

Reviewed Preprint
v1 • August 7, 2026
Not revised

✉ For correspondence:
shuler@jhmi.edu

Competing interests: No
competing interests declared

Funding: See [page 23](#)

Reviewing editor: Alicia Izquierdo,
University of California, Los Angeles,
United States

© 2026, Sutlief et al. This article is
distributed under the terms of the
[Creative Commons Attribution
License](#), which permits unrestricted
use and redistribution provided that
the original author and source are
credited.

Accumulation of neural state transitions in dorsomedial striatum predicts patch foraging decisions

Elissa Sutlief¹, Shichen Zhang¹, Kate Forsberg¹, Marshall G Hussain Shuler^{1,2}✉

¹Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, United States • ²Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, United States

eLife Assessment

The authors use a novel patch-leaving task to reveal a reward-reset strategy when mice choose to leave a depleting resource, and find that accumulated step-like activity in the dorsomedial striatum is correlated with the timing of these decisions. These **important** behavioral and neural findings are supported by substantial and **convincing** data. Additional control analyses would strengthen the evidence that these signals are specifically related to timing and patch-leaving decisions, rather than alternative task-related processes.

<https://doi.org/10.7554/eLife.111930.1.sa4>

Abstract

Activities with temporally distributed returns require deciding not only what to do, but when to stop. Although optimal foraging theory predicts that patch departure should depend on the total time invested in a patch, behavior across species often reflects sensitivity to recent rewards. The dorsomedial striatum (DMS) mediates both goal-directed decision-making and interval timing, two functions that converge during patch foraging, where animals must decide when to exit patches to maximize reward rates. We recorded extracellular activity from neurons in DMS while freely moving mice performed a patch-foraging task. Mice employed a reward-reset strategy, primarily basing their exit decisions on the time since the last reward, with systematic adjustments for patch residence time and environmental reward-rate context. Individual neurons in DMS underwent discrete firing rate transitions at characteristic delays following each reward. These transition times tiled the post-reward interval across the population, producing an accumulation-to-threshold signal that reset with each reward and whose rate was sensitive to the cost of time, increasing with higher environmental reward rates and as rewards occurred later in the patch. This accumulation reached a consistent threshold at the moment of patch exit, encoding the animal's intended stopping time. Together, these results identify a time-cost-variable, reward-reset, accumulation-to-threshold computation in DMS that integrates environmental reward rate, elapsed time from pursuit engagement, and elapsed time from reward occurrence to determine when to abandon a depleting pursuit.

1 Introduction

Decision-making is often framed as the selection among discrete options. However, many real-world decisions are continuous in time and instead require determining how long to persist in an ongoing activity before disengaging. This class of decisions, when to stop, is central to adaptive behavior, shaping how organisms allocate time across competing pursuits. Failures of this process are implicated in neuropsychiatric conditions such as addiction, depression, and OCD (Addicott et

al., 2017 [↗](#); Huys et al., 2015 [↗](#); Marton et al., 2019 [↗](#)). Despite its importance, the neural mechanisms that govern the timing of disengagement from ongoing behavior remain poorly understood.

1.1 Ecological Models of Time Allocation

These types of continuous, temporal decisions are common within ecological foraging paradigms. Specifically, patch foraging presents animals with an environment where resources are clustered in patches. The utility of searching a patch declines over time as its resources are harvested, requiring the animal to decide when to abandon it in search of a new one. Many ethological studies have characterized various factors that affect patch occupancy time, such as patch richness (Keller and Sullivan, 2023 [↗](#); Marshall et al., 2013 [↗](#); Utsumi et al., 2009 [↗](#)), inter-patch distance (Searle et al., 2006 [↗](#); Utsumi et al., 2009 [↗](#)), predation risk (Brown, 1988 [↗](#); Eccard et al., 2020 [↗](#); Ward et al., 2000 [↗](#)), and competition (Brown, 1988 [↗](#); Laguë et al., 2012 [↗](#)). To understand how long an animal should continue searching a given patch, we use the normative principle of reward-rate maximization, which posits that the best strategy is the one that results in the highest overall rate of reward. For typical patch-foraging tasks, where the patch's reward probability decays monotonically over time, the optimal policy is to leave a patch when the rate of reward inside the patch falls below the overall rate of reward achieved in the environment, as delineated by the Marginal Value Theorem (MVT) (Charnov, 1976 [↗](#)). The MVT provides the mathematically optimal time to exit a patch, yet patch occupancy across organisms is consistently reported to be over-patient compared to optimal (Baum, 1983 [↗](#); Bettinger and Grote, 2016 [↗](#); Guldener et al., 2024 [↗](#); Hall-McMaster et al., 2021 [↗](#); Harhen and Bornstein, 2023 [↗](#); Kilpatrick et al., 2021 [↗](#); Lloyd et al., 2023 [↗](#); Marshall et al., 2013 [↗](#); Nonacs, 2001 [↗](#); Wajnberg et al., 2000 [↗](#)).

1.2 Neural Substrates of Decision-Making and Timing

The basal ganglia play a key role in action selection and option evaluation during decision-making (Cisek, 2007 [↗](#); Mink, 1996 [↗](#); Redgrave et al., 1999 [↗](#)). The primary input nucleus of the basal ganglia, the striatum, receives cortical projections from a wide variety of regions, integrating information about current goals and shifting environmental variables to guide behavior (Balleine et al., 2007 [↗](#); Ferenczi et al., 2016 [↗](#); Frank et al., 2004 [↗](#)). Subregions of the striatum engage with different aspects of optimizing behavior, with the ventral striatum processing reward-related information (Berridge and Robinson, 2003 [↗](#); Salamone and Correa, 2012 [↗](#)), the dorsolateral striatum mediating habit formation (Packard and Knowlton, 2002 [↗](#); Smith and Graybiel, 2013 [↗](#); Yin et al., 2004 [↗](#)), and DMS engaged in goal-directed decision-making (Balleine et al., 2007 [↗](#); Yin et al., 2005 [↗](#)). Lesions of DMS severely impair behavioral flexibility in animals, causing them to lose their ability to adapt to shifting action-outcome contingencies (Castañé et al., 2010 [↗](#); Pisa and Cyr, 1990 [↗](#); Ragozzino, 2007 [↗](#); Torres et al., 2016 [↗](#); Yin et al., 2005 [↗](#)), while electrophysiological recordings show that DMS neurons dynamically encode decision variables, including action values, predicted outcomes, and choice signals (Burton et al., 2015 [↗](#); Ito and Doya, 2015 [↗](#); Kim et al., 2013 [↗](#); Lau and Glimcher, 2008 [↗](#); Parker et al., 2016 [↗](#); Thorn et al., 2010 [↗](#)).

However, much of the decision-making research in DMS uses tasks in which subjects choose between discrete options, such as two-alternative forced choice and multi-armed bandit tasks (Balleine et al., 2007 [↗](#); Bogacz et al., 2006 [↗](#); Samejima et al., 2005 [↗](#); Stott and Redish, 2014 [↗](#); Tang et al., 2022 [↗](#)). Patch foraging presents a different type of decision where the choice measured is a continuous variable. Rather than deciding what action to take, subjects must determine when to take action. The striatum has also been strongly implicated as a region central to timing, with lesions causing impairments in interval timing and temporal discrimination (Matell and Meck, 2004 [↗](#); Meck, 2006 [↗](#); Merchant et al., 2013 [↗](#)). DMS neurons have been shown to encode temporal information through activity patterns such as ramping (Emmons et al., 2017 [↗](#); Ponzi and Wickens, 2022 [↗](#); Vandaele et al., 2021 [↗](#)), and temporal receptive fields (Gouvea et al., 2015 [↗](#); Jin et al., 2009 [↗](#); Mello et al., 2015 [↗](#)). The convergent role of the DMS in both reward-related decision-making and interval timing makes it well-suited for the temporal decisions required during patch foraging.

1.3 Study Overview and Objectives

Here, we investigate how DMS supports decisions about when to abandon a depleting pursuit using a patch-foraging task in freely moving mice. In this task, the optimal policy depends only on total time invested in the patch, independent of individual reward times. Contrary to this prediction, we find that mice adopt a reward-reset timing strategy, in which the decision to leave is governed primarily by the time elapsed since the most recent reward, with systematic modulation by patch residence time and environmental reward rate. This strategy reflects a temporally structured process in which each reward event resets an internal timing process that determines how long the animal is willing to continue waiting.

At the neural level, we identify a corresponding population code in DMS. Individual neurons undergo discrete, reward-triggered transitions between firing-rate states at characteristic delays, collectively tiling the post-reward interval. Across the population, the number of transitioned units increases over time following each reward, forming an accumulation signal that resets with each new reward. The rate of accumulation scales with environmental reward rate and invested patch residence time, and predicts the animal's exit timing, reaching a threshold at the moment of patch departure.

These findings reveal a mechanism by which DMS encodes a dynamically rescaled signal encoding the time to leave, integrating invested patch residence time and environmental reward rate into a unified signal that governs when to stop. More broadly, they suggest that temporal decisions to disengage may be implemented through a time-cost-variable reward-triggered accumulation process, providing a bridge between interval timing and decision-making in naturalistic settings.

2 Results

2.1 Behavioral Characterization of Patch-Foraging Strategy

We began by characterizing time investment behavior in a patch-foraging task that required subjects to decide how to invest time in rewarding pursuits in their effort to maximize the overall reward rate reaped from an environment. Mice were placed in custom behavioral chambers with access to two nose-poke ports, the 'time-investment port' and the 'context port' (Figure 1a). The time-investment port served as the 'patch', where water rewards were randomly delivered from a lick spout with an exponentially decreasing probability over time. Mice invested time in the patch by maintaining their heads in the time-investment port, licking continuously as they sought to harvest rewards. As the likelihood of reward decreased with time, mice could choose to abandon the time-investment port at any moment in favor of the context port. The context port provided a fixed number of regularly spaced rewards over a fixed interval, after which a light cue turned on to indicate that the context port was depleted and the time-investment port had been replenished. As soon as the mouse returned to the time-investment port, the light cue turned off, signaling that the context port was replenished and available, posing the challenge of deciding how much time to invest in the time-investment port. Mice learned to alternate back and forth between the two ports throughout each behavioral session, investing a self-determined amount of time in the time-investment port.

Sessions were broken into blocks where the context port switched between being 'high' value (four rewards in five seconds) and 'low' value (four rewards in ten seconds), while the reward probability distribution in the time-investment port remained constant (Figure 1a). By changing the reward rate and time spent in the context port, we altered the cost of time by manipulating the overall reward rate of the environment (Charnov, 1976; McNamara and Houston, 1985; Namboodiri et al., 2014; Sutlief et al., 2025). Changing the cost of time incentivizes reward-rate maximizing agents to leave the time-investment port earlier during high value blocks and later during low value blocks. Critically, this experimental design shifted the reward-rate optimizing patch-exit strategy solely by manipulating factors external to the patch (Figure 1h). The central behavioral measure of interest was the amount of time invested in the time-investment port during each trial. We found that, on average, mice invested significantly more time in the time-

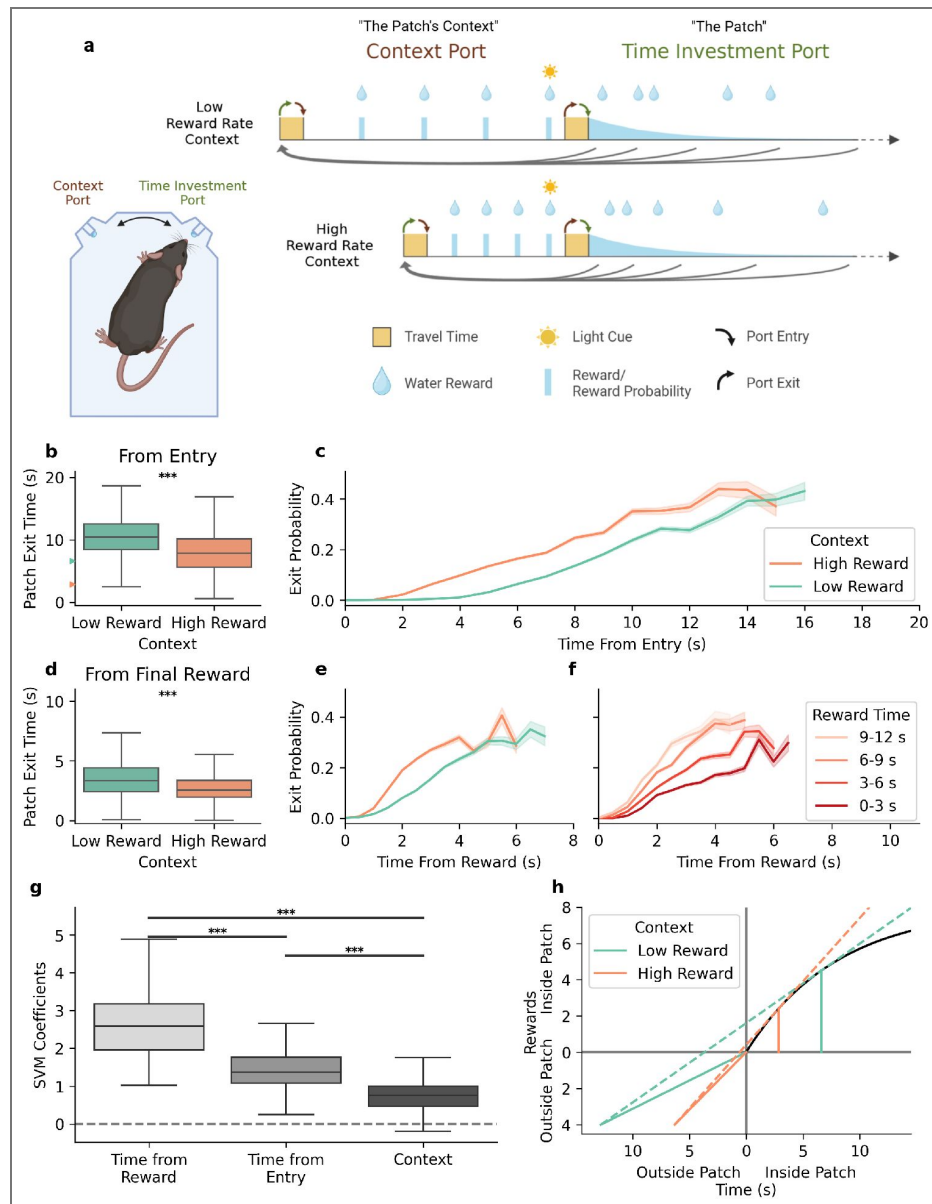


Figure 1. Task structure and behavioral strategy.

(a) Schematic of the behavioral box (left) and structure of the task (right). Mice alternate between collecting rewards from the "context port" and the "time-investment port". In the context port, four rewards are delivered at regular intervals before the port becomes inactive, indicated by a light cue turning on. In the time-investment port rewards are delivered stochastically with exponentially decreasing probability. The context port is reset when the mouse enters the time-investment port, while the time-investment port is reset after the mouse receives all four rewards in the context port. Created in BioRender. Sutlief, E. (2025) <https://BioRender.com/PLACEHOLDER>. (b) Observed exit times from the time-investment port across mice ($n = 10$ mice, 134 sessions, 6,970 trials), measured from the time of entry into the port (two-sample t-test, $p = 1.9 \times 10^{-232}$). Arrows indicate the theoretically optimal exit times for each context (see panel h). (c) The probability (hazard rate) of exiting the time-investment port based on the time since port entry (x-axis) and the context port reward rate (hue), (d) Same as b, except measured from the time of the final reward received in the time-investment port (two-sample t-test, $p = 8.6 \times 10^{-111}$). (e) The probability (hazard rate) of exiting the time-investment port based on the time since the most recent reward (x-axis) and the context port reward rate (hue), (f) The probability (hazard rate) of exiting the time-investment port based on the time since the most recent reward (x-axis) and the time of the most recent reward relative to time of port entry (hue), (g) Support vector machine (SVM) coefficients for predicting time-investment port occupancy. All three predictors are significantly different from zero (one-sample t-tests, all $p < 0.001$) and all pairwise comparisons are significant (paired t-tests, all $p < 0.001$). (h) The theoretically optimal times to exit the time-investment port (vertical lines) for each context block based on the maximum achievable overall reward rates (dashed lines) in accordance with the Marginal Value Theorem (MVT).

investment port during ‘low’ context port blocks compared to ‘high’ context port blocks (Figure 1b), progressively increasing their likelihood of departure as time passed in the time-investment port (Figure 1c). While the direction of this context-dependent shift aligns with theoretical expectations (Charnov, 1976; McNamara and Houston, 1985; Sutlief et al., 2025) and prior observations (Constantino and Daw, 2015; Kilpatrick et al., 2021; Marshall et al., 2013), mice typically stayed in the patch longer than is reward-rate optimal, consistent with the over-patience commonly reported in patch-foraging tasks (Charnov, 1976; McNamara and Houston, 1985; Constantino and Daw, 2015; Hayden et al., 2011; Hutchinson et al., 2008; Lenow et al., 2017; Nonacs, 2001; Wilke et al., 2009). Across sessions, mice achieved an overall reward rate of 28.7 ± 0.2 rewards/min (mean $\pm$ SEM, $n = 134$ sessions), with higher rates during high-context blocks (34.2 ± 0.3) than low-context blocks (23.3 ± 0.2 ; paired t -test, $p = 1.1 \times 10^{-91}$). This corresponded to 88.2 and 93.1 of the reward-rate-optimal policy for high and low blocks, respectively.

2.2 Time-cost-variable, Reward-Triggered Exit Time Policy

Since the probability of reward depends only on the time since entering the port, the optimal departure time is not affected by the occurrence and timing of rewards on any given trial. Nonetheless, the departure strategy exhibited by mice was very sensitive to the timing of rewards received in the time-investment port. While mice were extensively trained on the underlying reward probability distribution, the time at which the most recent reward was received relative to port entry was the strongest factor determining port exit (Figure 1g), followed by patch residence time, and finally by environmental reward-rate context. To better understand how reward events shaped exit policy, we examined the hazard rate of leaving the patch during the interval following rewards, split by context (Figure 1e) and the time of the reward relative to port entry (Figure 1f). We found that all three of these factors contributed to exit behavior, where each reward received would “reset” the amount of time the mouse would be willing to wait for a subsequent reward prior to departure, being determined by the patch occupancy time thus far and the environmental reward-rate context. Mice thus appeared to predominantly time from the last reward, though the time they were willing to wait following each reward decreased the later the reward was received and was shorter under the high than under the low-reward-rate context. This strategy of timing from the last reward rather than patch entry, though suboptimal, is improved by the willingness to wait progressively shorter amounts of time the longer they stay in the patch.

2.3 DMS Neurons Undergo Firing Rate State Transitions Following Rewards

Given that behavior was organized around reward-triggered timing, and that this timing was sensitive to both patch dwell time and time’s cost, we asked whether DMS activity expresses a corresponding representation of the post-reward interval that similarly exhibits these sensitivities, and whether that representation predicts time investment decisions.

Many DMS units showed reward-locked dynamics that resembled ramping when averaged across rewards (Figure 1c, bottom). However, inspection of single-trial responses revealed that these apparent ramps arose from discrete transitions between low and high firing-rate states following reward. This structure became evident when responses were aligned to each neuron’s inferred transition time rather than to the reward itself (Figure 2d). Figure 2a illustrates an example neuron exhibiting a reward-triggered suppression followed by activation, a pattern that recurred across the population.

We developed a classification procedure to identify step-like responses and estimate transition times on individual reward intervals (see Methods). Of 1,849 total units, 898 met criteria for reward-triggered state transitions (Figure 2e); validation with simulated flat-rate Poisson neurons confirmed a false positive rate of 1.35% (Figure 2-figure supplement 3). Across these step-like units, mean transition times were distributed across the post-reward interval, effectively tiling the time axis (Figure 2f). Within units, transition times across rewards were approximately normally distributed, with variability increasing for later mean transition times

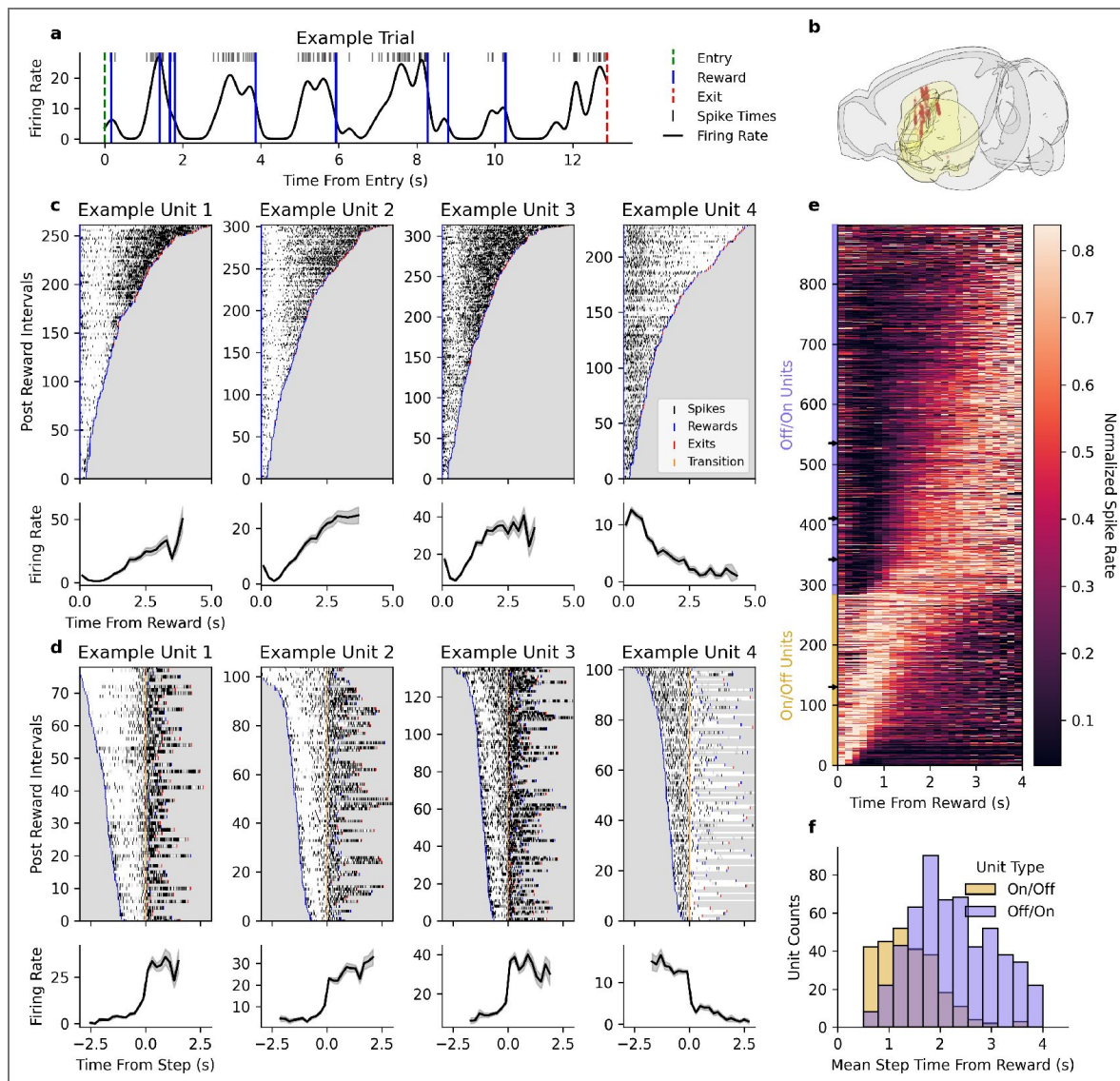


Figure 2. Dorsomedial striatum neurons exhibit step-like activity patterns following rewards.

(a) Example activity of a recorded unit during a single trial in the time-investment port. Top: Example unit spike raster. Bottom: Firing rate trace calculated by binning spikes (10 ms bins) and smoothed with a Gaussian kernel ($\sigma = 100$ ms), (b) Anatomical localization of recorded units in DMS for all recorded neurons (1,849 units across 10 mice), (c) Four example units showing reward responsive patterns. Top: Spike rasters for each reward-reward and reward-exit interval, ordered by the duration of the interval. Bottom: Corresponding mean firing-rate traces aligned to reward delivery (time = 0). (d) Same as c, except showing intervals where each unit exhibits a state-transition, a discrete step in activity post reward, identified by the inflection point of a sigmoid function fit to the firing rate. Rasters and mean firing rate traces are aligned to the detected step times (yellow), (e) Heatmap showing the normalized firing rates of all DMS units that make step transitions following rewards ($n = 898$ step-like units). Units are ordered by the direction of their activity step and their mean step time. Example units are marked with arrows, (f) Distribution of the mean activation times of the units displayed in e.

(Figure 2-figure supplement 1 [↗](#)). Apart from this reward-triggered state-switching activity pattern, several other identifiable task-related activity patterns were observed across the population, with units exhibiting, in varying combinations, responses to a variety of different task aspects, primarily to movement between ports and to licking (Figure 2-figure supplement 2 [↗](#)). Additionally, 320 units were characterized as non-responsive and did not show any identifiable task-related activity patterns.

2.4 Neural State Switches Correlate with Behavioral Exit Policy

We next examined whether variability in neural transition times tracked the animal's departure policy. Transition times covaried with several task parameters, most prominently the timing of reward relative to patch entry and context, with larger parameter correlations among units with later mean transition times (Figure 3-figure supplement 1 [↗](#)).

To relate neural dynamics to the mouse's intended leave time following each reward, we derived a "projected exit time" for each reward event: an estimate of how long the mouse would have waited after that reward if a subsequent reward had not occurred (see Methods section 4.9 [↗](#)). Although exit timing is variable when pooling across sessions (Figure 1b), within-session policy is comparatively stable (Figure 1-figure supplement 1a [↗](#)). We examined several related possible predictors from the reward dynamics in the time-investment port and found that the time of the last reward relative to port entry was the most correlated with leave time from the final reward for individual sessions (Figure 1-figure supplement 1g [↗](#)). We therefore fitted a session-specific regression model predicting leave time from the last reward as a function of context and the time of the last reward relative to port entry (Figure 1-figure supplement 2 [↗](#)), and used the fitted coefficients to project exit times on reward-reward intervals.

Neural transition times tracked these projected exit times. For the four example neurons in Figure 3 [↗](#), the probability of being in the high-firing state shifted later when the projected exit time was later and earlier when the projected exit time was earlier (Figure 3a [↗](#)). For each example neuron, transition time increased approximately linearly with projected exit time (Figure 3b [↗](#)). At the population level (n 898 step-like units), cumulative transition dynamics diverged most strongly when grouped by projected exit time: accumulation was steeper when projected exit time was earlier and shallower when projected exit time was later (all pairwise Kolmogorov-Smirnov tests significant, $p \leq 0.008$; Figure 3c [↗](#)). This grouping produced greater separation than grouping by time since entry (Figure 3d [↗](#)) or by context alone (Figure 3e [↗](#)), consistent with projected exit time being the most behaviorally aligned regressor (Figure 3-figure supplement 1 [↗](#)).

2.5 Accumulation of Neural State Transitions Predicts Exit Time

Finally, we asked whether reward-reset accumulation of state transitions provides a quantitative, trial-by-trial predictor of patch exit. We analyzed 53 sessions with a minimum of seven simultaneously recorded step-like units (see Methods). Figure 4a [↗](#) shows representative trials from an example session with 19 step-like units. Following each reward, units transitioned progressively into their post-step states, producing an accumulation of transitioned units that reset at each reward.

Across reward-to-exit intervals, accumulation dynamics scaled with exit timing: accumulation ramped faster when the mouse exited sooner and slower when the mouse exited later (Figure 4b [↗](#)). Notably, leading up to each exit, the accumulated count approached a similar level (15 transitioned units), consistent with a threshold mechanism. For each reward-to-exit interval, we fit a line to the accumulation trajectory and found that the slope was significantly negatively correlated with exit time (Figure 4c [↗](#)), whereas the intercept was not (Figure 4d [↗](#)).

This pattern generalized across sessions. Accumulation slopes were significantly negatively correlated with exit time (mean $r = -0.551$, one-sample t-test vs. 0: $p = 1.4 \times 10^{-26}$, 48 of 53 sessions individually significant; Figure 4e [↗](#)), while intercepts showed no significant correlation (mean $r = 0.013$, $p = 0.73$; Figure 4f [↗](#)), consistent with a variable accumulation rate from a stable starting point. Using a threshold-crossing rule applied to accumulation trajectories, we generated neurally predicted exit times that explained 39% of variance in observed exit times ($R^2 = 0.386$, $p = 3.1 \times 10^{-258}$;

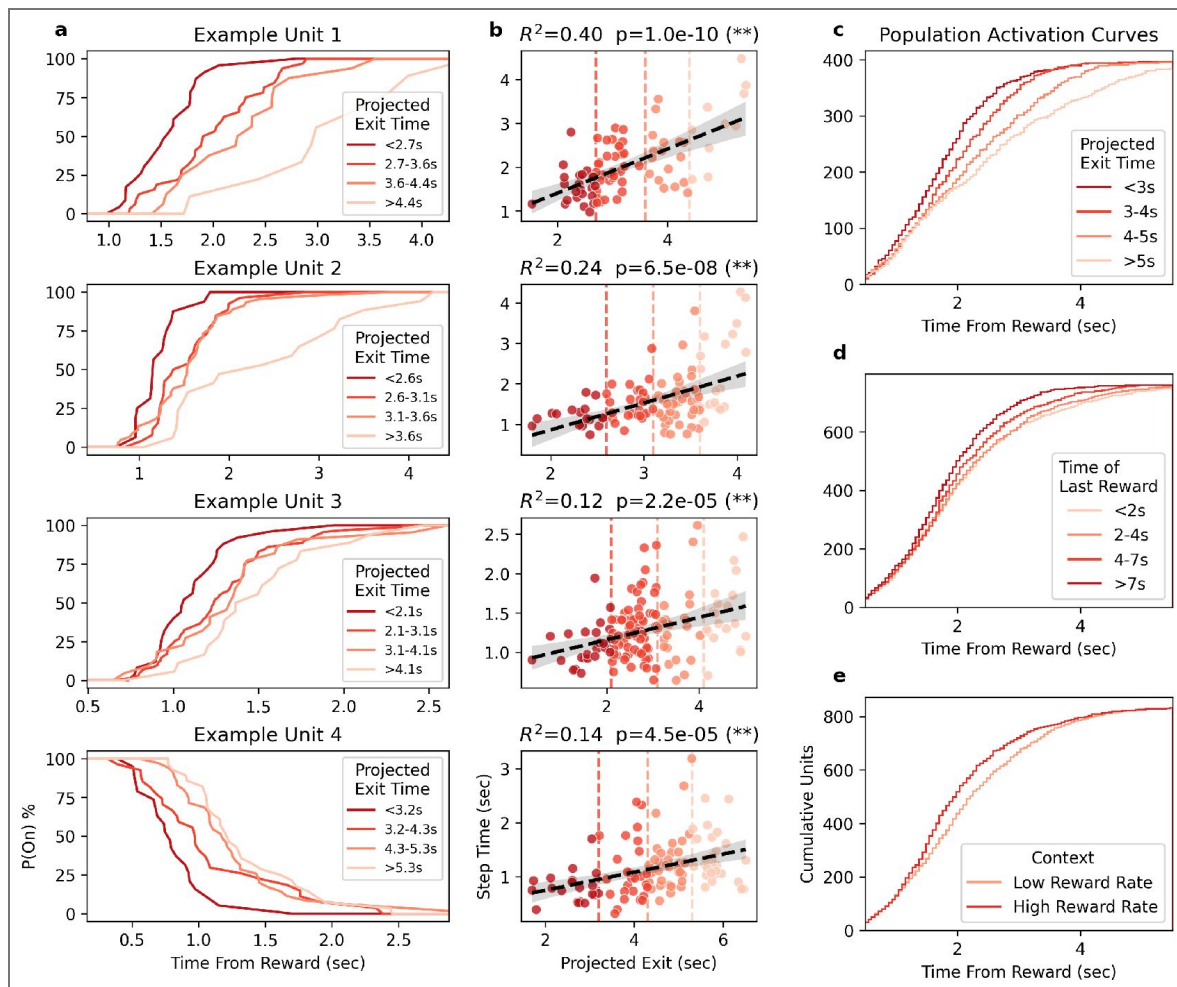


Figure 3. Neural state-transition times correlate with projected exit times.

(a) Plots for each example DMS unit (same as Figure 1c,d) showing the probability of the unit being in the “on” state in the time following reward, separated by quartiles of the projected behavioral exit time (see Figure 1-figure supplement 2). (b) Correlation between unit state-transition times and projected exit times for each example unit. Hue and vertical dashed lines indicate the quartile groupings used in a. (c) Population mean state-transition times ($n = 898$ step-like units) shown as a cumulative function for each of the four projected exit time ranges, including only units with measurements in every time range. All groups differ significantly (Kolmogorov-Smirnov, all $p < 0.008$). (d) Same as c, except separated by the time of the last reward. Only comparisons involving the “>7s” group differ significantly (Kolmogorov-Smirnov; “<2s” vs “>7s”: $p = 1.4 \times 10^{-5}$; “2-4s” vs “>7s”: $p = 3.7 \times 10^{-5}$; “4-7s” vs “>7s”: $p = 0.021$), while remaining comparisons are not significant (all $p > 0.06$). (e) Same as c, except separated by the reward rate in the context port. Groups differ significantly (Kolmogorov-Smirnov, $p = 5.3 \times 10^{-5}$).

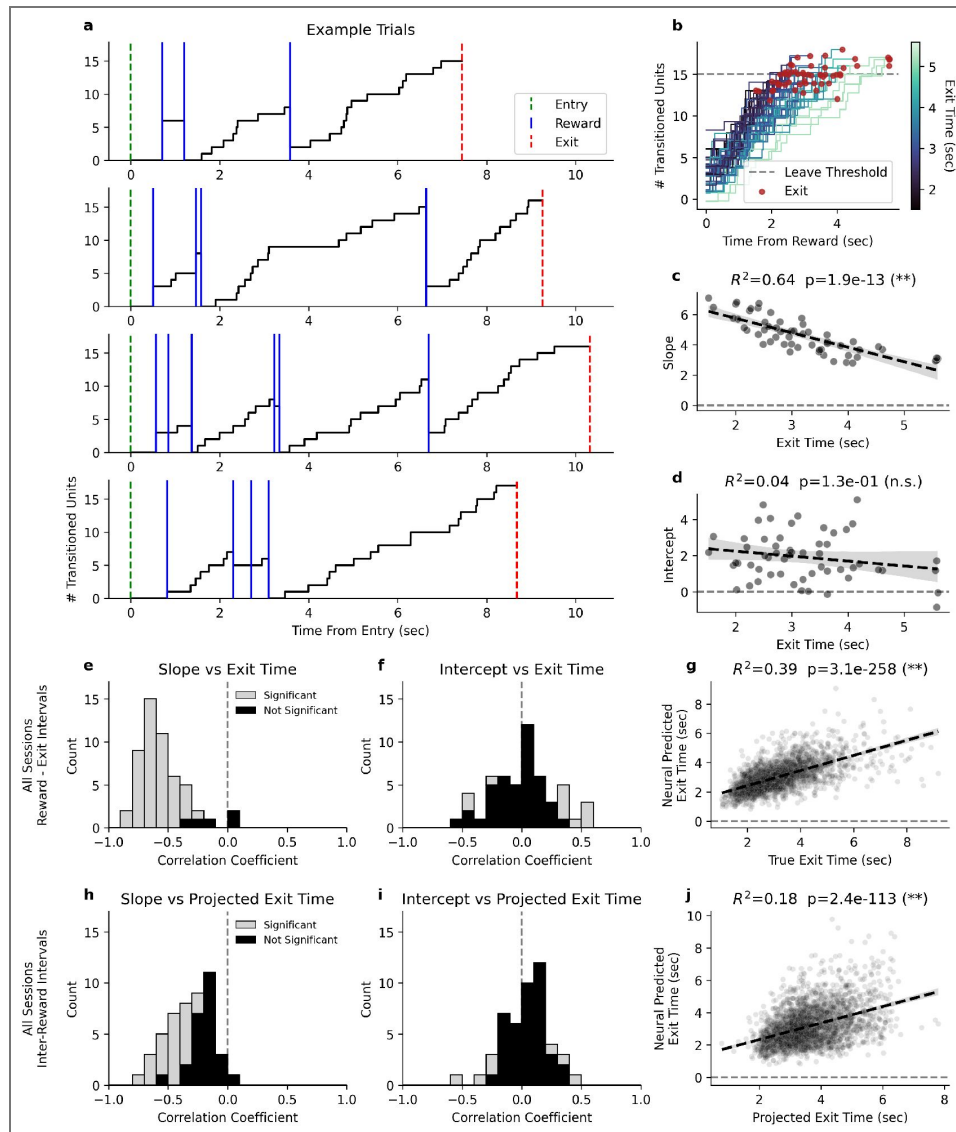


Figure 4. Accumulation of neural state transitions in DMS predicts exit timing.

(a) Four example trials from a single session (ES025, 56 trials, 19 step-like units), showing the cumulative step transitions of simultaneously recorded units. Vertical lines indicate entry (green), exit (red), and reward events (blue), (b) The accumulation lines for all reward-to-exit intervals from the same session as the examples in a. Lines are colored based on the duration of time between the reward and exit events. Exits are marked by maroon dots capping each line. The gray dashed horizontal line indicates the session-specific leave threshold (median final cumulative count at exit), (c) The slope of each accumulation line in b fit with a linear model, plotted against the exit time ($R^2 = 0.64$, $p = 1.9 \times 10^{-13}$, OLS regression), (d) The intercept of each accumulation line in b fit with a linear model, plotted against the exit time ($R^2 = 0.04$, $p = 0.13$, n.s.). (e) Histogram of per-session Pearson correlation coefficients between accumulation slopes and exit times for all sessions with at least 7 step-like units ($n = 53$ sessions from 6 mice; mean $r = -0.551$, one-sample t-test vs. 0: $t = -20.5$, $p = 1.4 \times 10^{-26}$; 48 of 53 sessions individually significant at $p < 0.05$). (f) Histogram of per-session Pearson correlation coefficients between accumulation y-intercepts and exit times ($n = 53$ sessions; mean $r = 0.013$, $t = 0.35$, $p = 0.73$; 11 of 53 individually significant), (g) The relationship between exit time predicted by the accumulate-to-threshold model and true exit times for all reward-to-exit intervals pooled across sessions ($n = 2,418$ intervals; $R^2 = 0.386$, $p = 3.1 \times 10^{-258}$, OLS regression), (h) Histogram of per-session correlation coefficients between accumulation slopes and behaviorally projected exit times during inter-reward intervals ($n = 48$ sessions; mean $r = -0.325$, $t = -12.0$, $p = 5.8 \times 10^{-16}$; 23 of 48 individually significant), (i) Histogram of per-session correlation coefficients between accumulation y-intercepts and behaviorally projected exit times during inter-reward intervals ($n = 48$ sessions; mean $r = 0.039$, $t = 1.38$, $p = 0.17$). (j) The relationship between exit time predicted by the accumulate-to-threshold model and behaviorally projected exit times for all inter-reward intervals ($n = 2,622$ intervals; $R^2 = 0.178$, $p = 2.4 \times 10^{-113}$, OLS regression).

Figure 4g). Applying the same analysis framework to inter-reward intervals (excluding reward-to-exit intervals) and comparing accumulation dynamics to behaviorally projected exit times yielded similar results (Figure 4h-j). Together, these results identify a reward-reset accumulation-to-threshold computation in DMS that integrates elapsed time, reward history, and environmental context to predict when the animal will abandon a depleting pursuit.

3 Discussion

Decisions about when to disengage from an ongoing activity are fundamental to adaptive behavior, yet the neural mechanisms that govern such temporal decisions remain poorly understood. Here, we show that during patch foraging, mice adopt a time-cost-variable, reward-reset timing policy, and that this policy is reflected in a corresponding neural computation in dorsomedial striatum (DMS). Specifically, we identify a population-level signal in which discrete, reward-triggered neural state transitions accumulate over time and reach a threshold at the moment of patch exit. The rate of this accumulation varies systematically with the cost of time and predicts exit timing on a trial-by-trial basis. Together, these findings show how DMS implements an accumulation-to-threshold computation that integrates elapsed time from reward and elapsed time from pursuit initiation with environmental reward rate to determine when to depart.

3.1 Neural Mechanisms of Temporal Decision-Making

At the individual neuron level, we observed step-like responses in DMS neurons, differing from the ramping or sparse temporal response dynamics typically associated with timing processes. Previous studies on interval timing in the striatum have predominantly reported firing rates that increase steadily over the course of the interval (Gouvea et al., 2015) or show sparse encoding of time fields (Mello et al., 2015). At the population level, our findings indicate that the striatum exhibits activity patterns that accord with the features of the behavioral strategy employed, where the intended interval to depart from a patch appears as timed from the last reward, though modified by the context in which the patch is embedded, and by the amount of time already spent in the patch. Specifically, the number of transitioned units resets and accumulates linearly after each reward, with the slope of this accumulation predicting the actual exit time from the patch. As with behavior, the time already spent in the patch as well as the patch context impact timing, affecting the rate of accumulation of transitioned units. When rewards occur earlier in the patch, the slope of accumulation is shallower, predicting greater time investment, whereas if rewards occur later in the patch, the slope of accumulation is steeper, predicting early departure times. When the time spent outside the patch in the context port is greater, the reward rate of the context port is lower, so that the cost of time is lower, resulting in a greater willingness to invest time in the patch. This is reflected in the correspondingly lower rates of accumulation of transitioned units in comparison to when the patch's context is 'high'. Collectively, these observations indicate a neural mechanism in which the DMS conveys the intended waiting interval after a reward, integrating elapsed time since last reward, elapsed time since patch entry, and the context of the patch.

3.2 Relationship to Accumulation-to-Bound Frameworks

The accumulation-to-threshold dynamics we observe in DMS share formal features with drift-diffusion and related accumulation-to-bound models, which have been widely applied to perceptual and valuebased decisions (Shadlen and Newsome, 1996; Gold and Shadlen, 2007; Brody and Hanks, 2016). In those frameworks, noisy evidence accumulates until a bound is reached, triggering a commitment to action. Here, the accumulated quantity is not sensory evidence but the cumulative count of discrete, reward-triggered neural state transitions representing the passage of time, with the bound corresponding to the threshold at which the animal departs the patch. This parallel suggests that accumulation-to-threshold computations may serve as a general motif across decision domains, extending from perceptual discrimination to self-timed decisions about behavioral persistence. Theoretical work has explored this connection deeply (Church and Deluty, 1977; Gibbon, 1977; Gibbon et al., 1984; Killeen and Fetterman,

1988 [\[1\]](#); Matell and Meck, 2000 [\[2\]](#); Namboodiri et al., 2014 [\[3\]](#); Simen et al., 2011 [\[4\]](#); Treisman, 1963 [\[5\]](#)), with varying models framing interval timing as an accumulation-to-threshold process that differ mainly in the form and generation of time's representation and how it is read-out. Additionally, models such as TOPDDM (Simen et al., 2013 [\[6\]](#)) and TIMERR-DDM (Namboodiri et al., 2016 [\[7\]](#); Namboodiri and Hussain Shuler, 2016 [\[8\]](#)), inspired by the Behavioral Theory of Timing (BeT) (Killeen and Fetterman, 1988 [\[9\]](#)), adapt their timing process to the temporal structure of reward within an environment. The linear accumulation to a singular threshold we observe here accords with TOPDDM, for whereas TOPDDM encodes interval timing by adjusting the rate of linear accumulation to a threshold, TIMERR-DDM encodes interval timing by adjusting the threshold of a non-linear accumulation process. Finally, the rate-scaling we observe, in which accumulation is faster under higher environmental reward rates and longer patch residence times, is also consistent with the proposal that the average reward rate impacts the cost of time, modulating the vigor of behavioral and neural processes (Niv et al., 2007 [\[10\]](#)).

Several prior findings in the dorsal striatum converge with our observations. Gouvea et al. (2015 [\[11\]](#)) showed that striatal population dynamics predict duration judgments, and Mello et al. (2015 [\[12\]](#)) identified a scalable population code for time in striatum. Emmons et al. (2017 [\[13\]](#)) reported ramping activity in DMS during interval timing and argued that frontostriatal temporal signals are consistent with drift-diffusion models. Regarding decision-making, Yartsev et al. (2018 [\[14\]](#)) demonstrated a causal role for anterior dorsal striatum in evidence accumulation, and Bolkan et al. (2022 [\[15\]](#)) showed that DMS pathway contributions to behavior depend on cognitive demand, with strong effects when tasks require evidence accumulation. Together, these studies establish that the dorsal striatum supports both temporal processing and accumulation-dependent decision-making. Our findings extend this work by identifying, in a naturalistic foraging context, a reward-reset accumulation signal in DMS for self-timed patch departure, implemented not as continuous rate changes but as the cumulative recruitment of neurons undergoing discrete post-reward state transitions.

3.3 Reward-Reset Strategy and Behavioral Optimality

In contrast to the behavioral predictions of classical optimal foraging theory, we observed the use of a reward-reset strategy, where the time spent in the port before a reward predicted how long the mouse would wait following that reward. According to the Marginal Value Theorem, the optimal patch departure time occurs when the instantaneous reward rate falls below the environmental average (Charnov, 1976 [\[16\]](#); Kolling et al., 2012 [\[17\]](#)), where the timing of individual rewards in a fully stochastic environment should have no effect. However, mice consistently used the time since the most recent reward as their main decision variable, similar to the strategies previously observed in parasitic wasps (Haccou et al., 1991 [\[18\]](#); van Alphen et al., 2003 [\[19\]](#); Wajnberg et al., 2000 [\[20\]](#)). Other variables, such as the reward-rate context and patch occupancy time, had modest but significant effects on the time to exit following each reset, demonstrating sensitivity to the time-evolving value of the patch, and to the environmental reward-rate context.

The suboptimality we observed (mice staying in patches longer than predicted by MVT) is consistent with previous reports of “over-patience” in patch-foraging tasks (Constantino and Daw, 2015 [\[21\]](#)). This pattern may reflect some inherent cognitive constraints in temporal decision-making (Cuthill et al., 1990 [\[22\]](#); Kacelnik and Krebs, 1985 [\[23\]](#); Krebs et al., 1974 [\[24\]](#)) or could represent an adaptive response to environmental uncertainty (Kacelnik and Todd, 1992 [\[25\]](#); Kacelnik and Bateson, 1996 [\[26\]](#); Olsson and Brown, 2006 [\[27\]](#)), where a longer occupancy time allows for more information collection before a certainty threshold is crossed. The reward-reset strategy may be particularly advantageous in environments where patch quality varies unpredictably, since recent rewards would provide more reliable information about the patch's current reward rate than a previously accumulated estimate of environmental statistics. While this is notably not the case for the task described in this study, the observation that mice use a reward-reset strategy nonetheless may suggest a generalized approach to similar foraging environments.

3.4 Conclusion

In summary, we identify a neural mechanism in which discrete, reward-triggered state transitions in DMS accumulate to a threshold that determines when to abandon a pursuit. This time-cost-variable, reward-reset accumulation provides a parsimonious account of how temporal information, reward history, and environmental context are integrated to guide decisions about when to stop. We suggest that DMS plays a central role in generating the temporal signals that underlie time-investment strategies during foraging, with downstream circuits likely converting this timing representation into the motor action to depart. Our findings are correlational, and causal manipulations will be needed to determine whether the accumulation dynamics we observe are necessary for patch departure decisions or instead reflect timing signals inherited from cortical or other inputs to DMS. Additionally, the generality of the reward-reset accumulation mechanism across different foraging environments and task structures remains to be tested. Nonetheless, by linking behavioral strategy to a specific population-level neural computation, these findings bridge interval timing and decision-making, and offer a framework for understanding how the brain controls persistence and disengagement in dynamic environments.

4 Methods

4.1 Compliance Statement

All animal procedures were conducted in accordance with the National Institutes of Health guidelines and were approved by the Johns Hopkins School of Medicine Institutional Animal Care and Use Committee (IACUC).

4.2 Subjects

Ten C57BL/6J mice (4 male, 6 female) were used in this study. Mice were purchased from The Jackson Laboratory and housed in standard vivarium cages. All animals had ad libitum access to food and water, except during periods of behavioral training and electrophysiological recording, when water was restricted. All animals maintained 85-90% of their original body weight throughout water restriction.

4.3 Task Design

Mice were trained on a patch-foraging task designed to specifically observe and measure the moment when a mouse decides to give up on a depleting patch (Figure 1a). The timing of that decision optimally depends on the relative expected value of remaining in the patch compared with the overall expected value of the environment. In our task, mice are required to harvest from a patch under low and high reward-rate environments while the underlying reward probability function within the patch remains consistent. Broadly, the task requires the mouse to switch back and forth between two ports as it harvests water rewards. The two ports are the time-investment port, which functions as a patch with reward delivered stochastically according to an exponentially decaying distribution, and the context port, which delivers four evenly spaced rewards over either a short or a long interval, depending on block.

The task is performed in a custom-built 7" x 5" box where the mouse has access to two ports into which it may poke its head. The ports are equipped with lick spouts that deliver 1 μ L water rewards, infrared sensors that detect head entry and licks, and LED light indicators that signal whether or not the port is active (LED off = port is active/rewards are available, LED on = port inactive/rewards are not available). The mouse begins in the first of the two ports, the context port, where it receives 4 water rewards regularly spaced over either 5 or 10 seconds, depending on the current block. After the final reward is delivered, the context port light turns on and the time-investment port light turns off, indicating the mouse may then switch ports. Travel between the ports takes approximately 300 ms. Upon entering the time-investment port, the context port light turns off, signaling that the mouse may abandon the time-investment port in favor of the predictable context port rewards at any time. For as long as the mouse chooses to stay, the time-

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Biological Samples	C57BL/6J mice	The Jackson Laboratory	RRID: IMSR_JAX:000664	10 mice (4 male, 6 female) used for electrophysiology. Water restricted during training/recording.
Equipment	Neuropixel 2.0 (multi-shank) probes	IMEC	2.0 multishank probes	Used for extracellular activity recording in dorsomedial striatum (DMS).
Equipment	Custom 3D printed headcap and probe insertion assembly	Custom	N/A	Used for Neuropixel implantation surgery.
Equipment	Custom-built behavioral chambers	Custom	N/A	7" x 5" box with nose-poke ports, lick spouts, infrared sensors, and LEDs.
Equipment	Raspberry Pi	Amazon	Model 4b	Used for custom task logic and control of sensors, LEDs, and solenoids.
Software	Custom Python scripts	Custom	N/A	Used for controlling task logic and recording time-stamped events.
Software	DataJoint	<i>Yatsenko et al. (2021)</i>	RRID: SCR_014543	Cloud-based data pipeline for preprocessing and spike sorting.

Key resources table.

investment port delivers rewards probabilistically, with the likelihood of reward decreasing exponentially as time passes. The cumulative probability of the exponential decay curve sums to eight, meaning that on average, staying indefinitely in the time-investment port would yield eight rewards, while the actual number of rewards that can be delivered is unbounded. Eventually, the moment of experimental interest arrives, being when the mouse decides to exit the time-investment port and return to the context port. As soon as it reenters the context port, the time-investment port light turns on and a single cycle (trial) is complete. Each time the mouse consumes a set of four context port rewards, the time-investment port reward probability is refreshed to its starting value, allowing the mouse to continuously harvest by alternating between the ports.

Mice learned to operate the task without the need for any shaping steps. To consider a mouse fully trained, we required that they did not leave the context port prematurely on more than 20% of the trials and that they move to the time-investment port within 3 seconds of the light cue turning on in the context port. This ensured that they experienced the intended reward rate when deciding when to leave the time-investment port. Mice typically reached criterion within one to two months of daily training (two sessions per day). No requirements are placed on their strategy in the time-investment port.

Each behavior session lasted 18 minutes and was divided into six three-minute blocks. Across blocks, the dynamics of the context port alternated between providing a high reward-rate context (4 rewards in 5 seconds) and a low reward-rate context (4 rewards in 10 seconds). The value of the starting block ('low' or 'high') was randomly determined for each session. The reward probability distribution of the time-investment port remained consistent throughout all trials across all mice. All of the logic of the task, as well as control of the sensors, LEDs, and solenoids, was custom-written in Python and deployed on a Raspberry Pi. All time-stamped events were saved to a text file for later analysis.

4.4 Analysis - Support Vector Machine (SVM) Model

To understand the factors contributing to the decision of when to leave the time-investment port, a Support Vector Machine (SVM) classifier was used to identify the conditions under which a mouse would either stay or leave. Three factors were supplied to the SVM: The time since entry into the time-investment port, the time since the most recent reward in the time-investment port, and the reward-rate context (either low or high). Each trial from a given session was divided into 10 ms time bins and assigned to the 'Stay' group. The five seconds following each exit was similarly divided and assigned to the 'Leave' group. The SVM then used a linear kernel to calculate the plane that best separated the two groups. The coefficients of each normalized feature were then compared across sessions to determine the relative importance of the time from reward, the time from entry, and the reward-rate context to the patch-leave decision strategies used by mice across sessions.

4.5 Surgery for Neuropixel Implantation

Mice underwent stereotaxic surgery to implant Neuropixel 2.0 (multi-shank) probes using a custom 3D printed headcap and probe insertion assembly. All surgeries were performed in aseptic conditions and followed an IACUC approved protocol. Animals were anesthetized with isoflurane and their body temperature was maintained at 37°C. Ophthalmic ointment was applied to the eyes and Lidocaine (0.0021 mL/g) was injected subcutaneously under the scalp. Fur was removed from the scalp before it was disinfected with alternating Betadine and 70% isopropyl alcohol. A midline incision exposed the skull, with excess skin removed to expose a large surface area of skull to ensure the headcap could be securely attached. The periosteum was cleared with alternating 3% hydrogen peroxide and 70% isopropyl alcohol. The skull was then leveled and the craniotomy target (dorsomedial striatum, AP: 0.7 mm, ML: 1.2 mm) was measured from bregma. The skull surface was thoroughly dried and a thin layer of Metabond dental cement was applied to the entire exposed surface, which was given 10 minutes to fully set. The craniotomy target was then marked on the Metabond, and an oval section (approximately 1.2 mm × 1 mm) was drilled through both metabond and skull. A separate small hole was drilled for a ground pin posteriorly on the

contralateral side. A small section of dura was carefully removed from the craniotomy. The custom headcap was then placed over the craniotomy, with an oval lip fitted inside to prevent skull regrowth where the Neuropixel probe would be inserted. A gold grounding pin was placed in the posterior hole. Metabond was used to secure both headcap and pin to the skull, ensuring the craniotomy and ground pin hole were fully enclosed. Buprenorphine (0.0083 mL/g) was then administered subcutaneously on the lower back, approximately 5 minutes prior to the end of the surgery. The Neuropixel 2.0 probe was then slowly inserted into DMS, using a built-in screw and shuttle system custom designed to interface with the headcap. The probe was lowered to a target depth of 5 mm at a rate of 1 mm/min. The ground pin was connected to the grounding pad of the Neuropixel probe via a soldered silver wire and the Neuropixel housing assembly was secured to the headcap with two external screws. The weight of the entire piece reached approximately 3 grams. Mice were supplied with hydration gel and monitored as they recovered in a heated cage until they were fully awake.

4.6 Electrophysiology

Neural activity was recorded from DMS of fully trained mice using Neuropixels 2.0 multi-shank probes. Each probe has four 10 mm shanks, with a total of 5120 recording sites across them. The acquisition system used included a National Instruments PXIe-1071 chassis, with a PXIe-8381 module. Neural signals were recorded with a sampling rate of 30 kHz across 384 channels using SpikeGLX software. Before recording, activity across the probe shanks was observed. The surface of the brain was located and used to estimate the band of recording sites most likely to cover the DMS, which the 384 channels were then assigned from which to record. Surface row information was saved to later validate histologically determined unit locations.

Probes were implanted using custom 3D printed fixtures for chronic, freely moving recordings, and the locations of all identified units were verified *post hoc* (Figure 2b). After recovery from surgical probe implantation, mice were retrained and DMS activity recorded for up to a month, depending on stability. Recordings were preprocessed using built-in methods from SpikeInterface (Buccino et al., 2020) and sorted using a consensus-based algorithm combining Kilosort 2.5 and Kilosort 3 (Pachitariu et al., 2016), all of which were orchestrated and managed with the DataJoint infrastructure (Yatsenko et al., 2021).

4.7 Histology

At the end of the experiment, mice were anesthetized with isoflurane and injected with sodium pentobarbital (200 mg/kg). Mice were then transcardially perfused with phosphate-buffered saline (PBS), then 4% paraformaldehyde (PFA). Brains were removed and placed in 4% PFA overnight, then transferred to PBS. Brains were sliced coronally at 50 μ m using a vibratome (Leica VTI000 S), then mounted on glass slides. Prior to surgery, the probes were coated in Vybrant DiI (V22885) and the tracks were then imaged using a fluorescence microscope (Keyence BZ-X800). Probe track locations were reconstructed within the Allen Mouse Brain Atlas using HERBS (<https://github.com/Whitlock-Group/HERBS>).

4.8 Neural Data Preprocessing and Spike Sorting

Recording sessions were processed using a cloud-based pipeline implemented with DataJoint (Yatsenko et al., 2021) and running on AWS cloud services. Code for the preprocessing and sorting steps is publicly available at https://github.com/dj-sciops/jhu_shulerab_element-array-ephys.

Preprocessing was conducted using the SpikeInterface toolbox. First, a bandpass filter was applied to the raw signals to remove frequencies below 300 Hz and above 6000 Hz. Broken channels were identified and interpolated over so they would not disrupt the spike sorting algorithm. Phase shift correction was applied to account for the phase delays between channels during the system's acquisition cycles. Lastly, common average referencing (CAR) was applied across all channels to reduce common noise artifacts.

Spike sorting was performed independently with Kilosort 2.5 and Kilosort 3. Because the two algorithms complement each other in detecting different signal types, we used the SpikeInterface consensus method to identify units that were reliably detected by both sorters, ensuring that only high-confidence units were carried forward. Quality metrics were calculated for all consensus units, then assessed during manual curation performed using Phy, where remaining noise units were removed. The final set of curated units for each session was stored in the cloud pipeline and made available for analysis. Software versions for all preprocessing, spike sorting, and analysis pipelines are specified in the accompanying code repositories (see Data Availability Statement). Statistical analyses were performed in Python 3.12 using SciPy 1.17.0, NumPy 1.26.4, scikit-learn 1.8.0, and statsmodels 0.14.6.

4.9 Analysis - Behaviorally Projected Exit Time Calculation

To predict how long a given mouse on a given session would be willing to wait following each reward, we fit the leave time following each final reward (the dependent variable) as a function of when that reward occurred relative to port entry (the predictor), using simple linear regression.

$$\text{Leave Time From Reward} \sim \text{Reward Time From Port Entry}$$

Separate models were fit to trials from each context block, to allow for shifts in both the slope and intercept of the correlation. We then used the regression models to predict how long following every reward the mouse would have been willing to wait, if no further rewards had been delivered on that trial.

4.10 Analysis - Neural Step-like Classification

To characterize neurons exhibiting step-like responses at varying temporal delays from reward, an algorithm was developed to identify high and low firing rate states across trials and precisely detect the timing of transitions between states. To do so, first the instantaneous firing rate of individual neurons was calculated as one over the interspike interval for 10 ms time bins. For each unit, all of the post-reward responses in the time-investment port were pooled together and a sigmoid curve was fitted to the unsmoothed firing rate data. The parameters (y-offset, height, slope, and center) of this session sigmoid were then used as initial parameters for sigmoid fits to every interval individually. Parameter bounds were set to force the individual interval (reward-to-reward and reward-to-exit) sigmoid fits to match the high and low firing-rate states identified in the session sigmoid fit. The absolute value of the sigmoid slope was required to be greater than 11 (in units of $\Delta(\text{spikes/second})/\text{second}$), preventing the sigmoid from fitting ramping activity with shallow curves. The center parameter (inflection point of the sigmoid) was bounded to the duration of the interval plus or minus 10%. The primary criterion for a successful fit was for the sigmoid to fit the center within the interval (at least 100 ms after the start and 100 ms before the end). On intervals where there was no clear step, the sigmoid's best fit pushed the center to one of the far edges of the interval, essentially fitting the interval as entirely low firing or entirely high firing, due to the strict bounds on the other parameters. Because intervals could be shorter than a given unit's typical transition latency, classification was restricted to intervals long enough for a transition to plausibly occur. Specifically, units were classified as step-like if the algorithm successfully identified a step on at least 50% of intervals whose duration exceeded the unit's mean transition time plus one standard deviation of its transition time distribution. To validate this pipeline, matched homogeneous Poisson process neurons (flat-rate, no step structure) were generated for each recorded unit and processed through the identical classification procedure, yielding a false positive rate of 1.35% (95% CI [0.85%, 2.04%]; Figure 2-figure supplement 3 [↗](#)).

4.11 Analysis - Neural Accumulation

Sessions with at least seven simultaneously recorded step-like units were included in the accumulation analysis. This minimum was chosen to ensure sufficient population sampling for reliable linear regression of cumulative transition counts; results were qualitatively consistent across a range of inclusion thresholds. The cumulative number of units in their post-transition state was measured over the course of each trial. The slope and intercept of the accumulation

across each reward-reward and reward-exit interval was calculated using simple linear regression. To predict the leave time from the neural activity, a single cumulative unit threshold per session was calculated as the median number of transitioned units at the time of each true exit within that session. This approach was selected after evaluating several threshold-setting methods; a single session-wide median threshold sufficiently captured the accumulation-to-threshold pattern observed in the data. The predicted leave time for each interval was then calculated as the intersection of its regression line with the session threshold.

Figure Supplements

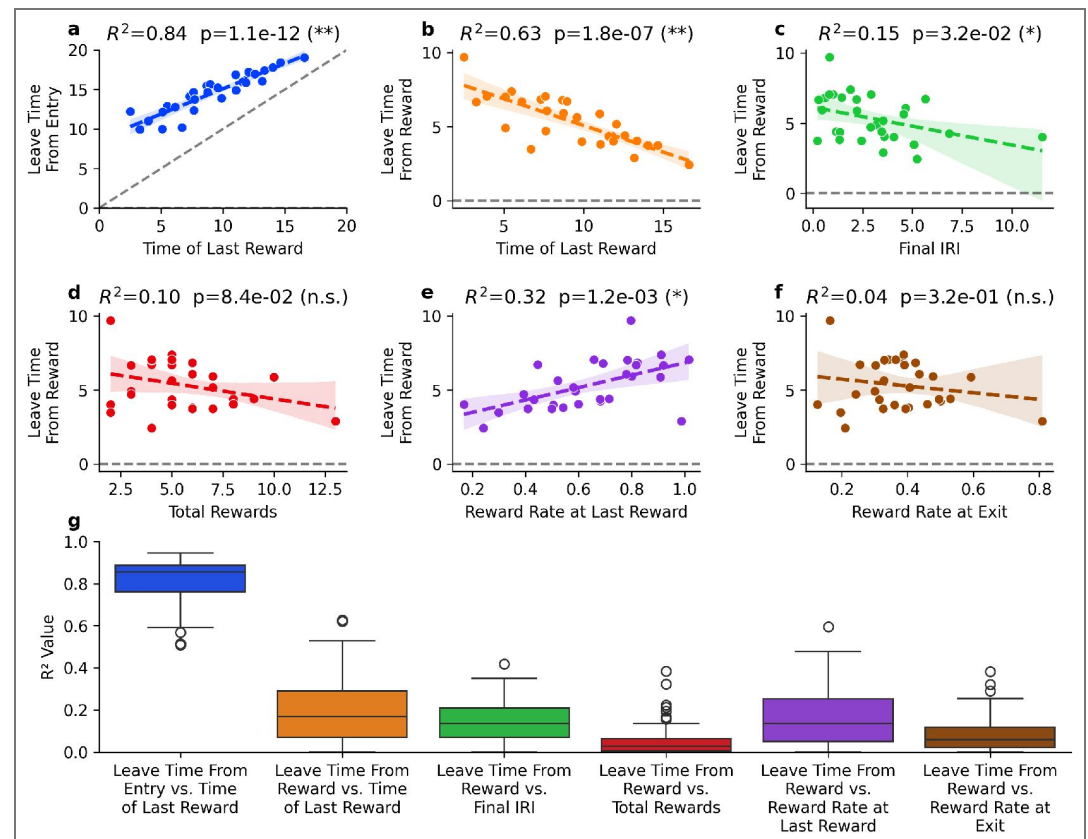


Figure 1-figure supplement 1. Session-specific behavioral correlates of patch exit timing. a-f) Linear regression analysis of potential predictors of patch exit behavior from an example session, a) Leave time from port entry plotted against the time of the last reward relative to entry, b) Leave time from the final reward plotted against the time of the last reward relative to entry, c) Leave time from the final reward plotted against the duration of the final inter-reward interval, d) Leave time from the final reward plotted against the total number of rewards received in the time-investment port during that trial, e) Leave time from the final reward plotted against the reward rate experienced in the time between port entry and the final reward, f) Leave time from the final reward plotted against the reward rate experienced in the time between patch entry and patch exit, g) Distribution of correlation strengths (R^2) across all recording sessions for each of the six behavioral predictors analyzed in panels a-f. Each data point represents the R^2 value from the linear regression for a single session. Box plots show median, quartiles, and range of correlation strengths.

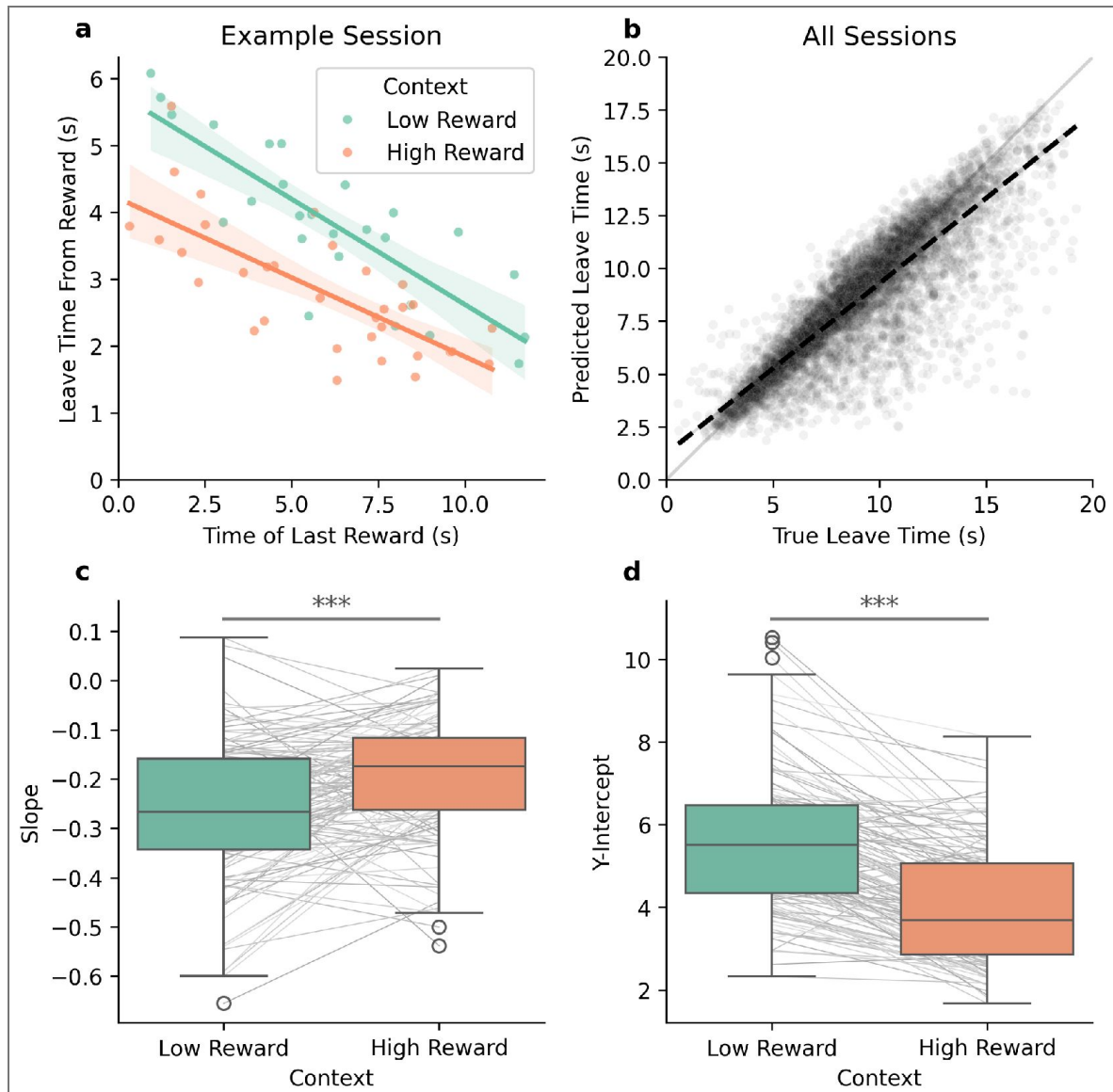


Figure 1-figure supplement 2. Session-specific linear regression models predict behavioral exit policy following rewards.

(a) Leave time from the final reward plotted against the time of the final reward relative to port entry for an example session. Data points are separated by context block reward rate with fitted linear regression lines, (b) Predicted leave times from port entry calculated using session-specific linear regression coefficients plotted against observed leave times across all sessions ($R^2 = 0.75$). (c) Comparison of regression line slopes between high and low reward-rate context blocks across sessions ($n = 134$ sessions, $p = 8.5 \times 10^{-8}$, paired t-test). (d) Comparison of regression line intercepts between high and low reward-rate context blocks across sessions ($n = 134$ sessions, $p = 2.6 \times 10^{-27}$, paired t-test).

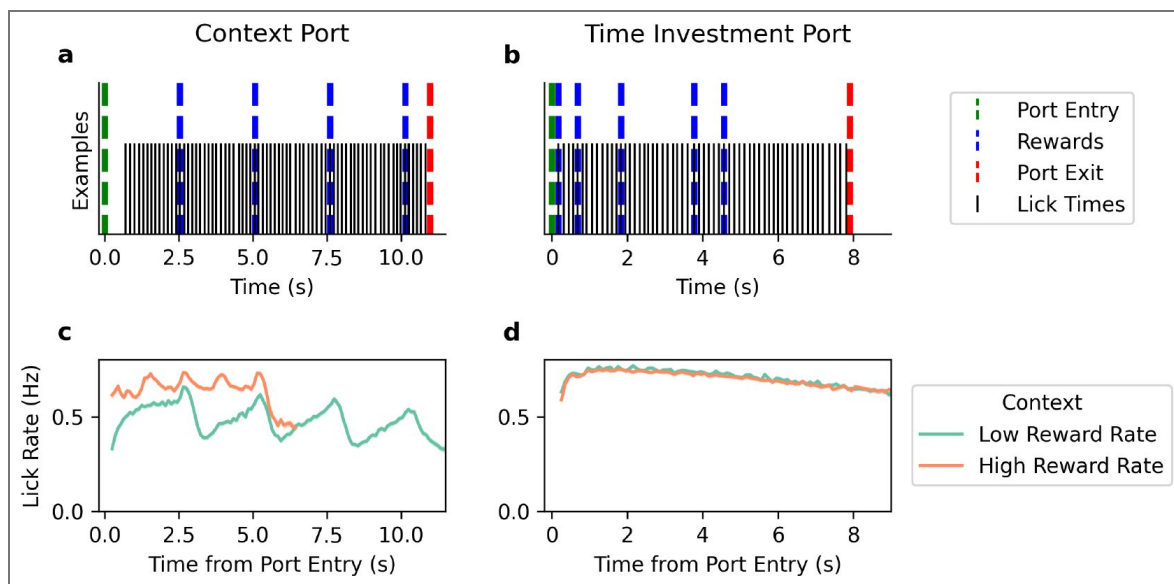


Figure 1-figure supplement 3. Lick patterns at the context and time-investment ports.

(a) Example single-trial lick pattern during a visit to the context port. Black tick marks indicate individual lick times. Dashed vertical lines mark port entry (green), reward deliveries (blue), and port exit (red), (b) Same as (a) for a visit to the time-investment port, (c) Mean lick rate as a function of time from port entry at the context port, plotted separately for low (green) and high (orange) reward-rate contexts. Shading indicates ± 1 SEM across intervals, (d) Same as (c) for the time-investment port. At both ports, licking begins shortly after entry and is sustained throughout the visit. At the context port, lick rate is higher and more sustained in the low reward-rate context, consistent with the longer visits observed in that condition.

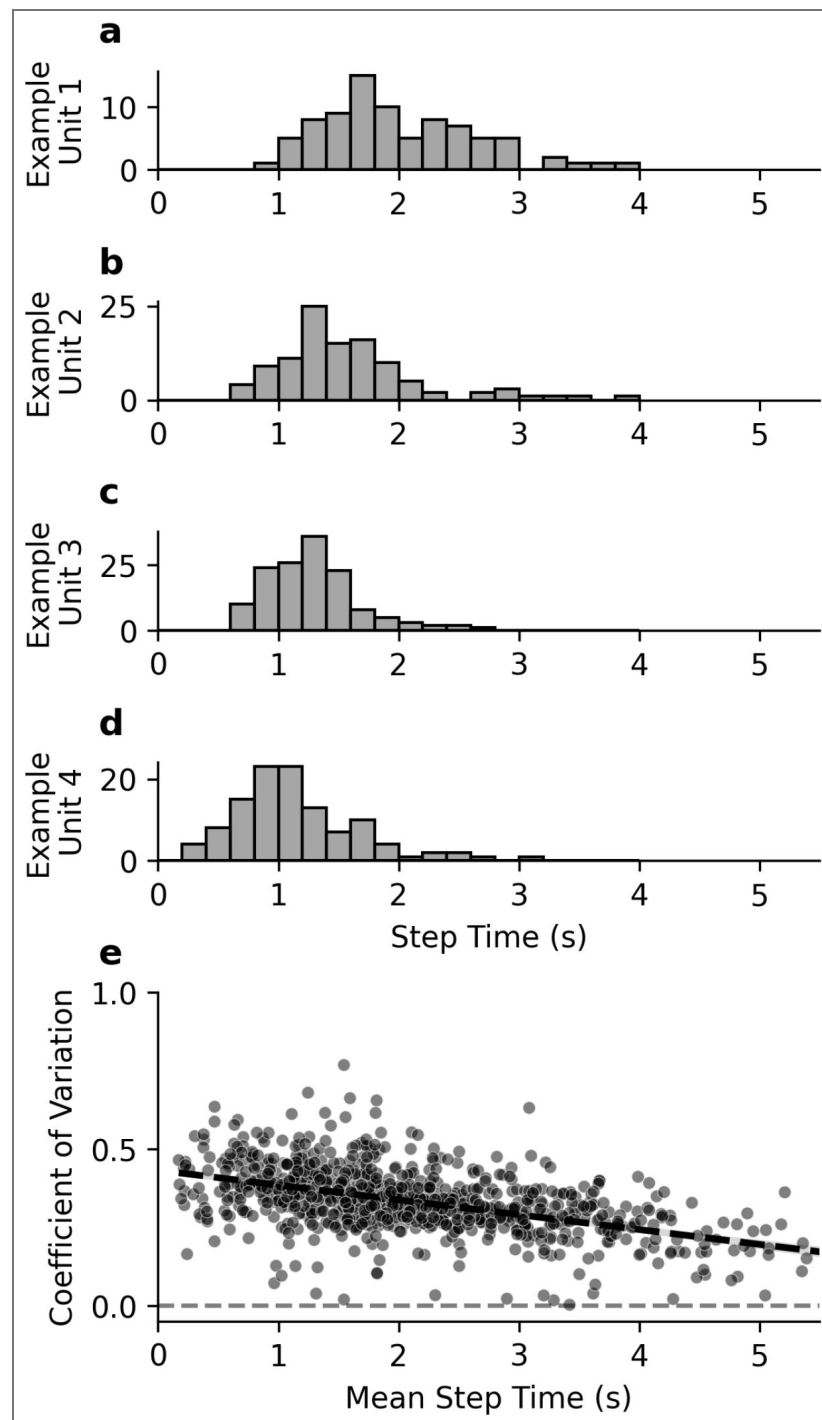


Figure 2-figure supplement 1. Step time distributions and variability scaling in DMS neurons.

(a-d) Histograms showing the distribution of step times for the four example DMS units from Figures 2 and 3. (e) Relationship between the mean step time and coefficient of variation (CV) of step times across all step-like units (n = 898 units). Linear regression line is shown in black ($R^2 = 0.281$, $p = 2.7 \times 10^{-66}$, OLS regression). Later-stepping units tend to have lower temporal variability relative to their mean step time.

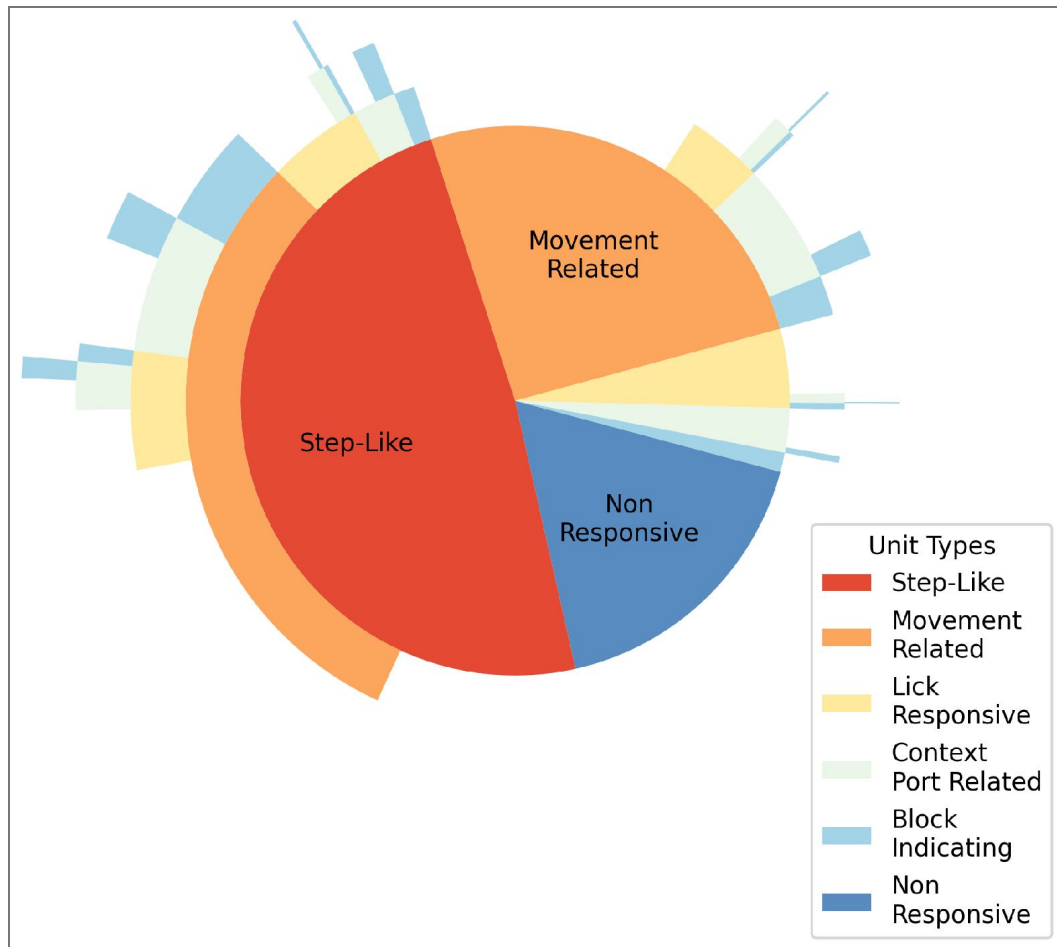


Figure 2-figure supplement 2. Classification and overlap of DMS unit response types during patch-foraging behavior.

Sunburst diagram showing the distribution of response types among all recorded DMS units ($n = 1,849$). Stop-like units exhibit detectable stop transitions in firing rate during the post-reward period in the time-investment port. Movement-related units show modulated activity during locomotion into or out of either port, which could be for general travel or port-specific entry/exit responses. Lick-responsive units display firing modulation phase-locked to the lick cycle during continuous licking behavior throughout port occupancy. Context port-related units exhibit activity modulation during the fixed inter-reward intervals in the context port or responses to the depletion cue. Block-indicating units modulate their firing rate during the initial 1.25 seconds following context port entry, before reward delivery, based on the expected reward-rate block from the previous trial. Non-responsive units show no detectable modulation to any measured task variables.

Figure 2-figure supplement 3. Poisson simulation validates step detection specificity.

(a) SVM classification scatter plot showing real step-like units (black, $n = 898$), real non-step units (gray, $n = 939$), and simulated flat-rate Poisson neurons (orange-red, $n = 1,627$) in CV-MCC feature space. The dashed line indicates the linear SVM decision boundary. Simulated neurons cluster below the boundary in the low-MCC, high-CV region. False positive rate: 1.35% ($22/1,627$; 95% CI [0.85%, 2.04%]). (b) Step-style histogram comparing MCC distributions across the three populations. Median MCC values: simulated = 0.29, real non-step = 0.34, real step-like = 0.45 (Mann-Whitney U, simulated vs. step-like: $p = 2.97 \times 10^{-273}$). For each of the 1,849 recorded units, a matched homogeneous Poisson process neuron was generated with the same mean firing rate and reward interval structure but zero modulation. Simulated spike trains were processed through the identical classification pipeline (sigmoid fitting, CV and MCC computation, SVM classification). Of 1,849 simulated units, 222 were excluded at the sigmoid fitting stage due to insufficient valid intervals (predominantly very low firing rate units, median 0.42 Hz).

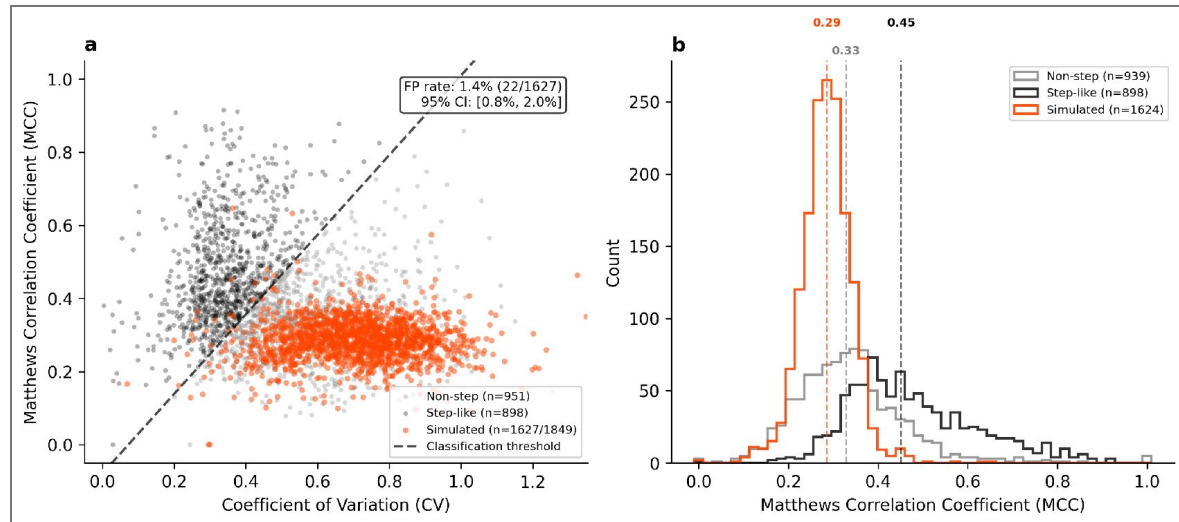
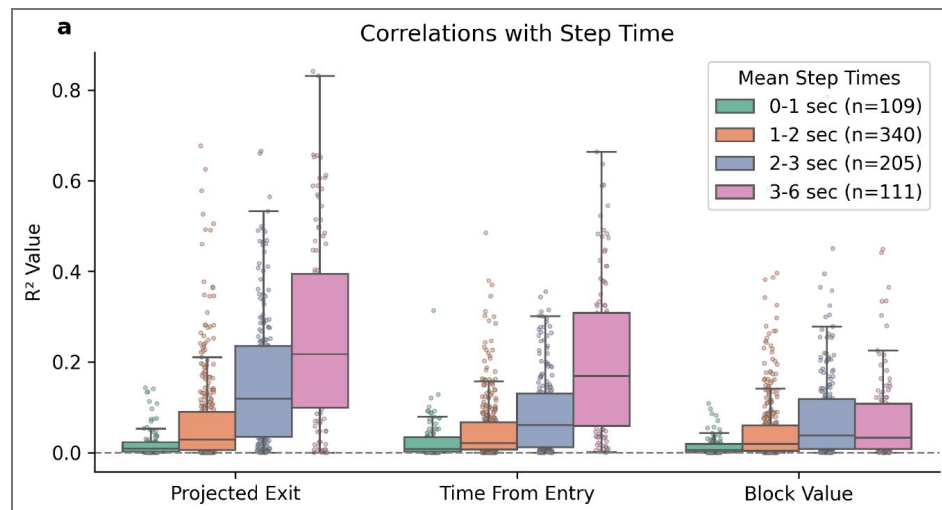


Figure 3-figure supplement 1. Correlation strength between neural step times and task variables varies by mean step time of DMS units.

Box plots showing the distribution of R^2 values quantifying the relationship between individual unit step times and three task variables: projected exit time (time the mouse would wait following each reward based on session-specific behavioral models), time of reward relative to port entry, and context block reward rate. Units are grouped by their mean step time. Individual R^2 values for each unit are overlaid as points using a swarm plot distribution to avoid overlap.



Data availability

All original data (including raw behavioral and neural recordings) and analysis code will be deposited and made publicly available on a DataJoint repository. The raw data, upon which all subsequent analyses were performed, will be available alongside the code required to reproduce the analyses and figures.

Acknowledgements

We thank Dr. Patricia Janak, Dr. Dan O'Connor, and Dr. Jeremiah Cohen for project guidance and valuable comments on the manuscript. We are also grateful to the DataJoint team for technical support. We also thank Rebekah Zhang, Dr. Simon Allard, and Dr. Tanya Martin, and the rest of the Hussain Shuler lab for their support.

Additional information

Author Contributions (CRediT)

Elissa Sutlief (First Author): Conceptualization, Investigation, Data Curation, Formal Analysis, Methodology, Project Administration, Software, Visualization, Writing - Original Draft, Writing - Review & Editing

Shichen Zhang (Author 2): Investigation, Methodology, Formal Analysis, Writing - Review & Editing

Kate Forsberg (Author 3): Investigation, Methodology

Marshall G Hussain Shuler (Senior Author / PI): Conceptualization, Supervision, Funding Acquisition, Project Administration, Writing - Review & Editing

Funding Sources

NIMH (5R01MH123446), NIA (R01AG063783, R01AG063783) to MGHS. NEI (T32EY007143) to ES.

Funding

Funder	Grant reference number	Author
HHS NIH National Institute of Mental Health (NIMH)	R01MH123446	Marshall G Hussain Shuler
HHS NIH National Institute on Aging (NIA)	RF1AG063783	Marshall G Hussain Shuler
HHS NIH National Institute on Aging (NIA)	R01AG063783	Marshall G Hussain Shuler
HHS NIH National Eye Institute (NEI)	T32EY007143	Elissa Sutlief
Kavli Neuroscience Discovery Institute		Marshall G Hussain Shuler

Author ORCID iDs

Marshall G Hussain Shuler:  <https://orcid.org/0000-0002-1927-0970>

References

Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML (2017) A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology* **42**:1931-1939 <https://doi.org/10.1038/npp.2017.108> | PubMed

van Alphen JJM, Bernstein C, Driessen G (2003) Information acquisition and residence time in parasitoids. *Trends in Ecology & Evolution* **18**:81-87 [https://doi.org/10.1016/S0169-5347\(02\)00035-6](https://doi.org/10.1016/S0169-5347(02)00035-6)

- Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. *The Journal of Neuroscience* **27**:8161-8165 <https://doi.org/10.1523/JNEUROSCI.1554-07.2007> | PubMed
- Baum WM (1983) Studying foraging in the psychological laboratory. In: Mellgren RL (Ed). *Advances in Psychology* **13** Elsevier. pp. 253-283 [https://doi.org/10.1016/s0166-4115\(08\)61799-9](https://doi.org/10.1016/s0166-4115(08)61799-9)
- Berridge KC, Robinson TE (2003) Parsing reward. *Trends in Neurosciences* **26**:507-513 [https://doi.org/10.1016/S0166-2236\(03\)00233-9](https://doi.org/10.1016/S0166-2236(03)00233-9) | PubMed
- Bettinger RL, Grote MN (2016) Marginal value theorem, patch choice and human foraging response in varying environments. *Journal of Anthropological Archaeology* **42**:79-87 <https://doi.org/10.1016/j.jaa.2016.03.002>
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review* **113**:700-765 <https://doi.org/10.1037/0033-295X.113.4.700> | PubMed
- Bolkan SS, Stone IR, Pinto L, Ashwood ZC, Garcia JM, Herman AL, Singh P, Bandi A, Cox J, Zimmerman GA, et al. (2022) Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nature Neuroscience* **25**:345-357 <https://doi.org/10.1038/s41593-022-01021-9> | PubMed
- Brody CD, Hanks TD (2016) Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology* **37**:149-157 <https://doi.org/10.1016/j.conb.2016.01.003> | PubMed
- Brown JS (1988) Patch use as an indicator of habitat preference, predation risk and competition. *Behavioral Ecology and Sociobiology* **22**:37-47 <https://doi.org/10.1007/bf00395696>
- Buccino AP, Hurwitz CL, Garcia S, Magland J, Siegle JH, Hurwitz R, Hennig MH (2020) SpikeInterface, a unified framework for spike sorting. *eLife* **9** <https://doi.org/10.7554/eLife.61834> | PubMed
- Burton AC, Nakamura K, Roesch MR (2015) From ventral-medial to dorsal-lateral striatum: neural correlates of reward-guided decision-making. *Neurobiology of Learning and Memory* **117**:51-59 <https://doi.org/10.1016/j.nlm.2014.05.003> | PubMed
- Castañé A, Theobald DEH, Robbins TW (2010) Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behavioural Brain Research* **210**:74-83 <https://doi.org/10.1016/j.bbr.2010.02.017> | PubMed
- Charnov EL (1976) Optimal foraging, the marginal value theorem. *Theoretical Population Biology* **9**:129-136 [https://doi.org/10.1016/0040-5809\(76\)90040-x](https://doi.org/10.1016/0040-5809(76)90040-x) | PubMed
- Church RM, Deluty MZ (1977) Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes* **3**:216-228 <https://doi.org/10.1037/0097-7403.3.3.216>
- Cisek P (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **362**:1585-1599 <https://doi.org/10.1098/rstb.2007.2054> | PubMed
- Constantino SM, Daw ND (2015) Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective & Behavioral Neuroscience* **15**:837-853 <https://doi.org/10.3758/s13415-015-0350-y> | PubMed
- Cuthill IC, Kacelnik A, Krebs JR, Haccou P, Iwasa Y (1990) Starlings exploiting patches: the effect of recent experience on foraging decisions. *Animal Behaviour* **40**:625-640 [https://doi.org/10.1016/S0003-3472\(05\)80692-X](https://doi.org/10.1016/S0003-3472(05)80692-X)
- Eccard JA, Liesenjohann T, Dammhahn M (2020) Among-individual differences in foraging modulate resource exploitation under perceived predation risk. *Oecologia* **194**:621-634 <https://doi.org/10.1007/s00442-020-04773-y> | PubMed
- Emmons EB, De Corte BJ, Kim Y, Parker KL, Matell MS, Narayanan NS (2017) Rodent medial frontal control of temporal processing in the dorsomedial striatum. *The Journal of Neuroscience* **37**:8718-8733 <https://doi.org/10.1523/JNEUROSCI.1376-17.2017> | PubMed

- Ferenczi EA, Zalocusky KA, Liston C, Grosenick L, Warden MR, Amatya D, Katovich K, Mehta H, Patenaude B, Ramakrishnan C, *et al.* (2016) Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science* **351** <https://doi.org/10.1126/science.aac9698> | PubMed
- Frank MJ, Seeberger LG, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**:1940-1943 <https://doi.org/10.1126/science.1102941> | PubMed
- Gibbon J (1977) Scalar expectancy theory and Weber's law in animal timing. *Psychological Review* **84**:279-325 <https://doi.org/10.1037/0033-295X.84.3.279>
- Gibbon J, Church RM, Meck WH (1984) Scalar timing in memory. *Annals of the New York Academy of Sciences* **423**:52-77 <https://doi.org/10.1111/j.1749-6632.1984.tb23417.x> | PubMed
- Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annual Review of Neuroscience* **30**:535-574 <https://doi.org/10.1146/annurev.neuro.29.051605.113038> | PubMed
- Gouvea TS, Monteiro T, Motiwala A, Soares S, Machens C, Paton JJ (2015) Striatal dynamics explain duration judgments. *eLife* **4** <https://doi.org/10.7554/eLife.11386> | PubMed
- Güldener L, Saravanakumar P, Happel MFK, Ohi FW, Vollmer M, Pollmann S (2024) Differential patch-leaving behavior during probabilistic foraging in humans and gerbils. *Communications Biology* **7** <https://doi.org/10.1038/s42003-024-06683-8> | PubMed
- Haccou P, Adas SJD, Van Alphen JJM, Visser ME (1991) Information processing by foragers: Effects of intra-patch experience on the leaving tendency of *Leptopilina heterotoma*. *The Journal of Animal Ecology* **60**:93-106 <https://doi.org/10.2307/5447>
- Hall-McMaster S, Dayan P, Schuck NW (2021) Control over patch encounters changes foraging behavior. *iScience* **24** <https://doi.org/10.1016/j.isci.2021.103005> | PubMed
- Harhen NC, Bornstein AM (2023) Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences* **120** <https://doi.org/10.1073/pnas.2216524120> | PubMed
- Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience* **14**:933-939 <https://doi.org/10.1038/nn.2856> | PubMed
- Hutchinson JMC, Wilke A, Todd PM (2008) Patch leaving in humans: can a generalist adapt its rules to dispersal of items across patches?. *Animal Behaviour* **75**:1331-1349 <https://doi.org/10.1016/j.anbehav.2007.09.006>
- Huys QJM, Daw ND, Dayan P (2015) Depression: a decision-theoretic analysis. *Annual Review of Neuroscience* **38**:1-23 <https://doi.org/10.1146/annurev-neuro-071714-033928> | PubMed
- Ito M, Doya K (2015) Distinct neural representation in the dorsolateral, dorsomedial and ventral parts of the striatum during fixed- and free-choice tasks. *The Journal of Neuroscience* **35**:3499-3514 <https://doi.org/10.1523/JNEUROSCI.1962-14.2015> | PubMed
- Jin DZ, Fujii N, Graybiel AM (2009) Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences* **106**:19156-19161 <https://doi.org/10.1073/pnas.0909881106> | PubMed
- Kacelnik A, Bateson M (1996) Risky theories: the effects of variance on foraging decisions. *American Zoologist* **36**:402-434 <https://doi.org/10.1093/icb/36.4.402>
- Kacelnik A, Krebs JR (1985) The ecology of foraging: decision-making and its mechanisms. In: *Perspectives in Ethology* **6** Springer. pp. 279-303
- Kacelnik A, Todd IA (1992) Psychological mechanisms and the Marginal Value Theorem: effect of variability in travel time on patch exploitation. *Animal Behaviour* **44**:91-104 [https://doi.org/10.1016/s0003-3472\(05\)80226-x](https://doi.org/10.1016/s0003-3472(05)80226-x)
- Keller JK, Sullivan PJ (2023) The importance of patch shape at threshold occupancy: functional patch size within total habitat amount. *Oecologia* **203**:95-112 <https://doi.org/10.1007/s00442-023-05453-3> | PubMed

- Killeen PR, Fetterman JG (1988) A behavioral theory of timing. *Psychological Review* **95**:274-295 <https://doi.org/10.1037/0033-295X.95.2.274> | PubMed
- Kilpatrick ZP, Davidson JD, El Hady A (2021) Uncertainty drives deviations in normative foraging decision strategies. *Journal of the Royal Society, Interface* **18** <https://doi.org/10.1098/rsif.2021.0337> | PubMed
- Kim H, Lee D, Jung MW (2013) Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *The Journal of Neuroscience* **33**:52-63 <https://doi.org/10.1523/JNEUROSCI.2422-12.2013> | PubMed
- Kolling N, Behrens TEJ, Mars RB, Rushworth MFS (2012) Neural mechanisms of foraging. *Science* **336**:95-98 <https://doi.org/10.1126/science.1216930> | PubMed
- Krebs JR, Ryan JC, Charnov EL (1974) Hunting by expectation or optimal foraging? A study of patch use by chickadees. *Animal Behaviour* **22**:953-964 [https://doi.org/10.1016/0003-3472\(74\)90018-9](https://doi.org/10.1016/0003-3472(74)90018-9)
- Laguë M, Tania N, Heath J, Edelstein-Keshet L (2012) The effects of facilitation and competition on group foraging in patches. *Journal of Theoretical Biology* **310**:88-96 <https://doi.org/10.1016/j.jtbi.2012.06.024> | PubMed
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* **58**:451-463 <https://doi.org/10.1016/j.neuron.2008.02.021> | PubMed
- Lenow JK, Constantino SM, Daw ND, Phelps EA (2017) Chronic and acute stress promote overexploitation in serial decision making. *The Journal of Neuroscience* **37**:5681-5689 <https://doi.org/10.1523/JNEUROSCI.3618-16.2017> | PubMed
- Lloyd A, Viding E, McKay R, Furl N (2023) Understanding patch foraging strategies across development. *Trends in Cognitive Sciences* **27**:1085-1098 <https://doi.org/10.1016/j.tics.2023.07.004> | PubMed
- Marshall HH, Carter AJ, Ashford A, Rowcliffe JM, Cowlshaw G (2013) How do foragers decide when to leave a patch? A test of alternative models under natural and experimental conditions. *The Journal of Animal Ecology* **82**:894-902 <https://doi.org/10.1111/1365-2656.12089> | PubMed
- Marton T, Samuels J, Nestadt P, Krasnow J, Wang Y, Shuler M, Kamath V, Chib VS, Bakker A, Nestadt G (2019) Validating a Dimension of Doubt in Decision-Making: A Proposed Endophenotype for Obsessive-Compulsive Disorder. *PloS One* **14**:e0218182 <https://doi.org/10.1371/journal.pone.0218182> | PubMed
- Matell MS, Meck WH (2000) Neuropsychological mechanisms of interval timing behavior. *BioEssays* **22**:94-103 [https://doi.org/10.1002/\(SICI\)1521-1878\(200001\)22:1<94::AID-BIES14>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1521-1878(200001)22:1<94::AID-BIES14>3.0.CO;2-E) | PubMed
- Matell MS, Meck WH (2004) Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Brain Research Cognitive Brain Research* **21**:139-170 <https://doi.org/10.1016/j.cogbrainres.2004.06.012> | PubMed
- McNamara JM, Houston AI (1985) Optimal foraging and learning. *Journal of Theoretical Biology* **117**:231-249 [https://doi.org/10.1016/S0022-5193\(85\)80219-8](https://doi.org/10.1016/S0022-5193(85)80219-8)
- Meck WH (2006) Neuroanatomical localization of an internal clock: a functional link between mesolimbic, nigrostriatal and mesocortical dopaminergic systems. *Brain Research* **1109**:93-107 <https://doi.org/10.1016/j.brainres.2006.06.031> | PubMed
- Mello GBM, Soares S, Paton JJ (2015) A scalable population code for time in the striatum. *Current Biology: CB* **25**:1113-1122 <https://doi.org/10.1016/j.cub.2015.02.036> | PubMed
- Merchant H, Harrington DL, Meck WH (2013) Neural basis of the perception and estimation of time. *Annual Review of Neuroscience* **36**:313-336 <https://doi.org/10.1146/annurev-neuro-062012-170349> | PubMed
- Mink JW (1996) The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology* **50**:381-425 [https://doi.org/10.1016/s0301-0082\(96\)00042-1](https://doi.org/10.1016/s0301-0082(96)00042-1)

- Namboodiri V, Mihalas S, Shuler M (2016) Analytical calculation of errors in time and value perception due to a subjective time accumulator: a mechanistic model and the generation of Weber's law. *Neural Computation* **28**:89-117 https://doi.org/10.1162/neco_a_00792 | PubMed
- Namboodiri VMK, Hussain Shuler MG (2016) The hunt for the perfect discounting function and a reckoning of time perception. *Current Opinion in Neurobiology* **40**:135-141 <https://doi.org/10.1016/j.conb.2016.06.019> | PubMed
- Namboodiri VMK, Mihalas S, Hussain Shuler MG (2014) Rationalizing decision-making: understanding the cost and perception of time. *Timing & Time Perception Reviews* **1**:1-40 <https://doi.org/10.1163/24054496-00101004> | PubMed
- Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**:507-520 <https://doi.org/10.1007/s00213-006-0502-4> | PubMed
- Nonacs P (2001) State dependent behavior and the Marginal Value Theorem. *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* **12**:71-83 <https://doi.org/10.1093/oxfordjournals.beheco.a000381>
- Olsson O, Brown JS (2006) The foraging benefits of information and the penalty of ignorance. *Oikos* **112**:260-273 <https://doi.org/10.1111/j.0030-1299.2006.13548.x>
- Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD (2016) Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv* <https://doi.org/10.1101/061481>
- Packard MG, Knowlton BJ (2002) Learning and memory functions of the Basal Ganglia. *Annual Review of Neuroscience* **25**:563-593 <https://doi.org/10.1146/annurev.neuro.25.112701.142937> | PubMed
- Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB (2016) Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nature Neuroscience* **19**:845-854 <https://doi.org/10.1038/nn.4287> | PubMed
- Pisa M, Cyr J (1990) Regionally selective roles of the rat's striatum in modality-specific discrimination learning and forelimb reaching. *Behavioural Brain Research* **37**:281-292 [https://doi.org/10.1016/0166-4328\(90\)90140-a](https://doi.org/10.1016/0166-4328(90)90140-a) | PubMed
- Ponzi A, Wickens J (2022) Ramping activity in the striatum. *Frontiers in Computational Neuroscience* **16** <https://doi.org/10.3389/fncom.2022.902741> | PubMed
- Ragozzino ME (2007) The contribution of the medial prefrontal cortex, orbitofrontal cortex and dorsomedial striatum to behavioral flexibility. *Annals of the New York Academy of Sciences* **1121**:355-375 <https://doi.org/10.1196/annals.1401.013> | PubMed
- Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem?. *Neuroscience* **89**:1009-1023 [https://doi.org/10.1016/s0306-4522\(98\)00319-4](https://doi.org/10.1016/s0306-4522(98)00319-4) | PubMed
- Salamone JD, Correa M (2012) The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**:470-485 <https://doi.org/10.1016/j.neuron.2012.10.021> | PubMed
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* **310**:1337-1340 <https://doi.org/10.1126/science.1115270> | PubMed
- Searle KR, Vandervelde T, Hobbs NT, Shipley LA, Wunder BA (2006) Spatial context influences patch residence time in foraging hierarchies. *Oecologia* **148**:710-719 <https://doi.org/10.1007/s00442-005-0285-z> | PubMed
- Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences* **93**:628-633 <https://doi.org/10.1073/pnas.93.2.628> | PubMed
- Simen P, Balci F, de Souza L, Cohen JD, Holmes P (2011) A model of interval timing by neural integration. *The Journal of Neuroscience* **31**:9238-9253 <https://doi.org/10.1523/JNEUROSCI.3121-10.2011> | PubMed
- Simen P, Rivest F, Ludvig EA, Balci F, Killeen PR (2013) Timescale invariance in the pacemaker-accumulator family of timing models. *Timing & Time Perception* **1**:159-188 <https://doi.org/10.1163/22134468-00002018>

- Smith KS, Graybiel AM (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* **79**:361-374 <https://doi.org/10.1016/j.neuron.2013.05.038> | PubMed
- Stott JJ, Redish AD (2014) A functional difference in information processing between orbitofrontal cortex and ventral striatum during decision-making behaviour. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **369** <https://doi.org/10.1098/rstb.2013.0472> | PubMed
- Sutlief E, Walters C, Marton T, Hussain Shuler MG (2025) The value of initiating a pursuit in temporal decisionmaking. *eLife* <https://doi.org/10.7554/elife.99957> | PubMed
- Tang H, Costa VD, Bartolo R, Averbeck BB (2022) Differential coding of goals and actions in ventral and dorsal corticostriatal circuits during goal-directed behavior. *Cell Reports* **38** <https://doi.org/10.1016/j.celrep.2021.110198> | PubMed
- 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 <https://doi.org/10.1016/j.neuron.2010.04.036> | PubMed
- Torres C, Glueck AC, Conrad SE, Morón I, Papini MR (2016) Dorsomedial striatum lesions affect adjustment to reward uncertainty, but not to reward devaluation or omission. *Neuroscience* **332**:13-25 <https://doi.org/10.1016/j.neuroscience.2016.06.041> | PubMed
- Treisman M (1963) Temporal discrimination and the indifference interval: implications for a model of the "internal clock". *Psychological Monographs* **77**:1-31 <https://doi.org/10.1037/h0093864> | PubMed
- Utsumi SA, Cangiano CA, Galli JR, McEachern MB, Demment MW, Laca EA (2009) Resource heterogeneity and foraging behaviour of cattle across spatial scales. *BMC Ecology* **9** <https://doi.org/10.1186/1472-6785-9-9> | PubMed
- Vandaele Y, Ottenheimer DJ, Janak PH (2021) Dorsomedial striatal activity tracks completion of behavioral sequences in rats. *eNeuro* **8** <https://doi.org/10.1523/ENEURO.0279-21.2021> | PubMed
- Wajnberg E, Fauvergue X, Pons O (2000) Patch leaving decision rules and the Marginal Value Theorem: an experimental analysis and a simulation model. *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* **11**:577-586 <https://doi.org/10.1093/beheco/11.6.577>
- Ward JF, Austin RM, Macdonald DW (2000) A simulation model of foraging behaviour and the effect of predation risk. *The Journal of Animal Ecology* **69**:16-30 <https://doi.org/10.1046/j.1365-2656.2000.00371.x>
- Wilke A, Hutchinson JMC, Todd PM, Czienskowski U (2009) Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science* **33**:497-529 <https://doi.org/10.1111/j.1551-6709.2009.01020.x> | PubMed
- Yartsev MM, Hanks TD, Yoon AM, Brody CD (2018) Causal contribution and dynamical encoding in the striatum during evidence accumulation. *eLife* **7**:e34929 <https://doi.org/10.7554/eLife.34929> | PubMed
- Yatsenko D, Nguyen T, Shen S, Gunalan K, Turner CA, Guzman R, Sasaki M, Sitonic D, Reimer J, Walker EY, et al. (2021) Datajoint Elements: Data Workflows for Neurophysiology. *bioRxiv* <https://doi.org/10.1101/2021.03.30.437358>
- Yin HH, Knowlton BJ, Balloine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience* **19**:181-189 <https://doi.org/10.1111/j.1460-9568.2004.03095.x> | PubMed
- Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning: Striatum and instrumental conditioning. *The European Journal of Neuroscience* **22**:513-523 <https://doi.org/10.1111/j.1460-9568.2005.04218.x> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

In this study, Shuler and colleagues record neurons from the DMS in mice performing a patch foraging task. In this task, mice had the choice between harvesting rewards from 2 ports - one the time-investment port where the rate of reward declined over time and the other a context port where the rate of reward was either high or low. Mice performed the task appropriately, switching between ports as the rate of reward declined in the time-investment port and switching more rapidly when the context port delivered high versus low rewards. The behavior of the mice was also strongly driven by time since the most recent reward receipt, in conflict with normative accounts of patch foraging. Individual DMS neurons showed bistable firing patterns, transitioning to high rates of activity at various times from reward. Overall, the population tiled the temporal space, and the accumulation of the number of neurons in the high firing state was predictive of patch exit. The rate of accumulation varied with things that also affected behavior.

Strengths:

Overall, the aims of the study were clear and important, the experiment directly addresses them, and the results are clear and provide compelling support for the authors' conclusions.

Weaknesses:

I have only a few comments and questions to consider, none of which are criticisms of what was done, really.

(1) Probably my chief question, alluded to in the discussion, is what the evidence is that DMS plays a causal role in generating these correlates and the resulting behavior, in light of the lack of causal evidence here. What are other options? Could such information depend on upstream areas such as OFC or mPFC, with DMS just a pass-through? And while I would not ask for causal data, is there a specific prediction? That is, if the area were inactivated, would mice stay longer or shorter? Not do the task? If I wanted to do a causal test of the authors' idea regarding the contribution of DMS to this behavior, what would be predicted, and what result would invalidate the hypothesis? Speculating on this a bit, beyond just saying DMS is involved, would be useful.

(2) Not much is said about the suboptimal strategy. Would DMS continue to play the same role if the mice showed no effect of recent reward and instead performed appropriately? Or is some other area doing that job? Or is this not important? I thought it was interesting that the mice basically did not treat the game quite like they were supposed to. Is it important to go back and look at what is happening in DMS under normative conditions to really know how this area contributes to proper foraging?

(3) Do these neurons also track time in the context port? Or do they only exhibit this behavior in the port where rewards are depleting? This seems like an interesting question. Do they show the same profile in different ports, if so?

<https://doi.org/10.7554/eLife.111930.1.sa3>

Reviewer #2 (Public review):

Summary:

Here, Sutlief et al. use a novel patch-foraging task to investigate the role of dorsomedial striatal (DMS) neurons in determining when animals disengage from a resource. They show that mice, contrary to canonical optimal-foraging predictions, adopt a strategy in which reward receipt resets timing behavior, with decisions further shaped by both cumulative time spent in a patch and the overall quality of the environment. The authors further demonstrate that a subset of DMS neurons exhibits step-like activity patterns during task performance. Importantly, the accumulation of these state transitions across the neuronal population predicts the timing of patch-leaving decisions on a trial-by-trial basis, providing a potential neural mechanism underlying decisions about when to abandon a currently exploited resource.

Strengths:

This study addresses an important question using a well-designed, interesting behavioral task. The finding that mice employ a reward-triggered exit-timing policy is particularly interesting, as it is pertinent to the many patch foraging-style tasks that have been developed for use in mice, where rewards are delivered as discrete events. The identification of step-like activity in DMS neurons is mostly compelling, and the authors' trial-by-trial analysis linking this activity to behavior provides some support for its relevance to patch-leaving decisions.

Weaknesses:

A key interpretational issue is whether the DMS signal reflects timing specifically, rather than movement initiation or other task-related factors. The authors argue that once a sufficient number of neurons transition, the animal exits the time-investment port. However, it remains unclear whether this population threshold reflects a timing computation that determines when to leave in the more abstract sense, or a signal more directly related to movement onset (that may also be initiated after some proportion of the population has changed its activity). An important control would be to examine neural activity while animals are engaged at the context port. In this epoch, animals presumably do not need to time their departure in the same way, but they still eventually initiate movement. If the DMS signal reflects timing rather than movement, one would not expect the same accumulation-to-threshold pattern of step-like transitions at the context port.

It would also be helpful for the authors to clarify the behavioral definition of the leaving decision. Can mice return to the time-investment port after exiting it if they do not subsequently enter the context port? How exactly is "exit" defined: as withdrawal from the time-investment port, entry into the context port, or some other behavioral event? Is there variability in the latency between time-investment port exit and context-port entry, and if so, is this latency related to DMS step-like activity? These details are important for interpreting whether the neural activity is aligned with a timing decision, movement initiation, or the execution of a transition between task states.

The classification approach for identifying step-like activity seems generally reasonable, and the low false-positive rate against homogeneous Poisson controls is reassuring. However, one potential issue is that the identification of trial-by-trial state transitions is not independent of the session-level characterization of each neuron. The algorithm first fits a sigmoid to the pooled session data and then uses the resulting high- and low-firing-rate states to constrain interval-level fits. This may bias the analysis toward finding step-like transitions in neurons whose activity is only approximately step-like at the session level, effectively reducing the space of alternative solutions available to the interval-level fits. As implemented, the approach therefore functions more as a detector of consistency with a session-defined step model than as an unbiased test of whether individual intervals are better described by discrete state transitions versus alternative dynamics such as ramps or gradual drifts. This concern could be addressed by comparing the constrained sigmoid model against alternatives, such as constant-rate or ramping models, on held-out intervals, or by deriving

state parameters from an independent subset of trials and testing classification on the remaining trials.

The inclusion threshold for the accumulation analysis is difficult to evaluate. Sessions were included if they contained at least seven simultaneously recorded step-like units, but this number is hard to interpret without knowing the total number of recorded units per session and the fraction classified as step-like. Seven units may be sufficient for fitting a population accumulation trajectory, but because the cutoff is based on an absolute number rather than a proportion of the recorded population, it is unclear whether included sessions reflect robust population-level step-like dynamics or a relatively small selected subset of DMS activity. Reporting the number and fraction of step-like units per session, as well as the sensitivity of the accumulation results across different inclusion thresholds, would help clarify this point.

<https://doi.org/10.7554/eLife.111930.1.sa2>

Reviewer #3 (Public review):

Sutlief and colleagues report behavioral and neural results from mice performing a patch foraging task. Behaviorally, they argue that time since last reward is a major determinant of when mice decide to leave a patch. In the brain, they find neurons in the dorsomedial striatum that show step-like changes in their firing rate at a range of times following reward. Population analyses show that the cumulative fraction of neurons that have undergone such a step-like change in firing rate can be used to predict patch-leaving times with impressive accuracy.

Overall, this is an interesting set of results that has been analyzed in a principled way. The manuscript is well written, the results are explained clearly, and the evidence supporting the authors' conclusions is strong. The manuscript is therefore a potentially valuable contribution to the growing literature assessing how the brain solves stopping problems like the patch foraging scenario. I have suggestions for the authors to consider that might further increase the rigor of their results, and a few suggestions for improving the clarity of the work for readers.

(1) I don't quite understand how the behavioral task works. Are mice rewarded for making discrete nose poke responses in the investment and context ports? Or are they required to nose poke and hold? Is reward given with some probability per response (which decreases with time in the patch), or is the reward probability a function of elapsed time in the patch, time since last response, or dwell time in the port? Also, exactly what equation defines how reward probability changes over time for the high- and low-value contexts? I couldn't find these details anywhere in the manuscript, and they would be helpful for better understanding the behavior and the later neural results.

(2) How was the optimal strategy determined? Several features of the author's task violate the assumption of the marginal value theorem, so computing the optimal residence time is not a straightforward application of the classic model. There's a diagram in Figure 1h that depicts an MVT-like graphical solution, but the conventions of the plot are not familiar to me, and there's no description of how it works in the results or methods. More detail here would be much appreciated. In a similar vein, the authors report that mice generally exceeded optimal residence times in patches, but no statistical comparison is provided to back up that statement. There should be some formal test of this if it is to be included in the results.

(3) The authors argue that time since last reward is the predominant determinant of patch leaving time. However, as the authors note, time since last reward is correlated with other task variables (patch reward rate, time in patch, etc.). I don't trust that SVM coefficients can be interpreted as straightforward measures of a variable's importance for classification

performance in the case of correlated predictors. A better approach would be to assess how well the model performs as subsets of variables are added or removed from the model.

(4) For the SVM analysis, I'm not quite understanding how or why the authors are using 5 s after mice left the patch as additional "Leave" examples. For instance, is time since entry computed for the investment patch, or the context patch that mice enter after they leave the investment patch? Similarly, is the time since the last reward relative to the investment patch, or the reward the mouse is likely to receive at the context patch? Moreover, I'm not sure it's safe to assume that because the mouse left at time t , time $t+1$ necessarily reflects conditions on which the mouse would definitely leave again. If we're thinking about the stay/leave decision as something that is being repeated sequentially on a fast time scale to determine how long mice stay in the patch, it doesn't follow that observing a mouse leave means that any patch conditions after that would necessarily result in the same decision. If that were the case, it would mean that seeing a mouse leave a patch after 2 s would preclude ever observing a residence time longer than 2 s, which is clearly not compatible with the authors' data. Ultimately, it's only possible to observe one decision to leave per trial; including data points beyond that as additional leave examples seems overly speculative to me.

(5) The authors validate their approach for quantifying step-like changes in firing rate using simulations of constant-rate Poisson spiking and observe a low false positive rate. This is encouraging, but it doesn't seem like the only way in which their method could go awry, or even the most concerning way. I would be much more interested in seeing the false positive rate for continuous, ramp-like changes in firing rate, which would be much more likely to trip up the authors' approach and are also the major relevant alternative hypothesis to step-like changes in firing rate. Random walks in firing rate might also be worth testing.

(6) The finding that cumulative "transitioned" neurons is predictive of patch leaving is interesting. However, I can't help but wonder how truly informative this variable is for predicting patch leaving. It seems as though neurons can only transition firing rates one time. That means that as time in the patch increases, the fraction of transitioned neurons naturally increases. Similarly, all visits must eventually end with the mouse leaving the patch, so the hazard rate of leaving increases with time in the patch. Given that, can the authors be certain that the cumulative transitioned neurons are really what's predicting patch leaving time, or would any generically increasing function perform roughly the same? An interesting test would be to mismatch the neural predictor and behavior at the level of trials. If this mechanism is really specific, rather than something that captures the general structure of an increasing hazard rate of leaving, then prediction of leaving time should work substantially better when the neural predictor is correctly matched to behavior on the trial for which it was recorded.

<https://doi.org/10.7554/eLife.111930.1.sa1>

Author response:

Reviewer #1

(1) Causality and the role of DMS; a specific, falsifiable prediction. We agree that the paper should not leave the causal question implicit, and we will expand the Discussion to state a concrete prediction rather than a general claim of involvement. Briefly, if the accumulation signal we describe carries the animal's intended departure time, then suppressing DMS during patch occupancy should not simply shift exit times in one direction but should degrade their structure: exit-time variability should increase, and exit timing should lose its systematic dependence on reward-rate context and on the time of the most recent reward. A plausible alternative outcome is disengagement from the task altogether, which would be uninformative and would need to be controlled for. The result that would falsify our

hypothesis is the one we will state explicitly: exit timing that remains as predictable, and as sensitive to context and reward history, under DMS suppression, as without it. We will also discuss the alternative the reviewer raises, that these signals are inherited from cortical inputs such as OFC or mPFC with DMS acting as a relay, and note that our data cannot presently distinguish this from a locally generated signal.

(2) Behavior under a normative strategy. This is an interesting question and we will address it in the Discussion. Our expectation, which we will frame as a prediction rather than a result, is that an animal timing from patch entry rather than resetting at each reward would show accumulation that begins at entry and proceeds to a context-dependent threshold at the reward-rate-optimal time, rather than the reward-triggered resets we observe. In this view, the reset structure of the neural signal is a reflection of the behavioral policy rather than a property of the region. We will make clear that this is a testable prediction that our current dataset does not address.

(3) Do these neurons also track time at the context port? We intend to answer with new analysis, and it converges with Reviewer #2's suggested analysis (below), so we treat the two together there.

Reviewer #2

(a) Timing versus movement initiation: activity at the context port. We take this to be a central interpretational concern. We will examine whether the step-like DMS activity extends to the context port, testing the interval between the final context-port reward and departure for the same step-like transitions and accumulation we observe in the time-investment port. We will apply the same comparison to the context-port inter-reward intervals, which addresses Reviewer #1's third point about whether these neurons also track time at the context port.

We want to flag one feature of the task that bears on how the outcome should be read. The context port is not a timing-free epoch. Its four rewards are delivered at predictable, regularly spaced intervals, and the interval between the final reward and the animal's departure is self-timed. Departure from the context port is therefore also a self-timed action, and observing accumulation there would not by itself indicate that the signal reflects movement initiation rather than timing. What the comparison can inform is whether the accumulation is specific to a decision about when to disengage from a depleting resource, or is a more general feature of self-timed departures. This is a meaningful distinction either way, and one we will report and interpret whichever direction the result falls.

(b) Operational definition of leaving; the exit-to-entry latency. We agree these details are necessary for interpretation and their absence is our omission. Exit is the final withdrawal from the time-investment port preceding the next context-port visit, and we will make that clear in the revised methods. Mice can and occasionally do re-enter the time-investment port without an intervening context-port visit (especially early in training). Such re-entries are not counted as exits. We will also examine whether the latency between time-investment-port exit and context-port entry relates to the accumulation slope on the corresponding interval, to test whether the neural signal relates to the decision or to the execution of the transition.

(c) Independence of interval-level fits from the session-level model. This is a fair characterization of the procedure, and we accept the distinction the reviewer draws between a detector of consistency with a session-defined step model and an unbiased test of discrete versus continuous dynamics. We will address it with a held-out validation: estimating each unit's state parameters and transition time from one half of its intervals and testing whether the transition times recovered from the withheld half agree. The discrete-versus-continuous comparison is addressed directly by the ramp simulations under Reviewer #3's point (5) below.

(d) The inclusion threshold for the accumulation analysis. We will add a supplementary figure reporting the total number of recorded units per session and the fraction classified as step-like, so that the seven-unit criterion can be evaluated against the recorded population rather than in the abstract. Yield varied substantially across sessions, from a handful of units to roughly one hundred, and we will show this distribution directly. We will also report the accumulation results across a range of inclusion thresholds spanning approximately five to eight simultaneously recorded step-like units, so that readers can assess sensitivity to the choice.

Reviewer #3

(1) Specification of the task. We agree the task description was insufficient, and we will correct this at the front of the Results and in the Methods. The time-investment port operates on a poke-and-hold basis: the mouse maintains its head in the port and rewards are delivered stochastically over time for as long as it remains, with no requirement to withdraw and re-poke. Reward delivery follows an exponentially decaying rate in time since port entry, with a time constant of eight seconds, integrating to an expected eight rewards of one microliter each (8 μ L total) for indefinite occupancy; we will give the explicit function. The reward probability function in the time-investment port is identical across blocks. The high- and low-reward-rate contexts are properties of the context port alone (four rewards over five seconds versus four rewards over ten seconds), and we will make this contrast unambiguous, since it is the manipulation on which the design rests.

(2) Derivation of the optimum, Figure 1h, and a formal test of overstaying. We appreciate this comment. The optimal residence time in our task is not obtained by the classical Charnov tangent construction. It is computed by explicit maximization of the overall reward rate over the full cycle, following the framework in Sutlief et al. (2025) and shown graphically in Figure 1h. We will expand the legend of Figure 1h so its conventions are stated explicitly, give the reward-rate-maximizing derivation as an explicit equation in the Methods, and reframe the surrounding text around reward-rate maximization as the normative principle, with MVT identified as the special case it is. We will also add the formal statistical comparison of observed residence times against the computed optimum, which the reviewer correctly notes was asserted rather than tested.

(3) Interpretation of SVM coefficients with correlated predictors. We accept this criticism. We will not rest the ordering of predictors on coefficient magnitudes alone. We will add a variable inclusion-and-ablation analysis, reporting cross-validated classification performance as each predictor is added to and removed from the model, so that the contribution of time since last reward is assessed by its effect on performance rather than by its normalized weight. We will additionally add a complementary analysis of the leave hazard that estimates the contribution of each variable without requiring the classification framework.

(4) The five-second post-exit window. The reviewer is right that we did not explain this choice, and right that it rests on an assumption. Our reasoning was that a single exit moment per trial leaves the decision boundary badly under-constrained, and that treating the moments immediately following an exit as conditions under which the animal would also have left is licensed by the fact that within-patch reward rate declines monotonically with time, so conditions in the counterfactual continued visit would have been strictly less favorable than those already rejected. We accept that this is an assumption rather than an observation and will state it as such. We will also report the analysis across a range of window durations so that the independence of the result on this choice is visible. To the reviewer's specific questions: both time since entry and time since last reward are computed with respect to the time-investment port throughout, and we will state this explicitly.

(5) False positive rate against ramps and random walks. We agree this is the more informative validation, and that continuous ramping is the most relevant alternative to ours.

We will generate simulated units with continuous ramp-like rate changes, matched to the firing rates and interval structure of our recorded units, and pass them through the identical classification pipeline to obtain false positive rates comparable to the Poisson analysis already reported. We will retain the flat-rate Poisson simulation and present the ramp results as additional panels of the same supplement. We will also explore random-walk dynamics. Together with the held-out validation of transition times described under Reviewer #2(c), this converts the step characterization from a single-null validation into a comparison against the relevant continuous alternatives.

(6) Specificity of the accumulation signal versus a generic increasing function. This is a valuable challenge and we will address it directly. We will implement the trial-mismatch control the reviewer proposes, randomly reassigning accumulation trajectories to reward-to-exit intervals within session and showing the extent to which predictive performance degrades relative to the correctly matched case.

We would also note two features of the existing results that speak to this concern, and which we will bring forward in the revision because we did not make them salient enough. First, a signal that merely tracked elapsed time would be expected to shift its starting level as well as its rate across trials with different exit times; instead, the accumulation slope is strongly related to exit time (mean $r = -0.551$) while the intercept is not (mean $r = 0.013$), indicating a variable rate from a stable origin. Second, and more to the point, the accumulation arrives at a common level at the moment of exit whether the animal leaves early or late. The rate of accumulation shifts with the animal's policy on that trial such that the threshold is met at the intended time.

What makes this predictor non-trivial is its trial-by-trial correspondence to behavior, not simply that it increases over time. We will make this argument explicitly alongside the shuffle control.

Summary

To summarize the planned additions: (i) analysis of step-like activity and its accumulation at the context port, with the interpretive caveat noted above; (ii) validation of step detection against ramping alternatives, together with held-out estimation of transition times; (iii) a trial-mismatch control for the specificity of the accumulation predictor; (iv) an inclusion-and-ablation analysis of the behavioral predictors and a complementary hazard model of the leave decision; (v) reporting of unit yield, step-like fraction, and sensitivity of the accumulation results to the inclusion threshold; (vi) analysis of the exit-to-context-entry latency in relation to the neural signal; (vii) a formal statistical test of overstaying relative to the computed optimum; and (viii) substantial clarification of the task specification, the operational definition of exit, and the derivation of the optimal residence time, including an expanded Figure 1h legend.

Our aim in the revision is to meet the specificity concern raised in the assessment as directly as the existing data allow, and we hope the revised manuscript will warrant reconsideration of the strength-of-evidence characterization.

We are grateful to the reviewers for the care evident in their reports, and to you both for handling the manuscript.

References

- Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136.
- Sutlief, E., Walters, C., Marton, T., & Hussain Shuler, M. G. (2025). The value of initiating a pursuit in temporal decision-making. *eLife*. <https://doi.org/10.7554/eLife.99957.2> .

<https://doi.org/10.7554/eLife.111930.1.sa0>