PREPRINT

RESEARCH ARTICLE PREPRINT

Action sequence dynamics rely on accessory hyperdirect pathway-related targets in the striatum

Lachlan A. Ferguson,[1]* Miriam Matamales,[1] Bernard W. Balleine,[1]
Jesus Bertran-Gonzalez[1]*

[1]Decision Neuroscience Laboratory, School of Psychology, University of New South Wales, Sydney, NSW, Australia

*lachlan_ferguson@hotmail.com (L.A.F); j.bertran@unsw.edu.au (J.B.G)

OPEN ACCESS

Citation: Ferguson LA, Matamales M, Balleine BW, Bertran-Gonzalez, J (2022) Action sequence dynamics rely on accessory hyperdirect-related targets in the striatum. bioRxiv, 2022/486040. doi: 10.1101/2022.03.28.486040

Uploaded: March 2022

Acknowledgements: We thank Jennifer Strempel, Anne Rowan and Lydia Williams for assistance with animal care.

Funding: This work was supported by an Australian Government Research Training Program Scholarship, a UNSW High Degree for Research Completion Scholarship and a UNSW Writing Scholarship to L.A.F., as well as by Australian Research Council (DP190102511, DP210102700) and the National Health and Medical Research Council (APP1165990) grants to J.B.G. and M.M.

Contributions: L.A.F., B.W.B. and J.B.G. conceived the study and designed the experiments. L.A.F. performed behavioural experiments, surgeries and viral manipulations. M.M. and L.A.F. performed behavioural and quantitative imaging analyses. J.B.G. and L.A.F. performed statistical analyses. M.M. and B.W.B. provided feedback on the manuscript. L.A.F. and J.B.G. wrote the paper.

Competing interests: The authors have no competing interests.

Abstract

Performing several actions in swift succession is often necessary to exploit known contingencies in the environment. However, after a change in the contingency rules, the ability to appropriately adapt rapid action sequences must be procured for continued success. By combining analyses of behavioural microstructure with circuit-specific tracing in mice, we report on a relationship between action timing-variability and successful adaptation that relies on post-synaptic targets of primary motor cortical (M1) projections to dorsolateral striatum (DLS). We found that M1 hyperdirect pathway projections to the STN also send dense axonal collaterals to external globus pallidus and dorsal striatum, with the highest synaptic volumes found in the DLS. Specific interruption of the M1→DLS circuit reduced the proportion of successful sequences while speeding-up and reducing action timing-variability, revealing a role for M1→DLS circuitry in setting the exploration/exploitation balance that is required for adaptively guiding the timing and success of instrumental action.

Introduction

For stable environmental contingencies that require more than a single action, animals can learn to perform a series of discrete responses that, as experience accrues, melds their internal boundaries into accurately timed streams of skilled behaviour, often expressed as a single unit or "chunk" (Graybiel, 1998; Lashley, 1951; Rosenbaum et al., 1983; Saling & Phillips, 2007; Sternberg et al., 1978; Terrace, 1987). However, in a fluctuating environment, both correctly acquiring these behavioural streams and enabling sufficient variation in them, even amongst the most well-learned actions, is essential for adapting to changed conditions (Sternad, 2018). During the development of overtrained actions, such as habits and skills, the priority of specific cortico-basal ganglia circuit function appears to be reorganised. Evidence for the serial transfer of goal-directed to habitual action indicates a shift in activity from medial prefrontal cortex and dorsomedial striatum (DMS) to sensorimotor cortices and dorsolateral striatum (DLS) (Balleine, 2019; Balleine et al., 2007; Balleine & O'Doherty, 2010; Ostlund & Balleine, 2005; Yin et al., 2004, 2006). In the context of motor skill learning, corticostriatal afferents targeting the DMS and DLS initially co-engage, but as skills develop, associative input strength declines more rapidly and to a greater degree than motor cortical inputs (Kupfer-

schmidt et al., 2017). In support of this, studies recording neuronal activity or assessing function using chemogenetics implicate medio-lateral striatal function in the regulation of action chunking and timing (D. Z. Jin et al., 2009; X. Jin et al., 2014; X. Jin & Costa, 2010; Matamales et al., 2017; Mello et al., 2015). Regulation of chunking has been associated with phasic shifts of neuronal activity in the striatum—particularly in the DLS—specific to either the first, last or sustained responding in a sequence of lever presses (X. Jin et al., 2014; X. Jin & Costa, 2010). Contributions to the temporal dynamics of well-learned action sequences have also been observed following manipulations of direct pathway projection neuron activity in the DLS; e.g., optogenetic stimulation extends ongoing action sequences (Tecuapetla et al., 2016), whereas chemogenetic inhibition during learning compresses sequences into briefer durations without impacting the total number of presses within them (Matamales et al., 2017). Thus, evidence strongly supports the contribution of DLS neurons in determining both the structure and timing of well-learned sequences.

Thus, the rapid processing of sensory stop-related cues in the STN suggests that the cessation of action induced by the hyperdirect pathway may rely on a bottom-up relationship with the slower activities found upstream in the striatum and GPe (Schmidt & Berke, 2017). Building on these temporal control functions, the STN itself has also been proposed to extend deliberation times when animals and humans face decision conflicts (Baunez et al., 2007; Baunez & Robbins, 1997; Frank, 2006; Frank et al., 2007). Clearly, evidence suggests that, beyond the control of immediate action, the hyperdirect pathway may interact with upstream, higher-order basal ganglia systems to influence the optimal span of action as an adaptive behavioural process. In terms of its connectivity, additional branches of the cortico-subthalamic pathway originating from layer V neurons are known to also send collaterals to the striatum (Kita & Kita, 2012), while other motor cortico-basal ganglia inputs with similar origin as the hyperdirect projection are capable of driving excitatory post-synaptic currents in the GPe (Karube et al., 2019) and striatum (Calabresi et al., 1990; Lovinger, 1991).

While evidence clearly implicates the classical hyperdirect pathway in cessation of ongoing action, recent neuroanatomy data and speculation based on circuit-wide functional assessments suggest a broader role for the hyperdirect pathway and its associated circuitry in regulating action span—a process that could be particularly important for the adaptation of mature action sequences to new requirements. Here we hypothesised that bottom-up hyperdirect pathway processing involving the striatum may play a role in shaping the temporal dynamics of action during the adaptation of action sequences. Using viral tracing studies, we sought to evaluate the relative densities of key cortico-subcortical projections, and found that relatively high volumes of collateral hyperdirect pathway projections reach both the DLS and GPe. By combining circuit-specific cell ablation with a novel instrumental paradigm that specifically promotes self-determined variation in action duration in mice, we found that a subtle depletion of M1→DLS post-synaptic connectivity disrupts action sequence chunking and timing in a way that compromises successful adaptation.

Results

As mice build skilled actions, the timing of their sequences adapts for success

The first experiment characterised behavioural adaptations that mice make to the timing and efficacy of their performance over instrumental training. To ensure that action timing and performance remained self-paced and uninfluenced by external cues common in instrumental conditioning procedures, we developed a novel self-paced chained-sequence task based

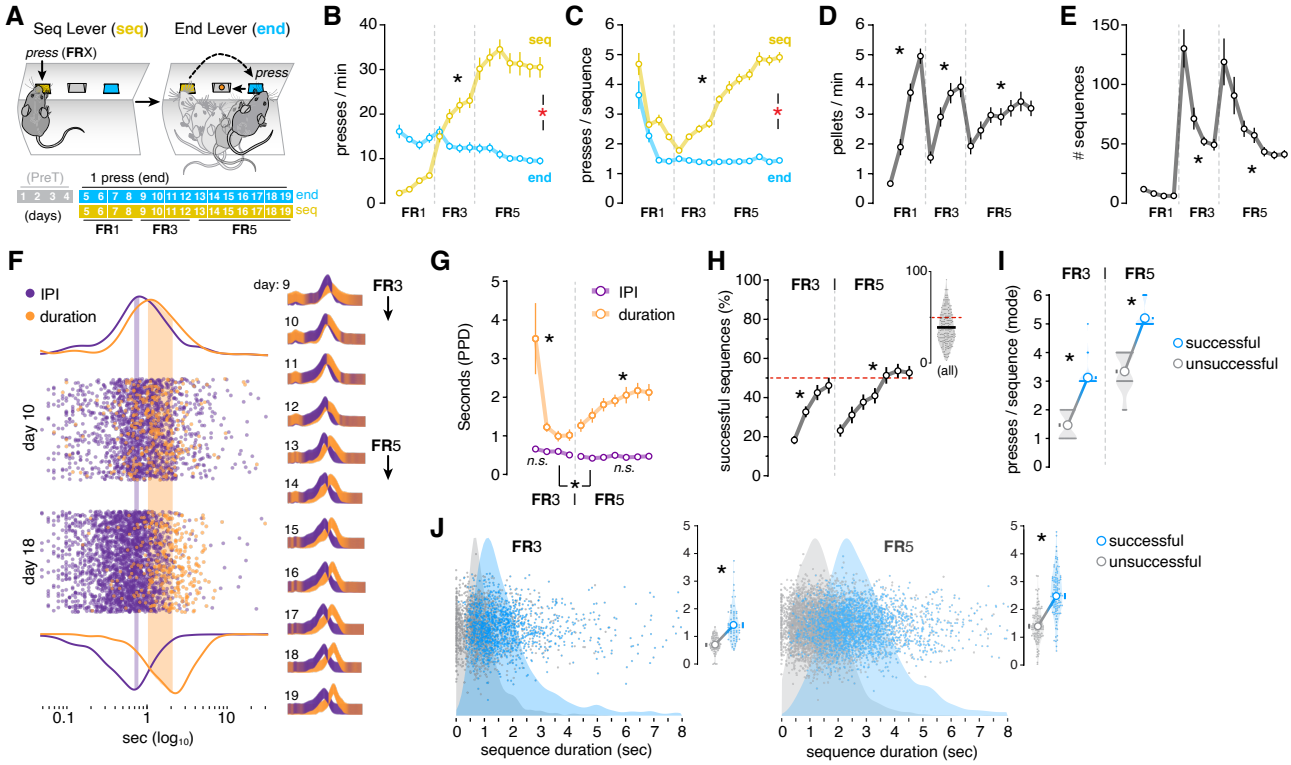

Figure 1. A self-paced sequence task reveals timing adaptations during training. (**A**) Animals were pre-trained with continuous reinforcement (1 press→1 reward) on the End lever for 4 sessions (sessions 1-4, PreT, see Figure S1A). Next, the Sequence lever was introduced on a fixed ratio (FR) 1 schedule for 4 sessions (sessions 5-8), whereby pressing on the Sequence lever must occur prior to pressing the End lever in order to receive reward. In the following 4 sessions (sessions 9-12), the press requirements on the Sequence lever increased to FR3. In the final 7 sessions (sessions 13-19), the press requirements on the Sequence lever increased to FR5. (**B**) Lever press rate measured as presses per minute on each lever type throughout FR1-FR5 training. (**C**) Sequence length measured as the number of presses per sequence in both the Sequence and End levers across FR1-FR3 training. (**D**) Reward rate measured in pellets per minute. See Figure S1C for the total rewards earned throughout training. (**E**) Total number of sequences performed throughout FR1-FR5 training. (**F**) Scatter plot of each IPI and sequence duration value (to log10) on the Sequence lever for all animals in an example FR3 (day 10) and FR5 (day 18) session, with probability density function curves indicating peak differences (shaded). Right diagrams show the probability density function curves on each day of FR3-FR5 training. See Figure S2A-B for individual days. (**G**) IPI and Duration expressed as averaged probability density peaks (PPD, seconds) across FR3 and FR5 training. (**H**) Percentage of sequences that successfully resulted in reward in FR3 and FR5 sessions across training and in all FR3 and FR5 training sessions collapsed (inset). Red dashed line denotes 50%. (**I**) Most frequently occurring (modal) number of presses in either unsuccessful or successful sequences for both FR3 or FR5 training. Truncated violin plots are fitted to data points (shaded). (**J**) Scatter plot with probability density function curves (shaded) of sequence duration for every unsuccessful (unrewarded) and successful (rewarded) sequence performed by the entire cohort during FR3 (left) and FR5 (right). Insets show PPD for each animal and day during FR3 (left) and FR5 (right) training. *, significant overall/simple effect (black) and interaction (red). N.S., not significant (Table S1).

on the Mechner Counting Task (Light et al., 2019; Mechner, 1958b). Mice were presented with two levers in tandem and earned a reward for a single press on the second 'End' lever, providing they had completed the required number of presses on the first 'Sequence' lever (Figure 1A, top). This allowed the mice to freely decide the duration of their action chains (performed on the Sequence lever) without relying on external cues associated with reward delivery. The requirements on the Sequence lever increased every four sessions from fixed ratio (FR) 1 to

FR3 to FR5 across training, whereas the End lever always required one single press (Figure 1A, bottom). General measures of performance in this task, such as lever press rate (Figure 1B) and number of presses per sequence (Figure 1C), significantly increased for the Sequence lever but not for the End lever over the course of training after pretraining (Figure S1A). This was supported by a significant session x lever interaction in both cases (Table S1), demonstrating that mice appropriately biased performance toward the Sequence lever. Mice also significantly increased the rate of rewards earned within each FR schedule (Figure 1D and Table S1), which was accompanied by a commensurate reduction in the number of sequences required to achieve them (Figure 1E), as well as reduced magazine entry rates (Figure S1B and Table S1). These results show that mice clearly distinguished between lever contingencies and became more effective within each phase of the task.

We then assessed if the improved effectiveness in earning rewards coincided with more efficient action sequence timing. To determine what changes in timing predominated in the adaptation of action, we measured the peak probability of both individual inter-press intervals (IPIs) within a sequence and whole sequence durations across training (FR3 and FR5). We found that the probability density distribution for IPIs remained stable throughout training, whereas the same function applied to sequence duration shifted to the right as training progressed (Figure 1F and S2A-B). Measures of the action timing peak distribution across training revealed that IPIs remained relatively stable, with only a moderate decline occurring between FR3 and FR5 phases. In contrast, the sequence duration initially declined during early acquisition (FR3), then steadily increased during FR5 training (Figure 1G and Table S1). Given the rise in reward rate, the related decline in the number of sequences and the elongation of the number of presses per sequence, we expected the likelihood of performing action sequences that ended in reward to increase as training progressed. We calculated the percentage of successful (rewarded) sequences relative to unsuccessful (unrewarded) sequences and found that the former significantly increased within each training phase (FR3 and FR5), reaching 39.11% on average across all training and, after five sessions, plateauing at approximately 50% success on FR5 (Figure 1H and Table S1). Mice were clearly capable of improving the efficacy of their sequences by increasing the chance of performing—at minimum—the required number of presses. It was unclear, however, if the adjustments to the number of actions from unsuccessful to successful trials coincided with adjustments in sequence timing; i.e., whether (i) more presses were added to a fixed period and executed at a faster rate, or (ii) sequence duration was extended with the addition of lever presses executed at a similar rate. We observed that the latter was the case: when a significantly greater number of presses was implemented for successful sequences (Figure 1I and S1D), the peak sequence duration of successful sequences shifted to significantly longer durations relative to unsuccessful attempts (Figure 1J and S1E, Table S1). These data suggest that, over and above changes in inter-press-intervals, the modulation of sequence duration seems to be the critical variable when adapting action for success.

New chunks are smoothly merged with previously established action scaffolds

We next examined how successful sequences are constructed when facing a change in schedule using the self-paced sequence task. Several strategies can be used to adapt established sequences to a new ratio requirement. A conservative scenario could involve incorporating a single additional press to previously successful sequences—a strategy that would prioritise previous history of success and engage limited action-exploration. Alternatively, a bolder approach could involve implementing more dramatic behavioural shifts by adding

multiple presses to historically successful sequences—deprioritising previous history of success and promoting action exploration. We investigated this by quantifying the frequency of every sequence category according to the number of presses per sequence and calculating the probability distribution of each category (Figure 2A and B). Interestingly, we found that successful responses were typically a 2-press chunk away from the most frequent unsuccessful sequence; i.e., responses jumped from unsuccessful 1-press actions to successful 3-press sequences in FR3 training (Figure 2A), and from unsuccessful 3-press sequences to successful 5-press sequences in FR5 training (Figure 2B). This analysis showed that, on experiencing variations in contingency, mice demonstrated an ability rapidly to modify their sequences to match the new requirement with low under- and overshooting.

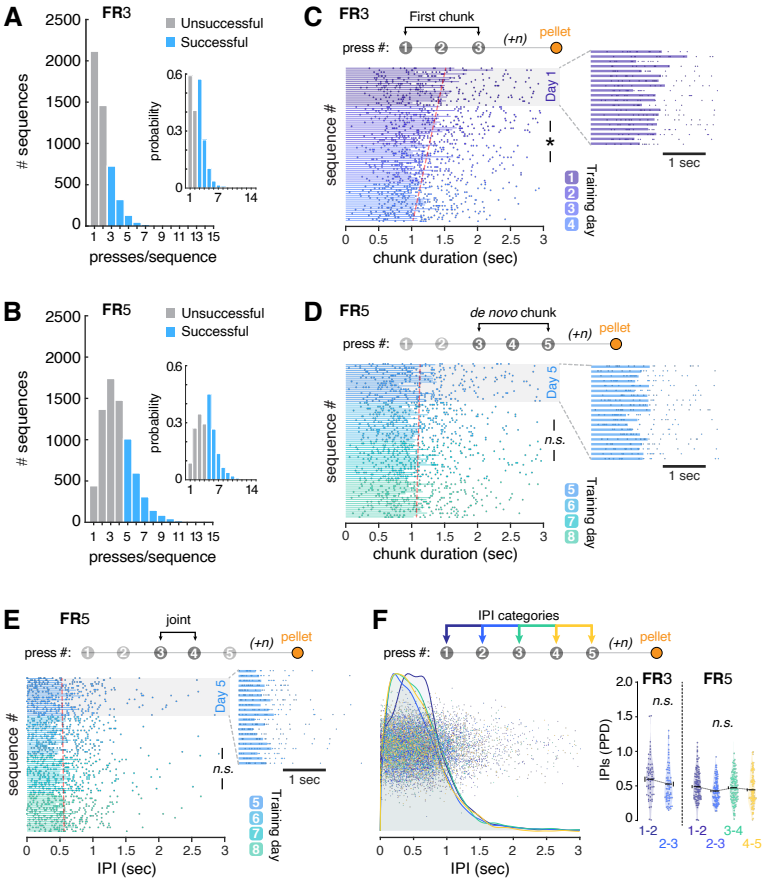

Figure 2. Successful action dynamics smoothly evolve as task requirements increase. (**A-B**) Frequency histograms showing the total number of sequences performed on the sequence lever according to the number of presses per sequence during FR3 (A) and FR5 (B) training. Insets show the probability distribution of the same sequence categories during FR3 (A, right) and FR5 (B, right) training. (**C-E**) Duration of successful subsequence intervals ranging from presses 1-3 (First chunk, C), presses 3-5 (*de novo* chunk, D) and presses 3-4 (joint, E) arranged chronologically across the first four sessions of FR3 and FR5. Data are the duration of each sequence by each mouse (dots) and the average across mice (bars). A linear regression model highlighting the chronological trend is fitted to the data (red dashed line). Insets are an enlarged view of the first session of the corresponding fixed ratio schedule. (**F**) Scatter plot with probability density function curves (shaded) of IPIs between 1-2, 2-3, 3-4 and 4-5 press transitions for every successful sequence performed by the entire mouse cohort during FR3 and FR5 training (left). Peak probability density (PPD) of relevant IPIs in successful sequences for each training session (dots) plotted for both FR3 and FR5 (right). n = any number of presses before reward. *, significant overall/simple effect (black) and interaction (red). N.S., not significant (Table S1).

Next, we explored the way lever press responses were chunked during action sequence learning and whether temporal gaps between chunks emerged as animals adapted their sequences to new ratio requirements (Rosenbaum et al., 1983). For this, the duration of each chunk within successful sequences was arranged chronologically, following the order in which each sequence occurred within a session. To observe the relationship between the chronological progression of successful sequences and the duration of their constituent chunks we analysed their linear relationship over both FR3 and FR5 training (Figure 2C and D, Figure S3A and B). We found that during FR3 training—when sequences are first acquired—the duration of the first chunk (time between presses 1-3) significantly declined over training (Figure 2C and S1A Table S1). In contrast, during FR5 training—when FR3 sequences have already been established and two extra presses are being added—the duration of the *de novo* chunk (time between presses 3-5) was the same as the first chunk in late FR3 training (~1 sec; Figure 2D), and remained constant throughout the rest of training (Figure 2D and S3B, Table S1). In light of the observed disparities in the evolution of the first and *de novo* chunks of successful sequences, we investigated if the two chunks were implemented as discrete units with a pause between them, or if they were smoothly integrated into an extended single sequence of action. We found that the time in-between the two chunks (i.e., the "joint" IPI; between presses 3 and 4) remained invariable as rewarded experience accrued across FR5 training (Figure 2E and S3C, Table S1). Furthermore, we found that the different IPI categories across successful sequences were indistinguishable from each other, including the joint IPIs connecting first and *de novo* chunks (Figure 2F, Table S1). Collectively, these data reveal that, in order to adapt to new contingencies, mice smoothly integrate new sub-sequence chunks into previously acquired sequence prototypes to form extended sequences immediately.

DLS and GPe are accessory targets of the hyperdirect pathway

Given the related roles played by the cortico-basal ganglia hyperdirect pathway (motor cortex to STN) and lateral striatal circuitry in action latency, and in light of recent neuroanatomical and neurophysiological links between these structures, we set out to explore the upstream connectivity of the canonical hyperdirect pathway using viral-based fluorescent tracing methods. We first aimed to explore the relative projection densities of the hyperdirect pathway to targets other than the STN using the Allen Mouse Brain Connectivity Atlas (Oh et al., 2014), which combines eGFP anterograde viral tracing with serial two-photon tomography throughout the entire brain. Three different cortical injection assays (spanning the M1, M2 and dorsal agranular insular area [AId] regions) identified the STN as the primary target (Figure 3A-C) (Table 1). EGFP-labelled axons in all three assays densely innervated several subcortical structures, particularly the lateral areas of the dorsal striatum (dorsolateral striatum; DLS) and lateral areas of the globus pallidus externa (lateral GPe) (Figure 3B). EGFP fluorescence quantification identified that, across major brain structures, the STN, striatum and GPe consistently had the three highest projection densities (amongst non-cortical regions) over the 3 assays (Figure 3C).

We next sought to investigate whether these regions (STN, striatum and GPe) were simply additional cortical targets or in fact collaterals of the M1→STN hyperdirect pathway. For this, we implemented a quantitative connectivity approach based on the retrograde transport of a Cre-expressing virus (AAV-hSyn-HI-eGFP-Cre-WPRE-SV40) injected at the hyperdirect pathway target (STN, Figure S4), followed by anterograde transport of a Cre-dependent reporter virus (AAV-hSyn1-FLEX-mGFP-2A-synaptophysin-mRuby) injected at the origin of the hyperdirect pathway (M1) (Figure 3D). Because the virus causes Cre-dependent anterograde

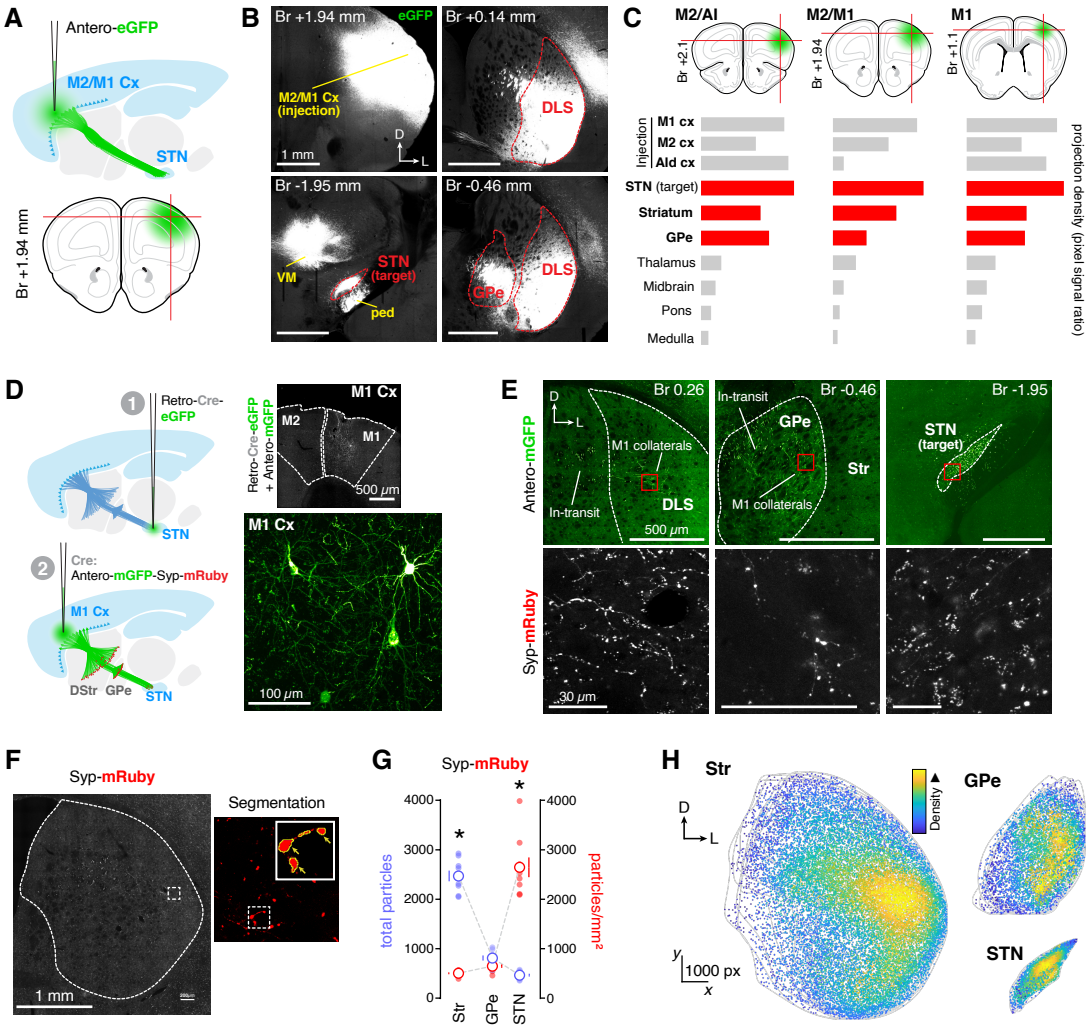

Figure 3. The hyperdirect pathway sends accessory shared collaterals to the DLS and GPe. (**A-C**) Anterograde tracing study using the Allen Mouse Brain Connectivity Atlas database (Oh et al., 2014). Three different cortical injection assays targeting the STN were identified (Source: M1 and M2; Target: STN; see methods). (A) Experimental diagram on one of the assays showing the injection site onto M2/M1 cortex and expected transport of an anterograde reporter virus reaching the STN (antero-eGFP). (B) Example two-photon tomography images of eGFP expression in the M2/M1 injection site (top-left); the DLS (top-right); the STN (bottom-left) and the GPe (bottom-right) (experiment number 180709942). (C) Schematic for M2/agranular insular cortex (AI) (top-left), M2/M1 (top-centre), and M1 (top-right) injection sites with corresponding projection density quantifications throughout various brain areas. Data are extracted from experiments 180719293 (left), 180709942 (centre) and 100141780 (right). Projection densities for cortical regions around the injection site ('Injection') are listed first, followed by the three highest density regions (red)—including the site of target search (STN)—followed by other representative high-density regions. (**D**) Schematic depicting the viral tracing strategy used to identify hyperdirect pathway accessory targets: (1) a retrograde AAV expressing Cre-eGFP was injected in the STN. (2) an anterograde AAV expressing Cre-dependent mGFP and synaptophysin (Syp)-mRuby (labelling presynaptic boutons) was injected in the M1. Right panels are confocal images showing GFP expression in the M1. See Figure S4 for quantification of STN targeting. (**E**) Spinning disk confocal images of anterograde-mGFP in the striatum (top-left), GPe (top-centre) and STN (top-right); and Syp-mRuby-labelled terminals in each region (bottom panels). (**F**) Spinning disk confocal image showing Syp-mRuby clusters segmented for particle analysis (see methods). (**G**) Total particles and Particle density (particles/mm2) quantification for DStr, GPe and STN regions [*p<0.05]. (**H**) Particle density maps overlaid for each region (3x slice/animal; n = 3) on side ipsilateral to STN and M1 injection sites. See Figure S5 for individual maps.

expression of mRuby in synaptic terminals (Fisher et al., 2020; Zhang et al., 2016), this method allowed us to quantify the distribution of any mRuby-labelled synaptic boutons in projections collateral to and within the mainstream hyperdirect targeting of the STN. Consistent with the previous experiment, we detected dense mRuby puncta in both DLS and lateral GPe, in addition to the final target STN (Figure 3E). Our particle density analysis showed that the total number of mRuby+ synaptic boutons was greater in the collateral projections to the striatum and GPe compared to the STN, with the greatest number occurring in the striatum (Figure 3G, purple trace) (Table S1). In contrast, the relative density of synaptic boutons within their regional space was greatest in the STN compared to the striatum and GPe (Figure 3G, red trace) (Table S1). The reconstruction of the distribution of mRuby+ terminals into particle density maps showed that the greatest synaptic densities occurred within the lateral segments of the striatum and GPe, whereas synaptic territories remained central in the final target STN (Figure 3H, Figure S5). These data extend recent reports showing upstream cortico-basal ganglia connectivity (Karube et al., 2019; Kita & Kita, 2012) and reveal that the hyperdirect pathway originating in M1 indeed sends shared projections to the DLS, thereby linking a candidate circuit to the temporal modulation of action during sequence learning.

Specific M1→DLS lesion impairs task success, speeds up action sequences, and reduces variability of action timing

Functional assays, such as lesion and chemogenetic suppression, indicate the DLS governs a variety of roles relevant to optimising task performance in sequence-based instrumental conditioning, ranging from skilled action kinematics, speed of action sequences, habit learning and the accurate acquisition of a serial order (Dhawale et al., 2021; Matamales et al., 2017; Yin, 2010; Yin et al., 2004). Our neuroanatomical data showed that the greatest volume of synaptic connectivity from collateral hyperdirect projections occurs in the striatum, particularly in its lateral subregion. Therefore, we investigated whether specific M1 projections related to hyperdirect pathway synaptic territories in the striatum contributed to modulating action time during sequence learning, and whether this influenced task success. We used a circuit-specific lesion approach in which DLS neurons that are post-synaptic to M1 projections were specifically lesioned in a well-defined area. A trans-synaptic anterograde travelling Cre virus (AAV2-EF1a-mCherry-IRES-WGA-Cre) was injected into the M1 (Figure S6A), and a Cre-dependent lesioning virus (AAV-flex-taCasp3-TEVp) into the DLS (Gradinaru et al., 2010; Yang et al., 2013). We carefully targeted the second injection to the high-density zones established in our previous hyperdirect pathway terminal mapping experiment (Figure 4A and S6B, see Figure 3H and S5), resulting in a significant decrease of postsynaptic Nissl+ cells limited to high-density zones (Figure 4B). When exposing these animals to the self-paced sequence task, we found that mice from both the DLS lesioned and Sham groups were capable of appropriately biasing their press rate (press/min) toward the sequence lever as sequence training progressed from FR3-FR7 (Figure 4C), as revealed by a strong session x lever interaction in both groups (Table S1). On the other hand, a summary of task success across all sequence training (FR3-FR7) showed that M1→DLS lesioned mice earned rewards at a slower rate (Figure 4D, Table S1) and performed sequences with a lower percentage of success relative to Sham controls (Figure 4E, Table S1), without impacting the total number of earned rewards or magazine entries per session (Figure S6D and E). Despite this, the M1→DLS lesioned group showed no difference in the number of presses per sequence relative to the Sham group when performing either unsuccessful or successful sequences (Figure 4F, Table S1). By contrast, the timing of these sequences was altered, such that both unsuccessful and successful sequences were faster following circuit-specific lesion (Figure 4G, Table S1).

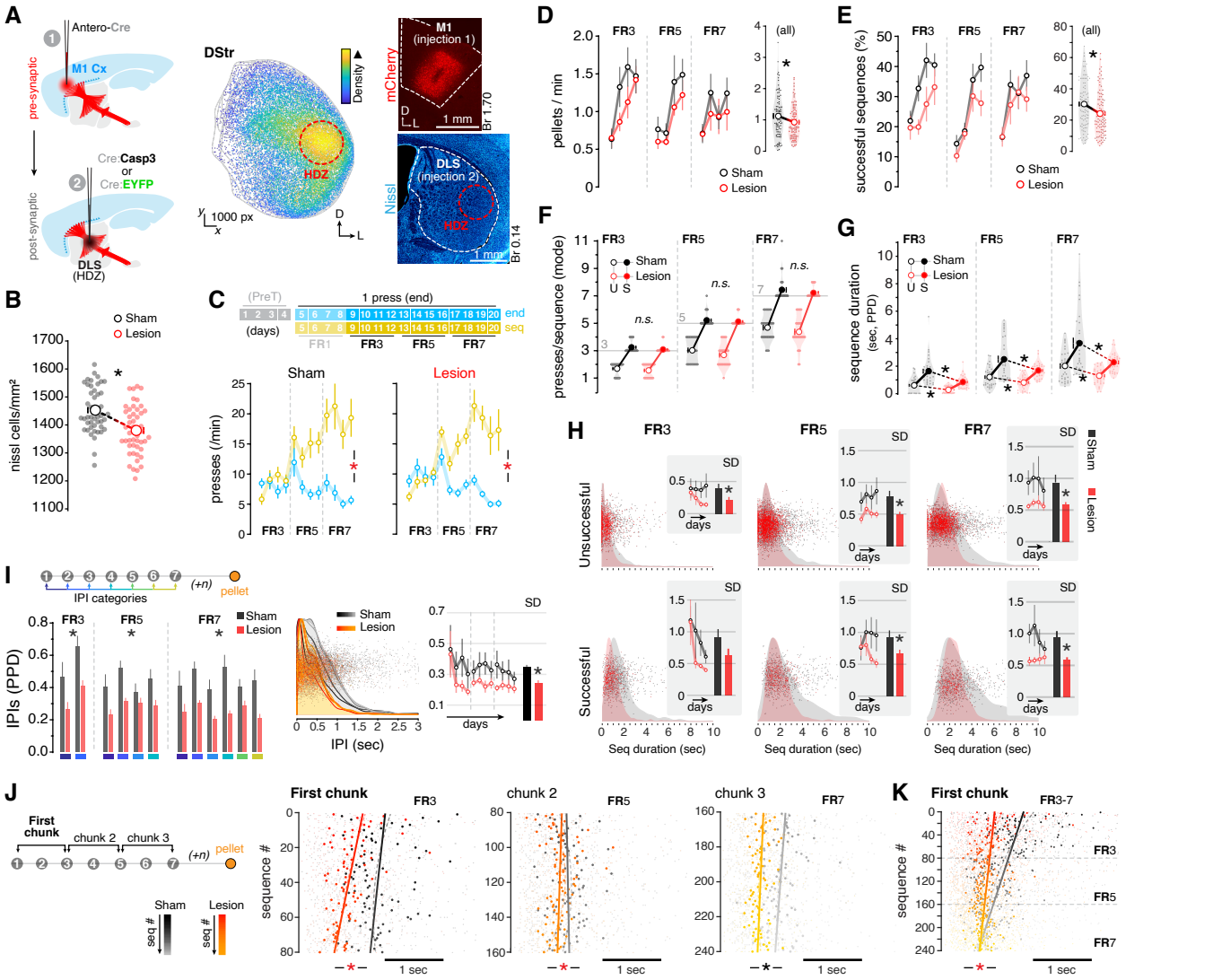

Figure 4. Specific interruption of the M1→DLS circuit alters the temporal dynamics of action sequences. (**A**) Schematic of the M1→DLS circuit lesion strategy. (1) An anterograde AAV expressing Cre (Antero-Cre) was injected into the M1. (2) An AAV expressing Cre-dependent Casp3 (Cre:Casp3 or Cre:EYFP in controls) was then injected into the High-Density Zone (HDZ) defined in the DLS (central panel, see Figure 3H). Right: confocal images of antero-Cre virus expression in the M1 injection site revealed by mCherry (top) and Nissl labelling with standardised HDZ outline in the DLS. (**B**) Nissl-based cell density quantification within the HDZ in the DLS. (**C**) Top: animals began pre-training with continuous reinforcement (CRF) on the End lever (Lend) for 4 sessions (PreT), then shifted to the tandem task on the Sequence lever (Lseq) progressing through FR1→FR3→FR5→FR7 schedules every 4 sessions (top). Bottom: lever press rate measured as presses per minute on each lever type throughout FR3→FR7 training in each group. For sessions 1-4 (PreT), see Figure S6C. (**D**) Reward rate (press/min) throughout FR3→FR7 training. Inset shows data from the 3 schedules collapsed. (**E**) Percentage of successful sequences across FR3→FR7 training. Inset shows data from the 3 schedules collapsed. (**F**) Most frequent (modal) number of presses in a sequence (press/sequence) for both Successful and Unsuccessful sequences in FR3, FR5 and FR7 schedules. (**G**) Sequence duration (peak probability distribution, PPD) for both Successful and Unsuccessful sequences in FR3, FR5 and FR7 schedules. (**H**) Scatter plot with PPD curves (shaded) of sequence duration for every unsuccessful (top) and successful (bottom) sequence produced by the entire cohort during FR3, FR5 and FR7 phases. Insets show standard deviation (SD) across training sessions and a summary bar graph of all sessions within the indicated schedule. (**I**) Left: schematic for quantification of IPIs within a sequence. Data (bottom) shows the PPD of the relevant IPIs for successful sequences in each schedule. Centre: Scatter plot with PPD functions (shaded) of IPIs between the indicated press transitions for every successful sequence performed by the entire cohort as FR3→FR7 training progressed (colour coded). Right: standard deviation (SD) across FR3→FR7 training. Inset shows a summary bar graph of all sessions collapsed. (**J**) Duration of successful subsequence intervals ranging

Figure 4 continued on next page

Figure 4 continued

from presses 1-3 (First chunk, left), presses 3-5 (chunk 2, centre) and presses 5-7 (chunk 3, right) arranged chronologically across FR3, FR5 and FR7 sessions. Data are the duration of each sequence by each mouse (small dots) and the average across mice (larger dots). A linear regression model highlighting the chronological trend is fitted to the data (line). The change in colour in each group's dataset reflects progression throughout training (see legend to the bottom-left). (**K**) Duration of the first successful subsequence segment (first chunk) throughout the entire FR3→FR7 training. n = any number of presses before reward. *, significant overall/simple effect (black) and interaction (red). N.S., not significant (Table S1).

Next, we explored the relationship between the sequence speed increases we observed and the variability of their execution as a source of explanation for the reduced success of M1→DLS lesioned animals. We found that the M1→DLS lesioned group maintained more consistent unsuccessful and successful sequence durations compared to Sham controls throughout sequence training (FR3-FR7) (Figure 4H), the latter group showing significantly more variable sequence durations of either type (Table S1). This effect was consistent with a significant reduction in the variability of the IPIs following lesion (Figure 4I-right panel, Table S1). We then sought to clarify if, in successful sequences, the timing changes induced by M1→DLS lesions generalised to the whole sequence or if action timing fluctuated at specific positions within the sequence. By sorting IPIs according to position in the sequence and comparing their differences within each training schedule, we found a general significant decrease in IPI time following M1→DLS lesions across training, with no differences between the IPIs according to position in the sequence for either FR5 and FR7 training (Figure 4I, Table S1). Similarly, our sequence structure analysis showed a reduced chunk duration for the first chunk (press 1-3) during FR3, for chunk 2 (press 3-5) during FR5 and for chunk 3 (press 5-7) during FR7 in the DLS lesion group, which significantly diverged from the Sham group as rewarded experience progressed in FR3 and FR5 (Figure 4J). Importantly, our analysis of the evolution of the first chunk—which we observed undergoes temporal change during initial action sequence acquisition in our previous study—revealed a highly supressed rate of change in M1→DLS lesioned mice, such that early training action speeds more closely resembled later training speeds relative to the significantly increasing speeds found in Sham controls (Figure 4K, Table S1). Notably, this effect was not observed during the later acquisition of chunk 2 (press 3-5), or the duration of a successful sequence as a whole (Figure S6F and G). Overall, our results showed that M1→DLS interruption interfered with the successful adaption of action sequences to an increasing response requirement by reducing the optimal range of action speed and variation.

Discussion

Leveraged adaptation of mature action sequences

A useful behavioural strategy for efficiently exploiting environmental contingencies often involves enacting accurately learned streams of swiftly executed actions. In fluctuating environments, however, sufficient variation in these streams of action must be generated to explore newly adaptive solutions, a process described as 'adaptive learning through variation and selection' (Burtsev, 2012) in which, contrary to the decrease in variability typically observed with skill mastery, increased behavioural noise promotes adaptive learning (Sternad, 2018). Our study found evidence, amongst the varied successful responses, for a balance between effective exploration and efficient exploitation of instrumental action during a shift in contingency rules. As training progressed, the proportion of successful sequences in the action stream improved, and the probability of producing the target sequence (at the exact required ratio) became greater than the probability of producing overextended sequences (overshooting the imposed ratio), demonstrating an impressive ability to accurately explore

and exploit the newly adaptive forms of action in mice. Similarly, the speed-based efficiency of actions that achieved reward also clearly improved with training, although this was somewhat restricted, such that the speed of the first chunk was the only structural segment to ever be compressed. Any new addition to this scaffold remained invariant, including the "joint" segment that melded the first scaffold sequence with *de novo* chunks. In our self-paced sequence task, adaptive responding was strictly dependent on an increasing fixed ratio schedule and not on any action timing demands, yet the consistency of action timing seen in later schedules indicates that skilled action timing is usefully maintained when transferred to updated contingencies. The high degree of internal cohesion (Mechner, 1958a), both within and between the chunked units of action in a sequence, also suggests that the integration of chunks into whole sequences occurs smoothly, without disruption to the consistent timing of multiple consecutive actions. Overall, our study demonstrates that, even if high levels of performance have been achieved, skilled action remains subject to controlled levels of variability, a phenomenon that appears critical for weighing the efficient exploitation of old resources against sufficient levels of exploration for adapting to newly emerged contingencies.

Anatomy and motoric function of hyperdirect pathway accessory collaterals

This study also uncovered new collaterals of the canonical hyperdirect pathway ramifying into different regions of the basal ganglia, particularly the lateral territories of the striatum and the GPe. We ensured the truly hyperdirect nature of these projections by implementing a dual-viral tracing approach by which M1 cortical neurons projecting to the STN were first labelled in isolation, and then their synaptic terminals throughout the basal ganglia were subsequently mapped. This method identified hyperdirect pathway collaterals to both the DLS and the GPe, with the highest synaptic volumes being in the DLS. While this study identified novel hyperdirect pathway collaterals to the basal ganglia, extending previous research (Karube et al., 2019; Kita & Kita, 2012), our method could not distinguish independent M1→STN projections with collaterals to DLS or GPe from collective M1→STN projections with collaterals to both. New trans-synaptic tracing technology able to disentangle this circuitry is forthcoming (Li et al., 2019). Nonetheless, the anatomical finding of shared connectivity suggests that the DLS and GPe may be responsible for processing some of the same information that has typically been attributed to the hyperdirect pathway's downstream links to the STN. It is possible, however, that the local integration of this timing information in downstream structures differs. For example, the local inter-cellular interactions between D2-SPNs and D1-SPNs within the striatum (Matamales et al., 2020) and between prototypic and arkypallidal cells within the GPe (Aristieta et al., 2021) have been implicated in adaptive learning and locomotion functions, respectively. The importance of these "upstream" targets of the accessory hyperdirect pathway projections align with similar descriptions of basal ganglia function described in the *center-surround model* (Nambu et al., 2002), in which information is processed in a feedback-loop that differentially recruits direct, indirect or hyperdirect pathways traversing "forward" and "backward" through the basal ganglia to control motor performance. Similarly, in rats, *race models* of basal ganglia-driven behavioural response inhibition describe competition between Go, Stop and Pause signals emerging from striatum, GPe and STN respectively, from which the timing of each competing signal is integral to the eventual behavioural output (Logan & Cowan, 1984; Mallet et al., 2016; Schmidt & Berke, 2017). Presently, using models of the basal ganglia to predict the impact of multiple hyperdirect inputs within the circuity is speculative, and requires further experimentation. For example, simultaneous *in vivo* recordings could demonstrate the temporal relationships of downstream firing that occurs in response to excitation from a hyper-

direct input ubiquitous to DLS, GPe and STN regions—further informing interpretations of *race models*. Just how the supply of shared motor cortical information to multiple basal ganglia nuclei governs function both locally and at the circuit level remains to be established; nevertheless, such broad projections suggest a widespread and coordinated integration of motor cortical information, a process that may be essential for adapting streams of behaviour throughout learning.

Accessory hyperdirect pathway circuitry and the injection of temporal variability to action

Based on the transfer and consistency of action timing across fixed ratio schedules observed in our initial behavioural study, and considering the anatomical evidence and arguments supporting meaningful functional interactions between the hyperdirect pathway and upstream basal ganglia centres, we propose that the upstream hyperdirect pathway destabilises the temporal boundaries of past sequence durations by allocating the minimum level of variability required for explorative performance. The primary evidence for these claims comes from the behavioural effects induced by specific ablation of the M1→DLS circuit. We found that this selective ablation did not alter the number of actions within successful or unsuccessful sequences *per se* but instead induced briefer sequences with less varied durations that were ultimately less effective at earning reward (i.e., a slower reward rate and a greater proportion of unsuccessful actions). The effect of M1-driven DLS lesions on sequence duration was particularly pronounced in the first action chunk (presses 1-3), in which the brevity of sequence duration relative to controls was greatest in the initial exploratory phases of sequence acquisition, but diminished as the more successful sham control animals sped-up their actions during training. The brevity and consistently reduced variability of sequence durations following the M1→DLS interruption, and the waning yet ongoing M1→DLS input observed in other studies of motor skill behaviour (Kupferschmidt et al., 2017), suggests that this circuitry modulates action sequence learning via its contribution to the degree of encoded variability (Neuringer, 2002) and its capacity to extend action sequence duration. Although in a different context, some prior evidence implicates the hyperdirect pathway (or at least its target in the STN) in similar forms of temporal extension to optimise behaviour. One such example is the proposed 'hold your horses' effect, which provides the STN with the capacity to "buy some time" when deliberating over difficult choices (Baunez et al., 2007; Baunez & Robbins, 1997; Frank, 2006; Frank et al., 2007). Here, in a similar way, hyperdirect accessory circuitry is well placed to act as an on-line "noise injector", providing the extra action time that is required for exploratory sequences to enhance their likelihood of success; perhaps directly through M1→DLS or through broader bottom-up subthalamo-cortico-basal ganglia processing. In any case, a greater range of possible action sequence durations, within which more varied action timing could fall, might be expected to facilitate the exploration process and the subsequent learning of the target sequence necessary for its exploitation—a kind of dual time-/variability-based contribution to the exploration-exploitation trade-off of action.

References

Adam, E. M., Johns, T., & Sur, M. (2020). Dynamic control of visually-guided locomotion through cortico-subthalamic projections. BioRxiv, 2020.02.05.936443.

Aristieta, A., Barresi, M., Azizpour Lindi, S., Barrière, G., Courtand, G., de la Crompe, B., Guilhemsang, L., Gauthier, S., Fioramonti, S., Baufreton, J., & Mallet, N. P. (2021). A Disynaptic Circuit in the Globus Pallidus Controls Locomotion Inhibition. Current Biology, 31(4), 707-721.e7.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and Subcortical Contributions to Stop Signal Response

Inhibition: Role of the Subthalamic Nucleus. Journal of Neuroscience, 26(9), 2424–2433.

Balleine, B. W. (2019). The Meaning of Behaviour: Discriminating Reflex and Volition in the Brain. Neuron, 104(1), 47–62.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. Journal of Neuroscience, 27(31), 8161–8165.

Balleine, B. W., & O'Doherty, J. P. (2010). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology, 35(1), 48.

Baunez, C., Christakou, A., Chudasama, Y., Forni, C., & Robbins, T. W. (2007). Bilateral high-frequency stimulation of the subthalamic nucleus on attentional performance: Transient deleterious effects and enhanced motivation in both intact and parkinsonian rats. European Journal of Neuroscience, 25(4), 1187–1194.

Baunez, C., & Robbins, T. W. (1997). Bilateral Lesions of the Subthalamic Nucleus Induce Multiple Deficits in an Attentional Task in Rats. European Journal of Neuroscience, 9(10), 2086–2099.

Burtsev, M. S. (2012). Adaptive Learning Through Variation and Selection. In N. M. Seel (Ed.), Encyclopedia of the Sciences of Learning (pp. 116–118). Springer US.

Calabresi, P., Mercuri, N. B., De Murtas, M., & Bernardi, G. (1990). Endogenous GABA mediates presynaptic inhibition of spontaneous and evoked excitatory synaptic potentials in the rat neostriatum. Neuroscience Letters, 118(1), 99–102.

Dhawale, A. K., Wolff, S. B. E., Ko, R., & Ölveczky, B. P. (2021). The basal ganglia control the detailed kinematics of learned motor skills. Nature Neuroscience, 1–14.

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-Signal Reaction-Time Task Performance: Role of Prefrontal Cortex and Subthalamic Nucleus. Cerebral Cortex, 18(1), 178–188.

Fife, K. H., Gutierrez-Reed, N. A., Zell, V., Bailly, J., Lewis, C. M., Aron, A. R., & Hnasko, T. S. (2017). Causal role for the subthalamic nucleus in interrupting behaviour. ELife, 6, e27689.

Fisher, S. D., Ferguson, L. A., Bertran-Gonzalez, J., & Balleine, B. W. (2020). Amygdala-Cortical Control of Striatal Plasticity Drives the Acquisition of Goal-Directed Action. Current Biology, 30(22), 4541-4546.e5.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. Neural Networks, 19(8), 1120–1136.

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: Impulsivity, deep brain stimulation, and medication in parkinsonism. Science, 318(5854), 1309–1312.

Gradinaru, V., Zhang, F., Ramakrishnan, C., Mattis, J., Prakash, R., Diester, I., Goshen, I., Thompson, K. R., & Deisseroth, K. (2010). Molecular and Cellular Approaches for Diversifying and Extending Optogenetics. Cell, 141(1), 154–165.

Graybiel, A. M. (1998). The Basal Ganglia and Chunking of Action Repertoires. Neurobiology of Learning and Memory, 70(1), 119–136.

Jin, D. Z., Fujii, N., & Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. Proceedings of the National Academy of Sciences, 106(45), 19156–19161.

Jin, X., & Costa, R. M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature, 466(7305), 457–462.

Jin, X., Tecuapetla, F., & Costa, R. M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nat Neurosci, 17(3), 423–430.

Karube, F., Takahashi, S., Kobayashi, K., & Fujiyama, F. (2019). Motor cortex can directly drive the globus pallidus neurons in a projection neuron type-dependent manner in the rat. ELife, 8, e49511.

Kita, T., & Kita, H. (2012). The Subthalamic Nucleus Is One of Multiple Innervation Sites for Long-Range

Corticofugal Axons: A Single-Axon Tracing Study in the Rat. Journal of Neuroscience, 32(17), 5990–5999.

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(2), 476-489.e5.

Lashley, K. S. (1951). The problem of serial order in behavior. In Cerebral mechanisms in behavior; the Hixon Symposium (pp. 112–146). Wiley.

Li, J., Liu, T., Dong, Y., Kondoh, K., & Lu, Z. (2019). Trans-synaptic Neural Circuit-Tracing with Neurotropic Viruses. Neuroscience Bulletin, 35(5), 909–920.

Light, K. R., Cotten, B., Malekan, T., Dewil, S., Bailey, M. R., Gallistel, C. R., & Balsam, P. D. (2019). Evidence for a Mixed Timing and Counting Strategy in Mice Performing a Mechner Counting Task. Frontiers in Behavioural Neuroscience, 13.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. Psychological Review, 91(3), 295–327.

Lovinger, D. M. (1991). Trans-1-aminocyclopentane-1,3-dicarboxylic acid (t-ACPD) decreases synaptic excitation in rat striatal slices through a presynaptic action. Neuroscience Letters, 129(1), 17–21.

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., & Berke, J. D. (2016). Arkypallidal Cells Send a Stop Signal to Striatum. Neuron, 89(2), 308–316.

Matamales, M., McGovern, A. E., Mi, J. D., Mazzone, S. B., Balleine, B. W., & Bertran-Gonzalez, J. (2020). Local D2- to D1-neuron transmodulation updates goal-directed learning in the striatum. Science, 367(6477), 549–555.

Matamales, M., Skrbis, Z., Bailey, M. R., Balsam, P. D., Balleine, B. W., Götz, J., & Bertran-Gonzalez, J. (2017). A corticostriatal deficit promotes temporal distortion of automatic action in ageing. ELife, 6, e29908.

Mechner, F. (1958a). Probability Relations Within Response Sequences Under Ratio Reinforcement1. Journal of the Experimental Analysis of Behaviur, 1(2), 109–121.

Mechner, F. (1958b). Probability Relations within Response Sequences under Ratio Reinforcement. Journal of the Experimental Analysis of Behavior, 1(2), 109–121.

Mello, G. B. M., Soares, S., & Paton, J. J. (2015). A Scalable Population Code for Time in the Striatum. Current Biology, 25(9), 1113–1122.

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. Neuroscience Research, 43(2), 111–117.

Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. Psychonomic Bulletin & Review, 9(4), 672–705.

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., Mortrud, M. T., Ouellette, B., Nguyen, T. N., Sorensen, S. A., Slaughterbeck, C. R., Wakeman, W., Li, Y., Feng, D., Ho, A., … Zeng, H. (2014). A mesoscale connectome of the mouse brain. Nature, 508(7495), 207–214.

Ostlund, S. B., & Balleine, B. W. (2005). Lesions of Medial Prefrontal Cortex Disrupt the Acquisition But Not the Expression of Goal-Directed Learning. Journal of Neuroscience, 25(34), 7763–7770.

Paxinos, G., & Franklin, K. B. J. (2007). The Mouse Brain in Stereotaxic Coordinates (3rd edition). Academic Press.

Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. Journal of Experimental Psychology: Human Perception and Performance, 9(1), 86–102.

Saling, L. L., & Phillips, J. G. (2007). Automatic behaviour: Efficient not mindless. Brain Research Bulletin, 73(1), 1–20.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. Nature Methods, 9(7), 676–682.

Schmidt, R., & Berke, J. D. (2017). A Pause-then-Cancel model of stopping: Evidence from basal ganglia neurophysiology. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1718), 20160202.

Schmidt, R., Leventhal, D. K., Mallet, N., Chen, F., & Berke, J. D. (2013). Canceling actions involves a race between basal ganglia pathways. Nature Neuroscience, 16(8), 1118–1124.

Sternad, D. (2018). It's not (only) the mean that matters: Variability, noise and exploration in skill learning. Current Opinion in Behavioral Sciences, 20, 183–195.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). 6 - The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting. In G. E. Stelmach (Ed.), Information Processing in Motor Control and Learning (pp. 117–152). Academic Press.

Tecuapetla, F., Jin, X., Lima, S. Q., & Costa, R. M. (2016). Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. Cell, 166(3), 703–715.

Terrace, H. S. (1987). Chunking by a pigeon in a serial learning task. Nature, 325(7000), 149–151.

Wessel, J. R., & Aron, A. R. (2017). On the Globality of Motor Suppression: Unexpected Events and Their Influence on Behavior and Cognition. Neuron, 93(2), 259–280.

Yang, C. F., Chiang, M. C., Gray, D. C., Prabhakaran, M., Alvarado, M., Juntti, S. A., Unger, E. K., Wells, J. A., & Shah, N. M. (2013). Sexually Dimorphic Neurons in the Ventromedial Hypothalamus Govern Mating in Both Sexes and Aggression in Males. Cell, 153(4), 896–909.

Yin, H. H. (2010). The Sensorimotor Striatum Is Necessary for Serial Order Learning. Journal of Neuroscience, 30(44), 14719–14723.

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(1), 181–189.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. Behavioural Brain Research, 166(2), 189–196.

Zhang, S., Xu, M., Chang, W.-C., Ma, C., Do, J. P. H., Jeong, D., Lei, T., Fan, J. L., & Dan, Y. (2016). Organization of long-range inputs and outputs of frontal cortex for top-down control. Nature Neuroscience, 19(12), 1733–1742.