Behavioral/Cognitive

Dorsolateral Striatal Task-initiation Bursts Represent Past Experiences More than Future Action Plans

Paul J. Cunningham,[1] **Paul S. Regier,**[2] **and A. David Redish**[1]

[1]Department of Neuroscience, University of Minnesota, Minneapolis MN 55455, and [2]Department of Psychiatry, University of Pennsylvania Philadelphia PA 19104

The dorsolateral striatum (DLS) is involved in learning and executing procedural actions. Cell ensembles in the DLS, but not the dorsomedial striatum (DMS), exhibit a burst of firing at the start of a well-learned action sequence ("task-bracketing"). However, it is currently unclear what information is contained in these bursts. Some theories suggest that these bursts should represent the procedural action sequence itself (that they should be about future action chains), whereas others suggest that they should contain representations of the current state of the world, taking into account primarily past information. In addition, the DLS local field potential shows transient bursts of power in the 50 Hz range (γ50) around the time a learned action sequence is initiated. However, it is currently unknown how bursts of activity in DLS cell ensembles and bursts of γ50 power in the DLS local field potential are related to each other. We found that DLS bursts at lap initiation in rats represented recently experienced reward locations more than future procedural actions, indicating that task-initiation DLS bursts contain primarily retrospective, rather than prospective, information to guide procedural actions. Furthermore, representations of past reward locations increased during periods of increased γ50 power in the DLS. There was no evidence of task-initiation bursts, increased γ50 power, or retrospective reward location information in the neighboring dorsomedial striatum. These data support a role for the DLS in model-free theories of procedural decision-making over planned action-chain theories, suggesting that procedural actions derive from representations of the current and recent past.

Key words: decision-making; habit; model-free; procedural learning; striatum; task bracketing

---

Significance Statement

While it is well-established that the dorsolateral striatum (DLS) plays a critical role in procedural decision-making, open questions remain about the kinds of representations contained in DLS ensemble activity that guide procedural actions. We found that DLS, but not DMS, cell ensembles contained nonlocal representations of past reward locations that appear moments before task-initiation DLS bursts. These retrospective representations were temporally linked to a rise in γ50 power that also preceded the characteristic DLS burst at task-initiation. These results support models of procedural decision-making based on associations between available actions and the current state of the world over models based on planning over action-chains.

Introduction

The dorsolateral striatum (DLS) plays a critical role in the development and maintenance of ballistically executed procedural actions. Bursts of activity in the DLS occur at the initiation and/or termination of well-learned action sequences and are believed to be a characteristic neurophysiological feature of procedural decision-making (Jog et al., 1999; Barnes et al., 2005; Jin and Costa, 2010; Smith and Graybiel, 2013; Jin et al., 2014; Regier et

al., 2015). Many theories suggest that procedural actions depend on Markovian representations of past events, such as previously experienced rewards, states, or actions, that allow the current situation to release a well-learned action sequence (Thorndike, 1932; Hull, 1942; Sutton and Barto, 1998; Daw et al., 2005). Other theories suggest that procedural actions depend on representations of a future action plan consisting of a chunked-together sequence of movements (Lashley, 1951; Bailey and Mair, 2006; Dezfouli and Balleine, 2012). This leads to the open question of what information is represented in bursts of DLS activity that occur when a procedural action is initiated: Do task-bracketed DLS bursts contain information about past and/or current events that release a well-learned action sequence, or do they contain information about future action plans, movement sequences, or expected terminal behavioral states?

In addition, there is evidence that transient oscillatory signals centered at ∼50 Hz appear in the DLS local field potential (LFP)

immediately preceding initiation of a rewarded action sequence (termed $\gamma50$) (Masimore et al., 2005). While both task-bracketed DLS bursts and $\gamma50$ events are most prominent around the time at which a learned action sequence is initiated, their temporal relation has never been directly explored. This leads to the open question of the degree to which DLS ensemble bursts and $\gamma50$ oscillations are related: What is the precise temporal relationship, if any, between task-bracketed DLS bursts and $\gamma50$ events on initiation of a procedural action sequence?

We addressed these questions by exploring the relation between $\gamma50$ oscillations in the DLS LFP and nonlocal representations of behaviorally meaningful events (i.e., representations of spatial locations on the maze when the rat is not at that location) around the time at which rats initiated a well-learned action sequence and DLS ensembles exhibited their characteristic firing burst. Our focus was therefore on the information contained within DLS ensembles specifically around the time they exhibited their "task-bracketed" burst of activity. We found that $\gamma50$ power rose and peaked just before the DLS burst that occurred around the time a learned action sequence was initiated. Much like the DLS burst, the rise in $\gamma50$ power developed with experience. Importantly, task-bracketed DLS bursts at movement initiation contained representations of recently visited reward sites (i.e., past-oriented, retrospective information), providing support for the model-free, Markovian hypothesis over the goal-directed, chunking hypothesis of procedural action sequences. There was no evidence of ensemble bursts of activity, experience-dependent changes in $\gamma50$ power, nor retrospective reward location representations in the neighboring but functionally distinct dorsomedial striatum (DMS). These results suggest that the relation between ensemble firing bursts at lap initiation, $\gamma50$ events, and retrospective reward location representations was likely a feature of procedural decision-making systems implemented within the DLS but not the DMS.

Materials and Methods

This study reanalyzed previously collected data, some of which were reported by Regier et al. (2015). Additional details regarding, subjects, surgery, and behavioral task can be found in Regier et al. (2015). All behavioral and neurophysiological analyses reported in this paper are novel.

Subjects and surgery

Subjects. Six male Fisher Brown Norway and Brown Norway (5× FBNF-1, 1× BN) rats (aged 9-16 months at time of experiment) served as subjects. All procedures were conducted in accordance with the National Institutes of Health guidelines for animal care and approved by the Institutional Animal Care and Use Committee at the University of Minnesota. Rats were housed individually and maintained on a 12 h light-dark cycle. Rats were provided with *ad libitum* water in their home cages throughout the experiment and received their daily food consumption within each session (Regier et al., 2015).
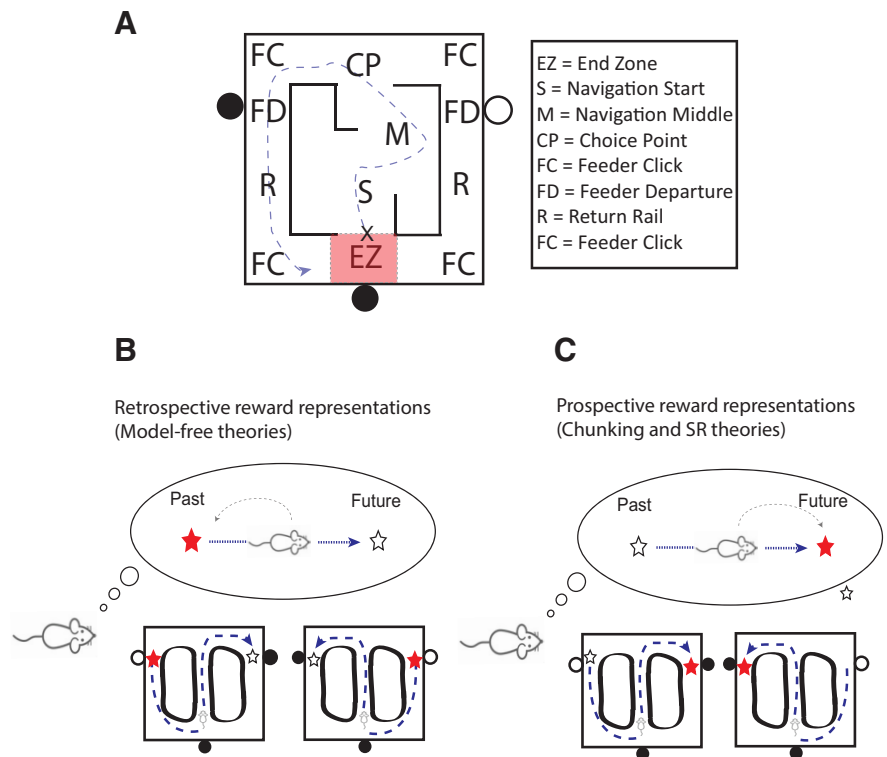
Figure 1. Hypotheses about the relation between ensemble activity bursts and representations in the DLS. *A*, Depiction of the Hebb-Williams maze used in the present study. Cross represents a hypothetic starting point of a rat's trajectory. Blue line indicates a hypothetical trace of a rat's trajectory through the maze to earn food (in this case, by turning left at the CP). Our focus was on $\gamma50$ oscillations and DLS cell ensemble representations while rats were in the EZ of the maze (red square). *B*, Model-free theories of procedural learning imply retrospective information should guide action selection, predicting representations of past rewards (or trajectories) around the time rats initiate a well-learned action sequence. Red stars represent feeder representations. Filled circles represent the correct feeder location for the upcoming lap. *C*, Chunking and successor-representation theories of procedural learning imply that prospective information should guide action selection, predicting representations of future trajectories around the time rats initiate a well-learned action sequence.

Surgery and striatum localization. After pretraining on the behavioral task, rats were implanted with unilateral 14-tetrode hyperdrives ($n=3$) or bilateral 28-tetrode hyperdrives ($n=3$) in the anterior DLS and posterior DMS. Following surgery, tetrodes were advanced 40-640 µm per day until reaching the dorsal striatum, as identified by the presence of medium spiny neurons (MSNs; i.e., relatively long interspike intervals [ISIs] and short firing bursts). Additional details regarding the surgical process and craniotomies can be found in Regier et al. (2015).

Histology. Once the experiment was complete, the location of each tetrode was identified by administering a 5 µA current for 10 s. Tetrode location was then confirmed histologically using coronal slices stained with cresyl violet (for more details, see Regier et al., 2015).

Experimental design and statistical analysis

Behavioral task. Additional details of the task can be found in Regier et al. (2015). Briefly, 6 rats traversed a Hebb-Williams maze (see Fig. 1A) to earn their daily food ration during 30 min sessions. One of three reward contingencies were presented to rats. The left contingency required rats to make repeated left turns to earn food. The right contingency required rats to make repeated right turns. The alternation contingency required rats to alternate between left and right turns across successive laps. Rats received 2 food pellets at the side feeder (indicated by circles next to feeder departure locations in Fig. 1A) and at the center feeder (indicated by a circle next to the end zone [EZ] location in Fig. 1A) for making the correct turn direction at the choice point (CP; see Fig. 1A). Rats did not receive food for incorrect turns. Approximately halfway through the session, the reward contingency changed. For example, the left contingency might be in effect for the first half of a
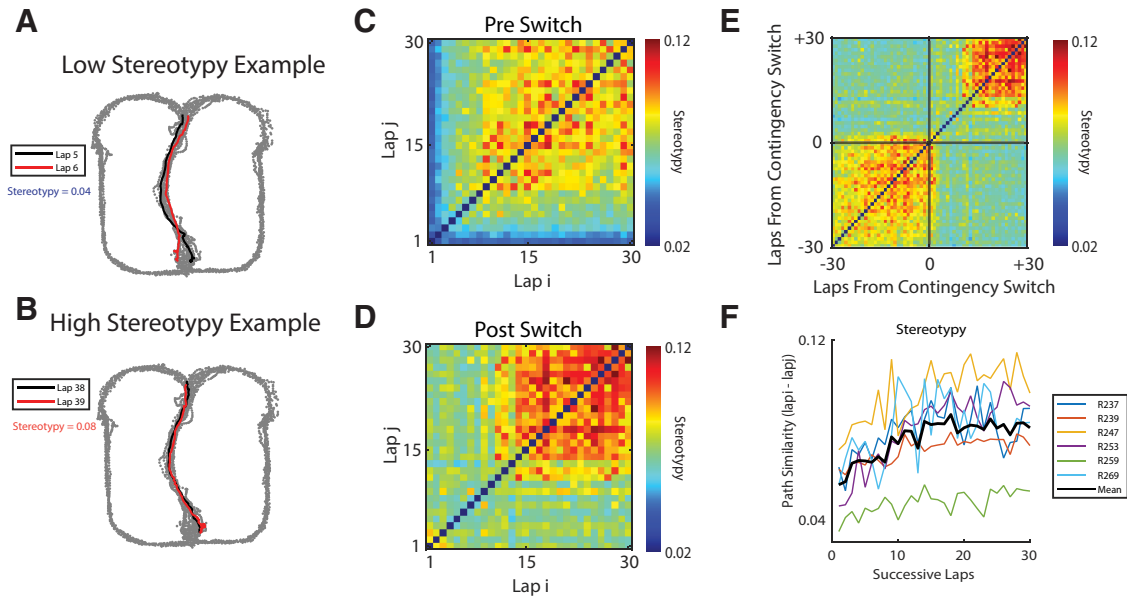
Figure 2. Action sequences became stereotyped across laps. **A**, An example illustrating relatively low similarity between two lap-adjacent path trajectories through the navigation sequence. There is relatively little overlap in path trajectories between the fifth and sixth laps. **B**, An example illustrating relatively high similarity between two lap-adjacent trajectories. Paths through the navigation sequence for the 38th and 39th laps were relatively similar. **C-E**, Mean lap$_i$-lap$_j$ comparison of path trajectories during the first 30 laps of the pre-switch (**C**) and post-switch (**D**) periods and aligned to the contingency switch (**E**). The highly consistent path trajectory between adjacent laps following the contingency switch was distinct from what it was before the switch. Nevertheless, during both pre- and post-switch periods, trajectories between adjacent laps became more similar to each other as laps progressed ($n = 32$ sessions across rats). **F**, Stereotypy across successive laps for each rat (colored lines) and averaged across rats (black line).

session whereas the right contingency would be in effect for the second half. Rats could freely navigate the maze throughout the session, and there were no explicit signals or cues indicating the start and end of the current lap. Rats were first trained to turn either left or right at the CP using a barrier that prevented them from turning the wrong direction. Next, rats were trained with only one contingency in effect throughout the entire session. Finally, rats were given sessions in which the contingency changed approximately halfway through the session with simultaneous DLS and DMS recordings.

A total of 35 recording sessions were obtained. Three of these recording sessions from 1 rat (R269) were excluded from analysis because for each session there were laps for which center feeder fires were not recorded and timestamped (leaving 32 sessions for analysis). For LFP analyses, all data from R269 were excluded because of noise in the LFP signal (leaving 29 sessions for analysis).

Quantification and statistical analysis

Movement velocity. The rat's location on the maze was monitored at 60 Hz by an overhead camera that detected an LED located on the rat's head stage. Velocity was computed using the Janabi-Sharifi et al. (2000) discrete-time adaptive windowing method, which estimates momentary speed (dx, dy) based on the rat's position at time t and the rat's subsequent positions at $t + 1$ s. This algorithm is well suited to identify the moment at which velocity changes and can therefore be used to estimate the time at which an action sequence was initiated.

Behavioral stereotypy. We quantified stereotypy of path trajectories through the navigation sequence of the maze (identified in Fig. 1A as navigation start [S] and navigation middle [M]) by comparing path trajectory on lap i to lap j. For each lap, we obtained the position of the rat along the navigation sequence of the maze and interpolated these positions using 1000 samples to measure path trajectory for that lap. For lap-by-lap stereotypy comparisons (see Fig. 2C–F), we found the reciprocal of the difference in path trajectory between lap i and lap j for each session (1) from the first to 30th lap during pre- and post-switch periods, and (2) during the 60 laps surrounding the contingency switch. Path stereotypy (i.e., the reciprocal of the difference in path trajectory between lap$_i$ and lap$_j$) was averaged across sessions ($n = 32$; sessions over all rats). To assess experience-dependent changes in stereotypy across rats, we

found the slope describing the change in lap-adjacent path stereotypy across the first 30 laps during pre- and post-switch periods using a linear regression. A paired-samples t test was used to determine whether the mean slope across rats was significantly >0, indicating an increase in lap-adjacent path stereotypy across laps ($n = 6$; rats).

Bursts of ensemble activity in the EZ. The maze was broken up into eight locations (depicted in Fig. 1A). Firing rates during the ± 1 s period (500 ms bins) around each maze location were determined for each cell on each lap. Firing rates for each maze location were normalized relative to mean firing rate across the maze for that cell (following analyses in Barnes et al., 2005; Thorn et al., 2010; Smith and Graybiel, 2013). Specifically, for each cell, we found the mean firing rate across all maze locations throughout the session. We then normalized firing rates in each bin corresponding to a given maze location (four bins per location) relative to the cell's overall mean firing rate for each lap: $Z_{bin\ i} = (S_{bin\ i} - S_{mean})/S_{SD}$ (Thorn et al., 2010). The result was a laps × binned-zones × cells matrix of normalized firing rate, which served as the basis for characterizing DLS and DMS firing patterns across maze locations. We then found for each lap the time at which mean firing rate was greatest while rats were in the EZ of the maze for both DLS and DMS cell ensembles. An "activity burst" was defined as the time at which the normalized firing rate of DLS or DMS cell ensembles were highest while rats were in the EZ of the maze for each lap.

To assess the reliability of between-region differences in task-bracketing across rats, we compared mean "task-bracketing scores" between DLS and DMS ensembles for each rat. Task-bracketing scores were found by subtracting normalized firing rate in the navigation sequence and CP of the maze from normalized firing rate in the EZ of the maze (following analyses in Barnes et al., 2005; Thorn et al., 2010; Thorn and Graybiel, 2014; Smith and Graybiel, 2013). Values >0 indicate relatively strong firing in the EZ of the maze compared with the subsequent navigation sequence. A paired-samples t test ($n = 6$; rats) was used (1) to compare mean task-bracketing scores between DLS and DMS across rats and (2) to determine whether mean task-bracketing scores in DLS or DMS were significantly >0 across rats. In addition, a Wilcoxon sign rank test with a Bonferroni correction for multiple comparisons ($n = 1061$; total laps) was used to compare the magnitude of activity bursts between DLS and DMS ensembles across laps, and a paired-samples t test ($n = 6$; rats) was used to compare mean burst size in DLS and

DMS across rats. Wilcoxon sign rank tests were used for this and subsequent comparisons between DLS and DMS because it allowed for matched-sample comparisons of simultaneously recorded neural activity without making assumptions about the nature of the distribution of any given neurophysiological measure (e.g., ensemble burst size).

γ50 power at neural ensemble activity bursts. We used a short-time Fourier transform to assess spectral power across the 1-100 Hz range of the LFP. For each tetrode in DLS and DMS, we obtained a spectrogram of power across the entire session, using 0.5 s windows with 50% overlap. For each lap within a session, we found mean (averaged across tetrodes) spectral power in the 1-100 Hz range during the -5 to 2 s period (100 ms bins) around ensemble activity bursts in the EZ. Power for each frequency was normalized relative to mean power at that frequency across the session. To assess experience dependent changes in $γ50$ power, we averaged normalized power in the 45-55 Hz range ($γ50$) across 5 lap bins within each session and then averaged normalized power in the first to fifth 5 lap bin across sessions ($n = 29$ sessions over all rats). To assess reliability of experience-dependent changes in $γ50$ power in DLS compared with DMS across rats, we found the slope describing the change in $γ50$ power across 5 lap bins in both DLS and DMS LFPs using a linear regression. $γ50$ power during each 5 lap bin was averaged across the 1 s preceding ensemble burst times. A paired-samples t test ($n = 5$; rats) was used to assess whether the slope relating $γ50$ power to lap bins was different between DLS and DMS, and whether the slopes for either region were significantly >0.

Cell categorization. Striatal cell-type categorization was based on methods described by Schmitzer-Torbert and Redish (2008). Each cell was determined to be phasic firing or nonphasic firing based on the proportion of time spent in relatively long ISIs. For each cell, we found all ISIs which exceeded 2 s and summed these ISIs across the entire session. This sum was then divided by total session time. If the proportion of time spent in ISIs longer than 2 s was >0.4, then the cell was classified as phasic firing (i.e., if a cell spent $>40\%$ of the session in ISIs > 2 s). Phasic firing cells are believed to be MSNs, the principal cell type within the dorsal striatum. If this proportion was <0.4, then the cell was classified as nonphasic firing. We further classified nonphasic firing cells into either high firing or tonic firing cells based on their post-spike suppression index. The post-spike suppression index quantifies the time it takes for a cell to return to its mean firing rate following a spike. Cells with a post-spike suppression index longer than 100 ms were classified as tonic firing, whereas cells with a post-spike suppression shorter than 100 ms were classified as high firing. The distinction between high firing and tonic firing cells was based on previous reports indicating bimodal distributions of post-spike suppression indices in nonphasic firing DLS cells (Schmitzer-Torbert and Redish, 2004, 2008). TFNs are believed to be tonically active cholinergic interneurons that have a relatively long afterhyperpolarization compared with HFNs, which are putative high firing interneurons (Kawaguchi, 1993; Schmitzer-Torbert and Redish, 2008; Thorn and Graybiel, 2014).

Because datasets needed to include both clean LFP and cellular activity, not all cells from Regier et al. (2015) were included in this analysis. Of the cells used in this analysis, $\sim$84% of DLS cells were classified as phasic firing while the remaining 16% were classified as high firing. For DMS, $\sim$67% of cells were phasic firing while the remaining 33% were high firing. We only found one tonic firing cell in the DMS, which was excluded from subsequent analyses.

Entrainment and pairwise phase consistency (PPC) analysis. We used PPC (Vinck et al., 2010, 2012) to quantify entrainment of cell activity to oscillations in the LFP. The phase ($φ$) of a given oscillation at the time of every spike from each cell was represented as a vector: $U_k = ($cos $(φ_k),$sin$(φ_k))$, where k refers to the number of spikes from a given cell. We then computed the dot product of U_k and U_j for each possible spike pair (k,j) for each cell (Vinck et al., 2012). The resultant value ranges from -1 to 1 and reflects the coincidence of $φ$ between each spike pair. $φ$ at spike times becomes maximally inconsistent as PPC approaches -1 and becomes maximally consistent as PPC approaches 1. Thus, for each cell, we obtained a single PPC score indicating mean consistency of $φ$ at pairwise spike times for that cell. We then obtained a distribution of PPC scores at randomly selected points in time (k = # of spikes = # of

random times) across 100 shuffles. Mean PPC for each cell was normalized relative to the distribution of PPC scores at randomly selected points in time. A cell was considered to be significantly entrained if its mean PPC at spike times was >2 SDs above the mean of the distribution of PPC from randomly selected points in time. A $χ^2$ test was used to compare distributions of significantly entrained cells between DLS and DMS for each frequency analyzed.

Finally, we assessed preferential spiking activity of cells by computing a tuning curve of spike count as a function of phase ($φ$) for a given oscillation. $φ$ was broke up into 16 bins, and mean spike count for each bin was obtained for each cell. Tuning curves were averaged across cells for each session. Tuning curves for each session were normalized by taking the ratio of spike counts in each bin to mean spike counts across all bins for that session. This normalization allowed for easier visualization and comparison of phasic firing and high firing cells and does not affect the patterns of activity as a function of $φ$. A Rayleigh test was used to assess nonuniformity in normalized firing rate across $φ$. The resultant vector length of obtained data was compared with a distribution of resultant vector lengths from shuffled spike-phase relations (500 shuffles) to assess whether normalized firing rate across $φ$ was not uniform for a given frequency.

Bayesian decoding. We used a spatial Bayesian decoding algorithm that estimates the animal's position on the maze at time, t, given ensemble spiking activity at that time (Zhang et al., 1998). This Bayesian algorithm leverages spatial tuning curves and overall spike rates across the session for each cell to construct posterior probability distributions over maze locations at any given time (i.e., the probability that the rat, at time t, is located at a given position in the maze given spatial tuning curves and overall spike rates). Both x and y positions of the maze were broken up into 64 bins, and for each cell we obtained a tuning curve of firing rate across the maze. In addition, we obtained firing rate during 100 ms bins across the entire session for each cell. We applied this decoding algorithm, assuming a uniform spatial prior, to cell ensemble activity while rats were in the EZ of the maze during the window covering -5 s to $+2$ s (100 ms bins) around the time of DLS/DMS bursts for each lap. Rats typically remained in the EZ throughout this window for the majority of laps. However, to ensure our reward location decoding estimates were nonlocal, we excluded from analysis any times for which the rat was not in the EZ. In this way, we use the term "nonlocal" to refer to representations of locations on the maze where the rat was not currently located (i.e., any maze location outside of the EZ).

Because reward locations were the only nonlocal representations observed while rats were in the EZ, we focused exclusively on nonlocal reward location representations for subsequent analyses. We used occupancy (i.e., time spent at any given point in the maze) to determine the precise locations where left and right rewards were delivered on the maze and averaged decoding probability across spatial bins within those localized areas to quantify reward location representations (termed "pReward"). Reward locations were determined separately for each session to ensure an individualized, accurate identification of reward location.

Reward location representations were collapsed across turn sequences (i.e., LL, LR, RL, RR) and contingencies (i.e., left, right, alternate) to specify "past" and "future" reward locations. Past representations were defined by pLeftReward for LL laps for the left contingency and LR laps for the alternate contingency, combined with pRightReward for RR laps for the right contingency and RL laps for the alternate contingency. For future representations, we used pLeftReward for LL (Left) and RL (Alternate) laps combined with pRightReward for RR (Right) and LR (Alternate) laps. Thus, past representations reflect the just-visited reward site, whereas future representations reflect the about-to-be-visited reward site (regardless of left vs right location). To measure retrospective bias, we found mean pReward representations during the period of maximum $γ50$ power (see section below for how the $γ50$ epoch was defined) for each lap. [LFP data for R269 were noisy and were therefore not included in LFP analyses. For this reason, we could not use $γ50$ epochs to average pReward representations. Thus, For R269, we found mean pReward in the 2 s prior to each DLS ensemble burst to calculate retrospective bias. Our main results regarding nonlocal reward location

representations remain when R269 was excluded from analyses.] We then took the log10 (pRewardPast/pRewardFuture) to measure biases in past versus future reward location representations for each lap. Values of 0 indicate no bias in reward location representations. Values >0 indicate a retrospective bias, whereas values <0 indicate a prospective bias in reward location representations. A Wilcoxon rank-sign test with a Bonferroni correction for multiple comparisons ($n = 1061$; total laps) was used to determine whether there was a significant retrospective bias in reward location representations for either DLS or DMS ensembles and to compare the degree of retrospective bias between regions. In addition, for each rat, we found mean retrospective bias in reward location representations in DLS and DMS ensembles. A Wilcoxon sign rank test ($n = 6$; rats) was used to determine whether between-region differences in retrospective bias were consistent across rats.

Finally, we found the number of laps for which the bias measure indicated that past reward location representations were twice as strong as future reward location representations, which defined laps with a retrospective bias. We also found the number of laps for which the bias indicated that future reward location representations were twice as strong as past reward location representations, which defined laps with a prospective bias. We then compared the proportion of laps with a retrospective and prospective bias in DLS and DMS ensembles.

Reward location representations during γ50 events. To assess the relation between reward location representations in DLS cell ensembles and γ50 oscillations in the LFP, we found the temporal epoch during which γ50 power was highest. To this end, for each lap, we found the time of the first and last burst of γ50 power while rats were in the EZ of the maze. A γ50 burst was defined as any point in time for which γ50 power was at least 2 SDs greater than the session-wide mean. Thus, for each lap, we obtained times for the first and last burst of γ50 power and defined three adjacent temporal epochs corresponding to the period before, during, and after these two times. The "during" epoch for each lap was defined as the period of time between the first and last γ50 burst. The "before" epoch was defined as the period, of equal duration as the "during" period for that lap, before the first burst in γ50 power. The "after" epoch was defined as the period (of equal duration as "during") following the last γ50 burst. We compared past and future reward location representations between these three epochs to explore their relation to γ50 events at lap initiation. A Wilcoxon sign rank test with a Bonferroni correction for multiple comparisons ($n = 1061$; total laps) was used for each pairwise epoch comparison to assess significant differences in reward location representations across laps. In addition, for each rat, we found mean past/future reward location representations during γ50 events and subtracted from it past/future reward location representations averaged across before and after γ50 events (i.e., "non-γ50" epochs). This resulted in a single value for each rat indicating the degree to which past/future reward location representations increased during γ50 events, with values >0 indicating stronger representations during γ50 events. A paired-samples *t* test ($n = 5$; rats) was used to assess whether past/future reward location representations were strongest during γ50 across rats.

Results

Rats navigated a Hebb-Williams maze (Fig. 1A) to earn food rewards by either turning left, right, or alternating between left-right turns. The contingency for earning food (i.e., left, right, or alternate) was randomly selected at the start of each session and changed approximately halfway through the session. Rats quickly recognized the correct turn sequence within each session and adapted that turn sequence following the mid-session contingency switch (Regier et al., 2015). In addition, the proportion of DLS ensemble activity in the EZ (Fig. 1A) relative to the rest of the maze increased as rats adopted their behavioral strategy (Regier et al., 2015), replicating the typical DLS task-bracketing effect (Jog et al., 1999; Barnes et al., 2005; Jin and Costa, 2010; Thorn et al., 2010; Smith and Graybiel, 2013; Jin et al., 2014). Here we expand on these findings through novel analyses

examining DLS representations and γ50 oscillations around the time at which DLS ensembles exhibited their activity bursts in the EZ of the maze, which corresponded roughly to the point at which the previous lap ended and the next lap began. As a control, we applied all neurophysiological analyses to the DMS to assess whether task-bracketed ensemble bursts, nonlocal representations, and γ50 oscillations were a feature of the DLS specifically or the dorsal striatum in general.

Behavioral stereotypy and bursts of DLS ensemble activity in the maze EZ

Procedural decision-making systems are characterized in part by the consistent and reliable topography of the behavior they govern (i.e., stereotypy). Rats developed a stereotyped path trajectory through the navigation sequence as they adopted the appropriate behavioral strategy (Fig. 2A–F). Both the pre-switch and post-switch periods were characterized by an increase in similarity of path trajectories between adjacent laps (Fig. 2C,D). A paired-samples *t* test revealed that the mean slope describing the change in lap adjacent path similarity across laps for each rat was significantly >0 ($t_{(5)} = 4.58$, $p = 0.003$, $d = 1.87$), indicating an increase in path stereotypy across laps (Fig. 2F). This stereotyped path trajectory was disrupted by the contingency switch (Fig. 2E) and was accompanied by an increase in vicarious trial-and-error (Regier et al., 2015), a behavioral marker of deliberative decision-making (Muenzinger and Gentry, 1931; Tolman, 1948; Redish, 2016). Interestingly, once rats adopted their new behavioral strategy to the changed contingency, a new stereotyped path trajectory developed (Fig. 2E). That is, path trajectory for any given lap during the post-switch period was similar to lap-adjacent paths but markedly different from path trajectories during the pre-switch period. Thus, rats adopted stereotyped path trajectories during both pre-switch and post-switch periods, but the stereotyped paths differed between these periods. These results suggest that scent following likely did not contribute to the stereotyped path trajectory and are consistent with the notion that the observed action sequences were governed by a procedural decision-making system.

As animals automate a rewarded action sequence, the DLS develops a characteristic firing pattern in which cell ensemble firing rates highlight boundaries of the learned action sequence (Jog et al., 1999; Barnes et al., 2005; Jin and Costa, 2010; Thorn et al., 2010; Smith and Graybiel, 2013; Jin et al., 2014; Regier et al., 2015). We assessed mean normalized firing rate at each of eight zones in the maze (Fig. 1A) across laps to assess changes in DLS and DMS ensemble activity patterns as rats adopted the appropriate behavioral strategy to the pre-switch contingency. During early laps, DLS ensembles fired at a relatively high rate throughout the maze. As laps progressed and rats adopted the appropriate behavioral strategy, DLS ensembles decreased their firing rate throughout the maze (evidenced by a significant linear regression across laps: $F = 10.25$, $p = 0.004$) but continued to exhibit bursts of firing in the EZ (Fig. 3A,C). This result is consistent with the notion that DLS ensembles shape their activity to highlight critical portions of an action sequence (or "task") rather than each action within the sequence (Jog et al., 1999; Barnes et al., 2005). DMS ensembles showed relatively heightened activity throughout the maze during early laps that disappeared as laps progressed (Fig. 3D,F; evidenced by a significant linear regression across laps: $F = 17.46$, $p = 0.0004$). However, DMS ensembles did not show heightened firing in the EZ that was sustained across laps in a manner similar to DLS ensembles. Thus, a burst of

firing at the start/end portion of the maze was characteristic only of DLS ensembles.

Between-region differences in firing patterns across the maze were also observed during the post-switch period. In the DLS, firing rate was highest when rats were in the EZ (as during the pre-switch period), although the contingency switch appeared to disrupt the size of these DLS bursts in the EZ (Fig. 3B). Interestingly, the DMS showed a relatively higher firing rate throughout the navigation sequence of the maze during the post-switch period compared with the pre-switch period (Fig. 3E).

Task-bracketing scores, averaged across pre- and post-switch periods, were significantly >0 in the DLS (Wilcoxon sign-rank test: $p = 0.047$, $d = 0.82$) but not DMS (Wilcoxon sign-rank test: $p = 0.281$), indicating that activity bursts at the start/end of a lap were characteristic of DLS, but not DMS, ensembles. In addition, task-bracketing scores across rats revealed greater task-bracketing in the DLS than DMS ($t_{(5)} = 1.20$, $p = 0.051$, $d = 0.75$) (Fig. 3G). These characteristic differences in firing patterns between DLS and DMS ensembles are similar to that reported by Thorn et al. (2010). These results are consistent with the notion that the DLS and DMS participate in functionally distinct decision-making systems, with the DLS playing a role in procedural decision-making systems and the DMS playing a role in more deliberative decision-making systems that activate when contingencies change and behavior must adapt accordingly.

Experience-dependent rise in γ50 power precedes DLS bursts at lap initiation

Bursts of DLS activity are often time-locked to the presentation of a cue indicating the opportunity to initiate movement (Barnes et al., 2005; Stalnaker et al., 2010) or the time at which a lever was pressed (Jin and Costa, 2010; Cui et al., 2013; Gremel and Costa, 2013; Jin et al., 2014). However, the temporal relation between these external events and the internal decision to initiate movement is unclear. Because laps in our task were uncued and self-initiated, we were able to more precisely measure how DLS bursts relate to self-initiation of a well-learned learned action sequence. We constructed peri-event averages of movement velocity around the time at which the normalized firing rate was highest while rats were in the EZ of the maze. As expected, activity bursts in the EZ were greater in the DLS than in the DMS (Fig. 4A–C; Wilcoxon sign-rank test: $p = 1.6 \times 10^{-12}$, $d = 0.218$). In addition, a paired-samples t test of mean burst size across rats revealed significantly larger EZ activity bursts in the DLS compared with the DMS ($t_{(5)} = 2.07$, $p = 0.046$, $d = 1.20$; Fig. 4C). Interestingly, rats initiated their learned action sequence ~500-1000 ms before the DLS burst (Fig. 4F). Thus, while bursts of DLS activity occurred in close temporal proximity to initiation of a learned action sequence, they did not appear to be causally related to movement initiation. Instead, this result suggests that
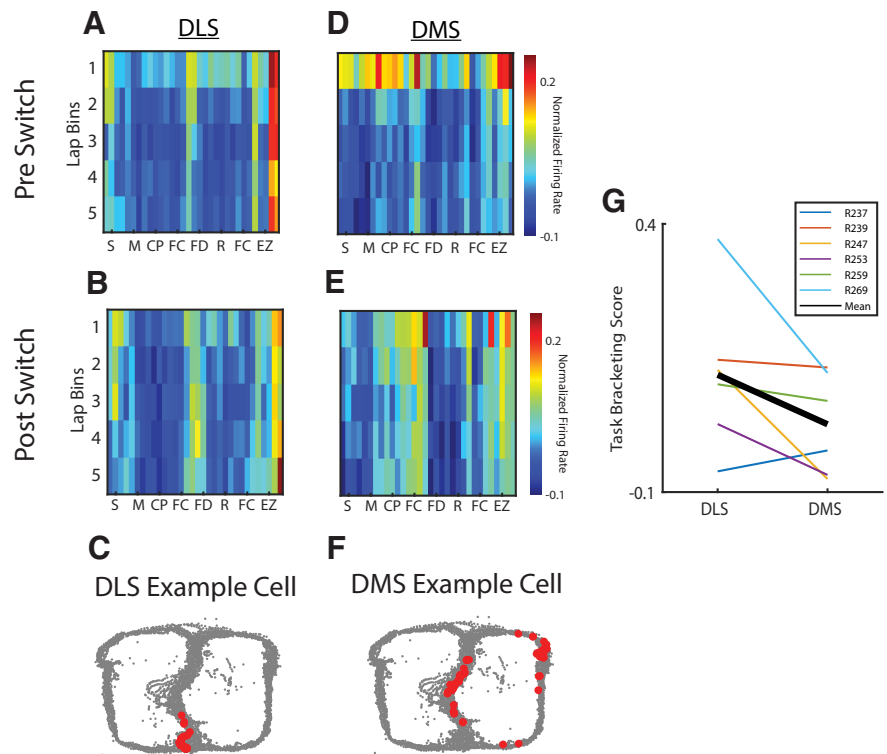

Figure 3. DLS cell ensemble bursts occurred in the EZ of the maze. **A, B, D, E,** Normalized firing rates in each of the eight maze locations averaged across 5 lap bins during the pre-switch period (**A,D**) and post-switch period (**B,E**) for the DLS (**A,B**) and DMS (**D,E**). DLS ensembles fired at a relatively high rate across the maze during early laps that decreased as rats adopted a stereotyped action sequence. However, DLS ensembles continued to show bursts of activity in the EZ of the maze. This firing pattern was absent in the DMS (n = 32 sessions across rats). **C,** Example DLS cell for a single lap that exhibited a burst of activity in the EZ of the maze. **F,** Example DMS cell for a single lap that showed activity during the navigation sequence and reward location without any activity in the EZ of the maze. **G,** Task bracketing scores for each rat (colored lines) and averaged across rats (black line).

DLS bursts at lap initiation might play a role in movement kinetics or sequence organization.

Given the temporal correspondence between DLS bursts and lap initiation, along with previous reports of a rise in γ50 power (45–55 Hz) at the time of self-initiated movement (Masimore et al., 2005), we next explored the relation between γ50 power and activity bursts in both the DLS and DMS. To this end, we assessed spectrograms ranging from 1 to 100 Hz in both the DLS and DMS. A transient rise in γ50 power was observed in the DLS but not the DMS (Fig. 4D,E). This rise in γ50 power in the DLS appeared ~500 ms before bursts of DLS ensemble activity in the EZ. Thus, a rise in γ50 power preceding bursts of ensemble activity (Fig. 3A) is a characteristic feature of the DLS and is largely absent in the DMS. This result is a replication of previous reports that found heightened γ50 power in the DLS when a rewarded action sequence was initiated (Masimore et al., 2005). In addition, these results extend this finding and suggest that heightened γ50 power is temporally linked to characteristic bursts of DLS ensemble activity that also occurs around the time a well-learned action sequence is initiated. Finally, we found that movement initiation at the start of a lap was closely aligned the moment of maximum γ50 power in the DLS (Fig. 4G), which is consistent with Masimore et al. (2005).

There was also a prominent rise in low-frequency power (2-20 Hz) following EZ activity bursts in both the DLS and DMS (Fig. 4D,E). This result is consistent with previous reports of increased θ power during periods in which rats traverse a maze to earn food (Berke et al., 2004; DeCoteau et al., 2007). Thorn
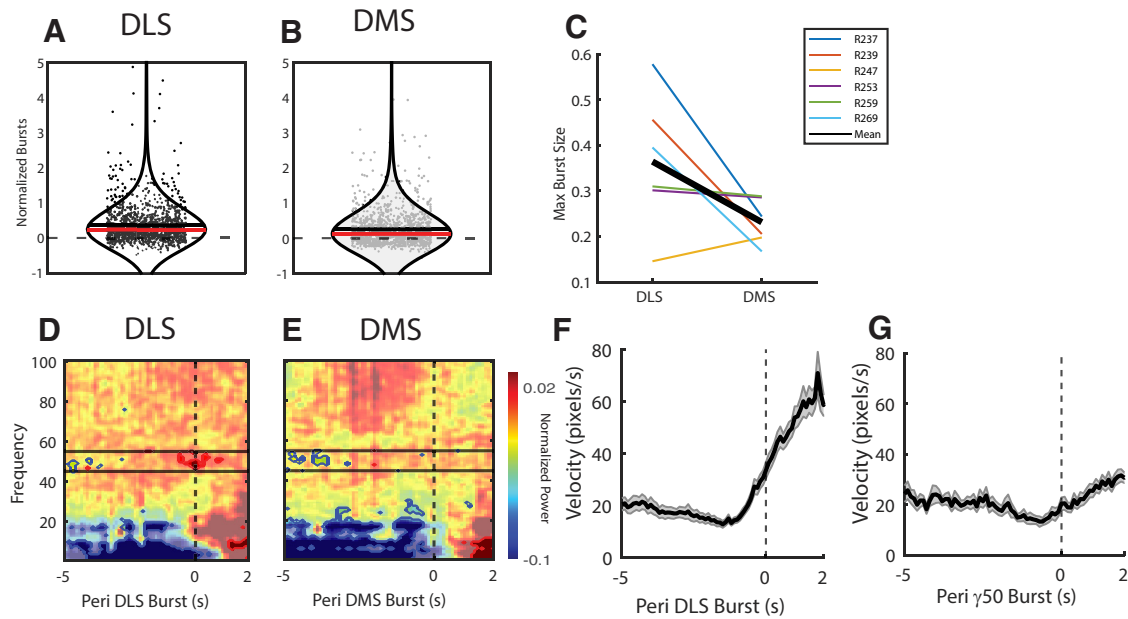
Figure 4. $\gamma50$ power in the DLS LFP peaked near DLS activity bursts. ***A***, ***B***, Violin plots represent the distribution of normalized EZ activity bursts in the DLS and DMS across laps ($n = 1061$; total laps). ***C***, Firing burst size (normalized firing rate) in DLS and DMS ensembles for each rat (colored lines) and averaged across rats (black line). Activity bursts were stronger in DLS ensembles compared with DMS ensembles. ***D***, ***E***, Normalized power across 1-100 Hz in the DLS (***D***) and DMS (***E***). Power at each 100 ms bin of the peri-event average for each frequency was normalized relative to average power at that frequency. Changes determined to be significant by Wilcoxon sign rank tests are outlined with contour lines and decreased transparency. Red contour lines indicate significant increases in power. Blue contour lines indicate significant decreases in power. Theta oscillations became more prominent following the neural activity bursts (and lap initiation) in both the DLS and DMS. However, a burst of $\gamma50$ power around the moment of the activity burst was observed only in the DLS ($n = 29$ sessions across rats). ***F***, Velocity rose moments before the burst of DLS ensemble activity in the EZ. Error bars indicate SEM. ***G***, Mean velocity around the time of maximum $\gamma50$ power. Error bars indicate SEM. Initiation of the action sequence is more closely linked in time to $\gamma50$ events (***G***) than ensemble bursts of activity (***F***).

and Graybiel (2014) reported more prominent θ_5 oscillations in the DLS and more prominent θ_{10} oscillations in the DMS. However, we did not find any differences in θ_5 versus θ_{10} power between DLS and DMS. Instead, normalized power for both θ_5 and θ_{10} was lower than average before lap initiation and increased around the time of lap initiation in both structures. It is possible that between-structure differences in θ_5 and θ_{10} power were not observed because DMS recordings in the present study were more posterior than in Thorn and Graybiel (2014). Nevertheless, our findings suggest that, while θ_5 and θ_{10} power increases following lap initiation in both DLS and DMS, a rise in $\gamma50$ power accompanying a burst of ensemble activity at lap initiation was seen only in the DLS.

Bursts of activity in DLS cell ensembles develop over time (Barnes et al., 2005; Thorn et al., 2010; Smith and Graybiel, 2013; Jin et al., 2014; Regier et al., 2015). While we found a relation between $\gamma50$ power and bursts of activity in the DLS, it is unclear whether this relation developed with experience. To address this, we assessed $\gamma50$ power averaged across 5 lap bins in both the DLS and DMS. The rise in $\gamma50$ power that preceded activity bursts developed across laps in the DLS (Fig. 5A,C). Normalized $\gamma50$ power in the DLS not only increased across laps but also became tightly centered around DLS burst time as rats adopted their behavioral strategy to the current reward contingency (Fig. 5A). This experience-dependent change in the relation between $\gamma50$ power and activity bursts was not observed in the DMS (Fig. 5B,D). A repeated-measures ANCOVA revealed a significant interaction in the slope of $\gamma50$ power across lap bins between DLS and DMS ($F_{(1,251)} = 7.59$, $p = 0.006$). In addition, the mean slope of $\gamma50$ power across lap bins for each rat was significantly greater in the DLS than DMS ($t_{(4)} = 2.74$, $p = 0.026$, $d = 0.931$; Fig. 5E).

In sum, $\gamma50$ power in the DLS LFP rose and peaked before DLS bursts as rats adopted a well-learned, rewarded action sequence. This rise in DLS $\gamma50$ power developed with experience, was tightly linked in time to the initiation of a well-learned action sequence, and was not observed in the neighboring DMS. These results suggest that $\gamma50$ events were a feature of DLS activity that were closely related to procedural action sequences and DLS ensemble bursts at lap initiation.

Phasic firing DLS cells were entrained to $\gamma50$ oscillations
A key facet of functional neuroanatomy is the entrainment of cell spiking to oscillations in the LFP (Buzsáki and Draguhn, 2004). Such phase-specific firing allows anatomically segregated brain regions to communicate information in a manner that allows adaptive behavior in real time (Gatev et al., 2006). Previous reports have found DLS cell entrainment to θ oscillations while rats earned reward in a navigation task (e.g., DeCoteau et al., 2007; Tort et al., 2008; Thorn and Graybiel 2014), although it is unclear whether DLS cells are also entrained to $\gamma50$ oscillations. To address this question, we did a PPC analysis to measure how consistently spike-pairs for a given cell fire at a similar phase of $\gamma50$ (Vinck et al., 2010, 2012; Thorn and Graybiel, 2014) for both phasic firing and high firing cells in the DLS and DMS.

Approximately 15% of DLS cells showed significant entrainment to $\gamma50$ compared with only 6% of DMS cells (Fig. 6A; $\chi^2 = 13.82$, $p = 0.0002$). In addition, there were significantly more phasic firing and high firing cells entrained to $\gamma50$ in DLS compared with DMS (Fig. 6B; phasic firing DLS vs DMS, $\chi^2 = 22.34$, $p = 2.3 \times 10^{-6}$; high firing DLS vs DMS, $\chi^2 = 4.83$, $p = 0.02$). Interestingly, phasic firing DLS cells showed a consistent spike-phase relation to $\gamma50$ oscillations (Fig. 6C; Rayleigh test, $p < 0.002$). Phasic firing cells in the DLS exhibited peak firing near the trough of $\gamma50$ oscillations (Fig. 6C,D), whereas high firing DLS cells showed no systematic preference for the $\gamma50$ phase
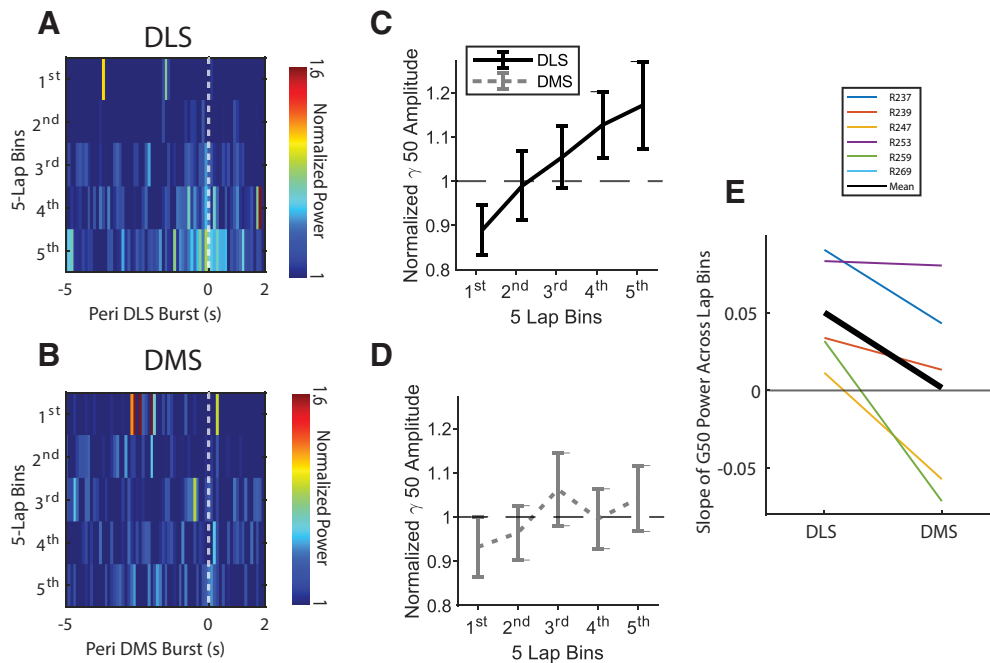
Figure 5. Rise in $\gamma 50$ power at the DLS burst developed with experience. Normalized $\gamma 50$ power during the $-5/2$ s surrounding bursts of ensemble activity averaged across 5 lap bins for DLS (*A,C*) and DMS (*B,D*). $\gamma 50$ power at each 100 ms bin was normalized with respect to average $\gamma 50$ power for that session. To obtain mean $\gamma 50$ power in *C* and *D*, normalized $\gamma 50$ power was averaged during the 1 s preceding activity bursts in the EZ across 5 lap bins. $\gamma 50$ power rose more steeply with experience in DLS compared with DMS ($n = 29$; sessions X rats). In addition, increased $\gamma 50$ power was concentrated around the time of the DLS burst ($n = 29$ sessions across rats). *E*, Mean slope of $\gamma 50$ power across lap bins for each rat (colored lines) and averaged across rats (black line).

at which they fired (though high firing DLS cells were not uniformly distributed; Rayleigh test, $p = 0.016$). Neither phasic firing (Rayleigh test, $p = 0.166$) nor high firing (Rayleigh test, $p = 0.106$) DMS cells showed a systematic spike-phase relation to $\gamma 50$ oscillations (Fig. 6*C*).

Previous work has found entrainment of DLS and DMS cells to θ_5 and θ_{10} oscillations, respectively (Thorn and Graybiel, 2014). We similarly found stronger entrainment to θ_5 compared with θ_{10} in the DLS (Fig. 6*A*). Nearly half (45%) of all DLS cells were significantly entrained to θ_5, but only ~14% of DLS cells were entrained to θ_{10}. Much like Thorn and Graybiel (2014), we found a large fraction of both phasic firing and high firing DLS cells entrained to θ_5. Both phasic and high firing cells in DLS showed preferential firing at the peak of θ_5 oscillations. However, unlike Thorn and Graybiel (2014), we found that DMS cells were also more strongly entrained to θ_5 than θ_{10} (Fig. 6*A*), with ~31% of DMS cells significantly entrained to θ_5 and only ~8% entrained to θ_{10}. Similar to DLS, both phasic and high firing cells in the DMS showed preferential firing at θ_5 peak. As noted earlier, it is possible we did not find differential entrainment of DLS and DMS cells to θ_5 and θ_{10} as did Thorn and Graybiel (2014) because DMS recordings from the present experiment were located more posterior than in Thorn and Graybiel (2014).

In sum, while both DLS and DMS cells were entrained to θ_5 oscillations, only DLS phasic firing cells were entrained to $\gamma 50$. To the extent that phasic firing striatal cells reflect activity of MSNs, these results suggest that entrainment and preferential firing of striatal activity to $\gamma 50$ oscillations are specific to MSNs in more lateral aspects of the dorsal striatum.

Nonlocal, retrospective reward location representations in the DLS

While there are many reports of heightened DLS activity at action boundaries and instruction-cue presentation, there have

not been attempts to examine whether nonlocal representations, either past- or future-oriented, are contained within these activity bursts. Addressing the question of whether DLS ensembles contain past- or future-oriented representations bears direct relevance to prominent, existing theories of DLS function as it relates to procedural decision-making. To address this question, we used a naive Bayesian spatial decoding algorithm (Zhang et al., 1998) to assess the extent to which behaviorally meaningful and potentially nonlocal sections of the maze were represented around the time at which rats initiated their learned action sequence and DLS ensembles exhibited their characteristic firing burst. Aligning decoding to the time of DLS and DMS bursts in the EZ (including only data while rats were in EZ) revealed nonlocal representations in DLS cell ensembles.

DLS cell ensembles represented nonlocal reward locations in the moments before lap initiation bursts in the EZ (Fig. 7*A,B*). This result suggests that nonlocal representations are not about movement trajectories or CPs but are instead about locations where rewards were collected. Critically, these reward location representations were consistently about past experiences more so than future ones. (Fig. 7*A–C*). The distinction between past- and future-reward location representations is best illustrated by the difference in left versus right reward location representations for from-left-to-right (LR) versus from-right-to-left (RL) turn sequences when the alternation contingency was in effect. For alternation turn sequences, the previously visited and about-to-be visited rewards are at separate locations, therefore allowing determination of whether nonlocal representations at DLS bursts are about the past or future. Left reward location representations were strongest during from-left-to-right (LR) turn sequences, whereas right reward location representations were strongest during from-right-to-left (RL) turn sequences (Fig. 7*A,B*). Thus, DLS cell ensembles represented the place where reward was just collected rather than the place where reward was about to be
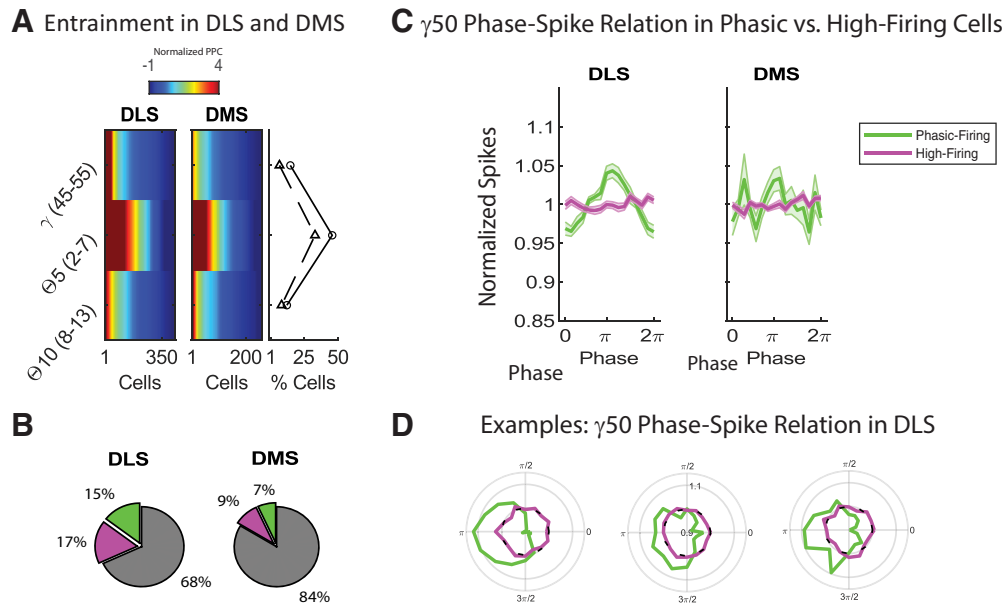
Figure 6. Phasic firing DLS cells were entrained to γ_{50}. **A**, Normalized PPC at γ_{50}, θ_5, and θ_{10} oscillations were sorted in descending order for cells in the DLS (left) and DMS (middle). The percent of cells entrained to each frequency was determined for DLS (circles) and DMS (triangles; right). DLS and DMS cells were most strongly entrained to θ_5. However, a larger percentage of cells in the DLS were entrained to γ_{50} compared with DMS. **B**, Percent of phasic firing and high firing cells entrained to γ_{50} in the DLS (left) and DMS. **C**, Mean (SEM) normalized spike rates in DLS (left) and DMS (right) as a function of γ_{50} phase for phasic firing (green) and high firing (magenta) cells. Spike rates were normalized relative to each cell's average spike rate. Only phasic firing cells in the DLS showed phase-dependent firing during γ_{50} oscillations, with peak firing rate near the γ_{50} trough ($n = 29$ sessions across rats). **D**, Examples of phase-dependent firing of DLS phasic firing and high firing cells for three separate sessions illustrating preferential firing of phasic firing DLS cells at the trough of γ_{50}.

collected. Importantly, past-reward location representations were not decaying remnants of location encoding from when the rat was at the reward location because the representations diminished between the previous reward experience and strengthened as time approached DLS bursts. Representations of past or future reward locations were largely absent in the DMS (Fig. 7A).

To further analyze the past- or future-orientation of information encoded in ensemble bursts, decoded reward locations were collapsed across turn sequences and contingencies based on whether the reward location was from the previous (past) or subsequent (future) lap. We computed a "retrospective bias" in reward location representations by taking the log ratio of mean decoding to past and future reward locations ($\log_{10}(pPastReward/pFutureReward)$) for each lap. Values of 0 indicate no bias, whereas values >0 and <0 indicate a bias for past or future reward location representations, respectively. There was a significant retrospective bias in reward location representations in the DLS (mean = 0.08; SD = 0.25; Wilcoxon sign-rank test: $p = 1.4 \times 10^{-35}$, $d = 0.40$), along with a weaker but significant retrospective bias in DMS (mean = 0.04; SD = 0.20; Wilcoxon sign-rank test: $p = 0.03 \times 10^{-9}$, $d = 0.23$). The retrospective bias was significantly stronger in DLS compared with DMS (Fig. 7C; Wilcoxon sign-rank test: $p = 1.4 \times 10^{-6}$, $d = 0.241$). A retrospective bias in DLS representations appeared on 18.2% of laps, with a prospective bias for only 3.1% of laps. In contrast, there was a retrospective bias in DMS representations on only 8.1% of laps, with 2.0% of laps showing a prospective bias. Further, past representations were significantly greater than future representations in the DLS but not DMS (Wilcoxon sign-rank test: $p = 0.00001$, $d = 0.135$ for DLS; $p = 0.13$, $d = 0.065$ for DMS). Finally, mean retrospective bias across rats was significantly stronger in DLS compared with DMS (Fig. 7D; Wilcoxon sign-rank test: $p = 0.003$, $d = 0.59$),

suggesting that the retrospective bias was consistent across rats.

Collectively, these results support the hypothesis that DLS more likely uses Markovian representations of past events rather than representing future action plans to regulate well-learned procedural action sequences. In addition, events represented in DLS ensembles during task-bracketed bursts at lap initiation were not about past or future action sequences but were instead about past reward locations. These past-reward representations in DLS appeared around the time at which an automated action sequence was initiated and DLS ensembles correspondingly exhibited a relatively high firing rate. These past-reward location representations were largely absent in the DMS.

Retrospective reward location representations in the DLS peaked during γ50 event

Given that past-reward location representations and rises in γ_{50} power occur at lap initiation, we assessed the degree to which retrospective representations were related to periods of high γ_{50} power in the DLS. Reward location representations were highest during the period in which γ_{50} power was also highest (Fig. 8A). Wilcoxon rank sum tests revealed that past and future reward location representations were significantly higher during γ_{50} events compared with before and after the period of γ_{50} power ($p < 0.0001$ for before vs during and during vs after comparisons). In addition, a repeated-measures ANOVA (epoch × past/future) revealed a main effect of epoch ($F_{(2,2772)} = 191.2$, $p < 0.0001$) but no interaction between epoch and past/future representations ($F_{(2,2772)} = 0.11$, $p = 0.89$). Follow-up paired-samples t tests revealed that mean past reward location representations were significantly greater during the γ_{50} event than before ($t_{(479)} = -7.87$, $p = 9.4 \times 10^{-12}$, $d = -0.107$) and after ($t_{(608)} = 14.52$, $p = 1.1 \times 10^{-40}$, $d = 0.184$) epochs. Future reward location representations were also greatest during the γ_{50} event
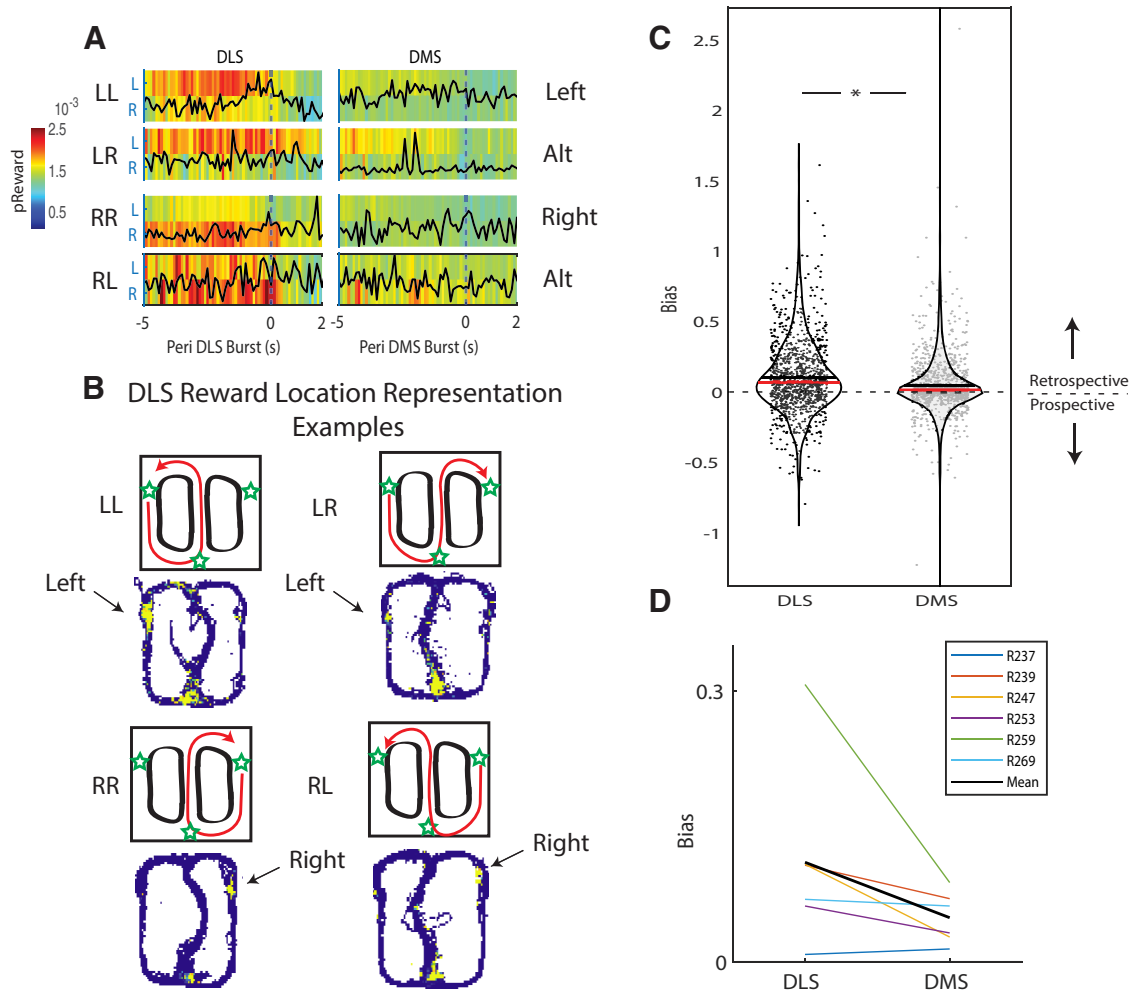
Figure 7. Nonlocal representations of past-reward locations appeared in DLS but not DMS. ***A***, Mean probability of decoding to left and right reward locations around neural activity bursts in the EZ of the maze for DLS (left) and DMS (right) on correct laps. Turn direction for the previous and subsequent lap for each lap sequence is indicated to the left of each row with the corresponding contingency indicated to the right (e.g., LL, left-past left-future turn sequence for the left contingency; RL, right-past left-future turn sequence for the alternate contingency). Solid lines indicate mean γ50 power during corresponding lap sequences. Reward location representations increased as time to the DLS burst approached, and the representations were stronger for past reward locations compared with future reward locations (n = 32 sessions across rats). ***B***, Examples of past-reward location representations in the DLS for each turn sequence taken 100 ms before the DLS burst. Note the relatively strong representations of current location (bottom of each maze) in addition to the nonlocal representations of the previously visited reward location. ***C***, Violin plots of retrospective bias in reward location representations for DLS and DMS across all laps (n = 1061; total laps). The retrospective bias in reward location representations was stronger in DLS than it was in DMS. ***D***, Retrospective versus prospective bias in reward location representations in DLS and DMS ensembles for each rat (colored lines) and averaged across rats (black line).

compared with before ($t_{(479)} = -8.21$, $p = 7.4 \times 10^{-13}$, $d = 0.101$) and after ($t_{(608)} = 14.93$, $p = 7.9 \times 10^{-43}$, $d = 0.163$). In addition, mean nonlocal reward location representations were significantly stronger during γ50 events than before or after for both past ($t_{(4)} = 3.75$, $p = 0.01$, $d = 0.633$) and future ($t_{(4)} = 4.18$, $p = 0.003$, $d = 0.734$) rewards across rats (Fig. 8B,C). These results suggest a close temporal correspondence between high-γ50 power and nonlocal reward location representations in the DLS around the time at which rats initiated a learned action sequence.

Discussion

Retrospective versus prospective representations in the DLS
Procedural actions are commonly believed to be governed by an association between antecedent stimuli and instrumental responses that develop through a history of reinforcement (Thorndike, 1932; Hull, 1942; Herrnstein and Prelec, 1991; Balleine and Dickinson, 1998). This notion is formalized with "model free" reinforcement learning algorithms, which learn the

value of taking an action in a given context by storing a cached value for state-action pairs based on previous experience (Sutton and Barto, 1998; Daw et al., 2005; Niv et al., 2007; Dayan and Daw, 2008). Model-free reinforcement learning algorithms therefore select actions based on representations of the current state. Under this theory, the DLS should represent features of the current state that have been critical for earning reward, such as information about past events that allow the current state to trigger previously successful actions (Fig. 1B).

Other perspectives conceptualize procedural actions as action-chains that have been "chunked" together into a cohesive sequence that defines a single unit eligible for action selection (Graybiel, 1998; Barnes et al., 2005; Smith and Graybiel, 2013). Dezfouli and Balleine (2012) formalized a model in which the decision to initiate a procedural action sequence is governed by a goal-directed system that deliberates over which chunked action sequence should be executed (Dezfouli and Balleine, 2013; Dezfouli et al., 2014). Thus, rather than deliberating over each constituent action within the sequence, the action-chunking
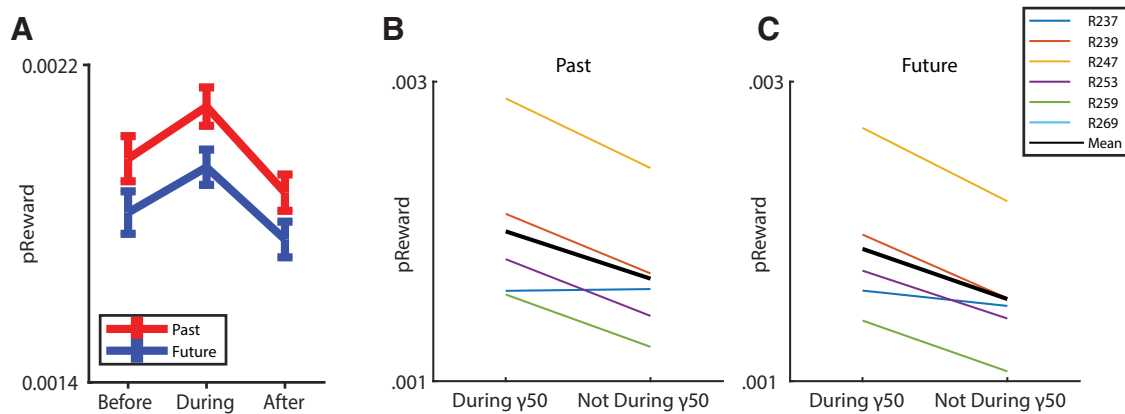
Figure 8. Retrospective reward representations in DLS were coupled to $\gamma 50$ events. **A**, Mean (SEM) probability of decoding to past and future reward locations before, during, and after the epoch of maximum $\gamma 50$ power. Both past and future reward representations were highest during the $\gamma 50$ event, and both significantly decreased once the $\gamma 50$ event was over ($n = 1061$; total laps). **B, C**, Mean past reward (**B**) and future reward (**C**) representations during $\lambda 50$ events compared with not during $\lambda 50$ events (i.e., pReward averaged across before and after periods) for each rat (colored lines) and averaged across rats (black line).

process allows the goal-directed system to deliberate over the sequence as a whole. This perspective suggests that the DLS should represent future action plans at the initiation of a procedural action sequence, perhaps by transiently representing the entire sequence or just the start and end of the to-be-executed sequence (Fig. 1C).

Similarly, successor representations offer a tractable solution to the problem of predicting future rewards without the need for computationally expensive world models. Successor representations learn a predictive map of the environment using representations of expected occupancy of the next state within a sequence of states (Dayan, 1993; Gershman et al., 2012; Momennejad et al., 2017; Gershman, 2018). Importantly, successor representations can be used to accurately predict reward when trajectories through a state space become consistent over time. Because a defining characteristic of procedural action sequences is their consistency and stereotypy, successor representations might play an important role in the computational basis of procedural decision-making. As animals learn to automate their action sequences, predictions about the next state (or location) in the form of a successor representation become more reliable and can therefore predict future rewards (Gershman, 2018). This perspective also suggests that the DLS should represent the future path of the animal in a compact representation.

Our results support a role for task-bracketed DLS bursts in "model-free" reinforcement learning over the other two theories (goal-directed action-chains and successor representations). Our data revealed that DLS representations around lap initiation encoded recent experiences more than future action sequences. Previous work has characterized task-bracketing DLS bursts as representing meaningful boundaries of a well-learned action sequence (Barnes et al., 2005; Thorn et al., 2010; Smith and Graybiel, 2013; Jin et al., 2014). However, it was unclear what information was represented within these task-bracketed bursts of activity. Our results replicate these findings and further suggest that the bursts at task-initiation are accompanied by representations of recently visited locations where reward was obtained rather than future action sequences, future reward locations, or successor representations. While it is certainly possible that other brain structures play roles in planned action-chains or successor representations, our data do not support a role for task-bracketed DLS bursts in these cognitive processes.

It is currently unclear what role representations of previous-visited reward locations play in procedural decision-making.

One possibility is that past-reward location representations reflect the consequences of a state identification process that allows animals to recognize the current situation (Redish et al., 2007) and subsequently execute the appropriate action sequence. In the current task, the location of previously acquired reward can serve as a cue informing the rat of the current contingency and can therefore guide action selection in the EZ of the maze.

Functional distinctions between DLS and DMS
Available evidence suggests that the DLS and DMS participate in functionally distinct decision-making systems (Featherstone and McDonald, 2005; Yin et al., 2006; Stalnaker et al., 2010; Thorn et al., 2010; Thorn and Graybiel, 2014; Murray et al., 2012; Kim et al., 2013; Ito and Doya, 2015; Regier et al., 2015; Vandaele et al., 2019). The DLS is believed to play a role in procedural decision-making systems that drive stabilized and automated behavior (Featherstone and McDonald, 2004; Yin et al., 2004, 2006). In contrast, the DMS is believed to play a role in more deliberative decision-making systems that adapt behavior to changing environmental contingencies (Featherstone and McDonald, 2005; Ragozzino, 2007; Ragozzino et al., 2002; Yin et al., 2006). Previous work has found that DLS and DMS ensembles exhibit systematic differences in firing patterns across maze locations as rats learn to traverse a maze to earn reward (Thorn et al., 2010; Thorn and Graybiel, 2014). While previous reports found these characteristic patterns across days (and weeks) of training (Thorn et al., 2010; Thorn and Graybiel, 2014), we found differences in activity patterns between DLS and DMS ensembles across maze locations within a single session, suggesting that these characteristic activity patterns can adjust to changing circumstances over relatively short timescales.

Interestingly, we found that task-initiation DLS bursts occurred moments after a self-initiated action sequence began (see also Yin et al., 2009). Some evidence suggests that the dorsal striatum plays a critical role in initiating a movement sequence (Neafsey et al., 1978; Hikosaka et al., 2000; Bailey and Mair, 2006; Jin and Costa, 2010; Cui et al., 2013; Jin et al., 2014, Yin and Knowlton, 2006). However, there is also evidence that the dorsal striatum plays a critical role in shaping performance-related parameters and kinematics once movement has been initiated (Anderson and Horak, 1985; Nowak et al., 2018; Rueda-Orozco and Robbe, 2015; Dudman and Krakauer, 2016; Markowitz et al., 2018; Crego et al., 2020). While it is

likely that the dorsal striatum plays a role in both action initiation and execution (e.g., Tecuapetla et al., 2016; Thura and Cisek, 2017; Klaus et al., 2019), our results suggest that past-reward location representations contained in bursts of DLS activity at lap initiation are more likely to contribute to movement parameters and kinematics of a stereotyped action sequence rather than its initiation per se. Perhaps past-reward location representations, by providing information about the current state of the world (e.g., turn direction required to earn reward), allow a well-learned action sequence to be executed in a ritualistic manner similar to the way it has been executed in the past (e.g., the duration, organization, vigor, or trajectory of the action sequence).

Importantly, while we found a significant bias for retrospective representations, our results do not preclude the existence of prospective information within DLS bursts. These occasional future-oriented representations might suggest a more nuanced view such that both retrospective and prospective information is contained in DLS bursts, and that variability in these representations may play a role in procedural decision-making systems. Thus, our results do not preclude prospective representations in the DLS, nor that such prospective representations may be important for procedural decision-making systems. Instead, our results show that retrospective representations were more prominent than prospective representations around the time at which a procedural action sequence was initiated.

Our results reveal the information represented within DLS bursts at lap initiation, which has implications for the functional role of these bursts in procedural action sequences. However, our results do not explicitly identify what this functional role might be. Retrospective representations in DLS bursts may play a role in regulating the degree to which the subsequent action sequence is ritualized and stereotyped. Retrospective representations may contribute to the rat's decision to turn either left or right at the CP, thereby contributing to reward-guided action selection. It is also possible that retrospective representations may not play a direct, functional role in upcoming action sequences and choices but instead contribute to the learning and development of procedural action sequences over time.

γ50 and DLS task-initiation bursts

Oscillations visible in the LFP organize spiking activity across brain regions to govern adaptive behavior (Buzsáki and Draguhn, 2004; Fries et al., 2007; Buzsáki et al., 2012; Nowak et al., 2018). Masimore et al. (2005) reported a transient (150 ms) rise in the power of LFP oscillations in the 45-55 Hz range, which they identified as γ50 events, in the DLS at the time that rats initiate movement while traversing a maze to earn reward. Replicating results from Masimore et al. (2005) with a new dataset from a new task, we also found rises in γ50 power at the moment rats initiated a well-learned action sequence. In addition, we found that this rise in γ50 power corresponded to the time at which DLS ensembles burst (i.e., "task-bracketed" DLS bursts) and developed with experience. Thus, our results link γ50 events, action sequence initiation, and task-bracketed DLS bursts. In addition, retrospective representations were strongest during γ50 events in the moments preceding lap initiation DLS bursts. As with past-reward location representations and task-bracketed activity bursts, γ50 events were largely absent in the DMS and were unaffected by experience. Thus, to the extent that the DLS, but not DMS, participates in procedural decision-making systems, our results suggest that retrospective state

information that peaks during γ50 events and grows with experience is a neurophysiological facet of procedural decision-making systems. Although we were unable to exclude the possible role of volume conduction in comparisons between DLS and DMS LFPs, results presented in Figure 6 found that DLS cells were phase-locked to the γ50 signal, whereas DMS cells were not.

It is currently unclear how γ50 events participate in DLS-based information processing during procedural decision-making. γ oscillations are prominent throughout the cortex and are believed to play a role in synchronizing information processing between cortical networks (Buzsáki and Draguhn, 2004; Fries et al., 2007). Interestingly, high γ (>30 Hz) oscillations in the human motor cortex are linked to self-initiated movement in humans (Cheyne et al., 2008; Huo et al., 2010; Santarnecchi et al., 2017). Given the dense projections from motor cortex to DLS, it is possible that γ50 oscillations and the temporally coupled past-reward representations play a role in organizing information flow between the motor cortex and basal ganglia to learn and/or execute appropriate action sequences within a given context (state). If retrospective reward representations in the DLS indeed reflect the current state, it is possible that these representations are coupled to γ50 oscillations to assist motor circuits in organizing and/or selecting the appropriate action sequence to be initiated.

Prominent γ50 oscillations have also been found in the ventral striatum related to reward-guided behaviors (Berke et al., 2004; van der Meer and Redish, 2009; Kalenscher et al., 2010; van der Meer et al., 2010; Howe et al., 2011). Berke et al. (2004) found γ50 oscillations in ventromedial striatum while rats traversed a maze to earn reward. van der Meer and Redish (2009) found post-reward bursts of γ50 power in ventral striatum that became more stable with experience. That task-dependent γ50 oscillations are found throughout dorsolateral and ventromedial striatum raises interesting questions about their possible functional role in reward-guided learning and behavior. While some researchers have suggested that γ50 oscillations may originate outside of the striatum (Carmichael et al., 2017), perhaps γ50 oscillations help regulate information flow across the striatum to allow dorsal and ventral regions to properly organize and time their output to downstream motor structures. Indeed, available evidence suggests that while DLS plays a role in procedural decision-making, ventral striatum plays a role in more deliberative, goal-directed decision-making. Perhaps γ50 oscillations help these distinct decision-making systems within the striatum to communicate effectively to organize reward-guided behavior. Regardless of the potential relation between γ50 oscillations in the ventral and dorsal striatum, our results suggest that γ50 oscillations play a critical role in procedural decision-making systems realized in the DLS.

In conclusion, we found that task-bracketed bursts of DLS activity at lap initiation contained primarily representations of previously visited reward locations. This result is inconsistent with action-chunking theories of procedural decision-making, which suggest that DLS bursts should represent the to-be-implemented action sequence. This result is instead consistent with model-free theories of procedural decision-making that suggest task-bracketed DLS bursts should represent certain features of the current state that allow a procedural action sequence to be implemented (e.g., the location of previously collected rewards). These retrospective representations in DLS ensembles were temporally linked to γ50 events in the DLS LFP and were strongest around the time rats initiated their well-learned action sequence.

Collectively, these results offer novel insights into the manner in which DLS neurophysiology participates in procedural decision-making systems.

References

Anderson ME, Horak FB (1985) Influence of the globus pallidus on arm movements in monkeys: III. Timing of movement-related information. J Neurophysiol 54:433–488.

Bailey KR, Mair RG (2006) The role of striatum in initiation and execution of learned action sequences in rats. J Neurosci 26:1016–1025.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.

Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature 437:1158–1161.

Berke JD, Okatan M, Skurski J, Eichenbaum HB (2004) Oscillatory entrainment of striatal neurons in freely moving rats. Neuron 43:883–896.

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents: EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Carmichael JE, Gmaz JM, van der Meer MA (2017) Gamma oscillations in the rat ventral striatum originate in the piriform cortex. J Neurosci 37:7962–7974.

Cheyne D, Bells S, Ferrari P, Gaetz W, Bostan AC (2008) Self-paced movements induce high-frequency gamma oscillations in primary motor cortex. Neuroimage 42:332–342.

Crego AC, Stocek F, Marchuk AG, Carmichael JE, van der Meer MA, Smith KS (2020) Complementary control over habits and behavioral vigor by phasic activity in the dorsolateral striatum. J Neurosci 40:2139–2153.

Cui G, Jun SB, Jin X, Pham MD, Vogel SS, Lovinger DM, Costa RM (2013) Concurrent activation of striatal direct and indirect pathways during action initiation. Nature 494:238–242.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.

Dayan P (1993) Improving generalization for temporal difference learning: the successor representations. Neural Computation 5:613–624.

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cogn Affect Behav Neurosci 8:429–453.

DeCoteau WE, Thorn C, Gibson DJ, Courtemanche R, Mitra P, Kubota Y, Graybiel AM (2007) Oscillations of local field potentials in the rat dorsal striatum during spontaneous and instructed behaviors. J Neurophysiol 97:3800–3805.

Dezfouli A, Balleine BW (2012) Habits, action sequences and reinforcement learning. Eur J Neurosci 35:1036–1051.

Dezfouli A, Balleine BW (2013) Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized. PLoS Comput Biol 8:e1003364.

Dezfouli A, Lingawi NW, Balleine BW (2014) Habits as action sequences: hierarchical action control and changes in outcome value. Philos Trans R Soc Lond B Biol Sci 369:20130482.

Dudman JT, Krakauer JW (2016) The basal ganglia: from motor commands to the control of vigor. Curr Opin Neurobiol 37:158–166.

Featherstone RE, McDonald RJ (2004) Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference learning. Neuroscience 124:23–31.

Featherstone RE, McDonald RJ (2005) Lesions of the dorsolateral or dorsomedial striatum impair performance of a previously acquired simple discrimination task. Neurobiol Learn Mem 84:159–167.

Fries P, Nikolic D, Singer W (2007) The gamma cycle. Trends Neurosci 30:309–316.

Gatev P, Darbin O, Wichmann T (2006) Oscillations in the basal ganglia under normal conditions and in movement disorders. Movement Disorders 21:566–1577.

Gershman SJ (2018) The successor representation: its computational logical and neural substrates. J Neurosci 38:7193–7200.

Gershman SJ, Moore CD, Todd MT, Norman KA, Sederberg PB (2012) The successor representation and temporal context. Neural Comput 24:1553–1586.

Graybiel AM (1998) The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem 70:119–136.

Gremel CM, Costa RM (2013) Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nat Commun 4:2264.

Herrnstein RJ, Prelec D (1991) Melioration: a theory of distributed choice. J Econ Perspect 5:137–156.

Hikosaka O, Takikawa Y, Kawagoe R (2000) Role of the basal ganglia in the control of purposive saccadic eye movements. Physiological Reviews 80:953–978.

Howe MW, Atallah HE, McCool A, Gibson DJ, Graybiel AM (2011) Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. Proc Natl Acad Sci USA 108:16801–16806.

Hull CL (1942) Principles of behavior: an introduction to behavior theory. New York: Appleton-Century-Crofts.

Huo X, Xiang J, Wang Y, Kirtman EG, Kotecha R, Fujiwara H, Hemasilpin N, Rose DF, Degrauw T (2010) Gamma oscillations in the primary motor cortex studied with MEG. Brain Dev 32:619–624.

Ito M, Doya K (2015) Distinct neural representations in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. J Neurosci 35:3499–3514.

Janabi-Sharifi F, Hayward V, Chen CS (2000) Discrete-time adaptive windowing for velocity estimation. IEEE Trans Contr Syst Technol 8:1003–1009.

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

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

Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. Science 286:1745–1749.

Kalenscher T, Lansink CS, Lankelma JV, Pennartz CM (2010) Reward-associated gamma oscillations in ventral striatum are regionally differentiated and modulate local firing activity. J Neurophysiol 103:1658–1672.

Kawaguchi Y (1993) Physiological, morphological, and histochemical characterization of three classes of interneurons in the rat neostriatum. Journal of Neuroscience 13:4908–4923.

Kim H, Lee D, Jung MW (2013) Signals for previous goal choice persists in the dorsomedial, but not dorsolateral striatum of rats. J Neurosci 33:52–63.

Klaus A, da Silva JA, Costa RM (2019) What, if, and when to move: basal ganglia circuits and self-paced action initiation. Annu Rev Neurosci 42:459–483.

Lashley KS (1951) The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior: the Hixon symposium, pp. 112–146. Wiley.

Markowitz JE, Gillis WF, Beron CC, Neufeld SQ, Robertson K, Bhagat ND, Peterson RE, Peterson E, Hyun M, Linderman SW, Sabatini BL, Datta SR (2018) The striatum organizes 3D behavior via moment-to-moment action selection. Cell 174:44–58.

Masimore B, Schmitzer-Torbert NC, Kakalios J, Redish AD (2005) Transient striatal γ local field potentials signal movement initiation in rats. Neuroreport 16:2021–2024.

Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ (2017) The successor representation in human reinforcement learning. Nature Human Behaviour 1:680–692.

Muenzinger KF, Gentry E (1931) Tone discrimination in white rats. J Comp Psychol 12:195–206.

Murray JE, Belin D, Everitt BJ (2012) Double dissociation of the dorsomedial and dorsolateral striatal control over the acquisition and performance of cocaine seeking. Neuropsychopharmacology 37:2456–2466.

Neafsey EJ, Hull CD, Buchwald NA (1978) Preparation for movement in the cat: II. Unit activity in the basal ganglia and thalamus. Electroencephalogr Clin Neurophysiol 44:714–723.

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 191:507–520.

Nowak J, Zich C, Stagg CJ (2018) Motor cortical gamma oscillations: what have we learnt and where are we headed? Curr Behav Neurosci Rep 5:136–142.

Ragozzino ME (2007) The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. Ann NY Acad Sci 1121:355–375.

Ragozzino ME, Ragozzino KE, Mizumori SJ, Kesner RP (2002) Role of the dorsomedial striatum in behavioral flexibility for response and visual cue discrimination learning. Behav Neurosci 116:105–115.

Redish AD (2016) Vicarious trial and error. Nat Rev Neurosci 17:147–159.

Redish AD, Jensen S, Johnson A, Kurth-Nelson Z (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. Psychol Rev 114:784–805.

Regier PS, Amemiya S, Redish AD (2015) Hippocampus and subregions of the dorsal striatum respond differently to a behavioral strategy change on a spatial navigation task. J Neurophysiol 114:1399–1416.

Rueda-Orozco PE, Robbe D (2015) The striatum multiplexes contextual and kinematic information to constrain motor habits execution. Nat Neurosci 18:453–460.

Santarnecchi E, Biasella A, Tatti E, Rossi A, Prattichizzo D, Rossi S (2017) High gamma oscillations in the motor cortex during visuo-motor coordination: a tACS interferential study. Brain Res Bull 131:47–54.

Schmitzer-Torbert NC, Redish AD (2004) Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J Neurophysiol 91:2259–2272.

Schmitzer-Torbert NC, Redish AD (2008) Task-dependent encoding of space and events by striatal neurons is dependent on neural subtype. Neuroscience 153:349–360.

Smith KS, Graybiel AM (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. Neuron 79:361–374.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. Front Integrative Neurosci 4:12.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: Massachusetts Institute of Technology:

Tecuapetla F, Jin X, Lima SQ, Costa RM (2016) Complementary contributions of striatal projection pathways to action initiation and execution. Cell 166:703–715.

Thorn CA, Graybiel AM (2014) Differential entrainment and learning-related dynamics of spike and local field potential activity in the sensorimotor and associative striatum. J Neurosci 34:2845–2859.

Thorn CA, Atallah H, Howe M, Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. Neuron 66:781–795.

Thorndike EL (1932) The fundamentals of learning. New York: Bureau of Publications, Teachers College.

Thura D, Cisek P (2017) The basal ganglia do not select reach targets but control the urgency of commitment. Neuron 95:1160–1170.

Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189–208.

Tort AB, Kramer MA, Thorn C, Gibson DJ, Kubota Y, Graybiel AM, Kopell NJ (2008) Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a T-maze task. Proc Natl Acad Sci USA 105:20517–20522.

Vandaele Y, Mahajan NR, Ottenheimer DJ, Richard JM, Mysore SP, Janak PH (2019) Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. eLife 8:e49536.

van der Meer MA, Redish AD (2009) Low and high gamma oscillations in rat ventral striatum have distinct relationships to behavior, reward, and spiking activity on a learned spatial decision task. Front Integr Neurosci 3:9.

van der Meer MA, Kalensher T, Lansink CS, Pennartz CM, Burke J, Redish AD (2010) Integrating early results on ventral striatal gamma oscillations in the rat. Front Neurosci 4:1–12.

Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CM (2010) The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. Neuroimage 51:112–122.

Vinck M, Battaglia FP, Womelsdorf T, Pennartz C (2012) Improved measures of phase coupling between spikes and the local field potential. J Comput Neurosci 33:53–75.

Yin HH, Knowlton BJ (2006) The role of basal ganglia in habit formation. Nat Rev Neurosci 7:464–476.

Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of the dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181–189.

Yin HH, Knowlton BJ, Balleine BW (2006) Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res 166:189–196.

Yin HH, Mulcare SP, Hilario MRF, Clouse E, Holloway T, Davis MI, Hansson AC, Lovinger DM, Costa RM (2009) Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nature Neuroscience 12:333–341.

Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J Neurophysiol 79:1017–1044.