **(g-l)** Chemogenetic activation of DMS A2A⁺ neurons during overtraining. WT: N = 12 (7 males); A2A-cre: N = 12 (6 males). **(g)** Representative immunofluorescent image of cre-dependent hM3Dq expression in DMS. **(h)** Map of DMS cre-dependent hM3Dq expression. **(i)** Procedure. **(j)** Training press rate. 2-way ANOVA, Training: $F_{1.87, 41.22} = 15.50$, $P < 0.0001$; Genotype: $F_{1, 22} = 0.80$, $P = 0.38$; Training x Genotype: $F_{8, 176} = 0.25$, $P = 0.98$. **(k)** Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 22} = 0.03$, $P = 0.87$; Value: $F_{1, 22} = 0.20$, $P = 0.66$; Genotype: $F_{1, 22} = 1.30$, $P = 0.27$. **(l)** Devaluation index. 2-tailed t-test, $t_{22} = 0.89$; $P = 0.38$, 95% CI -0.26 - 0.10. **(m-r)** Chemogenetic inactivation of DMS A2A⁺ neurons during test of behavioral control after learning. WT: N = 16 (9 males); A2A-cre: N = 14 (8 males). **(m)** Representative immunofluorescent image of cre-dependent hM4Di expression in DMS. **(n)** Map of DMS cre-dependent hM4Di expression. **(o)** Procedure. **(p)** Training press rate. 2-way ANOVA, Training: $F_{1.78, 49.82} = 107.5$, $P < 0.0001$; Genotype: $F_{1, 28} = 1.30$, $P = 0.26$; Training x Genotype: $F_{4, 112} = 5.08$, $P = 0.008$. **(q)** Test press rate normalized to pre-test training baseline. 2-way ANOVA, Value x Genotype: $F_{1, 28} = 0.04$, $P = 0.84$; Value: $F_{1, 28} = 47.56$, $P < 0.0001$; Genotype: $F_{1, 28} = 0.50$, $P = 0.49$. **(r)** Devaluation index. 2-tailed t-test, $t_{28} = 0.22$; $P = 0.83$, 95% CI -0.09 to 0.11. Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

DISCUSSION

Here we investigated the activity and function of two major DMS neuron populations during instrumental learning. The data reveal DMS cell-type specific stability and reorganization for action-outcome learning and agency. An ensemble of DMS D1⁺ neurons stably encodes actions during learning and develops encoding of reward

outcomes. Accordingly, D1$^+$ neurons are fundamental for both action-outcome learning and the use of such agency for flexible, goal-directed decision making. Conversely, the ensemble of DMS A2A$^+$ neurons that encodes actions realigns from initial action-outcome learning to habit formation and A2A$^+$ neurons only transiently support action-outcome learning. Thus, early in learning both DMS neuron subtypes encode actions to support the development of the action-outcome model underlying agency. But, whereas D1$^+$ neurons maintain their encoding to support agency and model-based decision making, A2A$^+$ neurons reorganize their action encoding with habit formation.

DMS D1$^+$ neurons stably encode actions to support action-outcome learning and agency. Consistent with prior reports[50, 62, 70, 71, 82, 83], we found that DMS D1$^+$ neurons encode the initiation and termination of instrumental actions, as well as earned reward outcomes. This encoding occurs during both action-outcome learning and habit formation, consistent with evidence of DMS neuronal activity after habit formation[84]. Correspondingly, we found that the D1$^+$ neuron population conveys information about actions, action rate, and outcomes. Critically, we discovered a subpopulation of DMS D1$^+$ neurons that stably encode action initiation throughout action-outcome learning and habit formation. Most of these neurons also developed encoding of action termination, thereby bracketing the start and stop of action sequences, as detected in striatal neurons previously[85-87]. Many of them also developed encoding of reward outcomes, consistent with prior reports of DMS encoding of action-outcome contingency[71]. An important question for future investigation is the information conveyed by the activity of this ensemble prior to action initiation. Action value is a likely possibility[83, 88, 89]. Regardless, this encoding profile well positions DMS D1$^+$ neurons to support action-outcome learning and, therefore, agency. Indeed, we found that DMS D1$^+$ neurons mediate both action-outcome learning and the application of this learned agency for flexible, goal-directed decision making. This accords with prior evidence of enhanced DMS D1$^+$ post-synaptic plasticity with action-outcome learning[90], the necessity of direct pathway striatal projection neurons for action-outcome learning[91], and that DMS direct pathway[92] and D1$^+$[88] neurons can promote goal-directed choices. Moreover, we found that D1$^+$ neuron activity is sufficient to promote action-outcome learning to enable the dominance of flexible, goal-directed decision making over inflexible habitual behavioral control, consistent with evidence that activation of DMS D1$^+$ neurons prevents habit formation[93]. Thus, DMS D1$^+$ neurons drive the action-outcome learning that supports agency for adaptive decision making. They do so, at least in part, by stably encoding actions.

DMS A2A$^+$ neurons also encode actions, but reorganize their action encoding as habits form. Like D1$^+$ neurons and prior reports[62, 70, 83], DMS A2A$^+$ neurons encode action initiation and termination across phases of instrumental learning. Unlike D1$^+$ neurons, A2A$^+$ neurons show little encoding of reward outcomes, consistent with prior reports[82], and quiet their activity during earned reward. Information about each actions, action rate, and outcomes could be decoded from DMS A2A$^+$ neurons. This accords with prior evidence that both striatal outputs represent the action space[89]. Critically, we discovered that A2A$^+$ neurons do not stably encode actions. During early action-outcome learning an ensemble of DMS A2A$^+$ neurons encodes actions. Correspondingly, DMS A2A$^+$ neurons support early action-outcome learning. With continued training this ensemble ceases to encode actions. Instead, another ensemble of A2A$^+$ neurons gradually gains action encoding as habits form. This subpopulation realignment

of action encoding is associated with habit formation. Realignment was incomplete if habits did not form with overtraining. DMS A2A$^+$ neurons have been implicated in habit[94-96]. This reorganization of action encoding may, therefore, contribute to a shift in the function of DMS A2A$^+$ neurons. Indeed, although DMS A2A$^+$ neurons initially support action-outcome learning, they become unnecessary for the expression of such agency for goal-directed decision making. Their activity is also not sufficient to promote action-outcome learning or prevent habit formation. Thus, the DMS A2A$^+$ neuron subpopulation encoding actions depends on the strategy generating those actions, e.g., whether it is driven by an internal action-outcome model or habit.

The reorganization of DMS A2A$^+$ neuron encoding of action could reflect a shift of encoding from one A2A$^+$ neuron subtype to another. Our finding that DMS A2A$^+$ neurons are necessary for action-outcome learning contrasts with prior reports that projection-specific chemogenetic inactivation of DMS indirect pathway neurons does not affect such learning[91]. Although other differences (species, pretraining procedures, option space, transduction efficiency) likely contribute to this discrepancy, it could be explained by differences in targeting. If DMS indirect striatopallidal projections are not necessary for action-outcome learning, then the DMS A2A$^+$ neurons we find that function in early action-outcome learning may be distinct subtype[56, 97-99], perhaps even a non-canonical projection. A2A$^+$ neurons that encode actions during action-outcome learning could be a unique subtype from those that encode habitual actions. Such speculation requires further investigation coupling functional, molecular, and projection-target profiling. Regardless, the reorganization of A2A$^+$ neuron action encoding accords well with functional evidence that these neurons mediate second forms of learning, including action-outcome reversal[91, 100], extinction[100], and habit[94-96].

These data update our understanding of DMS function in decision making by revealing that both D1$^+$ and A2A$^+$ neurons support DMS function in action-outcome and model-based learning[44, 45, 50, 52, 101], but only D1$^+$ neurons mediate DMS function in prospective use of action-outcome models for goal-directed decision making[9, 52, 102]. Our results are consistent with the notion that striatal D1$^+$ and D2/A2A$^+$ neurons have coordinated function. Striatal D1$^+$ and A2A$^+$ neurons can have similar activity during trained and untrained movements[62, 103]. Correspondingly, we find that both are active prior to instrumental actions. One view is that striatal D1$^+$ neurons drive selected actions and D2/A2A$^+$ neurons permit such actions by inhibiting competing or unrewarded actions[58, 62, 63, 70, 72, 74, 87, 103-107]. Consistent with the former, we found that action rate can be decoded from the action-initiation excited DMS D1$^+$ neuronal ensemble, that D1$^+$ neurons drive action-outcome learning and agency, and an ensemble of them stably encode actions. Aligning with the latter, action rate can only be decoded from the A2A$^+$ neurons that are inhibited by action initiation and inhibition of DMS A2A$^+$ neurons causes more competing food-port checking behavior. By inhibiting competing behaviors, DMS A2A$^+$ neurons could help establish appropriate action-outcome relationships, underlying their early necessity for such learning. Consistent with this, DMS D2/A2A$^+$ neurons help to modify behavioral programs encoded in D1$^+$ neurons[100]. Once agency is established, DMS A2A$^+$ neurons become dispensable, a function maintained by DMS D1$^+$ neurons. Thus, agency may rely on early coordination of DMS D1$^+$ and A2A$^+$ neurons to establish the action-outcome model and then maintenance of such encoding by D1$^+$ neurons to support decision making. Habit formation is associated with the emergence of

a distinct action-encoding DMS A2A$^+$ subpopulation. This view extends the coordinated D1$^+$ and D2/A2A$^+$ function model beyond movement execution to learning, in this case, of agency.

The discoveries here open the door to many important future questions. At the top of this list are the mechanisms that underlie the stability and reorganization of action encoding in DMS D1$^+$ and A2A$^+$ neurons. Such mechanisms are likely to be multifaceted, including, molecular and epigenetic factors[108-110], cellular plasticity[90], and circuit-level regulation[50, 110-112]. Towards the latter, cortical input from the prelimbic and orbitofrontal cortex are likely involved, perhaps especially in the D1$^+$ activity underlying action-outcome learning[50, 110, 111, 113]. Input from the basolateral and central amygdala may also help shape DMS D1$^+$ and/or A2A$^+$ neuron activity stability and reorganization with learning and habit formation[112]. Dopamine input is also, undoubtedly involved[114, 115]. Another set of open questions is how other striatal cell types, including interneurons[116] and astrocytes[117], participate and interact with DMS D1$^+$ and A2A$^+$ neurons to support agency and habit. And, further, whether and how D1$^+$ and A2A$^+$ neuron collateral interactions are involved[95, 118, 119]. An important larger question is whether stable D1$^+$ neuron encoding and/or A2A$^+$ subpopulation encoding realignment is a principle of striatal contributions to learning. Would such a pattern also occur in the DMS for goal-directed avoidance learning? Or in the dorsolateral striatum for habit formation? Further studies are needed to test this potential broader implication.

Adaptive decision making often requires understanding your agency. Knowing that your actions can produce particular consequences and using this to make thoughtful, deliberate, goal-directed decisions. Here we find cell-type specific striatal stability and reorganization underlying action-outcome learning, agency, and habit. DMS D1$^+$ neurons stably encode actions to support action-outcome learning and agency. DMS A2A$^+$ neurons encode actions to enable the initial development of the action-outcome model, but reorganize their action encoding with habit formation. These data provide neuronal circuit insights into how we learn and, thus, how we decide. This helps understand how stress[112, 113], exposure to drugs and alcohol[27, 51, 120], neurodevelopmental factors[105, 121], and aging[122] can lead to the disrupted decision making and pathological habits that characterize substance use disorders and mental illness.

AUTHOR CONTRIBUTIONS

MM and KMW conceptualized and designed the experiments, interpreted the data, and wrote the paper. MM executed all experiments and analyzed the data. AL, NP, JYG, MDM, WG, Alicia W, MS, Anna W, JSP assisted with experiments. BSH conducted the on-devaluation-test chemogenetic inactivation experiments. SPB conducted the behavioral pilot experiment. JRG conducted the electrophysiological validation experiments with support from SMH, CC, and MSL. NKG managed the breeding colonies and assisted with histological verification. GJB, ACS, HTB, and AMW provided guidance and assistance with data analysis. AMW provided guidance on data interpretation.

ACKNOWLEDGEMENTS

This research was supported by NIH R01DA046679 (KMW), NIH R01DA058374 (KMW), NIH T32DA024635 (JRG), NIH F32DA056201 (JRG), A.P. Giannini Fellowship (JRG), NIH K99MH135177 (JRG), NIH TL4GM118977 (NP), NIH F32MH135680 (GJB), NSF NeuroNex 1707408 (HTB), and the Staglin Center for Behavior and Brain Sciences.

COMPETING FINANCIAL INTERESTS

The authors have no biomedical financial interests or potential conflicts of interest to declare.

METHODS

Subjects

Male wildtype C57/Bl6J mice (Jackson Laboratories, Bar Harbor, ME) were used for the behavioral pilot experiment. All other experiments were conducted with male and female Drd1a-Cre[123] and Adora2A-Cre[124] transgenic mice and their wildtype littermates bred in house. Mice were between 9 - 16 weeks old at the time of experiment onset/surgery. Mice were housed in a temperature (68 - 79 °F) and humidity (30 - 70%) regulated vivarium on a 12:12 hr reverse dark/light cycle (lights off at 7 AM). Mice were initially housed in same-sex groups of 3 - 4 mice/cage. Prior to experiment onset/surgery mice were single-housed to facilitate food deprivation and preserve implants. Unless noted below, mice were provided with food (standard rodent chow, Lab Diet, St. Louis, MO) and water *ad libitum* in the home cage. Mice were handled for 3 - 5 days prior to the start of behavioral training for each experiment. All procedures were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the UCLA Institutional Animal Care and Use Committee.

Surgery

Mice were anesthetized with isoflurane (3% induction, 1% maintenance), and positioned in a digital stereotaxic frame (Kopf, Tujunga, CA). Subcutaneous Rimadyl (Carprofen; 5 mg/kg; Zoetis, Parsippany, NJ) was given pre-operatively for analgesia and anti-inflammatory purposes. An incision was made along the midline to expose the skull. For viral infusions, after performing a small craniotomy, virus was injected using a 28-g infusion needle (PlasticsOne, Roanoke, VA) connected to a 1-mL syringe (Hamilton Company, Reno, NV) by intramedic polyethylene tubing (BD; Franklin Lakes, NJ) and controlled by a syringe pump (Harvard Apparatus, Holliston, MA). Virus was injected at a rate of 0.1 µl/min and the needle was left in place for 10 min post-injection. Further experiment-specific surgical details are provided below. After surgery, mice were kept on a heating pad maintained at 35 °C for 1 hr and then single-housed in a clean homecage for recovery and monitoring. Mice received chow containing the antibiotic TMS for 7 days following surgery to prevent infection, after which they were returned to standard rodent chow.

Behavioral procedures

Apparatus. Training took place in Med Associates (East Fairfield, VT) wide mouse operant chambers housed within sound- and light-attenuating boxes. Each chamber had metal grid floors and contained a retractable lever to the left of a recessed food-delivery port (magazine) on the front wall. A photobeam entry detector was positioned at the entry to the food port. Each chamber was equipped with 2 pellet dispensers to deliver either 20-mg grain or chocolate-flavored purified pellets (Bio-Serv, Frenchtown, NJ) into the food port when activated. A fan mounted to the outer chamber provided ventilation and external noise reduction. A 3-watt, 24-volt house light mounted on the top of the back wall opposite the food port provided illumination. To monitor subject behavior, monochrome digital cameras (Med Associates) were positioned with a top-down view of the conditioning chambers.

Food deprivation. 3 - 5 days prior to the start of behavioral training, mice were food-deprived to maintain 85% - 90% of their free-feeding body weight. Mice were given 1.5 - 3.0 g of their home chow at the same time daily at least 2 hrs after training.

Outcome pre-exposure. To familiarize subjects with the food pellet that would become the instrumental outcome, mice were given 1 session of outcome pre-exposure. Mice were placed in a clean, empty cage and allowed to consume 30 of the food pellets from a metal cup. If any pellets remained, they were placed in the home cage overnight for consumption.

Magazine conditioning. Mice received 1 session of training in the operant chamber to learn where to receive the food pellets (20-mg grain or chocolate-purified pellets). Mice received 30 non-contingent pellet deliveries from the food port with a fixed 60-s intertrial interval.

Instrumental Training. Mice next received 1 session/day consecutively of instrumental conditioning in which lever presses earned delivery of a single food pellet. Earned pellet type (grain or chocolate) was counterbalanced across subjects within each group of each experiment. Each session began with the illumination of the house light and insertion of the lever and ended with the retraction of the lever and turning off of the house light. Each training session ended after 30 outcomes had been earned or 30 min elapsed. In all cases, instrumental training began on a fixed-ratio 1 schedule (FR-1), in which each action was reinforced with one food pellet outcome. Once 90% of the maximum session outcomes were earned, the reinforcement schedule was shifted to random-interval (RI) in which a variable interval must elapse following a reinforcer for another press to be reinforced.

Mice received either 4 or 8 (overtrained) consecutive sessions of RI training starting first with 1 session on a RI-15s schedule, and then the remaining sessions on a RI-30s schedule.

Alternate outcome exposures. To equate exposure of the non-trained pellet, all mice were given non-contingent access to the same number of the alternate food pellets as the earned pellet type (e.g., chocolate pellets if grain pellets served as the training outcome) in a different context (clear plexiglass cage).

Sensory-specific satiety outcome devaluation test. Testing began 24 hr after the final instrumental conditioning session. Mice were given 1 - 1.5 hr access to either 4 g of the food pellets previously earned by lever pressing (Devalued condition) or 4 g of the non-trained pellets to control for general satiety (Valued condition). The remaining pellets were weighed following prefeeding to measure total consumption. Consumption did not significantly differ between the Devalued v. Valued conditions or between the control and experimental group for any experiment (Supplemental Table 2). Immediately after this prefeeding, lever pressing was assessed during a brief, 5-min, non-reinforced probe test. To assess the efficacy of the sensory-specific satiety, following the probe test, mice were given a 10-min consumption choice test with simultaneous access to 1 g of both pellet types. To ensure that lever-press test data was not confounded by incomplete devaluation, subjects that failed to fully reject the devalued food type (i.e., consumed more than our margin of measurement error, 0.09 g, of the pre-fed pellet type during post-test consumption) were not included in the analysis (see exclusions noted for each experiment below). The remaining mice consumed less of the prefed pellet than non-prefed pellet, indicating successful sensory-specific satiety devaluation (Supplemental Table 3). 24 - 48 hr after the first devaluation test, mice received 1 session of instrumental retraining (RI-30s), followed the next day by a second devaluation test in which they were prefed the opposite food pellet. Thus, each mouse was tested in both the Valued and Devalued conditions, with test order counterbalanced across subjects within each group for each experiment.

Establishing goal-directed behavioral control following limited instrumental random-interval training and habit following overtraining

Naïve, male C57BL/6J mice (8 weeks old, Jackson Laboratory) were used to establish the training protocols for action-outcome learning and goal-directed decision making (N = 8) or habit formation with overtraining (N = 8). No subjects were excluded. Mice were food restricted and then began instrumental training, as described above. Mice were habituated to intraperitoneal (i.p.) injections during the final day of FR-1 training. Then, all subjects received an i.p. injection of 0.9% saline following each of the RI training sessions. Following training, mice received a counterbalanced pair of sensory-specific satiety outcome-specific devaluation tests, as above. No injection was given on test.

Microendoscopic calcium imaging

Naïve, male and female Drd1a-cre and A2A-cre mice (D1-cre: Final N = 4, 1 male; A2A-cre habitual N = 4, 1 male; A2A-cre goal-directed N = 4, 3 male) were used in this experiment to monitor calcium activity in individual $D1^+$ or $A2A^+$ DMS striatal projection neurons during instrumental conditioning. 18 (D1-cre: 9, A2A-cre: 9) subjects with lack of viral expression or without detectable individual neurons were excluded from the experiment. 3 (D1-cre: 2, A2A-cre: 1) subjects with viral overexpression were excluded from the dataset. At surgery, mice were infused with an adeno-associated virus (AAV) encoding the cre-dependent genetically encoded calcium indicator jGCaMP7s (0.40 µl; AAV9-syn-Flex-jGCaMP7s) into the DMS (AP +0.15, ML ±1.9, DV −2.65 mm from bregma). In a separate surgery, 4 - 7 days later, a unilateral 1.1-mm craniotomy (left/right hemisphere counterbalanced across subjects) was performed, centered above the DMS (AP +0.20, ML ±1.70 mm from bregma). Cortex was aspirated with a 27 - 30-g needle to a depth of approximately 1.8 mm. A 1-mm diameter, 4-mm long gradient refractive index (GRIN) lens (Inscopix, Palo Alto, CA) was implanted above the DMS (DV -2.25 mm). The lens was fixed to the skull with cyanoacrylate glue and secured with C&B Metabond quick adhesive cement system (Parkell Inc., Edgewood, NY), followed by opaque dental cement (Lang Dental Manufacturing, Wheeling, IL). Approximately 3 - 4 weeks following implant, a baseplate was attached to the previously formed headcap, ensuring proper alignment between the GRIN lens and the miniscope.

Mice were habituated to restraint and fitted with the miniscope for 2 days prior to behavioral training. Mice received a second magazine training session to become accustomed to retrieving pellets with the head-mounted miniscope. After the first FR-1 instrumental session, mice were subsequently fitted with the miniscope prior to each of 8 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. We noticed that not all of the A2A-cre subjects developed habits (insensitivity to devaluation) with overtraining. We, therefore, used a mean split to separate these subjects into

those that formed habits (Devaluation index ≥ 0.483; N = 4) and those that remained sensitive to devaluation showing goal-directed decision making (Devaluation index < 0.483; N = 4). These groups did not differ in their acquisition of lever pressing during training (Training session: $F_{1.51, 9.04} = 8.89$, $P = 0.01$; Group: $F_{1, 6} = 0.48$, $P = 0.52$; Training x Group: $F_{8, 48} = 1.02$, $P = 0.43$). They also did not differ in the efficacy of devaluation (see Supplemental Tables 2 - 3).

Miniscope system. Calcium imaging videos were recorded using an open-source, head-mounted UCLA Miniscope imaging device, which interfaces with the UCLA open-source Miniscope DAQ hardware and software (http://miniscope.org/) to capture the neuronal calcium dynamics with corresponding frame timestamps. A webcam (Logitech, San Jose, CA) recorded the Med Associates interface LEDs, which display light pulses coincident with task events (lever presses, food-port entry, pellet delivery). This video was simultaneously saved on the same computer alongside the calcium imaging video via the miniscope software with synchronized frame timestamps.

Calcium trace extraction and spike inference. Videos were initially processed using custom Python code[125] to concatenate all the videos from a session into one tiff stack, spatially downsample by a factor of 2, and temporally downsample to an effective frame rate of ~7.5 frames/s to accelerate image processing. The Python implementation of the Calcium Imaging Analysis package, 'CaImAn'[126], was then used to process the videos. We applied the nonrigid motion correction algorithm to remove tissue motion artifacts and then applied constrained non-negative matrix factorization for microendoscopic data to extract individual spatial contours and demix temporal fluorescence traces for each detected neuron. The deconvolved traces, the estimation of spikes that generate the calcium traces, were derived from the denoised fluorescence traces using CaImAn's '*deconvolveCa*' function. Neurons were coregistered across sessions using the open source package CellReg[127]. Activity for each neuron was Z-scored to the whole session to allow comparison of baseline and phasic activity across cells and subjects with different levels of GCaMP expression and/or lens proximity.

Behavior extraction. All behavior-event videos from a session were concatenated and downsampled as above. Using custom MATLAB code, the brightness of each LED on the interface was measured for each frame. A task event was considered to have occurred if the LED brightness was >2x the background brightness threshold (determined individually per recording), thus binarizing task events in each frame of the video. The extracted frame-by-frame event information was then interpolated to the timestamps of the calcium imaging video.

Cell classification. We used area under the receiver operating characteristic (auROC) analysis to identify neurons tuned to key behavioral events: 'action initiation', first press in session, first press after earned reward, and first press after a non-reinforced food-port check; 'action termination', last press before earned reward collection and last press before a non-reinforced food-port check; 'food-port check', entries into the food delivery port, separated for those in which there was a reward present v. non-reinforced entries; 'reward consumption', consumption of the earned food pellet occurring at the time of exit from the magazine. After aligning each neuron's activity to key events, ROC curves were computed by comparing the calcium responses across the occurrences of the behavioral event versus the calcium responses at equated baseline periods in which subjects were not engaged in the task. Calcium data was circularly shifted by random increments 1000 times to generate a distribution of null auROC values. Neurons with true auROC values >95th percentile of the distribution of shuffled auROCs within 2.9 s before or after the event were classified as significantly modulated by a specific behavior event.

Behavior decoding. We used support vector machines (SVM) to decode key behavioral events (action initiation, action termination, food-port check, and reward consumption) using the activity of specific neuron populations. To predict behavior events prior to onset, observation data used to train and test the classifier consisted of the 0.4-s epoch preceding each of the 4 distinct event labels. Since there were large variations in the number of each type of behavioral event, we controlled for the number of each event label. First, for the behavior event labels that were overrepresented (usually food-port entries), we found the median number for all event labels, and limited those event labels to that value. Then, for the behavior event labels that were underrepresented, we used the synthetic minority oversampling technique (SMOTE)[128] to generate synthetic activity samples to match the event label with the highest number of events. For each session, the classifier was trained (MATLAB function *fitcecoc*) on ¾ of the data, randomly selected, to generate a model and behavior was predicted (MATLAB function *predict*) with the generated model on the remaining data. Decoding performance was averaged across 20 different random data splits. Decoding performance was assessed by computing the F-score, a standard measure of binary classification accuracy that weighs recall (proportion of actual positives that are correctly predicted) against precision (the proportion of positive predictions that are actually positive) of the classifier. It is

calculated as the harmonic mean of precision and recall. We performed a 1000-fold shuffling procedure to confirm that the average decoding of randomized data is at the theoretical chance level of ¼ correct.

To examine whether lever-press rate could be explained by the neuronal activity, we fit a linear model (MATLAB function: *fitglm*) using the temporal fluorescence traces as predictors and behavior rate as the response. The discrete lever presses were transformed into an ongoing behavior rate by convolving the lever presses with a half-gaussian-post event kernel width of 60 s and a standard deviation of the Gaussian distribution of 6. The predictor and response data were Z-scored. A model was generated using the initial ¾ of the data and behavior rates were predicted (MATLAB function *predict*) with the generated model on the remaining data. Decoding performance was assessed by calculating the correlation between the predicted behavior rate and the actual behavior rate. We performed a 1000-fold circular shuffling procedure to compare decoding accuracy to that of randomized data.

Modulation fidelity. To assess cross-session response fidelity of individual neuron or population responses peri-behavior event, we calculated the average response of each neuron or the average population response, respectively, across trials for each session and calculated the Spearman correlation of the responses between sessions. To assess the within-session reliability of the ensemble to a specific behavior event, we quantified the percentage of ensemble neurons as a function of the percentage of behavior trials to which they responded for each training phase, using a response threshold of 2x standard deviation of the trace during the 2 s prior to behavior event onset. The correlation of this response distribution between sessions was measured to assess consistency across training. To assess the within-session reliability of individual neuron responses, we calculated the average Spearman correlation between behavior trials. We performed a 1000-fold circular shuffling procedure to compare modulation fidelity to that of randomized data.

Chemogenetic inhibition of DMS D1$^+$ neurons during instrumental learning

Naïve, male and female Drd1a-cre mice (Final *N* = 9, 5 males) and wildtype littermate controls (*N* = 7, 2 males) served as subjects in this experiment to assess the necessity of DMS D1$^+$ neuron activity for the action-outcome learning that supports goal-directed decision making. 3 D1-cre subjects with off-target viral expression were excluded from the dataset. 3 (D1-cre: 2, WT: 1) subjects were excluded for ineffective sensory-specific satiety. At surgery, all mice received bilateral infusion of AAV encoding a cre-inducible inhibitory designer receptor human M4 muscarinic receptor (hM4Di; AAV2-Syn-DIO-hM4Di-mCherry; 0.3 µl) into the DMS (AP +0.2, ML ±1.8, DV −2.65 mm from bregma). After 2 weeks of recovery, mice were food restricted and then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Then, all subjects received an i.p. injection of water-soluble clozapine-n-oxide (CNO; 2.0 mg/kg; Hello Bio, Princeton, NJ) 30 min prior to each of the 4 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. No CNO was given on test. CNO was given prior to the retraining session (RI-30s) in between tests.

Chemogenetic activation of DMS D1$^+$ neurons during instrumental overtraining

Naïve, male and female Drd1a-cre mice (Final *N* = 6, 3 males) and wildtype littermate controls (*N* = 6, 4 males) served as subjects in this experiment to assess whether DMS D1$^+$ neuron activity is sufficient to promote action-outcome learning for goal-directed behavioral control and prevent the habit formation that normally occurs with overtraining. 6 D1-cre subjects with off-target viral expression were excluded from the dataset. 4 (D1-cre: 1, WT: 3) subjects were excluded for ineffective sensory-specific satiety. 1 WT subject that ate twice as much prior to one test than the other was excluded from the dataset to prevent the confound of differential general satiety. At surgery, all mice received bilateral infusion of AAV encoding a cre-inducible excitatory designer receptor human M3 muscarinic receptor (hM3Dq; AAV2-Syn-DIO-hM3Dq-mCherry; 0.3 µl) into the DMS (AP +0.2, ML ±1.8, DV −2.65 mm from bregma). After 2 weeks of recovery, mice were food restricted and then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Then, all subjects received CNO (0.2 mg/kg i.p.; Hello Bio, Princeton, NJ) 30 min prior to each of 8 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. No CNO was given on test days. CNO was given prior to the retraining session (RI-30s) in between tests.

Chemogenetic inhibition of DMS D1$^+$ neurons during goal-directed decision making

Naïve, male and female Drd1a-cre mice (Final *N* = 8, 2 males) and wildtype littermate controls (*N* = 12, 7 males) served as subjects in this experiment to assess whether DMS D1$^+$ neuron activity is necessary for the expression of goal-directed decision making. 7 D1-cre subjects with off-target viral expression were excluded from the

dataset. At surgery, all mice received bilateral infusion of AAV encoding cre-inducible hM4Di (AAV2-Syn-DIO-hM4Di-mCherry; 0.3 µl) into the DMS. After 2 weeks of recovery, mice were food restricted and then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Mice received an i.p. injection of 0.9% saline 30 min prior to each of 4 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. Immediately after the pre-feeding, all mice received CNO (2.0 mg/kg i.p.) and commenced to the non-rewarded probe test 30 min later. Saline was given prior to the retraining session (RI-30s) in between tests.

Chemogenetic inhibition of DMS A2A$^+$ neurons during instrumental learning

Naïve, male and female A2A-cre mice (Final *N* = 7, 2 males) and wildtype littermate controls (*N* = 10, 1 male) served as subjects in this experiment to assess whether DMS A2A$^+$ neuron activity is necessary for the action-outcome learning that supports goal-directed decision making. 3 A2A-cre subjects with off-target viral expression were excluded from the dataset. 3 (A2A-cre: 1, WT: 2) subjects were excluded for ineffective sensory-specific satiety. At surgery, all mice received bilateral infusion of AAV encoding cre-inducible hM4Di (AAV2-Syn-DIO-hM4Di-mCherry; 0.3 µl) into the DMS. After 2 weeks of recovery, mice were food restricted and then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Then, all subjects received CNO (2.0 mg/kg i.p.) 30 min prior to each of 4 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. No CNO was given on test. CNO was given prior to the retraining session (RI-30s) in between tests.

Chemogenetic activation of DMS A2A$^+$ neurons during instrumental overtraining

Naïve, male and female A2A-cre mice (Final *N* = 12, 6 males) and wildtype littermate controls (*N* = 12, 7 males) served as subjects in this experiment to assess whether DMS A2A$^+$ neuron activity is sufficient to promote action-outcome learning for goal-directed behavioral control and prevent the habit formation that normally occurs with overtraining. 8 A2A-cre subjects with off-target viral expression were excluded from the dataset. 1 WT subject was excluded due to mechanical lesion. 11 (A2A: 6, WT: 5) subjects were excluded for ineffective sensory-specific satiety. At surgery, all mice received bilateral infusion of AAV encoding a cre-inducible hM3Dq (AAV2-Syn-DIO-hM3Dq-mCherry; 0.3 µl) into the DMS. After 2 weeks of recovery, mice were food restricted at then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Then, all subjects received CNO (0.2 mg/kg i.p.) 30 min prior to each of 8 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. No CNO was given on test. CNO was given prior to the retraining session (RI-30s) in between tests.

Chemogenetic inhibition of DMS A2A$^+$ neurons during goal-directed decision making

Naïve, male and female A2A-cre mice (Final *N* = 14, 9 males) and wildtype littermate controls (*N* = 16, 7 males) served as subjects in this experiment to assess whether DMS A2A$^+$ neuron activity is necessary for the expression of goal-directed decision making. 6 A2A-cre subjects with off-target viral expression were excluded from the dataset. At surgery, all mice received bilateral infusion of AAV encoding cre-inducible hM4Di (AAV2-Syn-DIO-hM4Di-mCherry; 0.3 µl) into the DMS. After 2 weeks of recovery, mice were food restricted at then began instrumental training, as described above. Mice were habituated to i.p. injections during the final day of FR-1 training. Then, all subjects received saline (0.9% i.p.) 30 min prior to each of 4 RI training sessions. Following training, mice received a pair of sensory-specific satiety outcome-specific devaluation tests, as above. Immediately after the pre-feeding, all mice received CNO (2.0 mg/kg i.p.) and commenced to the non-rewarded probe test 30 min later. Saline was given prior to the retraining session (RI-30s) in between tests. A2A-cre and control mice had slightly different press rates on the last day of training, and so test data were normalized to baseline press rates.

Electrophysiological validation of chemogenetic manipulations

Whole-cell patch clamp recordings were used to validate the efficacy of chemogenetic manipulation of DMS D1$^+$ and A2A$^+$ neurons. Male and female naïve Drd1a-cre and A2A-cre mice served as subjects. At surgery, mice received bilateral infusion of AAV encoding cre-inducible hM4Di (AAV2-Syn-DIO-hM4Di-mCherry; 0.3 µl), cre-inducible hM3Dq (AAV2-Syn-DIO-hM3Dq-mCherry; 0.3 µl), or a fluorophore control (AAV2-Syn-DIO-mCherry; 0.3 µl) into the DMS. Recordings were performed from slices containing the DMS ~2 - 3 weeks following infusion. Mice were ~2 months old at the time slices were taken. Mice were deeply anesthetized with isoflurane and transcardially perfused with ice-cold, oxygenated NMDG-based slicing solution containing (in mM): 30 $NaHCO_3$,

20 HEPES, 1.25 NaH_2PO_4, 102 NMDG, 40 glucose, 3 KCl, 0.5 CaCl2-2H_2O, 10 $MgSO_4$-H_2O (pH adjusted to 7.3-7.35, osmolality 300-310 mOsm/L). Brains were extracted and placed in ice-cold, oxygenated NMDG slicing solution. Coronal slices (300 µm) were cut using a vibrating microtome (VT1000S; Leica Microsystems, Germany), transferred to an incubating chamber containing oxygenated NMDG slicing solution warmed to 32-34 °C, and allowed to recover for 15 min before being transferred to an artificial cerebral spinal fluid (aCSF) solution containing (in mM): 130 NaCl, 3 KCl, 1.25 NaH_2PO_4, 26 $NaHCO_3$, 2 $MgCl_2$, 2 $CaCl_2$, and 10 glucose) oxygenated with 95% O_2, 5% CO_2 (pH 7.2-7.4, osmolality 290-310 mOsm/L, 32-34°C). After 15 min, slices were moved to room temperature and allowed to recover for ~30 additional min prior to recording. All recordings were performed using an upright microscope (Olympus BX51WI, Center Valley, PA) equipped with differential interference contrast optics and fluorescence imaging (QIACAM fast 1394 monochromatic camera 765 with Q-Capture Pro software, QImaging, Surrey, BC, Canada). Patch pipettes (3-5 MΩ resistance) contained a K-gluconate-based solution containing (in mM): 112.5 K-gluconate, 4 NaCl, 17.5 KCl, 0.5 $CaCl_2$, 1 $MgCl_2$, 5 ATP (K^+ salt), 1 NaGTP, 5 EGTA, 10 HEPES, pH 7.2 (270–280 mOsm/L). Biocytin (0.2%, Sigma-Aldrich, St. Louis, MO) was included in the internal recording solution for subsequent verification of fluorescence labeling in recorded cells. Recordings were obtained using a MultiClamp 700B Amplifier (Molecular Devices, Sunnyvale, CA) and the pCLAMP 10.7 acquisition software. Fluorescence-guided whole-cell patch clamp recordings in current-clamp mode were obtained from DMS $D1^+$ and $A2A^+$ neurons (Control, N = 19 cells, 4 D1-cre mice, 1 male; D1-cre, hm4Di: N = 8 cells, 4 mice, 3 males, hM3Dq: N = 10 cells, 5 mice, 3 males; A2A-cre, hm4Di: N = 8 cells, 4 mice, 3 males, hM3Dq: N = 10 cells, 5 mice, 3 males). After breaking through the membrane, recordings were obtained from cells while injecting suprathreshold depolarizing current (1 s). Current injection intensities between 100 - 800 pA, resulting in 5 - 15 action potentials, were selected for recordings. Electrode access resistances were maintained at <30 MΩ throughout recordings. Number of action potentials generated were recorded both prior to and after bath application of CNO (10 or 100 µM) for >10 minutes. Percent change in the number of action potentials before and after CNO was calculated for both hM4Di and hM3Dq expressing cells.

Histology

At the conclusion of instrumental training and testing, mice were euthanized and brain tissue was processed to assess viral expression location and spread and GRIN lens placement. Mice from miniscope experiments were anesthetized with isoflurane and transcardially perfused with ice-cold PBS followed by cold 4% paraformaldehyde. The brains were removed, post-fixed in 4% paraformaldehyde, then cryoprotected in 30% sucrose in PBS. 30-µm coronal slices were taken on a cryostat and collected in PBS. Sections were mounted on slides using ProLong Gold antifade reagent with DAPI (Invitrogen). Mice from chemogenetic experiments were anesthetized with isoflurane, brains were removed and rapidly frozen. 20-µm coronal slices were taken on a cryostat and dry mounted onto subbed slides. Slides were submerged in 4% paraformaldehyde, and coverlipped with ProLong Gold antifade reagent with DAPI (Invitrogen). All images were acquired using a Keyence (BZ-X710) microscope with 4X, 10X, and 20X objectives (CFI Plan Apo), CCD camera, and BZ-X Analyze software, to confirm viral expressions and GRIN lens placement.

Statistical Analysis

Datasets were analyzed by 2-tailed t-tests, one-tailed Bayes Factor, or 1-, 2-, or 3-way repeated-measures analysis of variance (ANOVA), or analysis of co-variance (ANCOVA), as appropriate (GraphPad Prism, GraphPad, San Diego, CA; MATLAB, SPSS, IBM, Chicago, IL). ANCOVA controlling for subject was used for all datasets in which cell rather than subject was the unit N. All data sets were checked for normality. Bonferroni post hoc tests corrected for multiple comparisons were performed to clarify statistical interactions. Chemogenetic validation data used uncorrected planned comparisons. Given single-session decoding results, corrected planned comparisons were used for cross-session decoding. Greenhouse-Geisser correction was applied to mitigate the influence of unequal variance between conditions. Alpha levels were set at $P < 0.05$.

Sex as a biological variable

Male and female mice were used in approximately equal numbers, where possible based on breeding and mouse availability, but the N per sex was underpowered to examine sex differences. Sex was therefore not included as a factor in statistical analyses, though individual data points are visually disaggregated by sex.

Rigor and reproducibility

See Supplemental Table 4 for key reagents. Group sizes were estimated based on prior work with this behavioral task[108] and to ensure counterbalancing of virus, pellet type, and devaluation test order. Investigators were not blinded to viral group because they were required to administer virus. All behaviors were scored using automated software (Med Associates). Each experiment included at least 1 replication cohort and cohorts were balanced by Viral group or hemisphere (for imaging) prior to the start of the experiment. Investigators were blinded to group when performing histological validation and determining exclusions based on viral spread or mistargeted implant. Calcium analyses included cross-validation and comparison to shuffled data.

Data and code availability

All data that support the findings of this study are available from the corresponding author upon request. Custom-written MATLAB code will be accessible via Dryad repository and available from the corresponding author upon request.

REFERENCES

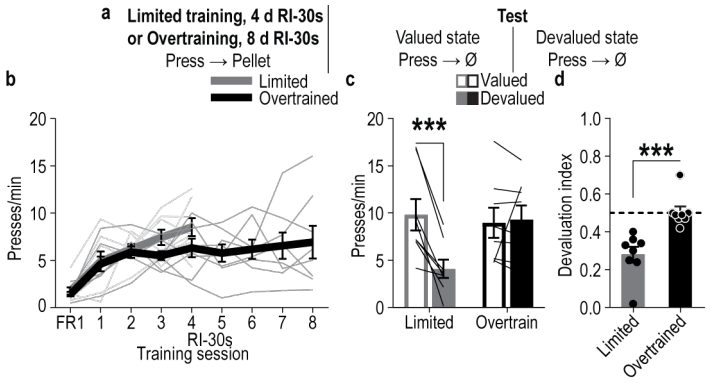
1. Wassum, K.M. Amygdala-cortical collaboration in reward learning and decision making. *Elife* **11** (2022).
2. Balleine, B.W. The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron* **104**, 47-62 (2019).
3. Balleine, B.W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407-419 (1998).
4. O'Doherty, J.P., Cockburn, J. & Pauli, W.M. Learning, Reward, and Decision Making. *Annu Rev Psychol* **68**, 73-100 (2017).
5. Doll, B.B., Simon, D.A. & Daw, N.D. The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol* **22**, 1075-1081 (2012).
6. Dickinson, A. & Balleine, B.W. Actions and responses: the dual psychology of behaviour. in *Spatial representation* (ed. N. Eilan, R. McCarthy & M.W. Brewer) 277-293 (Basil Blackwell Ltd, Oxford, 1993).
7. Dickinson, A. & Balleine, B.W. Animal Learning & Behavior. (1994).
8. Daw, N.D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience* **8**, 1704-1711 (2005).
9. Malvaez, M. & Wassum, K. Regulation of habit formation in the dorsal striatum. *Current Opinion in Behavioral Sciences* **20**, 67-74 (2018).
10. Graybiel, A.M. Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* **31**, 359-387 (2008).
11. Adams, C.D. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology* **34**, 77-98 (1982).
12. Smith, K.S. & Graybiel, A.M. Habit formation. *Dialogues Clin Neurosci* **18**, 33-43 (2016).
13. Dickinson, A. Actions and Habits: the development of behavioural autonomy. *Philosphical Transactions of the Royal Society of London* **B308**, 67-78 (1985).
14. Dolan, R.J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312-325 (2013).
15. Redish, A.D., Jensen, S. & Johnson, A. A unified framework for addiction: vulnerabilities in the decision process. *Behav Brain Sci* **31**, 415-437; discussion 437-487 (2008).
16. Vandaele, Y. & Ahmed, S.H. Habit, choice, and addiction. *Neuropsychopharmacology* **46**, 689-698 (2021).
17. Voon, V.*, et al.* Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry* **20**, 345-352 (2015).
18. Gillan, C.M., Robbins, T.W., Sahakian, B.J., van den Heuvel, O.A. & van Wingen, G. The role of habit in compulsivity. *Eur Neuropsychopharmacol* **26**, 828-840 (2016).
19. Corbit, L.H. & Janak, P.H. Habitual Alcohol Seeking: Neural Bases and Possible Relations to Alcohol Use Disorders. *Alcohol Clin Exp Res* (2016).
20. Ostlund, S.B. & Balleine, B.W. On habits and addiction: An associative analysis of compulsive drug seeking. *Drug Discov Today Dis Models* **5**, 235-245 (2008).
21. Zapata, A., Minney, V.L. & Shippenberg, T.S. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J Neurosci* **30**, 15457-15463 (2010).
22. Hogarth, L. & Chase, H.W. Parallel goal-directed and habitual control of human drug-seeking: Implications for dependence vulnerability. *J Exp Psychol Anim Behav Process* (2011).
23. Belin, D., Belin-Rauscent, A., Murray, J.E. & Everitt, B.J. Addiction: failure of control over maladaptive incentive habits. *Curr Opin Neurobiol* (2013).
24. Hogarth, L., Balleine, B.W., Corbit, L.H. & Killcross, S. Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Ann N Y Acad Sci* **1282**, 12-24 (2013).
25. Leblanc, K.H., Maidment, N.T. & Ostlund, S.B. Repeated cocaine exposure facilitates the expression of incentive motivation and induces habitual control in rats. *PLoS One* **8**, e61355 (2013).
26. Furlong, T.M., Jayaweera, H.K., Balleine, B.W. & Corbit, L.H. Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. *J Neurosci* **34**, 5012-5022 (2014).
27. Renteria, R., Baltz, E.T. & Gremel, C.M. Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nat Commun* **9**, 211 (2018).
28. Hogarth, L., Attwood, A.S., Bate, H.A. & Munafò, M.R. Acute alcohol impairs human goal-directed action. *Biol Psychol* **90**, 154-160 (2012).
29. Ray, L.A.*, et al.* Capturing habitualness of drinking and smoking behavior in humans. *Drug Alcohol Depend* **207**, 107738 (2020).
30. Groman, S.M., Massi, B., Mathias, S.R., Lee, D. & Taylor, J.R. Model-Free and Model-Based Influences in Addiction-Related Behaviors. *Biol Psychiatry* **85**, 936-945 (2019).
31. Gillan, C.M.*, et al.* Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry* **168**, 718-726 (2011).
32. Gillan, C.M.*, et al.* Enhanced avoidance habits in obsessive-compulsive disorder. *Biol Psychiatry* **75**, 631-638 (2014).
33. Vaghi, M.M.*, et al.* Action-Outcome Knowledge Dissociates From Behavior in Obsessive-Compulsive Disorder Following Contingency Degradation. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 200-209 (2019).
34. Horstmann, A.*, et al.* Slave to habit? Obesity is associated with decreased behavioural sensitivity to reward devaluation. *Appetite* **87**, 175-183 (2015).
35. Morris, R.W., Cyrzon, C., Green, M.J., Le Pelley, M.E. & Balleine, B.W. Impairments in action-outcome learning in schizophrenia. *Transl Psychiatry* **8**, 54 (2018).
36. Griffiths, K.R., Morris, R.W. & Balleine, B.W. Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. *Front Syst Neurosci* **8**, 101 (2014).
37. Morris, R.W., Quail, S., Griffiths, K.R., Green, M.J. & Balleine, B.W. Corticostriatal control of goal-directed action is impaired in schizophrenia. *Biol Psychiatry* **77**, 187-195 (2015).
38. Byrne, K.A., Six, S.G. & Willis, H.C. Examining the effect of depressive symptoms on habit formation and habit-breaking. *J Behav Ther Exp Psychiatry* **73**, 101676 (2021).
39. Alvares, G.A., Balleine, B.W. & Guastella, A.J. Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PLoS One* **9**, e94778 (2014).

40. Alvares, G.A., Balleine, B.W., Whittle, L. & Guastella, A.J. Reduced goal-directed action control in autism spectrum disorder. *Autism Res* (2016).
41. Malvaez, M. Neural substrates of habit. *J Neurosci Res* (2019).
42. Balleine, B.W. & O'Doherty, J.P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48-69 (2010).
43. Balleine, B.W., Delgado, M.R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *J Neurosci* **27**, 8161-8165 (2007).
44. Yin, H.H., Knowlton, B.J. & Balleine, B.W. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur J Neurosci* **22**, 505-512 (2005).
45. Corbit, L.H. & Janak, P.H. Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. *Eur J Neurosci* **31**, 1312-1321 (2010).
46. Lex, B. & Hauber, W. The role of dopamine in the prelimbic cortex and the dorsomedial striatum in instrumental conditioning. *Cereb Cortex* **20**, 873-883 (2010).
47. Liljeholm, M., Tricomi, E., O'Doherty, J.P. & Balleine, B.W. Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. *J Neurosci* **31**, 2474-2480 (2011).
48. McNamee, D., Liljeholm, M., Zika, O. & O'Doherty, J.P. Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate FMRI study. *J Neurosci* **35**, 3764-3771 (2015).
49. O'Hare, J., Calakos, N. & Yin, H.H. Recent Insights into Corticostriatal Circuit Mechanisms underlying Habits: Invited review for Current Opinions in Behavioral Sciences. *Curr Opin Behav Sci* **20**, 40-46 (2018).
50. Gremel, C.M. & Costa, R.M. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat Commun* **4**, 2264 (2013).
51. Corbit, L.H., Nie, H. & Janak, P.H. Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biol Psychiatry* **72**, 389-395 (2012).
52. Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci* **22**, 513-523 (2005).
53. Surmeier, D.J., Ding, J., Day, M., Wang, Z. & Shen, W. D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci* **30**, 228-235 (2007).
54. Durieux, P.F*., et al.* D2R striatopallidal neurons inhibit both locomotor and drug reward processes. *Nat Neurosci* **12**, 393-395 (2009).
55. Albin, R.L., Young, A.B. & Penney, J.B. The functional anatomy of basal ganglia disorders. *Trends Neurosci* **12**, 366-375 (1989).
56. Fang, L.Z. & Creed, M.C. Updating the striatal-pallidal wiring diagram. *Nat Neurosci* **27**, 15-27 (2024).
57. Graybiel, A.M. The basal ganglia. *Curr Biol* **10**, R509-511 (2000).
58. Freeze, B.S., Kravitz, A.V., Hammack, N., Berke, J.D. & Kreitzer, A.C. Control of basal ganglia output by direct and indirect pathway projection neurons. *J Neurosci* **33**, 18531-18539 (2013).
59. Goldberg, J.H., Farries, M.A. & Fee, M.S. Basal ganglia output to the thalamus: still a paradox. *Trends Neurosci* **36**, 695-705 (2013).
60. Deniau, J.M. & Chevalier, G. Disinhibition as a basic process in the expression of striatal functions. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res* **334**, 227-233 (1985).
61. Freeze, B.S., Kravitz, A.V., Hammack, N., Berke, J.D. & Kreitzer, A.C. Control of basal ganglia output by direct and indirect pathway projection neurons. *J Neurosci* **33**, 18531-18539 (2013).
62. Cui, G*., et al.* Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238-242 (2013).
63. Tecuapetla, F., Jin, X., Lima, S.Q. & Costa, R.M. Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell* (2016).
64. Aharoni, D. & Hoogland, T.M. Circuit Investigations With Open-Source Miniaturized Microscopes: Past, Present and Future. *Front Cell Neurosci* **13**, 141 (2019).
65. Dana, H*., et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat Methods* **16**, 649-657 (2019).
66. Adams, C.D. & Dickinson, A. Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology* **33**, 109-121 (1981).
67. Pnevmatikakis, E.A*., et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89**, 285-299 (2016).
68. Zhou, P*., et al.* Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *Elife* **7** (2018).
69. Pnevmatikakis, E.A. Analysis pipelines for calcium imaging data. *Curr Opin Neurobiol* **55**, 15-21 (2019).
70. Jin, X., Tecuapetla, F. & Costa, R.M. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nat Neurosci* **17**, 423-430 (2014).
71. Stalnaker, T.A., Calhoon, G.G., Ogawa, M., Roesch, M.R. & Schoenbaum, G. Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Front Integr Neurosci* **4**, 12 (2010).
72. Kravitz, A.V., Tye, L.D. & Kreitzer, A.C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* **15**, 816-818 (2012).
73. Yttri, E.A. & Dudman, J.T. Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* **533**, 402-406 (2016).
74. Kravitz, A.V*., et al.* Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622-626 (2010).
75. Pomrenze, M.B*., et al.* A Transgenic Rat for Investigating the Anatomy and Function of Corticotrophin Releasing Factor Circuits. *Front Neurosci* **9**, 487 (2015).
76. Tipps, M., Marron Fernandez de Velasco, E., Schaeffer, A. & Wickman, K. Inhibition of Pyramidal Neurons in the Basal Amygdala Promotes Fear Learning. *eNeuro* **5** (2018).

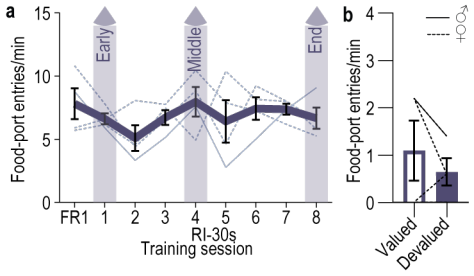
77. Tuscher, J.J., Taxier, L.R., Fortress, A.M. & Frick, K.M. Chemogenetic inactivation of the dorsal hippocampus and medial prefrontal cortex, individually and concurrently, impairs object recognition and spatial memory consolidation in female mice. *Neurobiol Learn Mem* **156**, 103-116 (2018).
78. Alexander, G.M.*, et al.* Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63**, 27-39 (2009).
79. Vazey, E.M. & Aston-Jones, G. Designer receptor manipulations reveal a role of the locus coeruleus noradrenergic system in isoflurane general anesthesia. *Proc Natl Acad Sci U S A* **111**, 3859-3864 (2014).
80. Qiu, M.H., Chen, M.C., Fuller, P.M. & Lu, J. Stimulation of the Pontine Parabrachial Nucleus Promotes Wakefulness via Extra-thalamic Forebrain Circuit Nodes. *Curr Biol* **26**, 2301-2312 (2016).
81. Zhu, H.*, et al.* Cre-dependent DREADD (Designer Receptors Exclusively Activated by Designer Drugs) mice. *Genesis* **54**, 439-446 (2016).
82. Nonomura, S.*, et al.* Monitoring and Updating of Action Selection for Goal-Directed Behavior through the Striatal Direct and Indirect Pathways. *Neuron* **99**, 1302-1314.e1305 (2018).
83. Shin, J.H., Kim, D. & Jung, M.W. Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. *Nat Commun* **9**, 404 (2018).
84. Vandaele, Y.*, et al.* Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *Elife* **8** (2019).
85. Jin, X. & Costa, R.M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457-462 (2010).
86. Smith, K.S. & Graybiel, A.M. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* **79**, 361-374 (2013).
87. Lee, J. & Sabatini, B.L. Striatal indirect pathway mediates exploration via collicular competition. *Nature* **599**, 645-649 (2021).
88. Tai, L.H., Lee, A.M., Benavidez, N., Bonci, A. & Wilbrecht, L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* **15**, 1281-1289 (2012).
89. Weglage, M.*, et al.* Complete representation of action space and value in all dorsal striatal pathways. *Cell Rep* **36**, 109437 (2021).
90. Shan, Q., Ge, M., Christie, M.J. & Balleine, B.W. The acquisition of goal-directed actions generates opposing plasticity in direct and indirect pathways in dorsomedial striatum. *J Neurosci* **34**, 9196-9201 (2014).
91. Peak, J., Chieng, B., Hart, G. & Balleine, B.W. Striatal direct and indirect pathway neurons differentially control the encoding and updating of goal-directed learning. *Elife* **9** (2020).
92. Ferguson, S.M., Phillips, P.E., Roth, B.L., Wess, J. & Neumaier, J.F. Direct-pathway striatal neurons regulate the retention of decision-making strategies. *J Neurosci* **33**, 11668-11676 (2013).
93. Yu, X., Chen, S. & Shan, Q. Depression in the Direct Pathway of the Dorsomedial Striatum Permits the Formation of Habitual Action. *Cereb Cortex* **31**, 3551-3564 (2021).
94. Li, Y.*, et al.* Optogenetic Activation of Adenosine A2A Receptor Signaling in the Dorsomedial Striatopallidal Neurons Suppresses Goal-Directed Behavior. *Neuropsychopharmacology* **41**, 1003-1013 (2016).
95. Tecuapetla, F., Koós, T., Tepper, J.M., Kabbani, N. & Yeckel, M.F. Differential dopaminergic modulation of neostriatal synaptic connections of striatopallidal axon collaterals. *J Neurosci* **29**, 8977-8990 (2009).
96. Furlong, T.M., Supit, A.S., Corbit, L.H., Killcross, S. & Balleine, B.W. Pulling habits out of rats: adenosine 2A receptor antagonism in dorsomedial striatum rescues meth-amphetamine-induced deficits in goal-directed action. *Addict Biol* (2015).
97. Gagnon, D.*, et al.* Striatal neurons expressing D1 and D2 receptors are morphologically distinct and differently affected by dopamine denervation in mice. *Sci Rep* **7**, 41432 (2017).
98. Saunders, A.*, et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015-1030.e1016 (2018).
99. Gokce, O.*, et al.* Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep* **16**, 1126-1137 (2016).
100. Matamales, M.*, et al.* Local D2- to D1-neuron transmodulation updates goal-directed learning in the striatum. *Science* **367**, 549-555 (2020).
101. Fermin, A.S.*, et al.* Model-based action planning involves cortico-cerebellar and basal ganglia networks. *Sci Rep* **6**, 31378 (2016).
102. Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D. & Daw, N.D. Model-based choices involve prospective neural activity. *Nat Neurosci* **18**, 767-772 (2015).
103. Cox, J. & Witten, I.B. Striatal circuits for reward learning and decision-making. *Nat Rev Neurosci* **20**, 482-494 (2019).
104. Delevich, K.*, et al.* Activation, but not inhibition, of the indirect pathway disrupts choice rejection in a freely moving, multiple-choice foraging task. *Cell Rep* **40**, 111129 (2022).
105. Ramírez-Armenta, K.I.*, et al.* Optogenetic inhibition of indirect pathway neurons in the dorsomedial striatum reduces excessive grooming in Sapap3-knockout mice. *Neuropsychopharmacology* **47**, 477-487 (2022).
106. Bakhurin, K.I.*, et al.* Opponent regulation of action performance and timing by striatonigral and striatopallidal pathways. *Elife* **9** (2020).
107. Carvalho Poyraz, F.*, et al.* Decreasing Striatopallidal Pathway Function Enhances Motivation by Energizing the Initiation of Goal-Directed Action. *J Neurosci* **36**, 5988-6001 (2016).
108. Malvaez, M.*, et al.* Habits Are Negatively Regulated by Histone Deacetylase 3 in the Dorsal Striatum. *Biol Psychiatry* (2018).
109. Shiflett, M.W. & Balleine, B.W. Molecular substrates of action control in cortico-striatal circuits. *Prog Neurobiol* **95**, 1-13 (2011).
110. Gremel, C.M.*, et al.* Endocannabinoid Modulation of Orbitostriatal Circuits Gates Habit Formation. *Neuron* **90**, 1312-1324 (2016).
111. Fisher, S.D., Ferguson, L.A., Bertran-Gonzalez, J. & Balleine, B.W. Amygdala-Cortical Control of Striatal Plasticity Drives the Acquisition of Goal-Directed Action. *Curr Biol* (2020).
112. Giovanniello, J.*, et al.* A dual-pathway architecture for stress to disrupt agency and promote habit. *Nature* (In Press).
113. Dias-Ferreira, E.*, et al.* Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* **325**, 621-625 (2009).

114. Seiler, J.L., *et al.* Dopamine signaling in the dorsomedial striatum promotes compulsive behavior. *Curr Biol* **32**, 1175-1188.e1175 (2022).
115. Hopf, F.W., Mailliard, W.S., Gonzalez, G.F., Diamond, I. & Bonci, A. Atypical protein kinase C is a novel mediator of dopamine-enhanced firing in nucleus accumbens neurons. *J Neurosci* **25**, 985-989 (2005).
116. Bradfield, L.A., Bertran-Gonzalez, J., Chieng, B. & Balleine, B.W. The thalamostriatal pathway and cholinergic control of goal-directed action: interlacing new with existing learning in the striatum. *Neuron* **79**, 153-166 (2013).
117. Kang, S., *et al.* Activation of Astrocytes in the Dorsomedial Striatum Facilitates Transition From Habitual to Goal-Directed Reward-Seeking Behavior. *Biol Psychiatry* **88**, 797-808 (2020).
118. Dobbs, L.K., *et al.* Dopamine Regulation of Lateral Inhibition between Striatal Neurons Gates the Stimulant Actions of Cocaine. *Neuron* **90**, 1100-1113 (2016).
119. Taverna, S., Ilijic, E. & Surmeier, D.J. Recurrent collateral connections of striatal medium spiny neurons are disrupted in models of Parkinson's disease. *J Neurosci* **28**, 5504-5512 (2008).
120. Yin, H.H., Park, B.S., Adermark, L. & Lovinger, D.M. Ethanol reverses the direction of long-term synaptic plasticity in the dorsomedial striatum. *Eur J Neurosci* **25**, 3226-3232 (2007).
121. Hayrapetyan, V., *et al.* Region-specific impairments in striatal synaptic transmission and impaired instrumental learning in a mouse model of Angelman syndrome. *Eur J Neurosci* **39**, 1018-1025 (2014).
122. Bertran-Gonzalez, J., Dinale, C. & Matamales, M. Restoring the youthful state of striatal plasticity in aged mice re-enables cognitive control of action. *Curr Biol* **33**, 1997-2007.e1995 (2023).
123. Lemberger, T., *et al.* Expression of Cre recombinase in dopaminoceptive neurons. *BMC Neurosci* **8**, 4 (2007).
124. Gong, S., *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J Neurosci* **27**, 9817-9823 (2007).
125. Blair, G.J., *et al.* Hippocampal place cell remapping occurs with memory storage of aversive experiences. *Elife* **12** (2023).
126. Giovannucci, A., *et al.* CaImAn an open source tool for scalable calcium imaging data analysis. *Elife* **8** (2019).
127. Sheintuch, L., *et al.* Tracking the Same Neurons across Multiple Days in Ca. *Cell Rep* **21**, 1102-1115 (2017).
128. Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* **16** (2002).

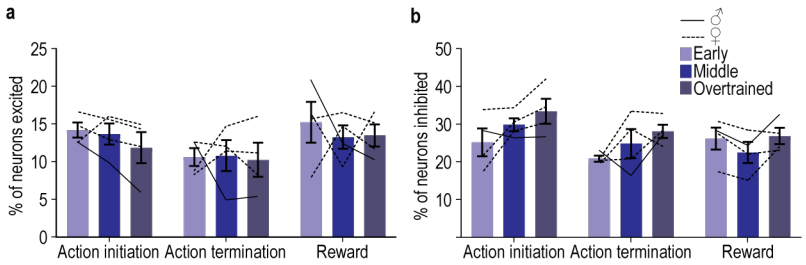
EXTENDED DATA FIGURES



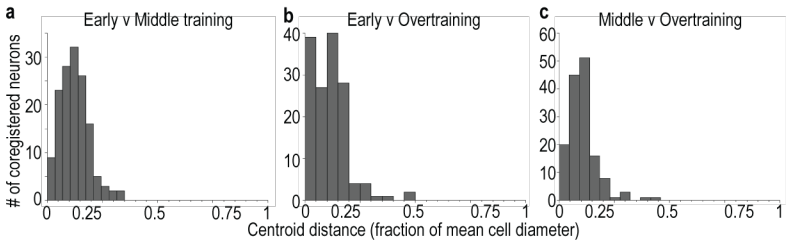
Extended Data Figure 1-1: Behavior is goal-directed following limited random-interval instrumental training and habitual following overtraining. (a) Procedure. Lever presses earned food pellet rewards on a random-interval (RI) 30-s reinforcement schedule. Mice received either limited training (4 RI sessions) or overtraining (8 sessions). Mice were then given a lever-pressing probe test in the Valued state, prefed on untrained food-pellet type to control for general satiety and Devalued state prefed on trained food-pellet type to induce sensory-specific satiety devaluation. Test order was counterbalanced across subjects within each group, with a single intervening retraining session. **(b)** Training press rate. 1-way ANOVA, Limited training: Training: $F_{2.75,\ 19.27} = 12.08$, $P = 0.0001$. Overtraining: Training: $F_{2.24,\ 15.69} = 4.56$, $P = 0.02$. **(c)** Test press rate. 2-way ANOVA, Value x Training duration: $F_{1,\ 14} = 14.69$, $P = 0.002$; Value: $F_{1,\ 14} = 11.69$, $P = 0.004$; Training duration: $F_{1,\ 14} = 1.31$, $P = 0.27$. **(d)** Devaluation index [(Devalued presses)/(Valued presses + Devalued presses)]. 2-tailed Mann-Whitney U test, U = 0, P = 0.002. N = 8/group (all male). Data presented as mean ± s.e.m. ***$P < 0.001$. Mice learn action-outcome relationships during instrumental conditioning on a random-interval schedule of reinforcement and use them for goal-directed decision making after limited training and form habits with overtraining.



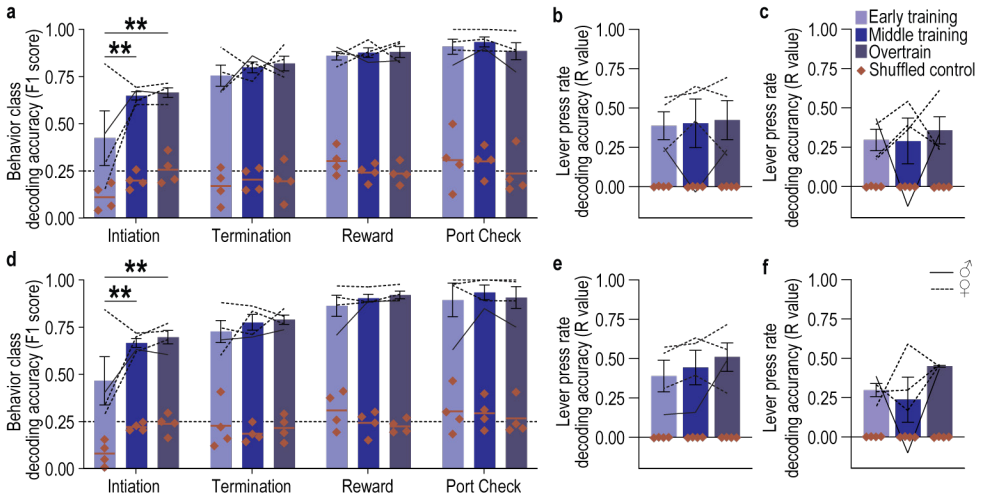
Extended Data Figure 1-2: Entries into the food-delivery port during training and test for DMS D1⁺ imaging experiment. (a) Training entry rate. 1-way ANOVA, Training: $F_{2.11, 6.31} = 0.75$, $P = 0.52$ **(b)** Test entry rate. 2-tailed t-test, $t_3 = 0.94$, $P = 0.42$, 95% CI -1.97 - 1.07. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



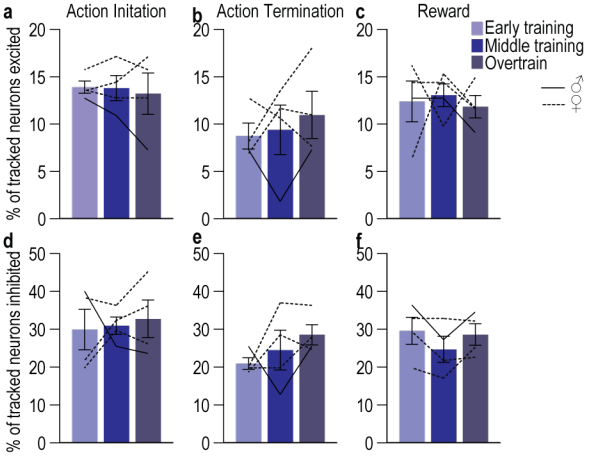
Extended Data Figure 1-3: Percent of DMS D1⁺ neurons activated or inhibited around actions and rewards. (a) Percent of all recorded neurons (Early: *N* = 870, Middle: *N* = 999; End *N* = 994) classified (auROC values >95th percentile of the distribution of shuffled auROCs within 2 s before or after event) as significantly excited around initiating lever presses, terminating lever presses, or reward consumption. Similar proportions of DMS D1⁺ neurons were activated around each type of event and across training. 2-way ANOVA, Training session: $F_{1.07, 3.21} = 0.24$, $P = 0.67$; Event: $F_{1.34, 4.03} = 2.76$, $P = 0.17$; Training x Event: $F_{1.71, 5.14} = 0.48$, $P = 0.62$. **(b)** Percent of all recorded neurons classified as significantly inhibited around initiating lever presses, terminating lever presses, or reward consumption. Similar proportions of the DMS D1⁺ neurons were inhibited around each type of event. With training there was a slight increase in the proportion of neurons inhibited around actions. 2-way ANOVA, Training: $F_{1.75, 5.26} = 5.34$, $P = 0.06$; Event: $F_{1.47, 4.40} = 4.32$, $P = 0.10$; Training x Event: $F_{1.66, 4.98} = 1.56$, $P = 0.29$. D1-cre: *N* = 4 (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



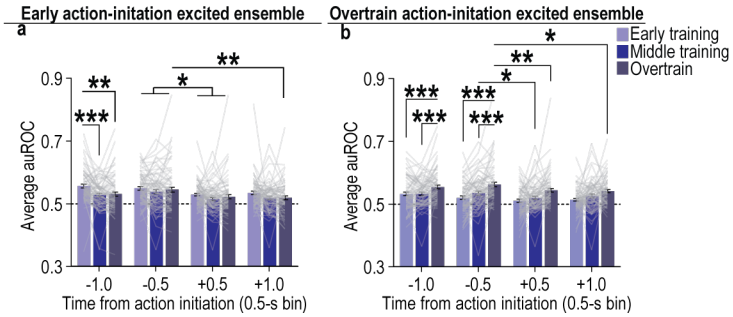
Extended Data Figure 2-1. Representative example of coregistration of DMS D1⁺ neurons across training. **(a)** Distribution of the distance between cell centroids (as a fraction of total cell diameter) of co-registered cell pairs in the 1st (early) and 4th (middle) training sessions. **(b)** Distribution of centroid distance of co-registered cell pairs in the 1st and 8th (overtrain) training sessions. **(c)** Distribution of centroid distance of co-registered cell pairs in the 4th and 8th training sessions.



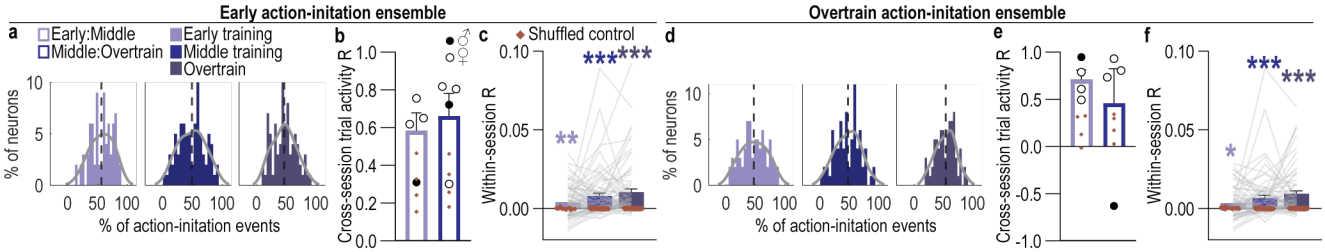
Extended Data Figure 2-2. Behavioral events can be decoded from the activity of the DMS D1⁺ action-initiation excited neuronal ensembles. (a-b) Decoding of behavioral events from the activity of DMS D1⁺ early action-initiation excited neurons. **(a)** Behavior class (initiating lever press, terminating press, reward collection, non-reinforced food-port check) decoding accuracy compared to shuffled control. Line at 0.25 = chance. 3-way ANOVA, Neuron activity (v. shuffled): $F_{1, 3} = 1118.27$, $P < 0.001$; Training session: $F_{2, 6} = 2.79$, $P = 0.14$; Behavior class: $F_{3, 9} = 16.94$, $P < 0.001$; Neuron activity x Training: $F_{2, 6} = 2.10$, $P = 0.20$; Neuron activity x Behavior Class: $F_{3, 9} = 3.53$, $P = 0.06$; Training x Behavior Class: $F_{6, 18} = 3.29$, $P = 0.23$; Neuron activity x Training x Behavior class: $F_{6, 18} = 0.30$, $P = 0.93$. **(b)** Lever-press rate decoding accuracy. R = correlation coefficient between actual and decoded press rate. 2-way ANOVA, Neuron activity: $F_{1, 3} = 12.70$, $P = 0.04$; Training: $F_{1.16, 3.49} = 0.08$, $P = 0.83$; Neuron activity x Training: $F_{1.20, 3.60} = 0.09$, $P = 0.82$. **(c)** Accuracy with which lever-press rate can be can be decoded from the activity of DMS D1⁺ early action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 53.88$, $P = 0.005$; Training: $F_{1.76, 5.29} = 0.09$, $P = 0.89$; Neuron activity x Training: $F_{1.73, 5.38} = 0.11$, $P = 0.88$. **(d-e)** Decoding of behavioral events from the activity of DMS D1⁺ overtrain action-initiation excited neurons. **(d)** Behavior class decoding accuracy compared to shuffled control. 3-way ANOVA, Neuron activity: $F_{1, 3} = 262.75$, $P < 0.001$; Training session: $F_{2, 6} = 2.78$, $P = 0.14$; Behavior class: $F_{3, 9} = 11.31$, $P = 0.002$; Neuron activity x Training: $F_{2, 6} = 2.16$, $P = 0.20$; Neuron activity x Behavior Class: $F_{3, 9} = 3.86$, $P = 0.05$; Training x Behavior Class: $F_{6, 18} = 3.27$, $P = 0.025$; Neuron activity x Training x Behavior class: $F_{6, 18} = 0.52$, $P = 0.79$. **(e)** Lever-press rate decoding accuracy. 2-way ANOVA, Neuron activity: $F_{1, 3} = 24.76$, $P = 0.02$; Training: $F_{1.02, 3.05} = 1.25$, $P = 0.35$; Population activity x Training: $F_{1.00, 3.04} = 1.24$, $P = 0.35$. **(f)** Accuracy with which lever-press rate can be can be decoded from the activity of DMS D1⁺ overtrain action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 84.88$, $P = 0.003$; Training: $F_{1, 3} = 1.25$, $P = 0.35$; Neuron activity x Training: $F_{1, 3} = 1.30$, $P = 0.34$. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. **$P < 0.01$. Actions, action rate, checking for and receiving reward, can be decoded from both the population activity of DMS D1⁺ neurons that are excited by action initiation early in training and those neurons that are excited by action initiation after overtraining. Decoding of action-initiation improves with limited training, but not further with overtraining. Action rate can also be decoded from the DMS D1⁺ neurons that are inhibited by action initiation.



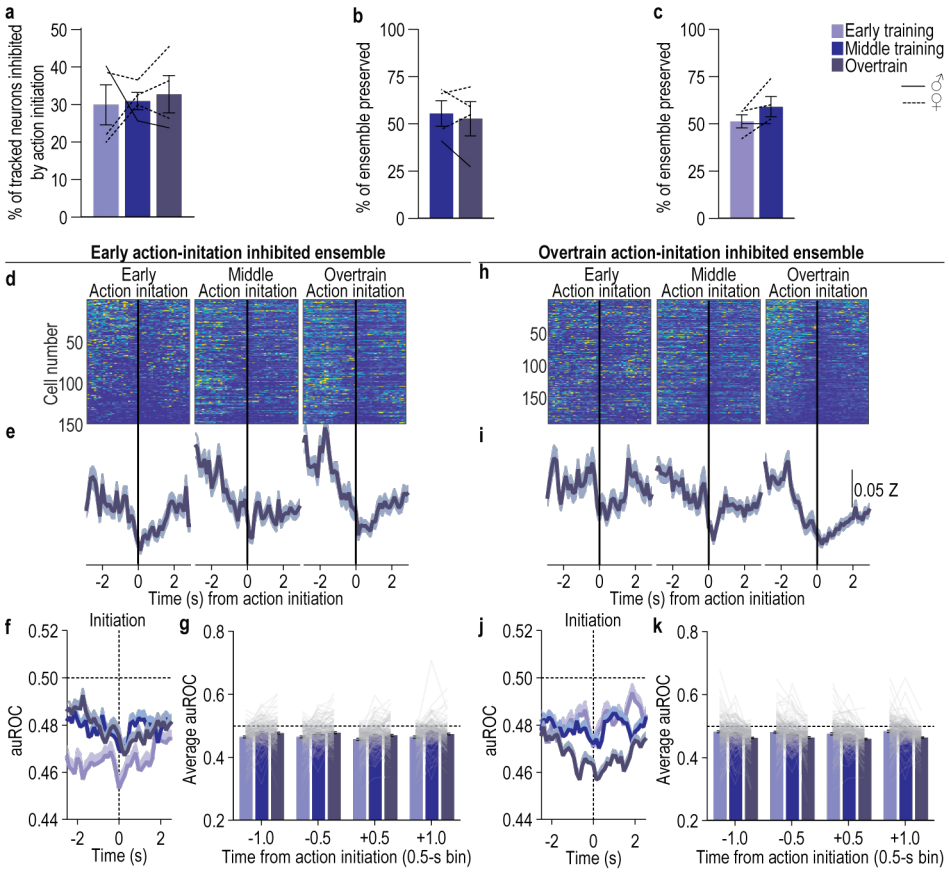
Extended Data Figure 2-3. Ensembles of DMS D1⁺ neurons are excited or inhibited by action initiation, action termination, and reward. (a-c) Percent of all recorded coregistered neurons (Average 137 coregistered neurons/mouse, s.e.m. 37.73) significantly excited by action initiation, action termination, and reward. **(a)** Approximately 13% of coregistered D1⁺ neurons were excited by action initiation. 1-way ANOVA, $F_{1.13, 3.63} = 0.14$, $P = 0.76$. **(b)** Approximately 8 - 10% of coregistered D1⁺ neurons were excited by action termination. 1-way ANOVA, $F_{1.62, 4.85} = 0.37$, $P = 0.67$. **(c)** Approximately 12 - 13% of coregistered D1⁺ neurons were excited by reward. 1-way ANOVA, $F_{1.46, 4.38} = 0.12$, $P = 0.83$. **(d-f)** Percent of all recorded coregistered neurons significantly inhibited by action initiation, action termination, and reward. **(d)** Approximately 30 - 33% of coregistered D1⁺ neurons were inhibited by action initiation. 1-way ANOVA, $F_{1.29, 3.86} = 0.14$, $P = 0.79$. **(e)** Approximately 21 - 29% of coregistered D1⁺ neurons were inhibited by action termination. 1-way ANOVA, $F_{1.17, 3.50} = 1.29$, $P = 0.34$. **(f)** Approximately 25 - 30% of coregistered DMS D1⁺ neurons were inhibited by reward. 1-way ANOVA, $F_{1.91, 5.76} = 2.52$, $P = 0.16$). D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. In no case did the proportion of excited or inhibited D1⁺ neurons change with training.



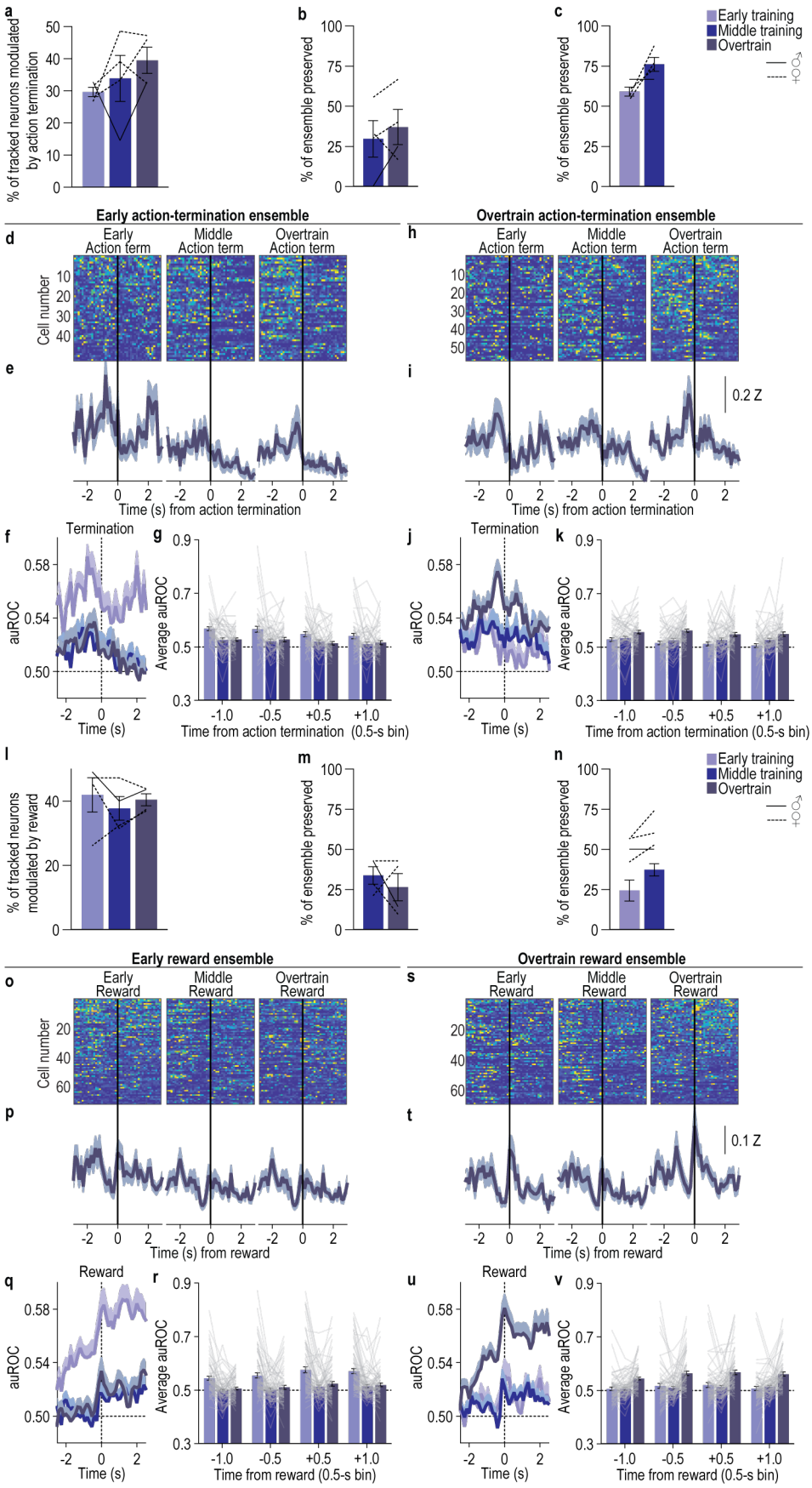
Extended Data Figure 2-4. Quantification of the modulation of DMS D1⁺ action-initiation excited neurons. (a) Modulation across training of DMS D1⁺ early action-initiation excited neurons (*N* = 77 neurons/4 mice; average 19.25 neurons/mouse, s.e.m. = 5.36). Modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.77, 133.28} = 1.24$, $P = 0.29$; Time bin: $F_{1.79, 133.96} = 13.96$, $P < 0.001$; Training x Time: $F_{3.40, 254.94} = 2.94$, $P = 0.028$. **(b)** Modulation across training of D1⁺ overtrain action-initiation excited neurons (*N* = 76 neurons/4 mice; average 19 neurons/mouse, s.e.m. 5.49). Modulation index around action initiation. 2-way ANCOVA, Training: $F_{1.8, 79.90} = 7.18$, $P = 0.002$; Time bin: $F_{1.64, 65.47} = 1.50$, $P = 0.23$; Training x Time: $F_{4.28, 171.34} = 1.50$, $P = 0.20$. Data presented as mean ± s.e.m.



Extended Data Figure 2-5. DMS D1⁺ neurons encode action initiation with high fidelity. (a-c) Fidelity with which D1⁺ early action-initiation excited neurons encode action initiation. **(a)** Distribution of the percentage of early action-initiation excited neurons as a function of the percentage of action-initiation events to which they respond for each training phase. **(b)** Cross-session correlation of the response distributions. 2-way ANOVA, Neuron activity distribution (v. shuffled): $F_{1, 3} = 16.06$, $P = 0.03$; Training session: $F_{1, 3} = 0.25$, $P = 0.65$; Distribution x Training: $F_{1, 3} = 0.11$, $P = 0.76$. Early action-initiation D1⁺ neurons tend to respond on more than half the action-initiation events across training and this is consistent across training. **(c)** Within-session correlation of the activity around action initiation of each early action-initiation excited neuron. 2-way ANCOVA, Neuron activity: $F_{1, 74} = 27.35$, $P < 0.001$; Training: $F_{2, 148} = 5.84$, $P = 0.004$; Activity x Time: $F_{2, 148} = 6.28$, $P = 0.002$. The activity of early action-initiation excited D1⁺ neurons around action initiation is correlated above shuffled control within a training session and becomes more correlated with training. **(d-e)** Fidelity with which D1⁺ overtrain action-initiation excited neurons encode action initiation. **(d)** Distribution of the percentage of overtrain action-initiation excited neurons as a function of the percentage of action-initiation events to which they respond for each training phase. **(e)** Cross-session correlation of the response distributions. 2-way ANOVA, Neuron activity distribution: $F_{1, 3} = 19.36$, $P = 0.02$; Training: $F_{1, 3} = 0.27$, $P = 0.64$; Distribution x Training: $F_{1, 3} = 0.35$, $P = 0.60$. Overtrain action-initiation D1⁺ neurons tend to respond on more than half the action-initiation events across training and this is consistent across training. **(f)** Within-session correlation of the activity around action initiation of each overtrain action-initiation excited neuron. 2-way ANCOVA, Neuron activity: $F_{1, 73} = 19.32$, $P < 0.001$; Training: $F_{2, 146} = 1.87$, $P = 0.16$; Activity x Time: $F_{2, 146} = 1.96$, $P = 0.15$. The activity of overtrain action-initiation excited D1⁺ neurons around action initiation is correlated above shuffled control within a training session and becomes more correlated with training. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = closed circles, Females = open circles.

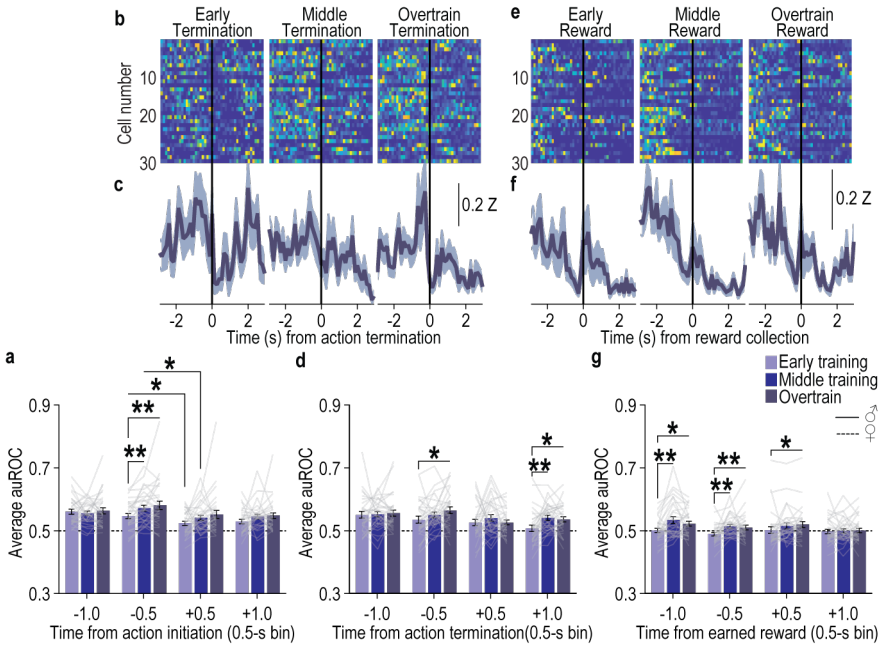


Extended Data Figure 2-6. An ensemble of DMS D1⁺ neurons is inhibited during action initiation across learning and as habits form. (a) Percent of all recorded coregistered DMS D1⁺ neurons (Average 137 coregistered neurons/mouse, s.e.m. 37.73) significantly inhibited by action initiation. Approximately 30 - 32% of DMS D1⁺ neurons were inhibited around action initiation and this did not change with training. 1-way ANOVA, $F_{1.29,\ 3.86} = 0.14$, $P = 0.79$. **(b)** Percent of D1⁺ early action-initiation inhibited neurons that continued to be significantly inhibited by action initiation on the 4th and 8th training sessions. Approximately 52 - 55% of the early action-initiation inhibited ensemble continued to be inhibited around action initiation during the middle and overtraining phases of training. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 0.56$, $P = 0.62$, 95% CI -19.03 - 13.37. **(c)** Percent of D1⁺ overtrain action-initiation inhibited neurons that were also significantly inhibited by action initiation on 1st and 4th training sessions. Approximately 51 - 59% of the overtraining action-initiation inhibited ensemble was also inhibited around action initiation during the preceding early and middle training phases. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 2.02$, $P = 0.14$, 95% CI -4.52 - 20.15. **(d-g)** Activity and modulation across training of DMS D1⁺ early action-initiation inhibited neurons. Heat map of minimum to maximum deconvolved activity (sorted by total activity) (d), Z-scored activity (e), and area under the receiver operating characteristic curve (auROC) modulation index (f) of these cells around action initiation across training. **(g)** auROC modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.65,\ 246.27} = 1.46$, $P = 0.24$; Time bin: $F_{2.16,\ 321.48} = 2.80$, $P = 0.06$; Training x Time: $F_{4.70,\ 699.66} = 0.27$, $P = 0.92$. This modulation of this early action-initiation inhibited ensemble did not significantly change with training. **(h-k)** Activity and modulation across training of DMS D1⁺ overtrain action-initiation inhibited neurons. Heat map (h), Z-scored activity (i), and auROC modulation index (j) of these cells around action initiation across training. **(k)** Modulation index around action initiation. 2-way ANCOVA, Training: $F_{1.82,\ 347.95} = 7.25$, $P = 0.001$; Time bin: $F_{1.91,\ 365.08} = 2.37$, $P = 0.10$; Training x Time: $F_{4.35,\ 831.34} = 0.67$, $P = 0.62$. This overtrain action-initiation inhibited ensemble became more inhibited by action initiation as training progressed. D1-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.

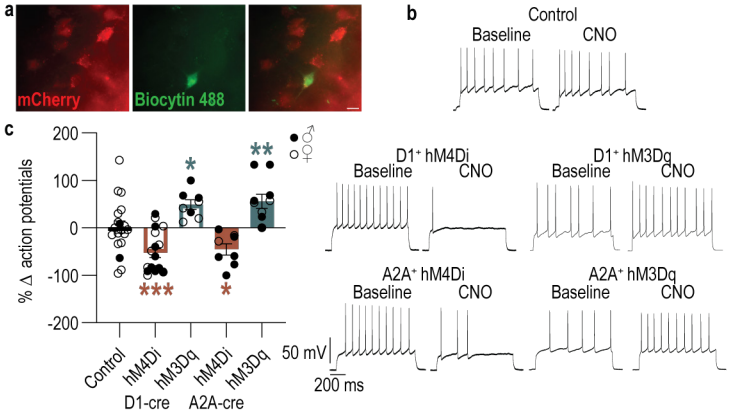


Extended Data Figure 2-7. Ensembles of DMS D1⁺ neurons are modulated by action termination and reward during learning and as habits form. (a) Percent of all recorded coregistered DMS D1⁺ neurons (Average 137 coregistered neurons/mouse, s.e.m. 37.73) significantly modulated (excited or inhibited, see

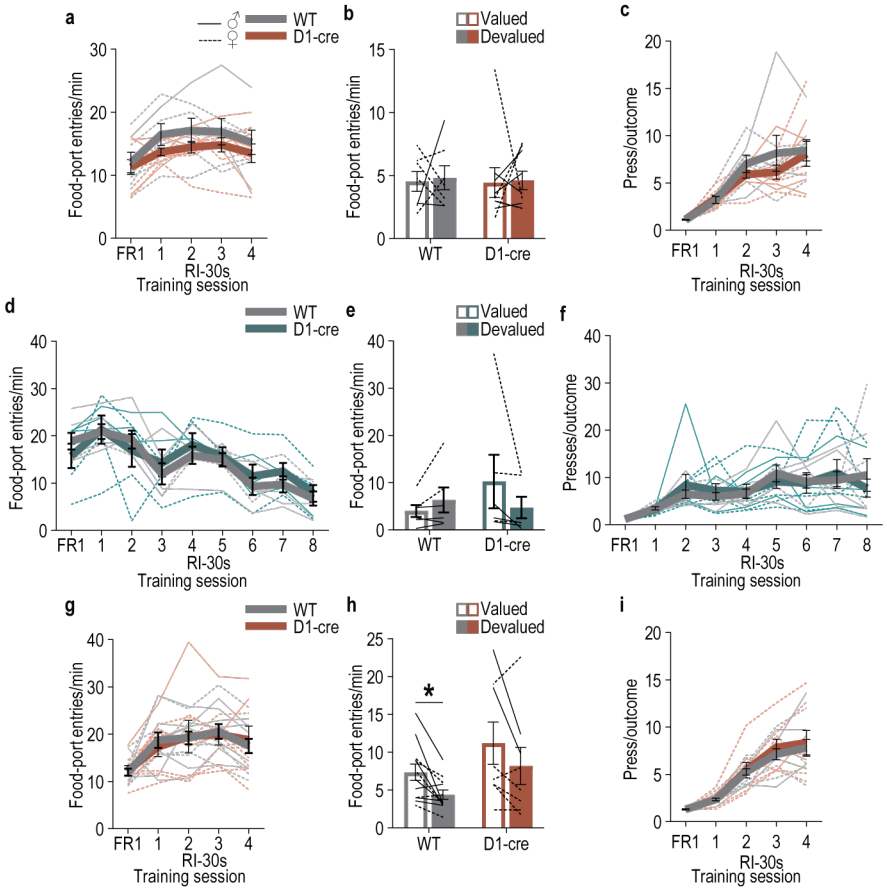
Extended Data Figure 2-3) by action termination. Approximately 30 - 39% of D1⁺ neurons were modulated around action termination and this did not change with training. 1-way ANOVA, $F_{1.53, 4.58} = 1.10$, $P = 0.38$. **(b)** Percent of DMS D1⁺ early action-termination excited neurons (N = 53 neurons/4 mice; average 13.25 neurons/mouse, s.e.m. 5.73) that were then also significantly excited by action termination on the 4th (middle) and 8th (overtrain) training sessions. Approximately 30 - 37% of the early action-termination ensemble continued to be excited by action termination across training. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 0.85$, P =0.46, 95% CI -20.34 - 35.07. **(c)** Percent of DMS D1⁺ overtrain action-termination excited neurons (N = 58 neurons/4 mice; average 14.5 neurons/mouse, s.e.m. 3.59) that were also significantly excited by action termination on the 1st (early) and middle training sessions. Approximately 59 - 76% of the overtraining action-termination ensemble was also excited by action termination during the preceding training phases. The proportion preserved did not significantly change with training. 2-tailed t-test, $t_3 = 2.52$, P =0.09, 95% CI -4.48 to 38.31. **(d-g)** Activity and modulation across training of DMS D1⁺ early action-termination excited neurons. Heat map of minimum to maximum deconvolved activity (sorted by total activity) (d), Z-scored activity (e), and area under the receiver operating characteristic curve (auROC) modulation index (f) of these neurons around action termination across training. **(g)** auROC modulation index averaged across 0.5-s bins around action termination. 2-way ANCOVA, Training: $F_{1.92, 97.81} = 1.08$, $P = 0.34$; Time bin: $F_{1.65, 84.19} = 0.48$, $P = 0.59$; Training x Time: $F_{4.20, 214.10} = 1.33$, $P = 0.26$. Modulation of this early action-termination ensemble did not significantly change with training. **(h-k)** Activity and modulation across training of DMS D1⁺ overtrain action-termination excited neurons. Heat map (h), Z-scored activity (i), and auROC modulation index (j) of these cells around action termination across training. **(k)** Modulation index around action termination. 2-way ANOVA, Training: $F_{1.77, 99.05} = 1.14$, $P = 0.32$; Time bin: $F_{1.78, 99.88} = 0.03$, $P = 0.96$; Training x Time: $F_{4.20, 235.08} = 0.71$, $P = 0.60$. Modulation of the overtrain action-termination ensemble did not significantly change as training progressed. **(l)** Percent of coregistered D1⁺ neurons significantly modulated (excited or inhibited) by earned reward. Approximately 39 - 42% of D1⁺ neurons were modulated by earned reward and this did not change with training. Friedman test, $x^2(2) = 1.20$, $P = 0.58$. **(m)** Percent of D1⁺ early reward excited neurons (N = 73 neurons/4 mice; average 18.25 neurons/mouse, s.e.m. 7.36) that continued to be significantly excited by reward during the middle and overtraining phases. Approximately 26 - 34% of the early reward ensemble continued to be excited by reward across training. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 0.70$, P =0.53, 95% CI -40.49 - 25.89. **(n)** Percent of DMS D1⁺ overtrain reward excited neurons (N = 70 neurons/4 mice; average 17.5 neurons/mouse, s.e.m. 6.34) that were significantly excited by reward on the early and middle training sessions. Approximately 24 - 37% of the overtraining reward ensemble was also excited by reward during the preceding training phases. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 1.31$, P =0.28, 95% CI -18.55 - 44.33. **(o-r)** Activity and modulation across training of DMS D1⁺ early reward excited neurons. Heat map (o), Z-scored activity (p), and auROC modulation index (q) of these cells around reward across training. **(r)** Modulation index around reward. 2-way ANCOVA, Training: $F_{1.74, 123.26} = 0.82$, $P = 0.42$; Time bin: $F_{2.45, 175.73} = 6.07$, $P = 0.01$; Training x Time: $F_{4.82, 342.52} = 1.59$, $P = 0.16$. This early reward ensemble was more excited after earned reward than before, suggesting modulation by reward experience. This did not significantly change as training progressed. **(s-v)** Activity and modulation across training of D1⁺ overtrain reward excited neurons. Heat map (s), Z-scored activity (t), and auROC modulation index (u) of these cells around reward on across training. **(v)** Modulation index around reward. 2-way ANCOVA Training: $F_{1.77, 120.87} = 2.26$, $P = 0.12$; Time bin: $F_{2.36, 160.42} = 1.30$, $P = 0.28$; Training x Time: $F_{5.37, 365.26} = 0.28$, $P = 0.93$. Modulation of the overtrain reward ensemble did not significantly change with training. D1-cre: N = 4 (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



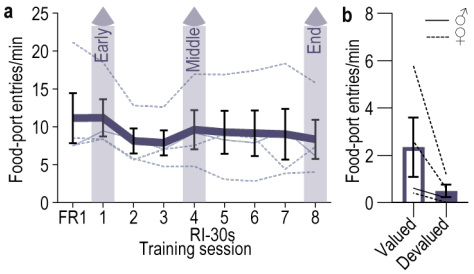
Extended Data Figure 2-8. A subensemble of stable DMS D1⁺ action-initiation excited neurons develop encoding of action termination and earned reward with training. We identified an ensemble (*N* = 31 neurons/4 D1-cre mice, 1 male; average 7.75 neurons/mouse, s.e.m. = 2.56) of D1⁺ neurons that stably encoded action-initiation across all phases of training. **(a)** Modulation, averaged in 0.5-s bins around action initiation of these neurons. 2-way ANCOVA, Training: $F_{1.70, 49.21} = 0.04$, $P = 0.94$; Time bin: $F_{1.73, 50.27} = 5.02$, $P = 0.01$; Training x Time: $F_{2.28, 123.97} = 2.78$, $P = 0.03$. Stable action-initiation excited D1⁺ neurons are more activated prior to action initiation than after and this modulation improves with training. **(b-c)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) (b) and Z-scored (c) activity around action termination across training of the D1⁺ stable action-initiation excited neurons. **(d)** Modulation, averaged in 0.5-s bins around action termination of these neurons. 2-way ANCOVA, Training: $F_{1.87, 54.10} = 1.01$, $P = 0.37$; Time bin: $F_{2.00, 57.95} = 1.48$, $P = 0.23$; Training x Time: $F_{3.62, 104.97} = 1.98$, $P = 0.11$. **(e-f)** Heat map (e) and Z-scored (f) activity around reward collection across training of D1⁺ stable action-initiation excited neurons. **(g)** Modulation, averaged in 0.5-s bins around earned reward of these neurons. 2-way ANCOVA, Training: $F_{1.99, 57.57} = 2.61$, $P = 0.08$; Time bin: $F_{1.84, 53.36} = 0.80$, $P = 0.45$; Training x Time: $F_{3.87, 112.10} = 1.74$, $P = 0.15$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



Extended Data Figure 3-1: Validation of chemogenetic approach to inhibit or activate DMS D1⁺ or A2A⁺ neurons. Fluorescence-guided, whole-cell patch clamp recordings in current-clamp mode were used to validate the efficacy of chemogenetic manipulation of DMS D1⁺ and A2A⁺ neurons. **(a)** Representative immunofluorescent image of a biocytin-filled, mCherry-positive cell. **(b)** Representative recordings of action potentials in single cells before (Baseline) and after CNO (10 μM for hM3Dq and 100 μM for hM4Di) bath application. Current injection intensity applied to induce action potential firing was 200 (Control, D1⁺ hM3Dq, A2A⁺ hM3Dq) or 250 pA (all others), with the same intensity used for baseline and CNO recordings. Cell membrane potentials were held at -80 mV using negative current to ensure consistent baseline conditions across recordings. **(c)** Percent change in action potentials under CNO, relative to pre-CNO baseline. 1-way ANOVA, $F_{4, 58} = 13.29$, $P < 0.0001$. Control, $N = 19$ cells, 4 D1-cre mice (1 male); D1-cre, hm4Di: $N = 8$ cells, 4 mice (3 males), hM3Dq: $N = 10$ cells, 5 mice (3 males); A2A-cre, hM4Di: $N = 8$ cells, 4 mice (3 males), hM3Dq: $N = 10$ cells, 5 mice (3 males). Data presented as mean ± s.e.m. Males = closed circles, Females = open circles. Scale bar = 10 μm. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, uncorrected. We were able to effectively chemogenetically inhibit and activate action potentials in both D1⁺ and A2A⁺ DMS neurons.



Extended Data Figure 3-2. Food-port entries during training and test with DMS D1⁺ neuron chemogenetic manipulation. (a-c) Chemogenetic inactivation of DMS D1⁺ neurons during learning. WT: N = 7 (2 males); D1-cre: N = 9 (5 males). **(a)** Training entry rate. 2-way ANOVA, Training: $F_{2.35, 32.91} = 7.40$, $P = 0.001$; Genotype: $F_{1, 14} = 1.30$, $P = 0.27$; Training x Genotype: $F_{4, 56} = 0.46$, $P = 0.77$. **(b)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 14} = 0.003$, $P = 0.96$; Value: $F_{1, 14} = 0.05$, $P = 0.83$; Genotype: $F_{1, 14} = 0.03$, $P = 0.86$. **(c)** Training average presses/earned reward outcome. 2-way ANOVA, Training: $F_{2.05, 28.68} = 32.52$, $P < 0.0001$; Genotype: $F_{1, 14} = 0.57$, $P = 0.46$; Training x Genotype: $F_{4, 56} = 0.77$, $P = 0.55$. **(d-f)** Chemogenetic activation of DMS D1⁺ neurons during learning. WT: N = 6 (4 males); D1-cre: N = 6 (3 males). **(d)** Training entry rate. 2-way ANOVA, Training: $F_{2.38, 23.79} = 12.75$, $P < 0.0001$; Genotype: $F_{1, 10} = 0.21$, $P = 0.74$; Training x Genotype: $F_{8, 80} = 0.74$, $P = 0.65$. **(d)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 10} < 0.0001$, $P > 0.99$; Value: $F_{1, 10} = 0.89$, $P = 0.37$; Genotype: $F_{1, 10} = 0.51$, $P = 0.49$. **(f)** Training average presses/earned reward outcome. 2-way ANOVA, Training: $F_{4.03, 60.47} = 9.42$, $P < 0.0001$; Genotype: $F_{1, 15} = 0.0001$, $P = 0.99$; Training x Genotype: $F_{8, 120} = 0.60$, $P = 0.78$. **(g-i)** Chemogenetic inactivation of DMS D1⁺ neurons at test of behavioral control strategy after learning. WT: N = 12 (7 males); D1-cre: N = 12 (5 males). **(g)** Training entry rate. 2-way ANOVA, Training: $F_{3.01, 54.19} = 13.85$, $P < 0.0001$; Genotype: $F_{1, 18} < 0.0001$, $P > 0.99$; Training x Genotype: $F_{4, 72} = 0.34$, $P = 0.85$. **(h)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 18} = 0.001$, $P = 0.97$; Value: $F_{1, 18} = 11.24$, $P = 0.004$; Genotype: $F_{1, 18} = 3.02$, $P = 0.10$. **(i)** Training average presses/earned reward outcome. 2-way ANOVA, Training: $F_{1.86, 44.59} = 96.52$, $P < 0.0001$; Genotype: $F_{2, 24} = 0.13$, $P = 0.88$; Training x Genotype: $F_{8, 96} = 0.13$, $P > 0.99$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. *$P < 0.05$. Neither chemogenetic inhibition nor activation of DMS D1⁺ neurons affected checks of the food-delivery port or altered the press-reward action-outcome relationship.



Extended Data Figure 4-1: Entries into the food-delivery port during training and test for DMS A2A$^+$ imaging experiment- habitual subjects. (a) Training entry rate. 1-way ANOVA, Training: $F_{2.69,\ 8.06} = 1.93$, $P = 0.20$ **(b)** Test entry rate. 2-tailed t-test, $t_3 = 1.87$, $P = 0.16$, 95% CI -5.00 - 1.30. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.

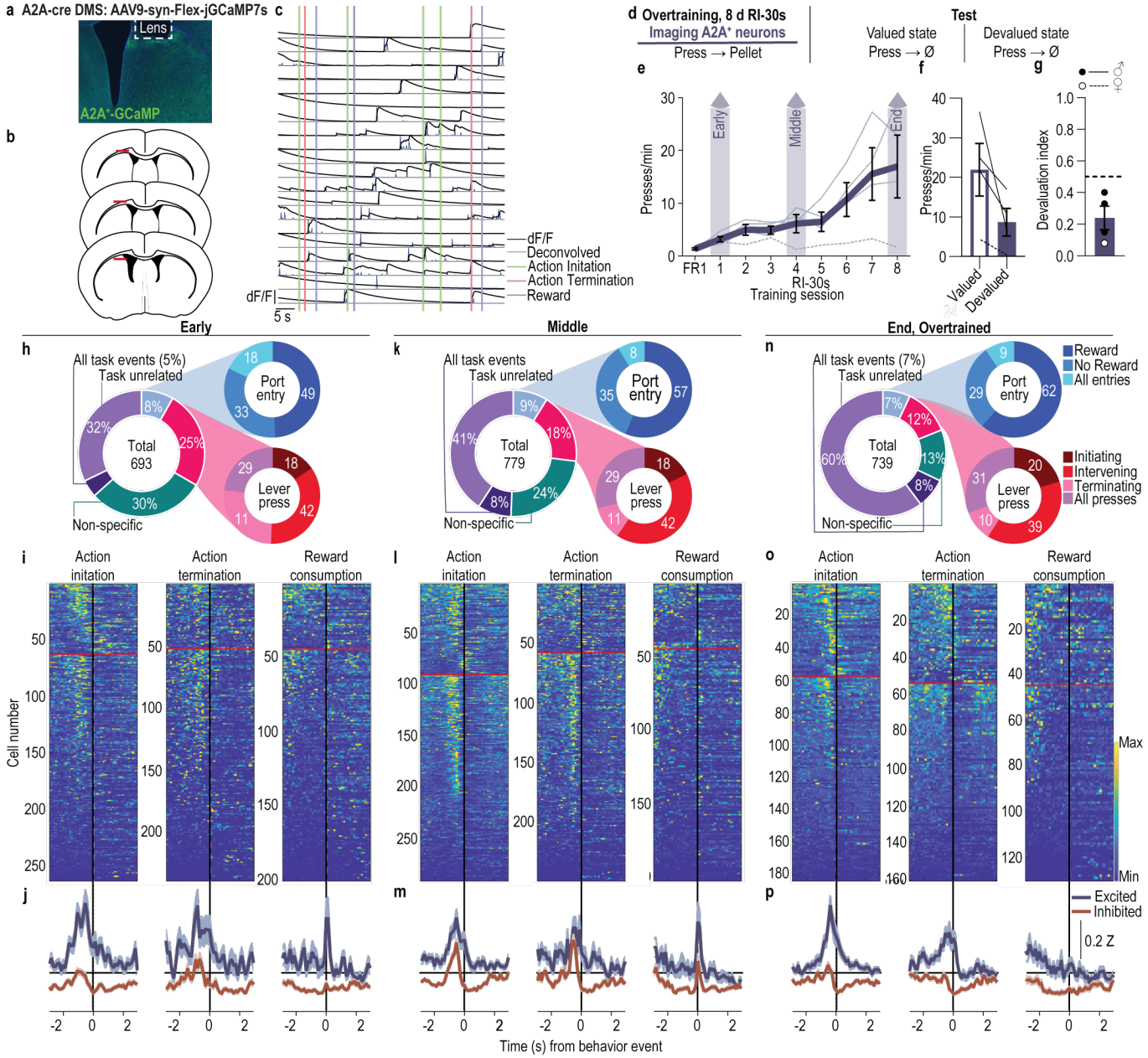
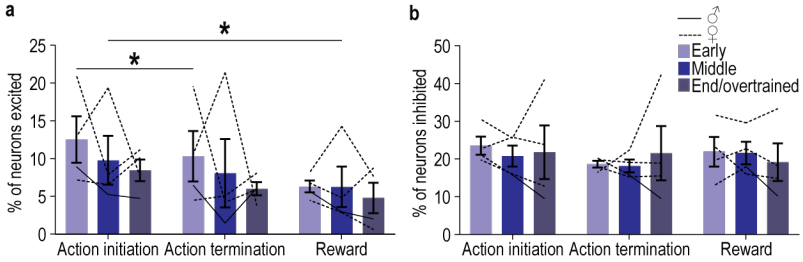
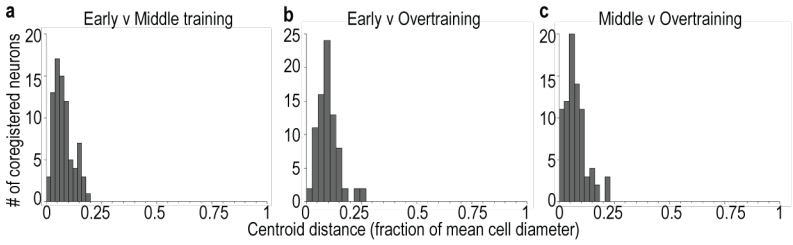


Figure 4-2: DMS A2A⁺ neurons are modulated by actions and rewards during instrumental learning and overtraining in subjects that did not form habits. (a) Representative image of cre-dependent jGCaMP7s expression in DMS A2A⁺ neurons. **(b)** Map of DMS GRIN lens placements for all subjects. **(c)** Representative dF/F and deconvolved calcium signal. **(d)** Procedure. RI, random-interval reinforcement schedule. **(e)** Training press rate. 2-way ANOVA, Training: $F_{1.49, 4.48} = 5.58$, $P = 0.06$. **(f)** Test press rate. 2-tailed t-test, $t_3 = 2.30$, $P = 0.11$, 95% CI -31.60 - 5.10. **(g)** Devaluation index [(Devalued condition presses)/(Valued condition presses + Devalued presses)]. One-tailed Bayes factor, $BF_{10} = 5.40$. **(h-j)** Activity of DMS A2A⁺ neurons on the 1st (Early) session of RI training. **(h)** Percent of all recorded neurons ($N = 693$) significantly modulated by lever presses, food-delivery port checks, and reward. **(i)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each DMS A2A⁺ neuron significantly modulated around lever-press action initiation (right), action termination (middle), or reward consumption (left). Above red line = excited, below = inhibited. **(j)** Z-scored activity of each population of modulated neurons. **(k-m)** Activity of DMS A2A⁺ neurons on the 4th (Middle) training session. **(k)** Percent of all recorded neurons ($N = 779$) significantly modulated by lever presses, food-delivery port checks, and reward. **(l)** Heat map of minimum to maximum deconvolved activity of each DMS A2A⁺ neuron significantly modulated around lever-press action initiation (right), action termination (middle), or reward (left). **(m)** Z-scored activity of each population of modulated neurons. **(n-p)** Activity of DMS A2A⁺ neurons on the 8th (End, Overtrain) training session. **(k)** Percent of all recorded neurons ($N = 739$) significantly modulated by lever

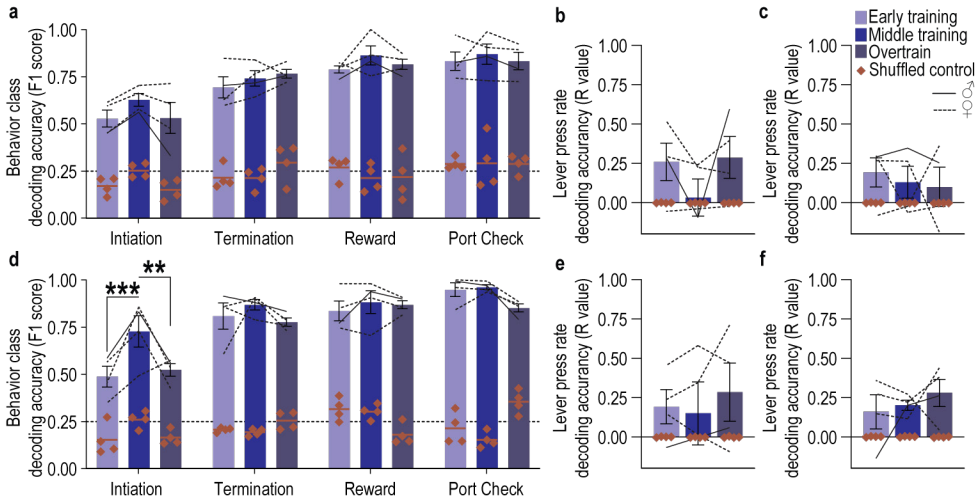
presses, food-delivery port checks, and reward. **(l)** Heat map of minimum to maximum deconvolved activity of each DMS A2A$^+$ neuron significantly modulated around lever-press action initiation (right), action termination (middle), or reward (left). **(m)** Z-scored activity of each population of modulated neurons. Data presented as mean ± s.e.m. A2A-cre: *N* = 4 (3 male). Males = closed circles/solid lines, Females = open circles/dashed lines. A2A$^+$ neurons in subjects that did not form habits with overtraining are activated prior to action initiation and termination and a subset are also activated by earned reward experience.



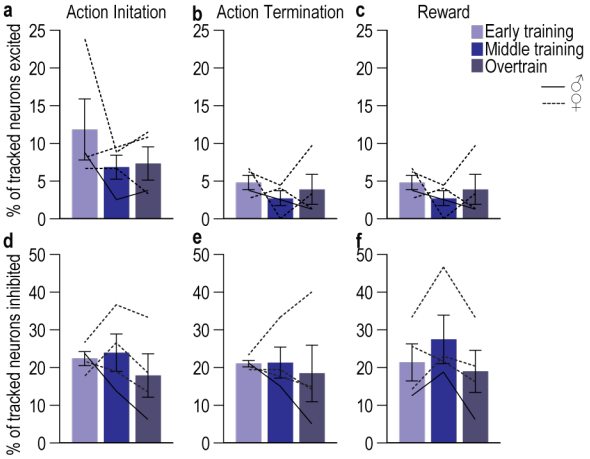
Extended Data Figure 4-3: Percent of DMS A2A⁺ neurons activated or inhibited around actions and rewards. (a) Percent of all recorded neurons (Early: N = 671, Middle: N = 697; End N = 658) classified (auROC values >95th percentile of the distribution of shuffled auROCs within 2 s before or after event) as significantly excited around initiating lever presses, terminating lever presses, or reward consumption. Action initiation excited more DMS A2A⁺ neurons than termination or reward. 2-way ANOVA, Event: $F_{1.12, 3.36} = 23.83$, $P = 0.01$; Training session: $F_{1.32, 3.97} = 0.05$, $P = 0.54$; Training x Event: $F_{1.86, 5.59} = 0.29$, $P = 0.77$. **(b)** Percent of all recorded neurons classified as significantly inhibited around initiating lever presses, terminating lever presses, or reward consumption. Similar proportions of the DMS A2A⁺ neurons were inhibited around each event type. 2-way ANOVA, Event: $F_{1.90, 5.69} = 2.03$, $P = 0.23$; Training session: $F_{1.12, 3.36} = 0.05$, $P = 0.86$; Training x Event: $F_{1.45, 4.33} = 0.90$, $P = 0.44$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



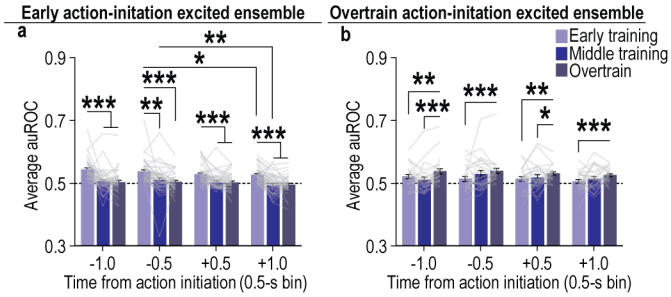
Extended Data Figure 5-1. Representative example of coregistration of DMS A2A^{+} neurons across training. (a) Distribution of the distance between cell centroids (as a fraction of total cell diameter) of co-registered cell pairs in the 1st (early) and 4th (middle) training sessions. **(b)** Distribution of centroid distance of co-registered cell pairs in the 1st and 8th (overtrain) training sessions. **(c)** Distribution of centroid distance of co-registered cell pairs in the 4th and 8th training sessions.



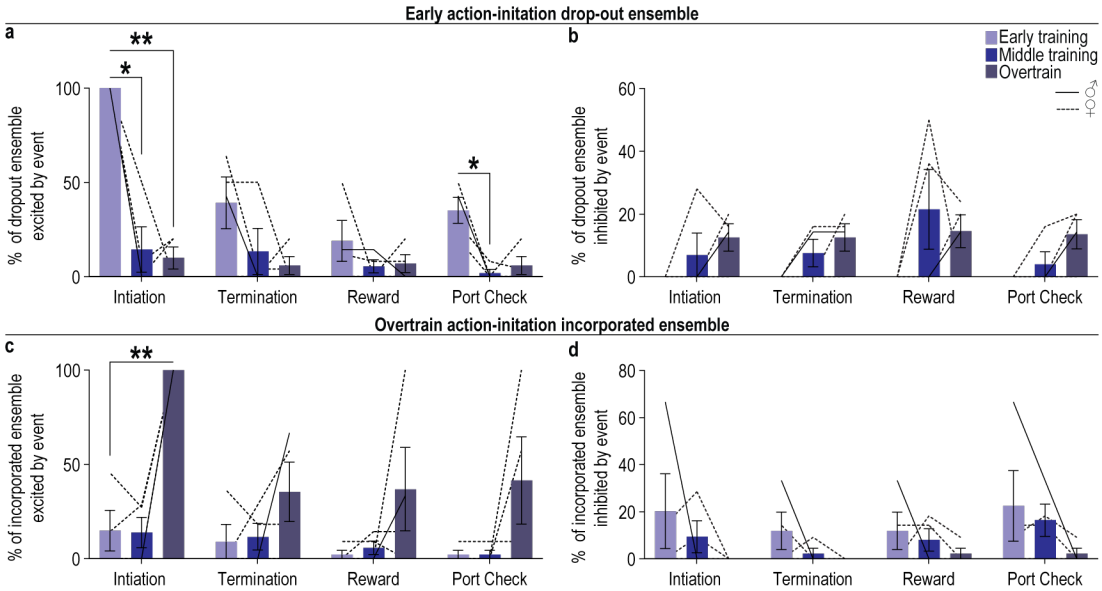
Extended Data Figure 5-2. Instrumental behavior can be decoded from the activity of DMS A2A⁺ action-initiation neuronal ensemble. (a-b) Decoding of behavioral events from the activity of DMS A2A⁺ early action-initiation excited neurons. **(a)** Behavior class (initiating lever press, terminating press, reward collection, non-reinforced food-port check) decoding accuracy compared to shuffled control. Line at 0.25 = chance. 3-way ANOVA, Neuron activity (v. shuffled): $F_{1, 3} = 169.03$, $P < 0.001$; Training session: $F_{2, 6} = 5.55$, $P = 0.04$; Behavior class: $F_{3, 9} = 50.84$, $P < 0.001$; Neuron activity x Training: $F_{2, 6} = 5.27$, $P = 0.05$; Neuron activity x Behavior Class: $F_{3, 9} = 4.63$, $P = 0.03$; Training x Behavior Class: $F_{6, 18} = 1.79$, $P = 0.16$; Neuron activity x Training x Behavior class: $F_{6, 18} = 0.34$, $P = 0.91$. **(b)** Lever-press rate decoding accuracy. R = correlation coefficient between actual and decoded press rate. 2-way ANOVA, Neuron activity: $F_{1, 3} = 3.76$, $P = 0.14$; Training: $F_{1.33, 3.99} = 4.49$, $P = 0.19$; Neuron activity x Training: $F_{1.32, 3.97} = 2.52$, $P = 0.819$. **(c)** Accuracy with which lever-press rate can be can be decoded from the activity of DMS A2A⁺ early action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 7.44$, $P = 0.07$; Training: $F_{1.40, 4.21} = 0.17$, $P = 0.77$; Neuron activity x Training: $F_{1.41, 4.22} = 0.16$, $P = 0.79$. **(d-f)** Decoding of behavioral events from the activity of DMS A2A⁺ overtrain action-initiation excited neurons. **(d)** Behavior class decoding accuracy compared to shuffled control. 3-way ANOVA, Neuron activity: $F_{1, 3} = 1296.08$, $P < 0.001$; Training session: $F_{2, 6} = 2.75$, $P = 0.14$; Behavior class: $F_{3, 9} = 32.57$, $P < 0.001$; Neuron activity x Training: $F_{2, 6} = 1.99$, $P = 0.22$; Neuron activity x Behavior Class: $F_{3, 9} = 11.86$, $P = 0.002$; Training x Behavior Class: $F_{6, 18} = 2.36$, $P = 0.07$; Neuron activity x Training x Behavior class: $F_{6, 18} = 2.91$, $P = 0.04$. **(e)** Lever-press rate decoding accuracy. 2-way ANOVA, Neuron activity: $F_{1, 3} = 1.92$, $P = 0.26$; Training: $F_{1.66, 4.96} = 0.57$, $P = 0.57$; Neuron activity x Training: $F_{1.63, 4.88} = 0.54$, $P = 0.58$. **(f)** Accuracy with which lever-press rate can be can be decoded from the activity of DMS A2A⁺ overtrain action-initiation inhibited neurons. 2-way ANOVA, Neuron activity: $F_{1, 3} = 30.03$, $P = 0.01$; Training: $F_{1.52, 4.57} = 0.45$, $P = 0.61$; Neuron activity x Training: $F_{1.51, 4.54} = 0.48$, $P = 0.60$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. $^{**}P < 0.01$, $^{***}P < 0.001$. Actions, checking for, and receiving reward, can be decoded from both the population activity of DMS A2A⁺ neurons that are excited by action initiation early in training and those neurons that are excited by action initiation after overtraining. Population decoding of action initiation improves with some training and then decreases with overtraining. Action rate can only be significantly decoded from overtrain action-initiation inhibited DMS A2A⁺ neurons, suggesting these neurons contribute to the decoding accuracy of the whole population of DMS A2A⁺ neurons.



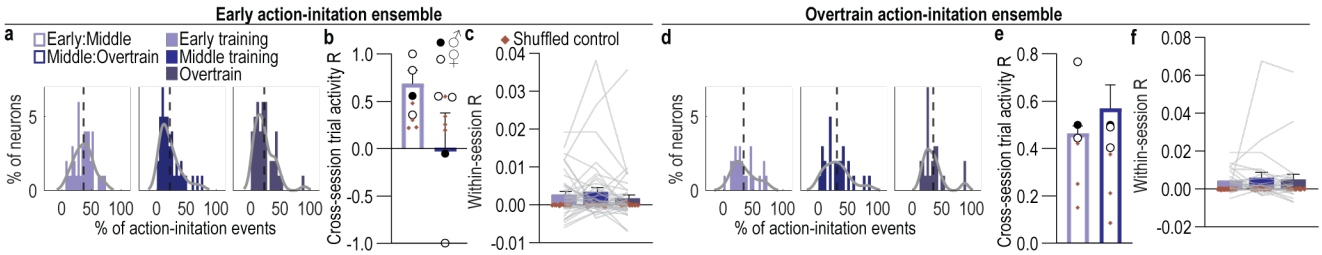
Extended Data Figure 5-3. Ensembles of DMS A2A⁺ neurons are excited and inhibited by action initiation, action termination, and reward. (a-c) Percent of all recorded coregistered neurons (Average 73.25 coregistered neurons/mouse, s.e.m. 17.09) significantly excited by action initiation, action termination, or reward. **(a)** Approximately 7 - 12% of coregistered DMS A2A⁺ neurons were excited by action initiation. 1-way ANOVA, $F_{1.16, 3.48} = 1.79$, $P = 0.27$. **(b)** Approximately 4 - 10% of coregistered DMS A2A⁺ neurons were excited by action termination. 1-way ANOVA, $F_{1.48, 4.43} = 1.12$, $P = 0.38$. **(c)** Only 3 - 5% of coregistered DMS A2A⁺ neurons were excited by reward. 1-way ANOVA, $F_{1.89, 5.68} = 0.75$, $P = 0.51$. **(d-f)** Percent of all recorded coregistered neurons significantly inhibited by action initiation, action termination, or reward. **(d)** Approximately 18-24% of coregistered DMS A2A⁺ neurons were inhibited by action initiation. 1-way ANOVA, $F_{1.05, 3.16} = 1.14$, $P = 0.37$. **(e)** Approximately 19 - 21% of coregistered DMS A2A⁺ neurons were inhibited by action termination. 1-way ANOVA, $F_{1.02, 3.05} = 0.20$, $P = 0.69$. **(f)** Approximately 19 - 27% of coregistered DMS A2A⁺ neurons were inhibited by reward. 1-way ANOVA, $F_{1.76, 5.28} = 3.52$, $P = 0.11$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. In no case did the proportion of excited or inhibited A2A⁺ neurons change with training.



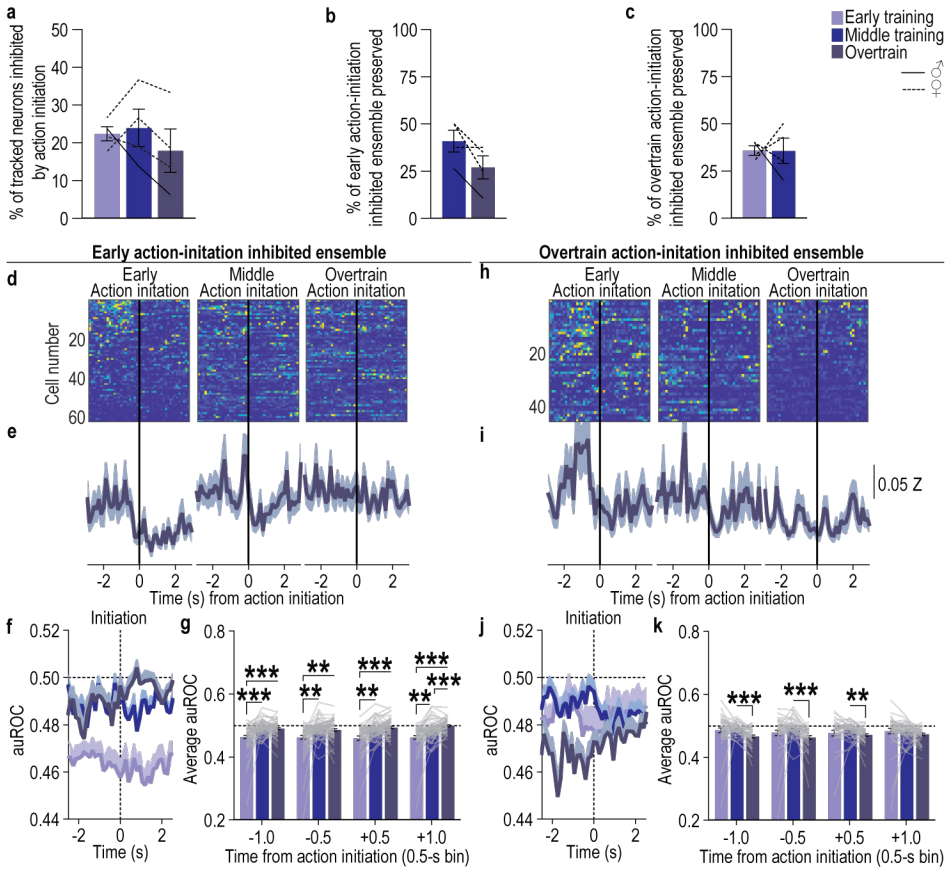
Extended Data Figure 5-4. Quantification of the modulation of DMS A2A⁺ action-initiation excited neurons. (a) Modulation across training of DMS A2A⁺ early action-initiation excited neurons (*N* = 42 neurons/4 mice; average 16.8 neurons/mouse, s.e.m. = 5.61). Modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.74, 69.82} = 4.10$, $P = 0.03$; Time bin: $F_{1.70, 68.14} = 0.18$, $P = 0.80$; Training x Time: $F_{4.70, 187.84} = 0.90$, $P = 0.68$. **(b)** Modulation across training of A2A⁺ overtrain action-initiation excited neurons (*N* = 25 neurons/4 mice; average 6.25 neurons/mouse, s.e.m. 2.69). Modulation index around action initiation. 2-way ANCOVA, Training: $F_{1.75, 40.22} = 1.15$, $P = 0.32$; Time bin: $F_{1.74, 39.90} = 0.10$, $P = 0.88$; Training x Time: $F_{4.23, 97.45} = 0.64$, $P = 0.70$. Data presented as mean ± s.e.m.



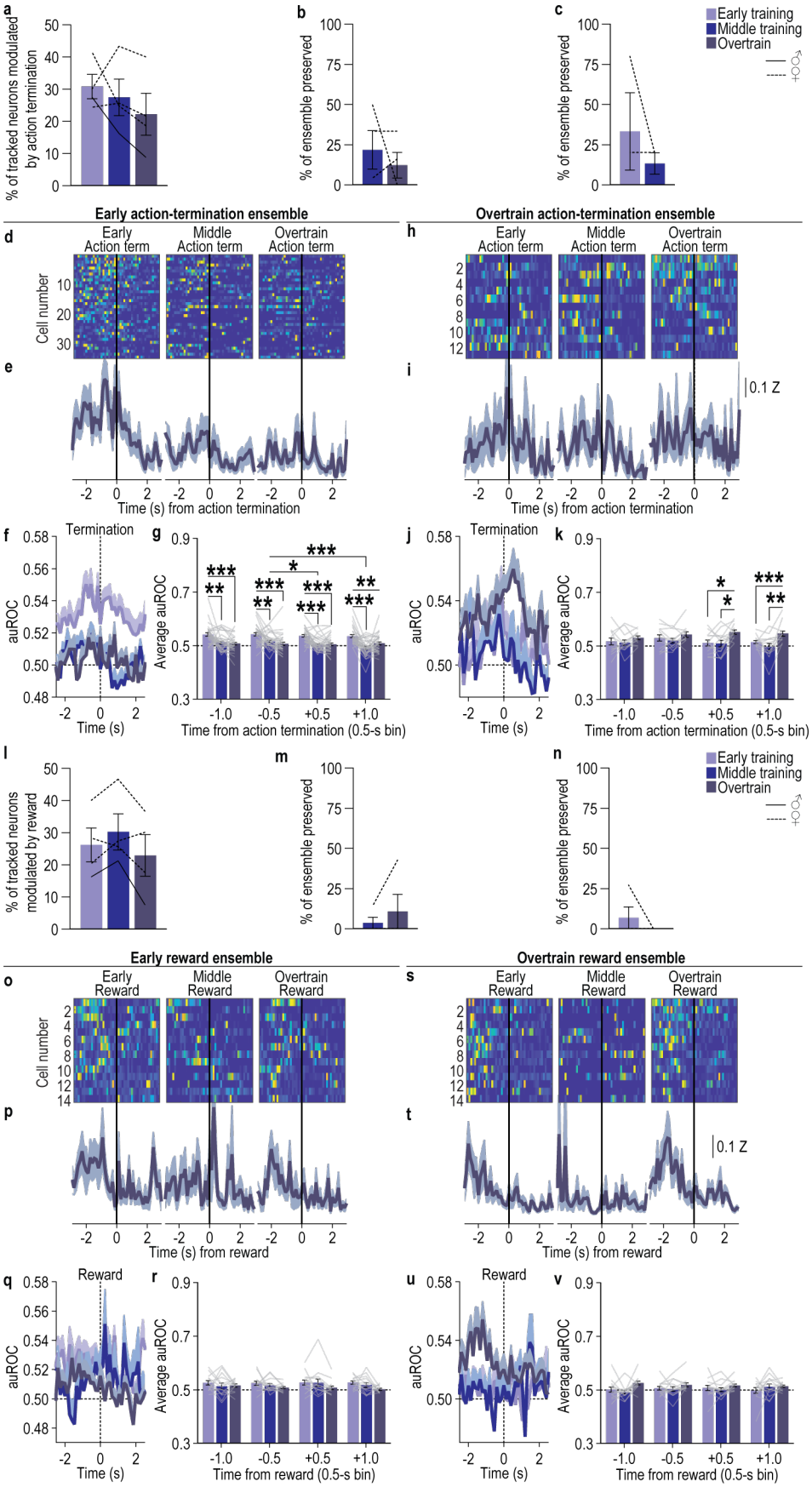
Extended Data Figure 5-5. Encoding of other task variables by A2A$^+$ neurons that drop out of the early action-initiation ensemble or get incorporated into the overtrain action-initiation excited ensemble. (a-b) Percentage of the neurons that dropout of the early action-initiation excited ensemble that are excited (a; 2-way ANOVA, Event x Training: $F_{1.79, 5.37} = 7.05$, $P = 0.03$; Event: $F_{1.21, 3.63} = 30.74$, $P = 0.006$; Training: $F_{1.13, 3.40} = 14.46$, $P = 0.02$) or inhibited (b; 2-way ANOVA, Event: $F_{1.98, 5.94} = 1.64$, $P = 0.27$; Training: $F_{1.50, 4.50} = 2.77$, $P = 0.17$; Event x Training: $F_{1.46, 4.37} = 1.25$, $P = 0.35$) by other events (terminating lever press, reward, food-delivery port check). By definition, these neurons stop being activated by action initiation with training. They do not begin to be activated by other task events with training. Instead, fewer of these neurons are activated by other task events and some of them become inhibited by tasks events. **(c-d)** Percentage of the neurons that are incorporated into the overtrain action-initiation excited ensemble that are excited (c; 2-way ANOVA, Training: $F_{1.1, 3.48} = 11.06$, $P = 0.03$; Event: $F_{2.24, 6.71} = 3.67$, $P = 0.08$; Event x Training: $F_{1.54, 4.62} = 2.41$, $P = 0.19$) or inhibited (d; 2-way ANOVA, Event: $F_{1.16, 3.48} = 1.95$, $P = 0.25$; Training: $F_{1.09, 3.28} = 1.24$, $P = 0.35$; Event x Training: $F_{1.89, 5.68} = 0.83$, $P = 0.48$) by other events (terminating lever press, reward, food-delivery port check). By definition, these neurons became activated by action initiation at overtraining. Earlier in training very few of these neurons were activated by other events. A small proportion of these neurons were inhibited by task events earlier in training and this proportion decreased to near 0 with training. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



Extended Data Figure 5-6. Fidelity with which DMS A2A⁺ neurons encode action initiation. (a-c) Fidelity with which A2A⁺ early action-initiation excited neurons encode action initiation. **(a)** Distribution of the percentage of early action-initiation excited neurons as a function of the percentage of action-initiation events to which they respond for each training phase. **(b)** Cross-session correlation of the response distributions. 2-way ANOVA, Neuron activity distribution (v. shuffled): $F_{1, 3} = 0.03$, $P = 0.88$; Training sessions: $F_{1, 3} = 1.34$, $P = 0.33$; Distribution x Training: $F_{1, 3} = 2.99$, $P = 0.18$. Early action-initiation A2A⁺ neurons tend to respond on less than half of the action-initiation events and this decreases with training and is not correlated across training sessions. **(c)** Within-session correlation of the activity around action initiation of each early action-initiation excited neuron. 2-way ANCOVA, Neuron activity: $F_{1, 36} = 2.71$, $P = 0.11$; Training: $F_{2, 72} = 2.47$, $P = 0.09$; Activity x Time: $F_{2, 72} = 2.45$, $P = 0.09$. The activity of early action-initiation excited A2A⁺ neurons around action initiation is not significantly correlated within a training session. **(d-e)** Fidelity with which A2A⁺ overtrain action-initiation excited neurons encode action initiation. **(d)** Distribution of the percentage of overtrain action-initiation excited neurons as a function of the percentage of action-initiation events to which they respond for each training phase. **(e)** Cross-session correlation of the response distributions. 2-way ANOVA, Neuron activity distribution: $F_{1, 2} = 32.64$, $P = 0.03$; Training sessions: $F_{1, 2} = 0.54$, $P = 0.54$; Distribution x Training: $F_{1, 2} = 0.45$, $P = 0.57$. Overtrain action-initiation A2A⁺ neurons tend to respond on less than half the action-initiation events across training and this is consistent across training. **(f)** Within-session correlation of the activity around action initiation of each overtrain action-initiation excited neuron. 2-way ANCOVA, Neuron activity: $F_{1, 22.00} = 0.40$, $P = 0.53$; Training: $F_{1.24, 27.30} = 0.63$, $P = 0.47$; Activity x Time: $F_{1.23, 27.09} = 0.65$, $P = 0.46$. The activity of overtrain action-initiation excited A2A⁺ neurons around action initiation is not correlated above shuffled control within a training session. A2A-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = closed circles, Females = open circles.



Extended Data Figure 5-7. The ensemble of DMS A2A⁺ neurons inhibited by action initiation shifts as habits form. (a) Percent of all recorded coregistered DMS A2A⁺ neurons (Average 73.25 coregistered neurons/mouse, s.e.m. 17.09) significantly inhibited by action initiation. Approximately 18 - 24% of coregistered DMS A2A⁺ neurons were inhibited by action initiation and this did not significantly change across training. 1-way ANOVA, $F_{1.05, 3.16} = 1.14$, $P = 0.37$. **(b)** Percent of DMS A2A⁺ early action-initiation-inhibited neurons that were then also significantly inhibited by action initiation on the 4th (middle) and 8th (overtrain) training sessions. Approximately 27 - 41% of the early action-initiation inhibited ensemble continued to be inhibited around action initiation during the middle and overtraining phases of training. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 2.70$, $P = 0.07$, 95% CI -30.41 - 2.52. **(c)** Percent of DMS A2A⁺ overtrain action-initiation-inhibited neurons that were significantly modulated by action initiation on prior the 1st (early) and 4th (middle) training sessions. Approximately 36% of the overtraining action-initiation inhibited ensemble was also inhibited around action initiation during the preceding early and middle training phases. The proportion preserved did not change with training. 2-tailed t-test, $t_3 = 0.01$, $P = 0.99$, 95% CI -29.03 - 28.80. **(d-g)** Activity and modulation across training of DMS A2A⁺ early action-initiation-inhibited neurons. Heat map of minimum to maximum deconvolved activity (sorted by total activity) (d), Z-scored activity (e), and area under the receiver operating characteristic curve (auROC) modulation index (f) of these cells around action initiation across training. **(g)** auROC modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.38, 83.91} = 1.38$, $P = 0.26$; Time bin: $F_{2.48, 151.16} = 3.28$, $P = 0.02$; Training x Time: $F_{5.10, 311.30} = 3.80$, $P = 0.001$. This early action-initiation inhibited ensemble became less inhibited by action initiation as training progressed. **(h-k)** Activity and modulation across training of coregistered DMS A2A⁺ overtrain early action-initiation-inhibited neurons. Heat map of minimum to maximum deconvolved activity (h), Z-scored activity (i), and area under the receiver operating characteristic curve (auROC) modulation index (j) of these cells around action initiation across training. **(k)** auROC modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.21, 53.15} = 5.67$, $P = 0.005$; Time bin: $F_{2.62, 115.18} = 2.16$, $P = 0.10$; Training x Time: $F_{4.48, 196.92} = 3.56$, $P = 0.006$. This overtrain action-initiation inhibited ensemble became more inhibited prior to action initiation as training progressed. A2A-cre: $N = 4$ (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.



Extended Data Figure 5-8. The ensembles of DMS A2A⁺ neurons encoding action termination and reward shift as habits form. (a) Percent of all recorded coregistered DMS A2A⁺ neurons (Average 73.25 coregistered

neurons/mouse, s.e.m. 17.09) significantly modulated (excited or inhibited) by action termination. Approximately 22 - 31% of DMS A2A$^+$ neurons were modulated around action termination and this did not change with training. $F_{1.00, 3.02} = 1.09$, $P = 0.37$. **(b)** Percent of DMS A2A$^+$ early action-termination excited neurons (N = 35 neurons/4 mice; average 8.75 neurons/mouse, s.e.m. 5.45) that were also significantly excited by action termination on the 4th (middle) and 8th (overtrain) training sessions. Only 12 - 22% of the early action-termination ensemble continued to be excited by action termination during the middle and overtraining phases of training. The proportion preserved did not change with training. Middle v. Overtrain $t_3 = 0.69$, P =0.54. **(c)** Percent of DMS A2A$^+$ overtrain action-termination excited (N = 13 neurons/4 mice; average 3.25 neurons/mouse, s.e.m. 1.18) neurons that were also significantly excited by action termination on the 1st (early) and 4th (middle) training sessions. Only 13 - 33% of the overtraining action-termination ensemble was also excited by action termination during the preceding early and middle training phases. The proportion preserved did not change with training. Early v. Middle $t_3 = 1.00$, P =0.42. **(d-g)** Activity and modulation across training of DMS A2A$^+$ early action-termination excited neurons. Heat map of minimum to maximum deconvolved activity (sorted by total activity) (d), Z-scored activity (e), and area under the receiver operating characteristic curve (auROC) modulation index (f) of these cells around action termination across training. **(g)** auROC modulation index averaged across 0.5-s bins around action termination. Training: $F_{1.66, 54.75} = 3.77$, $P = 0.03$; Time bin: $F_{1.82, 60.19} = 1.17$, $P = 0.33$; Training x Time: $F_{4.19, 138.28} = 1.01$, $P = 0.41$. This early action-termination ensemble became less modulated by action termination as training progressed. **(h-k)** Activity and modulation across training of DMS A2A$^+$ overtrain action-termination excited neurons. Heat map of minimum to maximum deconvolved activity (h), Z-scored activity (i), and area under the receiver operating characteristic curve (auROC) modulation index (j) of these cells around action termination across training. **(k)** auROC modulation index averaged across 0.5-s bins around action termination. Training: $F_{1.82, 19.98} = 3.20$, $P = 0.07$; Time bin: $F_{1.55, 17.10} = 2.28$, $P = 0.10$; Training x Time: $F_{3.25, 37.70} = 0.31$, $P = 0.83$. This overtrain action-termination ensemble became slightly more modulated by action termination as training progressed. **(l)** Percent of coregistered DMS A2A$^+$ neurons significantly modulated (excited or inhibited) by collection of the earned reward. Approximately 22 - 30% of DMS A2A$^+$ neurons were modulated by reward and this did not significantly change with training. $F_{1.32, 3.96} = 2.06$, $P = 0.23$. **(m)** Percent of DMS A2A$^+$ early reward excited neurons (N = 14 neurons/4 mice; average 3.50 neurons/mouse, s.e.m. 1.19) that were then significantly excited by reward on the 4th (middle) and 8th (overtrain) training sessions. Only 4 - 11% of the small early reward ensemble continued to be modulated by reward during the middle and overtraining phases of training. The proportion preserved did not change with training. Middle v. Overtrain 2-tailed Wilcoxon signed rank test, W = 1.00, P > 0.99. **(n)** Percent of DMS A2A$^+$ overtrain reward excited neurons (N = 14 neurons/4 mice; average 3.50 neurons/mouse, s.e.m. 2.50) that were significantly excited by reward on the 1st (early) and 4th (middle) training sessions. Only 0 - 7% of the small overtraining reward ensemble was also modulated by reward during the preceding early and middle training phases. The proportion preserved did not change with training. Early v. Middle 2-tailed Wilcoxon signed rank test, W = -1.00, P > 0.99. **(o-r)** Activity and modulation across training of DMS A2A$^+$ early reward excited neurons. Heat map of minimum to maximum deconvolved activity (o), Z-scored activity (p), and area under the receiver operating characteristic curve (auROC) modulation index (q) of these cells around reward across training. **(r)** auROC modulation index averaged across 0.5-s bins around reward. Training: $F_{1.53, 18.36} = 0.492$, $P = 0.57$; Time bin: $F_{1.54, 18.49} = 0.64$, $P = 0.50$; Training x Time: $F_{2.58, 30.98} = 0.41$, $P = 0.71$. **(s-v)** Activity and modulation across training of DMS A2A$^+$ overtrain reward excited neurons. Heat map of minimum to maximum deconvolved activity (s), Z-scored activity (t), and area under the receiver operating characteristic curve (auROC) modulation index (u) of these cells around reward on across training. **(v)** auROC modulation index averaged across 0.5-s bins around reward. Training: $F_{1.51, 18.10} = 1.07$, $P = 0.36$; Time bin: $F_{1.98, 23.76} = 2.78$, $P = 0.08$; Training x Time: $F_{2.94, 35.24} = 0.65$, $P = 0.58$. A2A-cre: N = 4 (1 male). Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines.

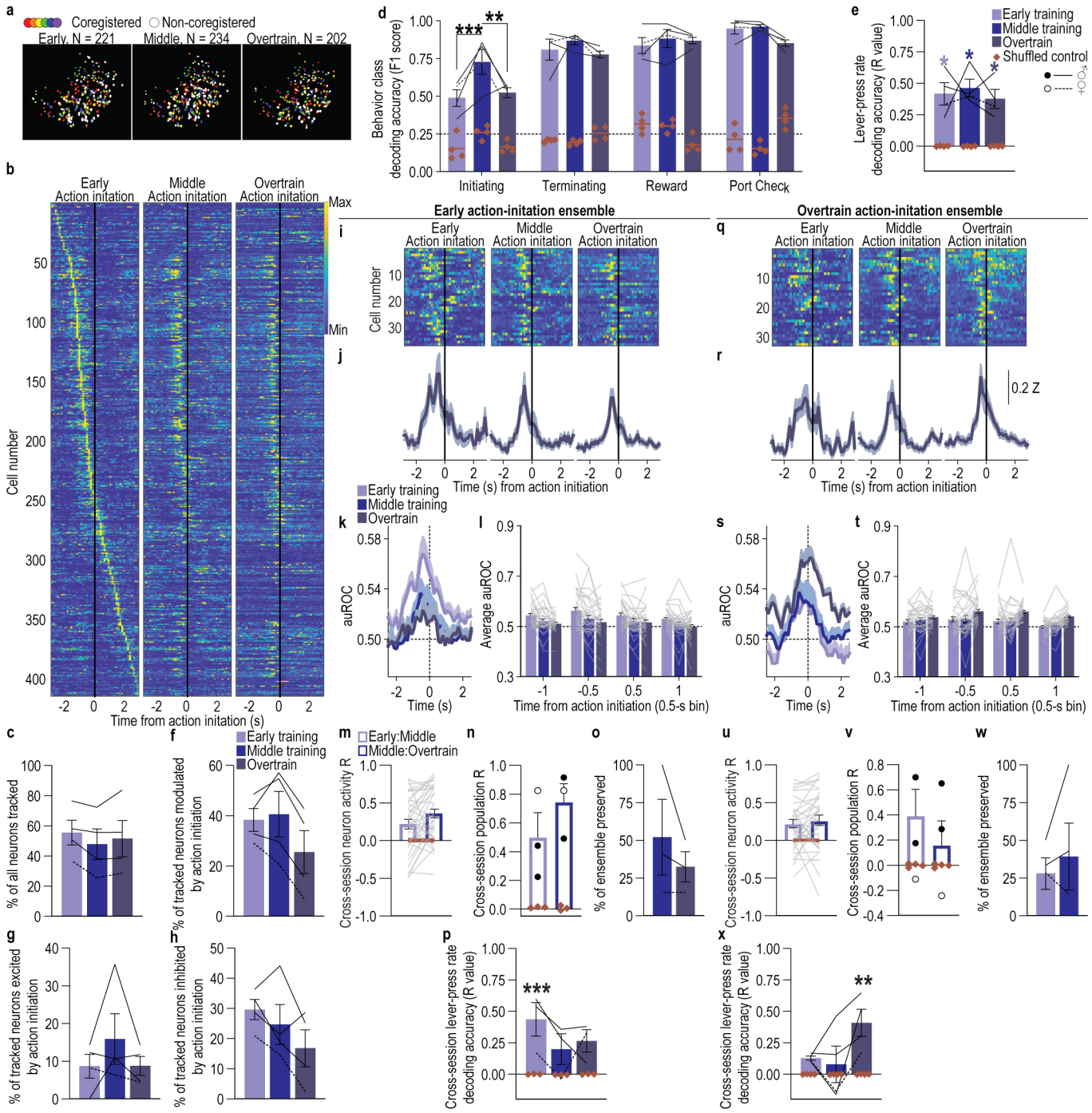
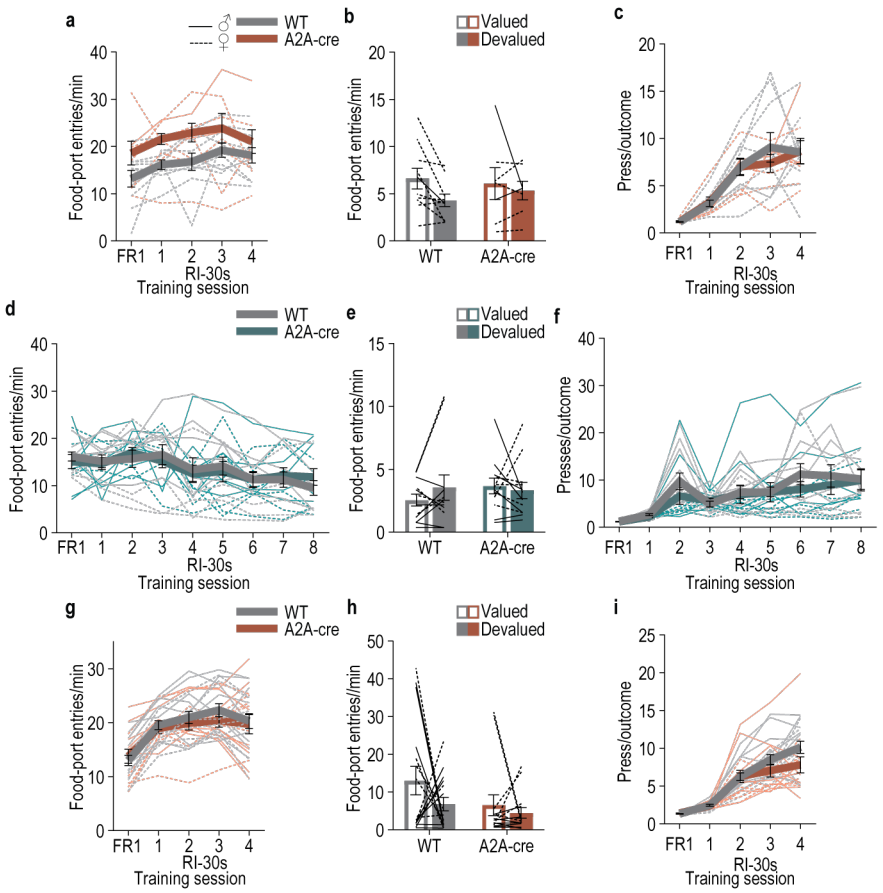
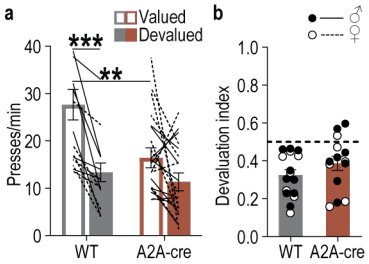


Figure 5-9: The ensemble of DMS A2A⁺ neurons that encodes action initiation is more stable if subjects remain goal-directed with overtraining. (a) Representative recorded A2A⁺ neuron spatial footprints during the first (early, left), 4th (middle), and 8th (overtrain) random-interval training session. Colored, co-registered neurons; white, non-co-registered neurons. **(b)** Heat map of minimum to maximum deconvolved activity (sorted by total activity) of each coregistered DMS A2A⁺ neuron around lever-press action initiation. **(c)** Percent of all A2A⁺ neurons coregistered across training in subjects that did not form habits with overtraining. 1-way ANOVA, $F_{1.32, 3.95} = 2.82$, $P = 0.17$. **(d)** Behavior class (initiating lever press, terminating press, reward collection, non-reinforced food-port check) decoding accuracy from A2A⁺ coregistered neuron activity compared to shuffled control. Line at 0.25 = chance. 3-way ANOVA, Neuron activity (v. shuffled): $F_{1, 3} = 548.47$, $P < 0.001$; Training session: $F_{2, 6} = 4.11$, $P = 0.08$; Behavior class: $F_{3, 9} = 40.4$, $P < 0.001$; Neuron activity x Training: $F_{2, 6} = 5.95$, $P = 0.038$; Neuron activity x Behavior Class: $F_{3, 9} = 16.24$, $P < 0.001$; Training x Behavior Class: $F_{6, 18} = 2.55$, $P = 0.02$; Neuron activity x Training x Behavior class: $F_{6, 18} = 5.11$, $P = 0.003$. **(e)** Lever-press rate decoding accuracy from A2A⁺ coregistered neuron activity. R = correlation coefficient between actual and decoded press rate. 2-way ANOVA, Neuron activity: $F_{1, 3} = 195.5$, $P = 0.0008$; Training: $F_{1.64, 4.92} = 0.23$, $P = 0.77$; Neuron activity x Training: $F_{1.62, 4.86}$

= 0.23, P = 0.76. **(f-h)** Percent of coregistered neurons (Average 106.5 coregistered neurons/mouse, s.e.m. 40.34) significantly modulated (f; 1-way ANOVA, $F_{1.05, 3.16}$ = 8.14, P = 0.06), excited (g; 1-way ANOVA, $F_{1.14, 3.43}$ = 1.72, P = 0.28), or inhibited (h; 1-way ANOVA, $F_{1.90, 5.70}$ = 3.37, P = 0.11) around action initiation. **(i-k)** Activity and modulation across training of DMS A2A$^+$ early action-initiation excited neurons (N = 47 neurons/4 mice; average 9.4 neurons/mouse, s.e.m. = 12.44). Heat map (i), Z-scored activity (j), and area under the receiver operating characteristic curve (auROC) modulation index (k) of early action-initiation excited neurons around action initiation across training. **(l)** Modulation index averaged across 0.5-s bins around action initiation. 2-way ANCOVA, Training: $F_{1.82, 64.00}$ = 0.48, P = 0.62; Time bin: $F_{2.18, 76.27}$ = 0.07, P = 0.94; Training x Time: $F_{3.78, 132.35}$ = 1.61, P = 0.18. **(m-n)** Cross-session correlation of the activity around action initiation of each early action-initiation excited neuron (m; 2-way ANCOVA, Neuron activity: $F_{1, 35}$ = 0.02, P = 0.90; Training: $F_{1, 35}$ = 0.001, P = 0.98; Activity x Time: $F_{1, 35}$ = 0.01, P = 0.91) or the population activity of these neurons (n; 2-way ANOVA, Neuron activity: $F_{1, 2}$ = 31.38, P = 0.03; Training: $F_{1, 2}$ = 1.25, P = 0.38; Activity x Time: $F_{1, 2}$ = 1.21, P = 0.39). **(o)** Percent of A2A$^+$ early action-initiation excited neurons that continued to be significantly excited by action initiation on the 4th and 8th training sessions. 2-tailed t-test, t_2 = 1.28, P =0.33, 95% CI -85.86 - 46.47. **(p)** Cross-session decoding accuracy of lever-press rate from the activity of A2A$^+$ early action-initiation-excited neuron population activity on the 1st training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: t_4 = 4.22, P = 0.04, 95% CI 0.03 - 0.85; Middle: t_4 = 1.97, P = 0.36, 95% CI -0.21 - 0.61; Overtrain: t_4 = 2.55, P = 0.19, 95% CI -0.15 - 0.67. **(q-s)** Activity and modulation across training of A2A$^+$ overtrain action-initiation excited neurons (N = 49 neurons/4 mice; average 9.80 neurons/mouse, s.e.m. 11.69). Heat map (q), Z-scored activity (r), and auROC modulation index (s) of overtrain action-initiation excited neurons around action initiation across training. **(t)** Modulation index around action initiation. 2-way ANCOVA, Training: $F_{1.65, 51.23}$ = 0.43, P = 0.43; Time bin: $F_{2.16, 67.03}$ = 00.17, P = 0.86; Training x Time: $F_{4.31, 133.64}$ = 0.59, P = 0.68. **(u-v)** Cross-session correlation of the activity around action initiation of each overtrain action-initiation excited neuron (u; 2-way ANCOVA, Neuron activity: $F_{1, 31}$ = 1.77, P = 0.19; Training: $F_{1, 31}$ = 0.28, P = 0.60; Activity x Time: $F_{1, 31}$ = 0.20, P = 0.66) or the population activity of these neurons (v; 2-way ANOVA, Neuron activity: $F_{1, 3}$ = 2.37, P = 0.22; Training: $F_{1, 3}$ = 1.20, P = 0.35; Activity x Time: $F_{1, 3}$ = 1.13, P = 0.37). **(w)** Percent of A2A$^+$ overtrain action-initiation excited neurons that were also significantly excited by action initiation on 1st and 4th training sessions. 2-tailed t-test, t_3 = 0.82, P =0.47, 95% CI -32.59 – 55.21. **(x)** Cross-session decoding accuracy of lever-press rate from the activity of A2A$^+$ overtrain action-initiation-excited neuron population activity on the 8th training session. Planned, Bonferroni corrected, 2-tailed t-tests, Early: t_6 = 1.48, P = 0.57, 95% CI -0.16 to 0.41; Middle: t_6 = 0.92, P > 0.99, 95% CI -0.21 - 0.36; Overtrain: t_6 = 4.73, P = 0.01, 95% CI 0.12 - 0.69. A2A-cre: N = 4 (3 male). Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. *P < 0.05, **P < 0.01, ***P < 0.001.



Extended Data Figure 6-1. Food-port entries during training and test with DMS A2A⁺ neuron chemogenetic manipulation. (a-c) Chemogenetic inactivation of DMS A2A⁺ neurons during learning. WT: *N* = 10 (1 male); A2A-cre: *N* = 7 (2 males). **(a)** Training entry rate. 2-way ANOVA, Training: $F_{2.87, 43.12} = 3.50$, $P = 0.02$; Genotype: $F_{1, 15} = 6.63$, $P = 0.02$; Training x Genotype: $F_{4, 60} = 0.31$, $P = 0.87$. **(b)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 15} = 1.08$, $P = 0.32$; Value: $F_{1, 15} = 4.06$, $P = 0.06$; Genotype: $F_{1, 15} = 0.03$, $P = 0.86$. **(c)** Training average lever presses/earned reward outcome. 2-way ANOVA, Training: $F_{2.38, 35.73} = 32.06$, $P < 0.0001$; Genotype: $F_{1, 15} = 0.08$, $P = 0.78$; Training x Genotype: $F_{4, 60} = 0.47$, $P = 0.76$. **(d-f)** Chemogenetic activation of DMS A2A⁺ neurons during learning. WT: *N* = 11 (7 males); A2A-cre: *N* = 10 (6 males). **(d)** Training entry rate. 2-way ANOVA, Training: $F_{4.22, 92.89} = 6.10$, $P = 0.0002$; Genotype: $F_{1, 22} = 0.00008$, $P = 0.98$; Training x Genotype: $F_{8, 176} = 0.54$, $P = 0.83$. **(e)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 22} = 1.50$, $P = 0.23$; Value: $F_{1, 22} = 0.35$, $P = 0.56$; Genotype: $F_{1, 22} = 0.27$, $P = 0.61$. **(f)** Training average lever presses/earned reward outcome. 2-way ANOVA, Training: $F_{2.42, 5334} = 16.33$, $P < 0.0001$; Genotype: $F_{1, 22} = 0.35$, $P = 0.56$; Training x Genotype: $F_{8, 176} = 0.76$, $P = 0.64$. **(g-i)** Chemogenetic inactivation of DMS A2A⁺ neurons during test of behavioral control after learning. WT: *N* = 16 (9 males); A2A-cre: *N* = 14 (8 males). **(g)** Training entry rate. 2-way ANOVA, Training: $F_{1.93, 53.92} = 28.28$, $P < 0.0001$; Genotype: $F_{1, 28} = 0.23$, $P = 0.63$; Training x Genotype: $F_{4, 112} = 0.70$, $P = 0.59$. **(h)** Test entry rate. 2-way ANOVA, Value x Genotype: $F_{1, 28} = 0.47$, $P = 0.50$; Value: $F_{1, 28} = 1.82$, $P = 0.19$; Genotype: $F_{1, 28} = 4.08$, $P = 0.05$. **(i)** Training average lever presses/earned reward outcome. 2-way ANOVA, Training: $F_{2.19, 61.42} = 81.93$, $P < 0.0001$; Genotype: $F_{1, 28} = 1.55$, $P = 0.22$; Training x Genotype: $F_{4, 112} = 2.25$, $P = 0.07$. Data presented as mean ± s.e.m. Males = solid lines, Females = dashed lines. Neither chemogenetic inhibition nor activation of DMS A2A⁺ neurons affected checks of the food-delivery port or altered the press-reward action-outcome relationship.



Extended Data Figure 6-2. Press rate during devaluation test with DMS A2A$^+$ neuron inhibition. (a) Test press rate. 2-way ANOVA, Value x Genotype: $F_{1, 28} = 6.46$, $P = 0.02$; Value: $F_{1, 28} = 28.86$, $P < 0.001$; Genotype: $F_{1, 28} = 5.15$, $P = 0.03$. **(b)** Devaluation index. 2-tailed t-test, $t_{28} = 1.36$; $P = 0.18$, 95% CI -0.03 - 0.16. Data presented as mean ± s.e.m. Males = closed circles/solid lines, Females = open circles/dashed lines. **$P < 0.01$, ***$P < 0.001$.

SUPPLEMENTAL TABLES

Subject ID	Genotype	Sex	Goal-directed or habitual	Total neurons			Co-registered neurons	% neurons coregistered		
				Early	Middle	End		Early	Middle	End
5330	D1cre	Female	Habitual	216	231	187	111	51.39%	48.05%	59.36%
5350	D1cre	Female	Habitual	204	236	250	146	71.57%	61.86%	58.40%
5399	D1cre	Female	Habitual	316	331	373	235	74.37%	71.00%	63.00%
6579	D1cre	Male	Habitual	134	171	154	55	41.04%	32.16%	35.71%
5543	A2Acre	Female	Habitual	111	137	148	74	63.06%	51.09%	47.30%
5567	A2Acre	Female	Habitual	85	98	78	30	35.29%	30.61%	38.46%
1932	A2Acre	Female	Habitual	352	329	285	113	32.10%	34.35%	39.65%
6620	A2Acre	Male	Habitual	123	133	147	80	65.04%	60.15%	54.42%
5562	A2Acre	Male	Goal-directed	38	55	49	14	36.84%	25.45%	28.57%
1786	A2Acre	Male	Goal-directed	221	234	202	169	76.47%	72.22%	83.66%
6608	A2Acre	Male	Goal-directed	308	322	321	179	58.12%	55.59%	55.76%
6619	A2Acre	Female	Goal-directed	126	168	167	64	50.79%	38.10%	38.32%

Supplemental Table 1: Neurons recorded and tracked for each subject.

		Prefeed Consumption		
		Valued Outcome mean +/- s.e.m.	Devalued Outcome mean +/- s.e.m.	Statistics
Figure 1	**D1-cre**	1.26 ± 0.23	1.20 ± 0.26	$t_3 = 0.54$ $P = 0.63$
Figure 1-1c	**Limited**	1.86 ± 0.14	2.28 ± 0.17	Training x Value: $F_{1, 14} = 0.35$, $P = 0.56$; Training: $F_{1, 14} = 12.95$, $P = 0.003$; Value: $F_{1, 14} = 3.40$, $P = 0.09$
	Overtrain	1.58 ± 0.14	1.79 ± 0.12	
Figure 3e	**WT**	0.96 ± 0.96	0.95 ± 0.19	Virus x Value: $F_{1, 14} = 0.27$, *P=0.61*; Virus: $F_{1, 14} = 0.026$, P=0.87; Value: $F_{1, 14} = 0.19$, P=0.68
	D1-cre	0.89 ± 0.14	0.94 ± 0.19	
Figure 3k	**WT**	1.50 ± 0.27	1.30 ± 0.23	Virus x Value: $F_{1, 11} = 0.92$, P=0.36; Virus: $F_{1, 11} = 0.50$, P=0.49; Value: $F_{1, 11} = 0.024$, P=0.88
	D1-cre	1.60 ± 0.25	1.70 ± 0.18	
Figure 3q	**WT**	1.36 ± 0.12	1.36 ± 0.14	Virus x Value: $F_{1, 18} = 0.032$, P=0.86; Virus: $F_{1, 18} = 0.087$, P=0.77; Value: $F_{1, 18} = 0.013$, P=0.91
	D1-cre	1.40 ± 0.20	1.43 ± 0.18	
A2A habitual vs Goal-directed	**A2A-cre habitual**	1.38 ± 0.24	1.25 ± 0.24	Group x Value: $F_{1, 6} = 0.06$, $P = 0.81$; Group: $F_{1, 6} = 0.13$, $P = 0.73$; Value: $F_{1, 6} = 0.83$, $P = 0.40$
	A2A-cre goal-directed	1.53 ± 0.07	1.30 ± 0.34	
Figure 4	**A2A-cre habitual**	1.38 ± 0.24	1.25 ± 0.24	$t_3 = 2.19$ $P = 0.12$
Figure 5-7	**A2A-cre goal-directed**	1.53 ± 0.07	1.30 ± 0.34	$t_3 = 0.59$ $P = 0.60$
Figure 6e	**WT**	1.35 ± 0.19	1.46 ± 0.22	Virus x Value: $F_{1, 16} = 1.05$, P=0.32; Virus: $F_{1, 16} = 0.03$, $P = 0.87$; Value: $F_{1, 16} = 0.0006$, P=0.98
	A2A-cre	1.50 ± 0.22	1.40 ± 0.19	
Figure 6k	**WT**	1.40 ± 0.18	1.30 ± 1.39	Virus x Value: $F_{1, 22} = 0.31$, P=0.58; Virus: $F_{1, 22} = 0.03$, P = 0.87; Value: $F_{1, 22} = 0.10$, P = 0.76
	A2A-cre	1.36 ± 0.09	0.16 ± 0.13	
Figure 6q	**WT**	1.82 ± 0.15	1.91 ± 0.14	Virus x Value: $F_{1, 28} = 0.56$, P=0.46; Virus: $F_{1, 28} = 3.5$, P=0.07; Value: $F_{1, 28} = 0.0056$, P=0.94
	A2A-cre	1.65 ± 0.14	1.53 ± 0.12	

Supplemental Table 2: Sensory-specific satiety prefeed consumption. Values reflect average amount consumed in grams ± s.e.m.

		Post-choice Consumption		
		Valued Outcome mean +/- s.e.m.	Devalued Outcome mean +/- s.e.m.	Statistics
Figure 1	**D1-cre**	0.14 ± 0.06	0.00 ± 0.02	$t_7 = 2.58$ $P = 0.04$
Figure 1-1c	**Limited**	0.32 ± 0.06	0.03 ± 0.02	Training x Value: $F_{1, 30} = 0.004$, $P = 0.95$; Training: $F_{1, 30} = 0.41$, $P = 0.53$; Value: $F_{1, 30} = 38.49$, $P < 0.001$
	Overtrain	0.29 ± 0.06	0.00 ± 0.00	
Figure 3e	**WT**	0.18 ± 0.04	0.06 ± 0.04	Group x Value: $F_{1, 30} = 0.038$, P=0.85; Group: $F_{1, 30} = 0.26$, P=0.62; Value: $F_{1, 30} = 12.22$, P=0.0015
	D1-cre	0.16 ± 0.06	0.03 ± 0.02	
Figure 3k	**WT**	0.14 ± 0.05	0.07 ± 0.03	Group x Value: $F_{1, 24} = 0.98$, P=0.33; Group: $F_{1, 24} = 0.21$, P=0.65; Value: $F_{1, 24} = 12.8$, P=0.0015
	D1-cre	0.15 ± 0.04	0.02 ± 0.02	
Figure 3q	**WT**	0.26 ± 0.04	0.05 ± 0.01	Group x Value: $F_{1, 38} = 0.047$, P=0.83; Group: $F_{1, 38} = 0.20$, P=0.66; Value: $F_{1, 38} = 57.41$, P<0.0001
	D1-cre	0.28 ± 0.05	0.06 ± 0.03	
A2A habitual vs Goal-directed	**A2A-cre habitual**	0.24 ± 0.10	0.01 ± 0.02	Group x Value: $F_{1, 14} = 0.11$, $P = 0.75$; Group: $F_{1, 14} = 0.26$, $P = 0.62$; Value: $F_{1, 14} = 12.22$, $P < 0.01$
	A2A-cre goal-directed	0.30 ± 0.11	0.02 ± 0.01	
Figure 4	**A2A-cre habitual**	0.24 ± 0.10	0.01 ± 0.02	$t_7 = 2.23$ P = 0.06
Figure 5-7	**A2A-cre goal-directed**	0.30 ± 0.11	0.02 ± 0.01	$t_7 = 2.72$ P = 0.030
Figure 6e	**WT**	0.15 ± 0.04	0.00 ± 0.01	Group x Value: $F_{1, 34} = 0.35$, P=0.56; Group: $F_{1, 34} = 0.012$, P=0.91; Value: $F_{1, 34} = 26.2$, P<0.0001
	A2A-cre	0.13 ± 0.03	0.01 ± 0.02	
Figure 6k	**WT**	0.19 ± 0.04	0.02 ± 0.04	Group x Value: $F_{1, 46} = 1.9$, P=0.18; $F_{1, 46} = 0.07$, P=0.79; Value: $F_{1, 46} = 29.2$, P<0.0001
	A2A-cre	0.16 ± 0.03	0.03 ± 0.02	
Figure 6q	**WT**	0.29 ± 0.04	0.07 ± 0.02	Group x Value: $F_{1, 58} = 0.55$, P=0.46; Group: $F_{1, 58} = 0.039$, P=0.85; Value: $F_{1, 58} = 72.62$; P<0.0001
	A2A-cre	0.31 ± 0.04	0.04 ± 0.02	

Supplemental Table 3: Average post-probe-test choice consumption. Values reflect average amount consumed in grams ± s.e.m.

Category	Item	Vendor	Catalog #	Lot #	Titer
Viruses	AAV2-hSyn-DIO-hM3D(Gq)-mCherry	Addgene	44361	v97910	2.0×10^{13} gc/mL
	AAV2-hSyn-DIO-hM4D(Gi)-mCherry	Addgene	44362	v68359	1.5×10^{13} gc/mL
	AAV9-Syn-FLEX-GCaMP7s	Addgene	50459	v107704	1.6×10^{13} gc/mL

Supplemental Table 4: Key reagents information.