RESEARCH REPORT

Spatially restricted inhibition of cholinergic interneurons in the dorsolateral striatum encourages behavioral exploration

Kenneth A. Amaya [iD] | Kyle S. Smith

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

Correspondence
Kenneth A. Amaya, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA.
Email: Kenneth.A.Amaya.GR@dartmouth.edu

Abstract

When pursuing desirable outcomes, one must make the decision between exploring possible actions to obtain those outcomes and exploiting known strategies to maximize efficiency. The dorsolateral striatum (DLS) has been extensively studied with respect to how actions can develop into habits and has also been implicated as an area involved in governing exploitative behavior. Surprisingly, prior work has shown that DLS cholinergic interneurons (ChIs) are not involved in the canonical habit formation function ascribed to the DLS but are instead modulators of behavioral flexibility after initial learning. To further probe this, we evaluated the role of DLS ChIs in behavioral exploration during a brief instrumental training experiment. Through designer receptors exclusively activated by designer drugs (DREADDs) in ChAT-Cre rats, ChIs in the DLS were inhibited during specific phases of the experiment: instrumental training, free-reward delivery, at both times, or never. Without ChI activity during instrumental training, animals biased their responding toward an "optimal" strategy while continuing to work efficiently. This effect was observed again when contingencies were removed as animals with ChIs offline during that phase, regardless of ChI inhibition previously, decreased responding more than animals with ChIs intact. These findings build upon a growing body of literature implicating ChIs in the striatum as gate-keepers of behavioral flexibility and exploration.

KEYWORDS
basal ganglia, DREADDs, habit, learning, reward

1 | INTRODUCTION

When pursuing goals, animals balance plans of action that have worked previously (exploitation) while exploring new actions for potentially greater benefit (exploration).

Behavioral exploration allows for the learning of new environmental conditions related to the occurrence of desirable events like rewards and can be adaptive when those conditions are unstable or change. Conversely, while some exploration is normally expected to persist as the possibility of

better reward may exist, the emergence of more exploitative behavior can be advantageous when learned behaviors can be repeated in stable conditions for continued goal achievement, allowing them to be performed in a semi-automatic manner (Cohen et al., 2007; Dickinson & Weiskrantz, 1985; Dolan & Dayan, 2013; Graybiel, 2008). These exploitative behaviors are thought to help mitigate the demand placed upon the neural circuitry governing decision-making, allowing for a rapid and cognitively undemanding response to occur. However, continuing to perform exploitative behaviors can be maladaptive if their utility changes (e.g., when environmental conditions change that favor a different behavioral strategy to more efficiently get a goal). In severe cases, a failure to alter a learned behavior when it is no longer useful can contribute to the pathology of behavioral compulsivity, such as occurs in obsessive-compulsive disorder, Tourette's syndrome, and addiction (Everitt & Robbins, 2005; Gillan et al., 2011; Voon et al., 2015).

A major hub in the brain for performing actions in a consistent manner is the dorsolateral aspect of the striatum (DLS; primate putamen homologue). For example, the DLS is critical for phenomena like positively and negatively reinforced skills, response sequence learning, outcome-insensitive habits, and even non-reinforced movement patterns like grooming behavior; one way of framing this globally is to consider the DLS as important for encouraging the honing and exploitation of behavioral repertoires that "work" for the organism (Amaya & Smith, 2018; Barnes et al., 2005; Graybiel, 2008; Malvaez & Wassum, 2018; Yin & Knowlton, 2006). Physiologically, while the striatum is predominantly comprised of GABAergic projection neurons known as medium spiny neurons (MSNs) and a smaller number of GABAergic interneurons, a small population of tonically active cholinergic interneurons (ChIs) have been reported to play a critical role in striatal function. They modulate dopaminergic inputs and encode salient reward-seeking events in a manner regarded as important for striatum-based learning (Aosaki et al., 1994; Cachope et al., 2012; Kawaguchi et al., 1995; Threlfell et al., 2012). Activation of ChIs in the DLS is sufficient to augment lever pressing on an alternative habit in a habit-substitution (reversal) task, where a learned habitual lever response must be abandoned for responding on a new now-rewarded lever (Aoki et al., 2018). However, in that study there was no observed effect of ChI activation on reducing the initially learned behavior (the "old habit") or was there an augmentation of responding for behaviors that did not involve a task–reward switch. Interestingly, the study also reported that neurochemical ablation of DLS ChIs had no effect on either habit development or substitution (Aoki et al., 2018). The latter finding is curious as it suggests that ChIs in the DLS are not necessary for making habits, unlike DLS as a whole. This reflects that (a) this is a limit on what these ChI neurons contribute to action performance in the DLS, and (b) that ChIs in the DLS promote behavioral flexibility. Thus, there is a paucity of knowledge on how ChIs could be involved in transitions between behavioral exploration and exploitation. Better understanding DLS ChI functionality in this regard would also shed light on the seemingly counterintuitive findings of previous work where ChI activity was not tied to habit formation but to behavioral flexibility. In other words, does DLS ChI activity promote the continued use of acquired behaviors (habit-like) or promote behavioral flexibility?

To examine this gap in knowledge, we took a rather simple approach: ChIs were transduced with inhibitory DREADD receptors in a small area of the DLS using ChAT-Cre rats (Witten et al., 2011) and ChI activity was inhibited at various points during behavioral training or testing as task rules were changed. As the DLS is involved in motor learning developing into skilled behavior (e.g., Kupferschmidt et al., 2017), we chose a task involving a brief training period whereby rats learned to press a lever for reward (FR1), then surprisingly shifted the task to require three presses (FR3) for reward over three sessions. Later, we changed the task environment again with a shift to a random time (RT) 60-s delivery of reward which required no lever presses. This design allowed us to pit behavioral exploitation against exploration because continuing to perform as before would still yield reward delivery but exploring and changing behavior based on the new contingencies can result in a more optimal performance (i.e., abandoning an FR1 one-press strategy in favor of a more efficient FR3 three-press strategy at the initial shift, and abandoning pressing entirely during RT-60). If ChI inhibition results in increased exploration of alternative strategies during FR3 training, then, under normal conditions, DLS ChIs may be promoting exploitative behaviors. If, on the other hand, inhibition results in animals fixating on previously learned strategies during FR3 training, it could be concluded that these neurons normally promote explorative behaviors. After training, RT60 testing allowed us to assess the pliability of the learned behavioral responses. To best capture the effect of ChI inhibition, we looked at two main measures of behavior, pressing rates and types of pressing bouts, as well as reward procurement behavior (magazine entry). This survey can illuminate whether animals continue pressing at learned rates despite a task shift, and also whether animals that continue pressing do so using an already-acquired strategy (e.g., one-press bouts during the shift to FR3 training) or form a new strategy (e.g., three-press bouts during the FR3 training). Importantly, there is no impetus to change behavior other than for the reason to shift responding to be the most energetically optimal; costs associated with behavioral perseveration (i.e., continuing to use the initially learned one-press strategy) include costs to energy and efficiency. In other words, reward will come if animals always keep using their one-press strategy but shifting this strategy can lead to more optimal

behavior. The findings we report are that animals normally continue the initially learned, exploitative strategy of pressing in one-press bouts during FR3 training; they kept doing what they had previously learned to do. However, with inhibition of DLS ChIs, animals exhibit a shift toward behavioral exploration that results in the use of more optimal response strategies. When contingencies were removed during an RT60 test, with animals switched off or onto a ChI inhibition status, or maintained on their prior designation, animals with inhibition of DLS ChIs again showed more flexible responding by reducing lever press behavior to a greater degree.

2 | MATERIALS AND METHODS

2.1 | Animals

Subjects were male heterozygous ChAT-Cre-positive Long-Evans rats (total $N = 50$, Witten et al., 2011), weighing 250-400 g upon surgery. Rats were housed in a climate-controlled colony room that was illuminated from 7:00 a.m. to 7:00 p.m. Rats were initially pair housed but were then housed individually following surgery for the entirety of the experiment. Rats were given ad libitum access to food and water before and continuing 2 weeks after surgery. Rats were then placed on a food restriction schedule in which they were maintained at 85% of their ad libitum weights for the duration of the experiment. Experiments were carried out in accordance with the National Institute of Health's Guide for the Care and Use of Laboratory Animals and protocols were approved by the Dartmouth College Institutional Animal Care and Use Committee.

2.2 | Surgical procedures

Surgery was performed under aseptic conditions with isoflurane anesthesia. Intracranial viral infusions were made with a 10-μl syringe equipped with a 33-gauge beveled needle (World Precision Instruments, Inc., Sarasota, FL) and a Quintessential Stereotaxic Injector (Stoelting Inc., Kiel, WI). Infusions were made into the DLS bilaterally at 0.5 mm anterior from bregma, 3.8 mm from the midline, and 4.3 mm ventral from the skull surface. Each infusion was 0.5 μl in volume and was made at a rate of 0.15 μl/min. Following infusion, the syringe was left in place for 3 min to allow for diffusion. Rats received infusions of the Cre-dependent inhibitory hM4D(Gi) DREADD ($N = 40$; AAV8-hSyn-DIO-hM4D(Gi)-mCherry; Addgene) or of a virus that contained Cre-dependent DNA for mCherry but not the hM4D(Gi) receptor ($N = 10$; AAV8-hSyn-DIO-mCherry; Addgene). Expression of the transgenes was allowed to take place over the course of 3 weeks before the beginning of behavioral training. Final group assignments are shown in Figure 1d.

2.3 | Apparatus

Behavioral procedures were carried out in 8 identical standard conditioning chambers (24 × 30.5 × 29 cm; Med Associates, Georgia, VT) enclosed in sound-attenuating chambers (62 × 56 × 56 cm) outfitted with an exhaust fan to provide airflow and background noise (~68 dB). The conditioning chambers consisted of aluminum front and back walls, clear acrylic sides and top, and grid floors. Each chamber was outfitted with a food cup recessed in the center of the front wall. Retractable levers were positioned to the left and right of the food cup. These levers were 4.8 cm long and positioned 6.2 cm above the grid floor. The levers protruded 1.9 cm when extended. The chambers were illuminated by a house light mounted 15 cm above the grid floor on the back wall of the chamber. Task events were controlled by computer equipment located in an adjacent room.

2.4 | Behavioral procedures

Training began 3 weeks following surgery with an initial 30-min magazine acclimation session where grain pellets were delivered at an average rate of one pellet every 30 s. Delivery of the pellet was accompanied by an audible click made by the food hopper. The following day, rats were advanced to a fixed-ratio-1 (FR1) schedule of delivery, where two levers were continuously presented to the animals, one active and the other inactive, and one active lever press earned one grain pellet reward (BioServ, Product #F0165, 45 mg dustless precision pellets: Protein 21.3%, Fat 3.8%, Carbohydrate 54.0%). Before being able to earn the next reward, animals were required to enter the magazine area. Lever designations remained constant throughout training and testing and were counterbalanced across animals. Successful completion of the FR1 training session was defined as earning 20 reinforcers within a 30-min session. There was no administration of CNO or Vehicle at this point in the experiment and rats were not advanced in training until completion of this FR1 task.

After successful completion of one FR1 training session, animals began FR3 training where three correct lever presses were required to earn a single reinforcer. FR3 training was chosen as it required learning of a new task requirement while having a defined response—three presses were consistently required to earn outcomes. Thirty minutes prior to each FR3 training session, animals were given an injection of either Vehicle (sterile water, 1 ml/kg, i.p.) or CNO (1 mg /mL / kg in sterile water, i.p.; National Institute of Mental Health Chemical Synthesis and Drug Supply Program). During the FR3 training phase, the 10 animals that received infusions of virus lacking the DNA for the hM4D(Gi) all received CNO injections and are designated as "CNO control." The 40 animals that received viral infusions that included the DNA for
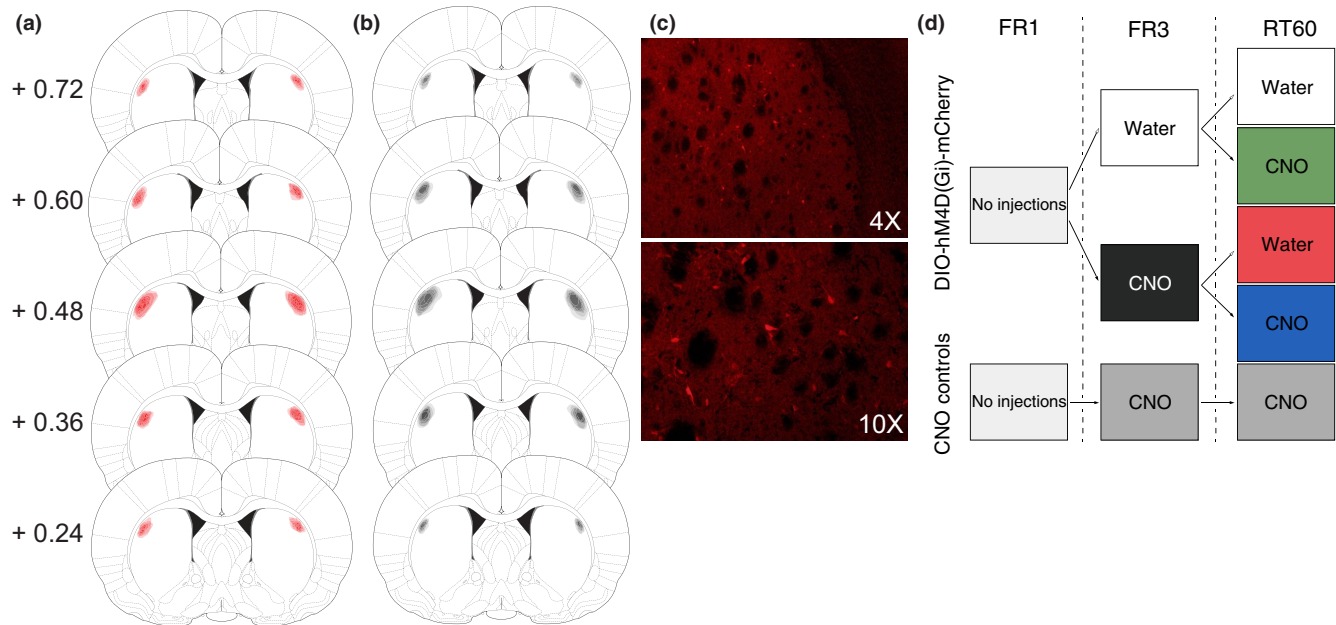
FIGURE 1 Histological results. (a) Schematic representation of DIO-hM4D(Gi)-mCherry expression in the dorsolateral striatum (DLS) of rats from Group Vehicle and Group CNO ($n = 40$). (b) Schematic representation of DIO-mCherry expression in the DLS of rats from Group CNO control ($n = 10$). Numbers indicate distance from bregma in mm. Coronal slices adapted from Paxinos and Watson (2009). (c) Representative image showing DIO-hM4D(Gi)-mCherry in the DLS. (d) Group assignments by experiment phase, with CNO administration noted. During FR1 training, injections were not given. Upon FR3 training, animals were split into three groups (Vehicle, CNO, and CNO control). Group Vehicle and Group CNO received hM4D(Gi) DREADDs tagged with mCherry while Group CNO control received infusions mCherry alone during surgery. At RT60, groups were split once again, to control for CNO-exposure history. CNO, clozapine N-oxide

the hM4D(Gi) receptor, Group Gi, were split into two groups of 20 animals each, one that received injections of water and the other that received injections of CNO ("Vehicle" and "CNO," respectively). A group summary is presented in Figure 1d.

After three consecutive FR3 training sessions were completed, animals advanced to RT60 testing. Levers were still presented to animals, but no action was required for reinforcers to be delivered and lever pressing had no effect on reinforcer delivery. Instead, reinforcers were simply delivered to food cups every 60 s, on average. The CNO control group continued to receive injections of CNO prior to each test session. The two groups that received either Vehicle or CNO injections during training were split into four groups during RT60 testing. Groups can be summarized as such: 10 animals received water injections for the entirety of the experiment (Veh-Veh), 10 animals received CNO injections for the entirety of the experiment (CNO-CNO), 10 animals received CNO injections during RT60 test only (Veh-CNO), and 10 animals received CNO injections during FR3 training only (CNO-Veh). The crossover design was used to control for prior CNO history. A summary of group identities is depicted in Figure 3a. RT60 testing concluded after three sessions. During this RT60 test, the optimal action would be to cease responding and simply wait for reinforcers to be delivered. This task was chosen as a measure of animals' sensitivity to

a change in the causal consequences of their actions. While previous studies (e.g., Balleine & Dickinson, 1998; Corbit & Balleine, 2000) used a contingency degradation procedure whereby rewards continue to be earned as was learned but additional rewards are noncontingently delivered, administering a full contingency removal here made for a more objectively clear optimal action strategy of cessation of lever pressing altogether. Furthermore, this free reinforcement procedure has been used previously following instrumental training, allows animals to act freely, and produces response decrements that are similar to extinction while reducing spontaneous recovery (Balleine & Killcross, 1994; Boakes & Halliday, 1975). However, we recognize that our strategy produces difficulties in linking behavioral responses to RT60 here to those resulting from the contingency degradation procedure in the literature.

2.5 | Data analysis

Lever deflections and pellet magazine entries were recorded through MedPC. Lever press rates are calculated in two ways: per reinforcer delivered and per minute. Presses per reinforcer show how efficient animals were in working while presses per minute informs how quickly an animal is working for reinforcers.

Additionally, a key measure in this study was pressing strategy or the type of lever press bouts used by animals. Bouts are defined as the number of lever presses made between magazine entries. While animals could use a variety of bout types, the most efficient bout strategy is implicitly defined by the task parameters. During the FR1 phase of training, a press-check method (one-press bout) would be optimal. However, during FR3 training, if this same strategy was employed, only one of every three checks would be successful. Therefore, the most optimal strategy would be a press–press–press–check strategy (three-press bout), where less energy is spent going between the lever and food cup.

All figures were constructed using R ("ggplot2") and stylized using Adobe Illustrator. All statistical tests were carried out using R, as previously described (R Core Team 2016; see Amaya et al., 2020; Smedley & Smith, 2018). Individual linear mixed models (R; "lme4") were used to analyze effects of dependent variable responding (ex. lever presses per reinforcer or minute (ppr, ppm)) by fixed effects of experimental group and session while accounting for random effects of differences in individual starting values for the dependent variable in Session 1. Initially, contrasts to compare the dependent variables of the two control groups were used. Once no difference was observed between the two control groups, zero sum contrasts were made for categorical variables (i.e., group) comparing the experimental group to both control groups together. Linear mixed models are fit by maximum likelihood and t-tests use Satterthwaite approximations of degrees of freedom (R; "lmerMod"). Linear mixed models were analyzed with package lme4 from CRAN (Bates et al., 2015). The reported statistics include parameter estimates (β values), 95% confidence intervals, and p-values (R; "lmerTest"). Linear mixed models were used because they consider aspects of the data structure that repeated measures ANOVA cannot and allows for safer generalization to larger populations of animals. For instance, mixed models expect a greater likelihood that repeated measures taken from one animal over time tend to be more similar than across animals and can account for these trends (Boisgontier & Cheval, 2016; Smedley & Smith, 2018).

For the analysis of bout data during training, we created generalized linear mixed models using family = binomial to analyze effects of dependent variable responding (bout-type probability) by fixed effects of session, an experimental (CNO) versus control (Vehicle + CNO control) contrast, and an interaction between these, along with random effects of individual rat starting points. For generalized linear mixed models assessing bout probabilities during training, session was re-centered such that Session 3 of training, the final session, is the comparison point for the main effect of Group. The generalized linear mixed model is similar to the linear mixed model as it factors in start points for each individual rat, but the parameter estimates given by this type of model are best interpreted as odds ratios, where the numerical value is the ratio of two probabilities: the probability of an event happening in one condition compared to the probability of the same event happening in another condition. As the parameter estimates of generalized linear mixed models are multiplicative, an odds ratio of 1.0 indicates no effect of treatment between groups while values greater than 1.0 show an increased probability of an event happening in a group that received treatment and a value less than 1.0 reflecting lower odds.

For the analysis of lever press data during RT60 testing, a linear mixed model was used to analyze effects of dependent variable responding (lever presses per reinforcer or minute (PPR)) by fixed effects of experimental group and session while accounting for random effects of differences in individual starting values for the dependent variable in Session 1. Reported statistics include parameter estimates, confidence intervals, and p-values. To analyze changes in responding during RT60, particularly Day 2 compared to Day 1 baseline responding (PPR or magazine entries made), the percent change in responding was calculated. After using a Shapiro–Wilk normality test to determine that the data were not normally distributed, effects of final Group assignment on percent responding were calculated using a Wilcoxon rank-sum test.

For the analysis of bout data between FR3 training and RT60 testing, generalized linear mixed models using family = binomial to analyze effects of bout-type probability by fixed effects of Session, a CNO-Switch contrast (was CNO or vehicle delivery consistent for both phases of the experiment or did animals switch from vehicle to CNO or vice versa between experimental phases), and the interaction between Session and Switch, along with random effects of individual rat starting points were created. Models were created to predict one-press bout probability as well as three-press bout probability. Meanwhile, for the analysis of within-bout-type change between testing Day 1 and Day 2, an approach similar to PPR and magazine entry changes was employed. Data normality was tested using a Shapiro–Wilk normality test, then Group effects were assessed using a Wilcoxon rank-sum test.

2.6 | Histological procedures

Following behavior, rats were anaesthetized with sodium pentobarbital (100 mg/kg) and perfused intracardially with 0.9% saline, followed by 10% formalin. Brains were removed and stored in 20% sucrose, and then sectioned at 60 μm. Sections were then mounted on microscope slides and cover slipped with a DAPI-containing hardset mounting medium (Vectashield; Vector Laboratories, Burlingame, CA, USA) for verification of hM4D(Gi)-mCherry or mCherry expression in the DLS using a fluorescent microscope (Olympus, Center Valley, PA, USA). To assess for bilateral expression of hM4D(Gi)-mCherry, the extent of the areas of

expression was mapped onto structural boundaries per Paxinos & Watson, 2009 (Figure 1a,b).

3 | RESULTS

3.1 | Histological results

Figure 1a,b shows a schematic representation of hM4D(Gi)–mCherry (1A, Group Gi) and mCherry (1B, Group mCherry) expression. The spread of virus was confined to the most dorsal and lateral aspects of the striatum, and representative images are presented in Figure 1c. Spread of DREADDs expression along the injector tract above DLS was not observed. No animals were excluded from the experiment for histological inaccuracies.

3.2 | DLS ChI inhibition during fixed-ratio training promotes optimality while preserving action efficiency

Prior to fixed-ratio-3 training (FR3), animals were required to complete magazine training, where pellets were freely delivered (no levers present) and a fixed-ratio-1 (FR1) training session. Animals from each group, Group Gi ($n = 3$) and Group mCherry ($n = 2$), were excluded from the experiment for failing to complete FR1 training after multiple sessions (≥ 14 sessions) of failed completion. No CNO was administered during these initial training phases; animals began receiving injections of CNO or vehicle during FR3 training. Therefore, neither inhibition of cholinergic interneurons nor administration of CNO could be an explanation for failure to advance to FR3 training.

The mean number of lever presses per reward (PPR) during FR3 training is presented in Figure 2a. Each group, over the 3 days of FR3 training, pressed approximately three times per reward delivered. A linear mixed model with fixed effects of Session, Group, and an interaction between these predictors, along with random effects of rat starting points revealed no significant effects of any of the predictors of the presses per reward data as all groups across the 3 days of training pressed the correct lever at similar rates. In other words, ChI manipulation did not impact the number of times the lever was pressed with respect to the number of rewards that were delivered; all animals were pressing very close to the minimum number of presses required to earn the maximum number of outcomes.

The mean number of lever presses per minute (PPM) during FR3 training is presented in Figure 2b. Presses per minute as a measure of performance consider time rather than the number of outcomes delivered. Animals were capable of pressing the lever the minimum number of times

to earn reinforcers, one potential measure of efficiency, as shown in Figure 2a. However, another measure of efficiency could include how quickly an action is performed.

To compare group press rates with respect to time, a linear mixed model used PPM as the dependent variable by fixed effects of Session, Group, and the interaction between Session and Group, with random intercepts for individual start points included. There was no main effect of Group (estimate: -3.07, CI: -0.38 to 6.53, $p = 0.084$), indicating that the experimental animals (CNO) did not differ from control animals (CNO controls + Vehicle) in their overall mean press rates. However, there was a significant effect of Session (estimate: 3.33, CI: 2.72–3.93, $p < 0.001$), showing that all animals increased their lever press rates as training progressed. There was also a significant interaction between Group and Session (estimate: 1.08, CI: 0.28–1.88, $p = 0.010$), indicating that the experimental group increased PPM over sessions to a greater degree than controls did, which suggests that ChI inhibition increased the speed at which experimental animals earned reinforcers over training sessions.

The mean number of magazine entries, presented as entries made per reward delivered, is presented in Figure 2c. A linear mixed model of these data revealed no significant Group effect or a significant interaction between Group and Session. However, a significant effect of Session was observed (estimate: -1.88, CI: -2.26 to $-(1.49)$, $p < 0.001$), suggesting that all animals slightly decreased the rate at which they entered the magazine area per each reward delivered.

The mean probabilities of employing one-press (white), two-press (orange), and three-press (purple) bouts during each FR3 training session, with group identities represented by shape, are presented in Figure 2d–f, respectively. For each bout type, separate generalized linear mixed models were created to analyze the effects of ChI inhibition (Group), training Session, and the interaction between Group and Session on the likelihood of employing each strategy, while accounting for individual rat starting probabilities. First, the one-press bout model revealed significant effects of Group (odds ratio (OR): 0.51, CI: 0.38–0.69, $p < 0.001$), Session (OR: 0.81, CI: 0.75–0.87, $p < 0.001$), and a significant interaction between Group and Session (OR: 0.72, CI: 0.65–0.80, $p < 0.001$). Next, the two-press bout model revealed significant effects of Group (OR: 1.31, CI: 1.05–1.64, $p = 0.017$), Session (OR: 1.13, CI: 1.06–1.20, $p < 0.001$), and a significant interaction between Group and Session (OR: 1.11, CI: 1.02–1.20, $p = 0.014$). The three-press bout model showed significant effects of Group (OR: 2.23, CI: 1.49–3.33, $p < 0.001$), Session (OR: 1.50, CI: 1.33–1.68, $p < 0.001$), and a significant interaction between Group and Session (OR: 1.45, CI: 1.25–1.68, $p < 0.001$). Additionally, four-press and greater bout probabilities were also considered (data not graphed) and were found to be increased. This model revealed significant effects of Group (OR: 3.83, CI: 1.63–9.00, $p = 0.002$), Session (OR: 1.84, CI:
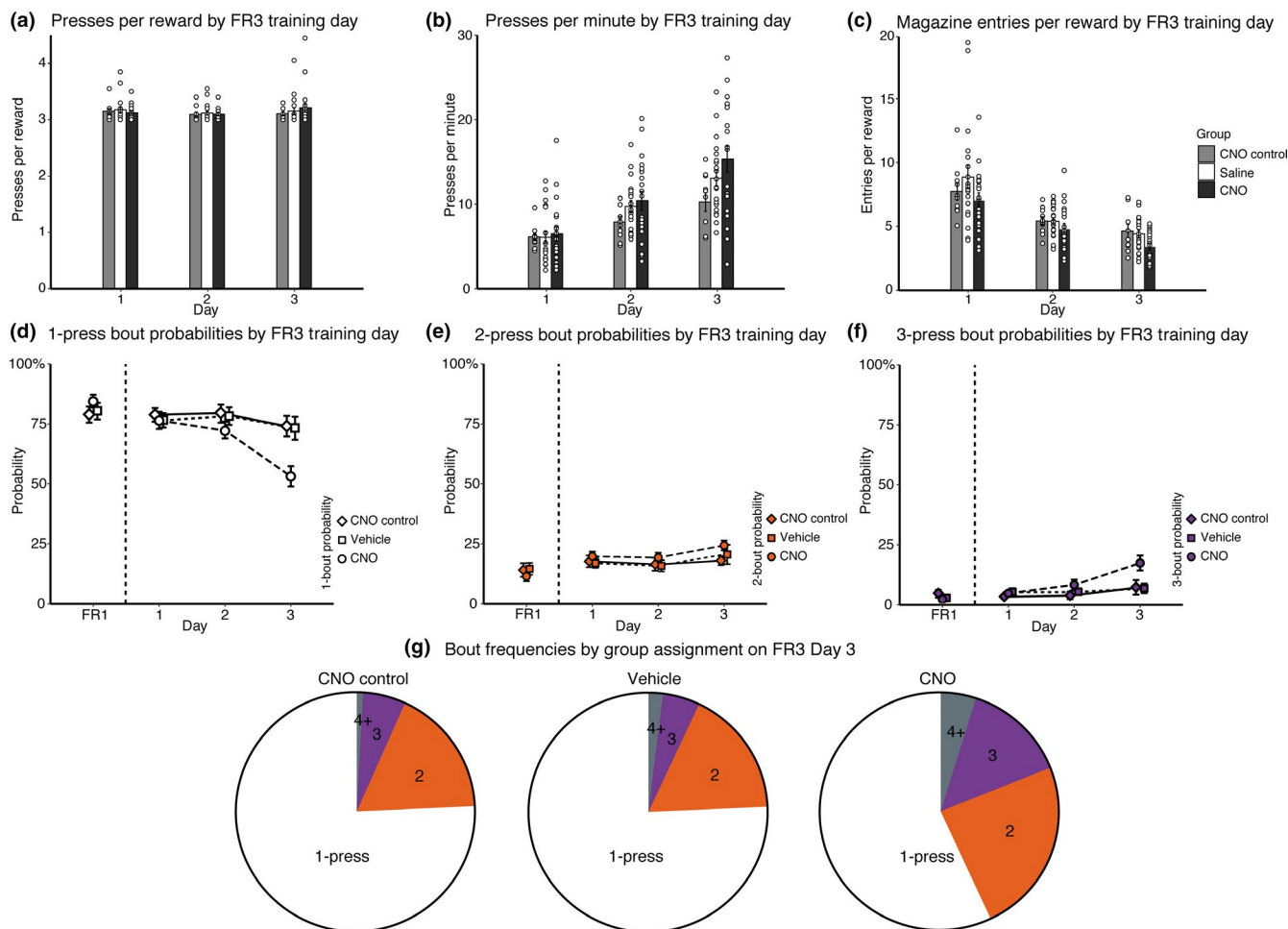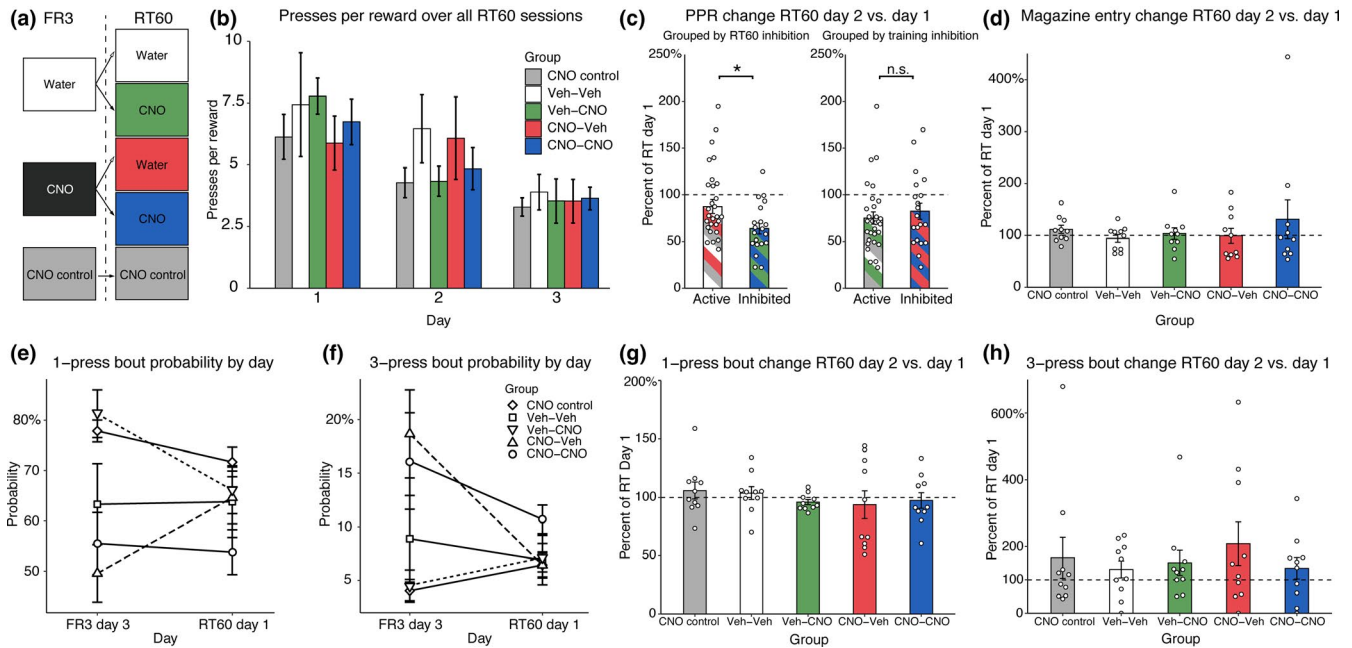
FIGURE 2 Behavioral training results. (a) Presses per reinforcer. All three groups of rats, Group CNO, Group Vehicle, and Group CNO control, pressed the active lever equivalently as measured by the number of presses made per reinforcer over FR3 training days. (b) Presses per minute. The average Group CNO press rate as a measure of presses made per minute increased compared to the two control groups over FR3 training days. (c) Magazine entries per reward. The overall average entries made per reward decreased over training days. (d–f) Bout-type probabilities by training day. Three types of bouts are presented, one-bout (d), two-bout (e), and three-bout (f) in white, orange, and purple, respectively. Shapes indicate group assignment (diamond = CNO control, square = Vehicle, circle = CNO). There is a drop in the average one-bout probability, and a concomitant rise in two- and three-bout probabilities, in Group CNO but not in control groups. FR1 bout probabilities, before animals were exposed to CNO, are presented to the left of the dashed line to show stability in tendencies between FR1 training and the 1st day of FR3 training. Error bars represent ± *SEM*. (g) Bout frequencies by group assignment on FR3 Day 3. Group Vehicle and Group CNO show similar one-press (white), two-press (orange), three-press (purple), and four-press and greater (gray) bout frequency tendencies on the final day of fixed-ratio training. Group CNO used the one-press bout strategy less than control animals while increasing the use of the other bout strategies

1.27–2.65, $p = 0.001$), and a significant interaction between Group and Session (OR: 2.72, CI: 1.55–4.76, $p < 0.001$).

Taken together, these data show that the ChI-inhibition group decreased the likelihood of employing the one-press bout strategy overall and as training progressed and increased the likelihood of employing other bout strategies. Importantly, while there were significant effects seen in the two-press bout model, indicating that the experimental animals were more likely to employ this strategy over time than their control counterparts, the effects in that model are modest in comparison to those seen in the three-press and four-plus-press bout models, meaning that the experimental

animals disproportionately increased their propensity to use the those strategies over the two-press bout strategy over training sessions. Despite that fact, overall, the two-press bout strategy was still used more frequently on the final training session. A breakdown of these strategy frequencies on the final day of FR3 training is presented in Figure 2g with the mean bout size of the experimental CNO group (1.67) being greater than the mean bout sizes of either control group (CNO control: 1.32; vehicle control: 1.30). We note that no animal fully adapted to the more optimal three-press bout strategy, which we suspect was due to the short period of time provided (three sessions) in which to change their behavior.

FIGURE 3 Behavioral testing results. (a) Schematic representing how animals with hM4D(Gi) were split into groups from FR3 training (two groups) to RT60 testing (4 groups). Splitting animals as such allowed to control for cholinergic interneuron inhibition history when assessing behavior in test. CNO control animals received a virus that lacked the DNA for the hM4D(Gi) DREADD but were given CNO injections during each phase of the experiment. (b) RT60 presses per reward (PPR). The average press rate of each group as a measure of PPR over all RT60 test sessions. (c) Percent change in PPR behavior among groups. Press behavior on Day 2 of RT60 testing was compared to the press behavior of Day 1 and is presented as a percent change. Animals with cholinergic interneurons off-line at the time of test (left) decreased lever press behavior more than controls, regardless of CNO-exposure history. This effect was not present if animals were grouped by ChI inhibition during training (right). Bars are colored to indicate which groups were pooled together for analysis. (d) Percent change in magazine entries made per minute across all groups. Similar to the results presented in c, magazine entry rates on Day 2 of test were compared to magazine entry rates on Day 1 and are presented as a percent change. There are no observed effects of group on magazine entry behavior, indicating that entry behavior remained steady across these two test sessions. (e) One-press bout probability changes between the final training day and the first test day by RT60 group assignment. (f) Three-press bout probability changes between the final training day and the first test day by RT60 group assignment. (g) Changes in one-press bout probabilities during the second RT60 test session as a percentage of the first RT60 test session, by group assignment. (h) Changes in three-press bout probabilities during the second RT60 test session as a percentage of the first RT60 test session, by group assignment. All error bars represent $\pm$ *SEM*

3.3 | Inhibition of DLS ChI activity during RT60 training promotes exploration (relation to ChI inhibition history)

The objectively logical strategy upon removal of action–outcome contingencies during RT60 training would thus be for animals to cease pressing and merely wait for reinforcers to be delivered, as no action is required in this phase of the experiment. Behavioral flexibility leading toward an optimal strategy would thus be seen as lever press reduction, as is seen with contemporary "contingency degradation" procedures (see above). In such studies, however, it is often unclear whether animals continue performing as they had but at a reduced rate (i.e., action bout type as defined here) or additionally change their performance structure as a potentially orthogonal measure of increased flexibility. Thus, here, analyses of RT60 data are centered on both types of measures: lever press rates with respect to rewards delivered, bout-types employed, and how each of these changed over sessions. A

review of how groups were split from training to testing is presented in Figure 3a. In Figure 3b, a summary of presses per reinforcer by group over three test days is presented. A linear mixed model using presses per reinforcer as the dependent variable by fixed effects of CNO assignment, and logarithmic Session, and the interaction between logSession and CNO assignment with random intercepts for individual rat was created. Logarithmic Session (logSession) was chosen as the data appear to asymptote, a key characteristic of an exponential decay curve. This model yielded no significant effects of CNO assignment or any significant interactions between CNO assignment and logSession. There was a significant effect of logSession (estimate: -8.49, CI: -13.59 to (-3.39), $p = 0.001$), indicating that lever presses per reinforcer decayed over test sessions. Complete statistics for this model can be found in Table 1.

The data presented in Figure 3b appear variable between groups before lever pressing tapers off and asymptotes by the end of testing. Therefore, focusing on only the first two test

TABLE 1 Complete statistics from a linear mixed model using presses per reinforcer as the dependent variable by fixed effects of CNO assignment, logarithmic Session, and the interaction between logSession and CNO assignment with random intercepts for individual rat

	PPR		
Predictors	Estimates	CI	p
(Intercept)	19.70	10.48–28.92	<0.001
CNO1	4.69	−8.70 to 18.09	0.494
CNO2	8.36	−4.38 to 21.10	0.201
CNO3	−2.49	−16.32 to 11.34	0.724
CNO4	1.86	−10.63 to 14.34	0.771
logDay	−8.49	−13.59 to −3.39	0.001
CNO1:logDay	−1.87	−9.28 to 5.54	0.622
CNO2:logDay	−4.33	−11.37 to 2.72	0.231
CNO3:logDay	1.74	−5.91 to 9.38	0.657
CNO4:logDay	−0.76	−7.66 to 6.15	0.831
Random effects			
σ^2		3.84	
$\tau_{00\ \text{Rat}}$		5.04	
ICC		0.57	
N_{Rat}		50	
Observations		150	
Marginal R^2/Conditional R^2		0.187/0.648	

sessions, the percentage of lever pressing during Session 2 as compared to Session 1 is presented in Figure 3c. This figure suggests that animals with ChI inhibition at the time of test, regardless of previous CNO exposure during training, act similarly when action–outcome contingencies are removed. To test this, animals were grouped together based on ChI inhibition during test into control (CNO control + Vehicle) and experimental (CNO) groups. A Shapiro–Wilk normality test reveals that these percent data reported are not normally distributed ($W = 0.939$, $p = 0.013$). Therefore, to analyze group effects, a Wilcoxon rank-sum test revealed a significant effect of group on percentage of baseline responding ($W = 202$, $p = 0.035$). Importantly, grouping the animals by their previous CNO exposure was not a significant predictor of percentage of baseline responding ($W = 342$, $p = 0.415$). Additionally, CNO delivery alone does not explain the effects because grouping animals by CNO administration, regardless of their DREADDs expression, was not significant either ($W = 329$, $p = 0.576$). Figure 3d compares magazine entries per minute on RT60 test day 2 relative to test day 1, similarly to how PPR data were analyzed. There was no group effect on magazine entries as rewards were still delivered during test ($W = 269$, $p = 0.549$).

Figure 3e shows how one-press bout mean probabilities changed from the final FR3 training session to the first RT60 test session, while Figure 3f shows how three-press bout probabilities changed in this way as well. Both figures show data split by final group assignment. One-press bout probabilities increased, decreased, or remained constant depending on whether or not ChI inhibition changed between the training and testing phases of the experiment. That is, if ChIs were inhibited only during test, then the probability of a one-press bout decreased on the first day of RT60 testing. Similarly, if ChIs were inhibited only during training, but their activity was undisrupted during test, the probability of employing a one-press bout strategy was increased. However, if ChI activity remained constant through both phases of the experiment, no change in one-press bout probability was observed. To capture this in the generalized linear mixed model, the factor "Switch" was created to represent these changes in CNO administration. Consider switches to be either up-switches (CNO–Veh animals) or down-switches (Veh–CNO animals) determined by whether ChIs are online or offline during test relative to training. This factor along with Session and the interaction between Switch and Session was used to examine potential effects on bout probability. Here, there were significant main effects of switching as compared to not switching ChI inhibition, be it inhibiting for the first time during test (odds ratio (OR): 0.20; CI: 0.09–0.45; $p < 0.001$) or not inhibiting ChIs during test (OR: 6.13; CI: 2.89–12.98; $p < 0.001$). There was no significant Session effect (OR: 0.86; CI: 0.74–1.01; $p = 0.07$). However, there were significant interactions between Switch and Session, indicating that the up-switch (OR: 2.34; CI: 1.66–3.29; $p < 0.001$) and down-switch groups (OR: 0.44; CI: 0.32–0.60; $p < 0.001$) changed the propensity with which the one-press bout strategy was employed differently over sessions than did the no-switch group.

Similar findings were observed when considering three-press bout probability changes between training and testing experimental phases. Using the same Switch factor, Session, and the interaction between Switch and Session, a generalized linear mixed model was created to analyze effects on three-press bout probability. The model revealed significant main effects of up-switch (OR: 5.81; CI: 2.15–15.72; $p = 0.012$) and down-switch (OR: 0.24; CI: 0.08–0.73; $p = 0.001$) groups compared to the no-switch group but no significant main effect of Session (OR: 0.88; CI: 0.68–1.14; $p = 0.34$). However, there were significant interactions between Switch and Session, indicating that the up-switch (OR: 0.42; CI: 0.25–0.70; $p = 0.001$) and down-switch groups (OR: 1.93; CI: 1.11–3.36; $p = 0.019$) changed the propensity with which the three-press bout strategy was employed differently over sessions than did the no-switch group. In other words, as ChIs activity changed between experimental phases, the likelihood of using particular bout strategy changed in "Switch" groups.

The bout-type effects described above are selective in that there are no such changes observed between RT60 test sessions, only between training and testing when ChI

activity switches from being online-to-offline or vice versa. Figure 3g,h illustrate this for one-press and three-press bout probabilities, respectively. These data are presented in a manner similar to press and magazine entry data (Figure 3c,d) for the ease of visualization and are analyzed in the same manner as well. There was no effect of switch on the change in one-press bout probability ($W = 238$, $p = 0.23$) or three-press bout probability ($W = 325$, $p = 0.63$) between the first two RT60 sessions. Overall, these bout probability results show that changing CNO administration (ChI inhibition) between training and testing had an effect on the strategies employed by animals while lever pressing persisted on the first RT60 test session. Generally, this suggests that the presence of a relationship between an action and its outcome had some bearing on whether cholinergic signaling altered the strategies being used.

4 | DISCUSSION

In environments where rewards occur contingent on certain behaviors, there is often a tradeoff between exploring new options of behavior (for potentially greater reward gain per level of energy expenditure) and behaviorally exploiting what has been learned (for a relatively easy continued use of what has worked in the past). In the brain, the DLS, including its projection neurons (MSNs) and GABAeric interneurons, is thought to be a critical component of behavioral exploitation in forms that can range from maintaining action sequences, expressing habits, and honing/executing complex motor skills (Amaya & Smith, 2018; Balleine, 2019; Barnes et al., 2005; Dudman & Krakauer, 2016; Graybiel, 2008; Kalueff et al., 2016; Klaus et al., 2019; Kubota et al., 2009; Malvaez & Wassum, 2018; Packard & Knowlton, 2002; Smith & Graybiel, 2013; Yin & Knowlton, 2006; Yin et al., 2004; Yttri & Dudman, 2018). Cholinergic interneurons (ChIs) are another potentially critical population of neurons for this DLS-related function. While relatively sparse in the striatum, ChIs have been shown to have a profound impact on striatal processes. Specifically, ChIs have been studied in the DMS and DLS with respect to acquiring learned action–outcome associations and monitoring habit-like action plans, respectively. ChIs in both the DMS and DLS were implicated in behavioral flexibility, with ChI ablation in DMS hampering flexibility and ChI stimulation in DLS promoting adoption of a substituted response plan for reward (Aoki et al., 2018; Bradfield et al., 2013). However, the role of DMS ChIs pertaining to flexibility is still debated with a report showing increased flexibility following ablation (Okada et al., 2014). Additional results from Aoki et al. (2018) indicated that chemogenetic stimulation of DLS ChIs did not accelerate initial habit formation.

The results we report here extend these findings to show that inhibiting DLS ChIs promotes behavioral exploration leading to more energetically optimal strategies of reward-seeking behavior. By extension, normally active ChI activity in the DLS would promote behavioral exploitation and acquisition of new action sequence strategies when the environment changes. Specifically, we show that animals develop a one-press bout strategy as would be expected of an FR1 reinforcement schedule. Animals continued using this strategy when the reinforcement contingency was suddenly switched to FR3, as evidence that they were exploiting their one-press bout learning and not behaving behaviorally exploratory. ChI inhibition during this FR3 training period led to more behavioral exploration, as indicated by a change in bout types being used; this exploration ultimately led to a more energetically optimal strategy (i.e., increased use of three-press bouts), although this was not exclusively the strategy used as all non-one-press bout-type frequencies (2, 3, and 4+) increased. While this effect was quite evident, we recognize that it was not a truly profound effect as the dominant type of bout in animals with ChI inhibition remained the one-press. This could be due to many factors, including a lack of complete ChI inhibition resulting from the DREADDs method, the small anatomical area of DLS coverage of the DREADDs receptor, and/or the fact that a small population of total DLS cells even within the small spatial area were inhibited. Given these factors, in fact, it remains notable how clearly such a limited perturbation of overall DLS activity did affect behavior.

When contingencies were removed during RT60 testing after FR3 training, inhibition of ChIs similarly facilitated the cessation of lever press behavior, arguably the optimal strategy to use. In other words, animals with ChI inhibition were more likely to decrease pressing, as would be the most appropriate response with respect to optimal energy expenditure. Magazine entries made per minute remained consistent across groups from test day 1 to test day 2. We must highlight that this could be because animals responded to the sound of the feeder clicking to deliver reinforcers during these sessions, as these sessions were fully rewarded. Pressing bout-type data were considered in this RT60 test phase as well. While the most straightforward measure to consider during RT60 test is lever press behavior plainly, bout probabilities could shed additional light on behavioral flexibility insofar as whether or not animals were fixed on a certain strategy or were willing to try other strategies when press behavior was divorced from outcome delivery. Indeed, flexibility of this sort in some groups was observed, but appeared not to be specific to ChIs being on- or offline but instead whether there was a change in CNO administration between experimental phases. Specifically, between the training and testing phases, the number of groups increased from 3 to 5, with some groups continuing to receive the same treatment (CNO-CNO, Veh-Veh, CNO control) and others switching

treatments (Veh-CNO, CNO-Veh) at that point in the experiment. Animals that received the same treatment for both training and test maintained the proportion of bout types employed between the two phases of the experiments while animals that switched designations showed changes in the odds that one strategy was used over another. This is interesting as it hints that not only is the simple binary of on- or offline insufficient to fully capture ChI contributions to behavioral flexibility, but flexibility can be expressed in multiple forms: lever press cessation and strategy shifts. Focusing on the latter, onset (or offset) of inactivation appears to play a role in bout-type selection. The Veh-CNO group specifically shows that inactivation mid-experiment buoyed the mean three-press bout probability modestly, but significantly, over time when compared to the group that did not experience a switch. A murkier effect is seen in the CNO-Veh group, one that previously had ChI inhibition during training but normal ChI activity during test, as these animals actually reverted back to reliance on a one-press bout strategy. Perhaps ChI inactivation promoted the adoption of an optimal strategy in real-time, during training, but this preference was not lasting. Additional support of the switch being a key factor in altering bout-type probabilities is that the probability shifts observed between training and the first test session were not observed between RT60 Session 1 and Session 2, when there were no CNO administration switches. To summarize, we note that bout changes were not as clearly related to ChI activity status during the RT60 switch as they were during the ratio schedule change earlier, but instead may relate to the interaction of ChI history and current ChI state. Also, any bout trends that were similar to those seen during ChI inhibition in the ratio schedule switch could be unrelated to the RT60 reinforcement condition per se. However, the drop in response rate during this RT60 period was notably different than behavior during the ration schedule switch, and thus likely did relate to the change in reinforcement schedule. Further work to fully understand these effects is required.

Some limitations of our work must be considered. Physiologically, cholinergic interneurons in the striatum are known to pause their tonic activity in response to salient events after learning (Aosaki et al., 1994; Zhang & Cragg, 2017). These pauses have also been shown to coincide with phasic dopaminergic firing, potentially signaling reward prediction errors, suggesting that they are involved in reinforcement learning (Morris et al., 2004). Characteristics of ChI activity pauses also may have distinct functional roles in learning: pause amplitude has been tied to spatial and temporal information about cues and rewards and pause rebound amplitudes vary with outcome probability (Apicella et al., 2009, 2011; Sardo et al., 2000). Our manipulation here is temporally blunt with respect to these ChI firing dynamics, as ChIs were essentially inhibited for the duration of the session in our study; therefore, if the ChI pause played a role

in behavior here, we fail to parse out how the details of ChI activity, pausing, and rebound may be most important for the behaviors observed here. Regardless, the accelerated emergence of explorative behaviors reported in this experiment could support a gain-of-function hypothesis, with the cholinergic pause at the crux. A second limitation concerns interpretation of the RT60 data. While it provides a clear readout of how animals do or do not reduce responding in a flexible manner, it still remains equivocal as to whether reductions in responding reflect (a) response inhibition, or (b) learning that the action–outcome contingency is degraded. Resolving this disparity would require comparing ChI inhibition during an extinction session to this RT60 session or using a traditional contingency degradation procedure (see Section 2). Furthermore, it might be argued that our task-long inhibition of ChI activity increased three-press bouts (in the FR3 switch) and reduced RT60 pressing (in the RT60 switch) because it disrupted DLS function and effectively impaired memory recall. In this view, rats with ChI inhibition are more flexible because there is less interference from the memory of prior learned behaviors (e.g., the FR1 learning). We believe this is a possible mechanistic explanation of why exploration was encouraged by ChI inhibition, and one that deserves further examination. However, it is unlikely that the prior memories were disrupted in any dramatic sense because one-press bouts were still favored in animals even beyond the FR1 learning stage of the experiment.

Our results are not in complete agreement with the two prior studies that are most related to our work. In one view, if inhibiting DLS ChIs promotes goal directedness, presumably by reducing overall DLS function and its ability to encourage habits as the strategy of behavior, then our results agree with those from Bradfield et al. (2013) where ChIs in DMS were critical for what is regarded as typical DMS function. However, this view is in disagreement with results from Aoki et al. (2018), where excitation of DLS ChIs promoted flexibility, a process not canonically ascribed to DLS function, and inhibition of DLS ChIs there also had no effect on the development of a habit response. Clearly, more work needs to be done to reconcile how aspects of ChI activity in different striatal areas contribute to the role of those areas in behavior. Additionally, concerning implications for DLS function, we note above the broad range of action- and habit-related phenomena that the DLS helps support. If we take the stance that DLS ChIs normally help behavioral exploitation and the use of acquired response sequences (since ChI inhibition promoted exploration here), then we can say their role is in line with these conclusions about DLS function. Striatum-wide ChI ablation promoted compulsive social behaviors, further supporting a role for these interneurons in perseverative behaviors (Martos et al., 2017). Speculatively, we extend a link to evidence that the DLS helps as well with controlling the vigor of action performance (Crego et al., 2020;

Dudman & Krakauer, 2016; Smith & Graybiel, 2013; Yttri & Dudman, 2018). Thus, reducing ChI function here might reduce vigor in behavior on our task, and a reduction in vigor can conceivably unmask or otherwise provide an inroad for cognitive processes to influence the ongoing behavior (Crego et al., 2020; Dudman & Krakauer, 2016). Conversely, it is also possible that inhibition of ChI activity here reduced performance vigor, suggesting ChIs might normally limit the contribution of other DLS cell types to promoting vigor. In this view, the increase in occurrence of larger pressing bout sizes (i.e., bouts above one-press) resulting from ChI inhibition could reflect a more vigorous engagement with the operant lever. Additional tests are warranted to explore the ChI role in vigor further.

The results presented here add to an emerging story of how cholinergic interneuron activity contributes to behavioral flexibility. While prior work examined how ChIs in the dorsal striatum contribute to substitution of established habits and behavioral flexibility after contingency changes, the present study adds that inhibition of DLS ChIs can promote behavioral flexibility, and thus optimality, both early in training, presumably before any habits were firmly established, and after action-outcome contingencies were suddenly removed. Considering the mechanisms underlying the transitions between explorative and exploitative behaviors will prove valuable in resolving complications that arise as a product of disorders in decision-making.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

KAA and KSS designed research; KAA performed research and analyzed data; KAA and KSS wrote the manuscript.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15117.

DATA AVAILABILITY STATEMENT

Data will be made available by the corresponding author by request.

ORCID

Kenneth A. Amaya https://orcid.org/0000-0002-4820-1528

REFERENCES

Amaya, K. A., & Smith, K. S. (2018). Neurobiology of habit formation. *Current Opinion in Behavioral Sciences*, *20*, 145–152. https://doi.org/10.1016/j.cobeha.2018.01.003

Amaya, K. A., Stott, J. J., & Smith, K. S. (2020). Sign-tracking behavior is sensitive to outcome devaluation in a devaluation context-dependent manner: Implications for analyzing habitual behavior. *Learning & Memory*, *27*, 136–149. https://doi.org/10.1101/lm.051144.119

Aoki, S., Liu, A. W., Akamine, Y., Zucca, A., Zucca, S., & Wickens, J. R. (2018). Cholinergic interneurons in the rat striatum modulate substitution of habits. *European Journal of Neuroscience*, *47*, 1194–1205.

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience*, *14*, 3969–3984. https://doi.org/10.1523/JNEUROSCI.14-06-03969.1994

Apicella, P., Deffains, M., Ravel, S., & Legallet, E. (2009). Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *European Journal of Neuroscience*, *30*, 515–526.

Apicella, P., Ravel, S., Deffains, M., & Legallet, E. (2011). The role of striatal tonically active neurons in reward prediction error signaling during instrumental task performance. *Journal of Neuroscience*, *31*, 1507–1515. https://doi.org/10.1523/JNEUROSCI.4880-10.2011

Balleine, B. W. (2019). The meaning of behavior: Discriminating reflex and volition in the brain. *Neuron*, *104*, 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*, 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Balleine, B., & Killcross, S. (1994). Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. *Behavioral Brain Research*, *65*, 181–193.

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*, 1158–1161. https://doi.org/10.1038/nature04053

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Boakes, R. A., & Halliday, M. S. (1975). Disinhibition and spontaneous recovery of response decrements produced by free reinforcement in rats. *Journal of Comparative and Physiological Psychology*, *88*, 436–446.

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience and Biobehavioral Reviews*, *68*, 1004–1005.

Bradfield, L. A., Bertran-Gonzalez, J., Chieng, B., & Balleine, B. W. (2013). The thalamostriatal pathway and cholinergic control of goal-directed action: Interlacing new with existing learning in the striatum. *Neuron*, *79*, 153–166. https://doi.org/10.1016/j.neuron.2013.04.039

Cachope, R., Mateo, Y., Mathur, B. N., Irving, J., Wang, H.-L., Morales, M., Lovinger, D. M., & Cheer, J. F. (2012). Selective activation of cholinergic interneurons enhances accumbal phasic dopamine

release: Setting the tone for reward processing. *Cell Reports*, *2*, 33–41. https://doi.org/10.1016/j.celrep.2012.05.011

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 933–942.

Corbit, L. H., & Balleine, B. W. (2000). The role of the hippocampus in instrumental conditioning. *Journal of Neuroscience*, *20*, 4233–4239. https://doi.org/10.1523/JNEUROSCI.20-11-04233.2000

Crego, A. C. G., Štoček, F., Marchuk, A. G., Carmichael, J. E., van der Meer, M. A. A., & Smith, K. S. (2020). Complementary control over habits and behavioral vigor by phasic activity in the dorsolateral striatum. *Journal of Neuroscience*, *40*, 2139–2153. https://doi.org/10.1523/JNEUROSCI.1313-19.2019

Dickinson, A., & Weiskrantz, L. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*, 67–78.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*, 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Dudman, J. T., & Krakauer, J. W. (2016). The basal ganglia: From motor commands to the control of vigor. *Current Opinion in Neurobiology*, *37*, 158–166.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*, 1481–1489.

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *American Journal of Psychiatry*, *168*, 718–726.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*, 359–387.

Kalueff, A. V., Stewart, A. M., Song, C., Berridge, K. C., Graybiel, A. M., & Fentress, J. C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nature Reviews Neuroscience*, *17*, 45–59. https://doi.org/10.1038/nrn.2015.8

Kawaguchi, Y., Wilson, C. J., Augood, S. J., & Emson, P. C. (1995). Striatal interneurones: Chemical, physiological and morphological characterization. *Trends in Neurosciences*, *18*, 527–535. https://doi.org/10.1016/0166-2236(95)98374-8

Klaus, A., Alves da Silva, J., & Costa, R. M. (2019). What, if, and when to move: Basal ganglia circuits and self-paced action initiation. *Annual Review of Neuroscience*, *42*, 459–483.

Kubota, Y., Liu, J., Hu, D., DeCoteau, W. E., Eden, U. T., Smith, A. C., & Graybiel, A. M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. *Journal of Neurophysiology*, *102*, 2142–2160. https://doi.org/10.1152/jn.00522.2009

Kupferschmidt, D. A., Juczewski, K., Cui, G., Johnson, K. A., & Lovinger, D. M. (2017). Parallel, but dissociable, processing in discrete corticostriatal inputs encodes skill learning. *Neuron*, *96*, 476–489.e5. https://doi.org/10.1016/j.neuron.2017.09.040

Malvaez, M., & Wassum, K. M. (2018). Regulation of habit formation in the dorsal striatum. *Current Opinion in Behavioral Sciences*, *20*, 67–74. https://doi.org/10.1016/j.cobeha.2017.11.005

Martos, Y. V., Braz, B. Y., Beccaria, J. P., Murer, M. G., & Belforte, J. E. (2017). Compulsive social behavior emerges after selective ablation of striatal cholinergic interneurons. *Journal of Neuroscience*, *37*, 2849–2858. https://doi.org/10.1523/JNEUROSCI.3460-16.2017

Morris, G., Arkadir, D., Nevet, A., Vaadia, E., & Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, *43*, 133–143. https://doi.org/10.1016/j.neuron.2004.06.012

Okada, K., Nishizawa, K., Fukabori, R., Kai, N., Shiota, A., Ueda, M., Tsutsui, Y., Sakata, S., Matsushita, N., & Kobayashi, K. (2014). Enhanced flexibility of place discrimination learning by targeting striatal cholinergic interneurons. *Nature Communications*, *5*, 3778.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annual Review of Neuroscience*, *25*, 563–593.

Paxinos, G., & Watson, C. (2009). *The rat brain in stereotaxic coordinates: Compact* (6th ed.) San Diego, CA.: Academic Press.

Sardo, P., Ravel, S., Legallet, E., & Apicella, P. (2000). Influence of the predicted time of stimuli eliciting movements on responses of tonically active neurons in the monkey striatum. *European Journal of Neuroscience*, *12*, 1801–1816.

Smedley, E. B., & Smith, K. S. (2018). Evidence of structure and persistence in motivational attraction to serial Pavlovian cues. *Learning & Memory*, *25*, 78–89.

Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, *79*, 361–374. https://doi.org/10.1016/j.neuron.2013.05.038

Threlfell, S., Lalic, T., Platt, N. J., Jennings, K. A., Deisseroth, K., & Cragg, S. J. (2012). Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron*, *75*, 58–64. https://doi.org/10.1016/j.neuron.2012.04.038

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*, 345–352.

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., Yizhar, O., Cho, S. L., Gong, S., Ramakrishnan, C., Stuber, G. D., Tye, K. M., Janak, P. H., & Deisseroth, K. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*, *72*, 721–733. https://doi.org/10.1016/j.neuron.2011.10.028

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*, 464–476.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*, 181–189.

Yttri, E. A., & Dudman, J. T. (2018). A proposed circuit computation in basal ganglia: History-dependent gain. *Movement Disorders*, *33*, 704–716. https://doi.org/10.1002/mds.27321

Zhang, Y.-F., & Cragg, S. J. (2017). Pauses in striatal cholinergic interneurons: What is revealed by their common themes and variations? *Frontiers in Systems Neuroscience*, *11*, 80. https://doi.org/10.3389/fnsys.2017.00080