Nature Communications

https://doi.org/10.1038/s41467-026-71048-0

Article in Press

Revealing abrupt transitions from goal-directed to habitual behavior

Received: 13 June 2025

Accepted: 9 March 2026

Cite this article as: Moore, S., Wang, Z., Zhu, Z. *et al.* Revealing abrupt transitions from goal-directed to habitual behavior. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-71048-0

Sharlen Moore, Zyan Wang, Ziyi Zhu, Joy Wang, Ruolan Sun, Yeonjae A. Lee, Adam Charles & Kishore V. Kuchibhotla

Revealing abrupt transitions from goal-directed to habitual behavior

Sharlen Moore[1,2,#], Zyan Wang[1,#], Ziyi Zhu[1,2,3], Joy Wang[1], Ruolan Sun[4], Yeonjae A. Lee[1], Adam Charles[2,4,5], Kishore V. Kuchibhotla[1,2,3,4,*]

[1]Department of Psychological and Brain Sciences, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD, USA.
[2]Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD, USA.
[3]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.
[4]Department of Biomedical Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland, USA.
[5]Center for Imaging Science, Johns Hopkins University, Baltimore, Maryland, USA.

[#]Equal contribution
*Correspondence: kkuchib1@jhu.edu

Keywords: goal-directed, habit, engagement, learning, dorsal striatum

Abstract

The speed of goal-directed to habit transitions has been debated since Clark Hull asked in 1943: is habit formation slow or sudden? To address this, male mice were given home-cage access to citric-acid water that reduced—without eliminating—reward-seeking for plain water in an auditory go/no-go task. Animals learned to discriminate quickly but exhibited ongoing state-like fluctuations in engagement. Strikingly, these fluctuations abruptly ceased (transition) long after discrimination stabilized, with HMM-GLM modeling pinpointing a ~3 trial transition. We confirmed this as a goal-directed to habit transition using sensory-specific outcome devaluation, DLS lesions, motor stereotypy, and pupillary responses. Dual-site fiber photometry showed equally abrupt DLS dynamics at the transition: outcome-related activity dropped and stimulus-response activity sharpened, suggesting a switch-like mechanism that recruits a readily available habit circuit rather than gradual changes across a threshold. Thus, habits can emerge suddenly, mediated by an abrupt dorsostriatal shift from outcome- to stimulus-driven processing.

Introduction

Humans and other animals are often thought to be creatures of habit. When driving, for example, we are initially told that the color of a traffic light should guide our actions: green to go and red to stop. Through practice, we learn purposefully and are driven by the conscious goal to follow the rules of the road. Over time, these rules become automatic; without deliberation, we will push the gas pedal on a green light and the brake pedal for a red light. More specifically, an initial goal-directed action (R) in response to a cue (S) yields a desired outcome (O) which then slowly evolves into a habit where the cue elicits the action (S-R) without necessarily having the goal in mind[1,2]. The automatization of decisions can be thought of as an efficient way to offload well-learned contingencies to free up resources for more flexible, goal-directed learning[3]. The expression and perseverance of habits, however, can also be maladaptive with neural circuits being co-opted in substance use disorders or compulsive behaviors[4–6]. Understanding the exact time course of habit development is critical to disentangling its neural basis and could help inform future interventional strategies for combating habit-related disorders.

An assumption of a slow, gradual shift from goal-directed to habitual control underpins most models of learning and informs relevant approaches to understanding the neurobiological basis of habitual behavior[7–14]. To date, however, the nature, timing, and properties of the transition in decision control has been challenging to pinpoint due to methodological constraints. In rodents, studies of habits often use distinct reinforcement schedules to bias goal-directed, or habitual behavior[15]. The gold standard, however, for assessing whether a behavior is under goal-directed or habitual control at a specific time point exploits the observation that goal-directed actions are sensitive to the outcome[16,17] (e.g., animals will only perform an action when the reward is desired) while habitual behavior is less sensitive to the outcome (e.g., animals will continue to perform said action even if the reward is not explicitly desired). This behavioral characterization relies on defining habitual behavior as the loss or absence of goal-directed control[18,19]. Sensitivity to the outcome has been successfully operationalized in laboratory testing with outcome devaluation[4,20] procedures in which a specific reward is devalued (through satiety or taste aversion). Outcome devaluation, or a related alternative called contingency degradation, is typically implemented at set time points outside of the normal training regimen (e.g., the middle and end of a multi-day training period). To date, no approach exists to disambiguate between goal-directed and habitual control in real-time, during training[21,22]

Although powerful, methods such as outcome devaluation and contingency degradation inherently limit assessing the nature, timing, and properties of the transition between goal-directed and habitual control in individual animals due to the discrete test sessions and cohort-level comparisons. Can we behaviorally identify habitual transitions in real-time and during training? Is the transition slow or sudden? What are the characteristics of these transitions in individual animals? What are the neural mechanisms that relate to the transition in real-time? Addressing these questions requires a new behavioral approach that assesses the decision mode en passant, without discrete test sessions, without explicitly biasing behavior to one or the other process, and within individual animals.

Here, we present such an approach. We reasoned we could reduce—but not eliminate—reward-seeking for plain water droplets in an auditory cued go/no-go task by allowing animals free access to a citric acid (CA) water in the home-cage. CA water largely fulfills hydration needs, but mice still show a strong preference for plain water when it becomes available. Thus, we reasoned that we could infer behavior as 'goal-directed' if engagement were to be weaker or fluctuate during a task that provides plain water (based on the underlying desire for plain water); in contrast, under habitual control (in which animals are less sensitive to the outcome), the S-R nature of the behavior would drive high and stable engagement despite ongoing changes in the underlying desire for the outcome.

Results

Free access to CA water reduces reward-seeking for plain water

We gave mice ad libitum access in the home cage to water laced with citric acid (CA) (Fig. 1a, CA yellow) which makes it slightly acidic to the taste. Mice continue to drink to provide ongoing hydration but gain less weight than mice given ad libitum access to plain water[23,24]. We refer to these mice as self-restricted as they individually balance their hydration needs against their mild distaste for CA water. Before instrumental training, self-restricted mice lost significantly less weight than mice on a standard water restriction (WR) protocol (Fig. 1b; WR85 = 17.8%±2.3%, CA = 8.9%±1.9% average and std weight loss respectively, p=0.000055, Wilcoxon rank sum test). This difference was maintained throughout training (Supplementary Fig. 1a) (Wilcoxon rank sum test, $p=1.0x10^{-8}$), while no differences were observed in the animals' initial weight (Supplementary Fig. 1b; Wilcoxon rank sum test, p=0.98). Before mice begin discriminative auditory training (Fig. 1a), they first experience 2 days of instrumental training in which they

learn to make an un-cued, self-paced instrumental lick to subsequently receive a small reward (lick training, 3 µl plain water droplet) (Fig. 1a; lavender). Self-restricted CA mice obtained significantly fewer rewards compared to WR mice during these sessions (Fig. 1c; Wilcoxon rank sum test, p=0.00079). We also assessed licking patterns as a proxy of reward-seeking. Self-restricted mice executed fewer licks per session (Supplementary Fig. 1c-1e; yellow; Wilcoxon rank sum test, p=0.021), and exhibited different lick patterns (Supplementary Fig. 1c-d), while maintaining the same lick frequency when engaged in licking (Supplementary Fig. 1f) (Wilcoxon rank sum test, p=0.67). This suggests that self-restricted CA mice exhibit reduced reward-seeking for plain water despite similar consummatory behavior during active licking.

Abrupt transition in engagement may reflect a change in decision control

Mice were then trained on a discriminative auditory go/no-go task in which they learned to lick to one tone (S+, rewarded) for a water reward (hit) and withhold licking to another tone (S-, non-rewarded) to avoid a timeout (Fig.1d; correct reject). Self-restricted CA mice exhibited lower response rates to the S+, consistent with reduced levels of reward-seeking, but surprisingly only minor differences in discrimination performance throughout learning. Specifically, when restricting performance assessment to blocks of high engagement (>50 trials with >90% hit rate), task performance during learning was similar to WR mice (Fig. 1e and 1f). In addition, self-restricted and WR mice exhibited similar, high discrimination performance (75%) within 1,500 trials (WR=1132 ± 87 and CA=1422 ± 234 trials, Wilcoxon rank-sum test p=0.93).

We next sought to explore in detail the impact of reduced reward-seeking on task engagement. To do that, we assessed changes in response-rate to the S+. These changes could be driven by a continuously lower or, alternatively, a fluctuating response rate. In WR mice, animals initially increase their action rates to both tones, followed by a slow reduction in response to the S- (Fig. 2a; left example animal). The S+ response (hit rate) stayed consistently high with minimal variability. We observed a striking contrast in self-restricted CA mice (Fig. 2a; right), where for thousands of trials, they exhibit a fluctuating hit rate, regularly shifting from epochs of high engagement (high hit rate) to low engagement (low hit rate), suggesting that mice are intermittently engaging in the task. We thus focused on the hit rate as the response rate of interest (Fig. 2b). Self-restricted CA mice showed significantly fewer blocks of high engagement (hit rate > 90%) compared to WR mice (Fig. 2c; Wilcoxon rank sum test, p=0.0028).

Surprisingly, after this high-variability phase, most self-restricted CA mice abruptly transitioned to a low-variability phase (Fig. 2d$_i$; red line=transition), an effect rarely seen in WR85 mice (Supplementary Fig. 2a). We observed that 10 out of 12 (83.3%) of the CA mice exhibited high hit rate variability, of which 8 out of 10 (80%) transitioned to a low-variability phase (Supplementary Fig. 2b). Transitions occurred more than 1,000 trials after animals exhibited expert discrimination performance, suggesting this transition was not due to ongoing contingency learning (d'>2 expert=1436±454 trials vs. transition=3832±385 trials) (Supplementary Fig. 2c). To confirm that this change in hit rate was not due to a sudden change in underlying motivation, we measured the weights of the animals daily. Importantly, we observed (1) no differences in discrimination performance around the transition (paired t-test, p=0.69) (Supplementary Fig. 2d) (2) no evidence of changes in weight pre- versus post-transition (Supplementary Fig. 2e) (paired t-test, p=0.20), and (3) no changes in consumption in the home cage based on measurements of post-task and next-day weights (Supplementary Fig. 2f) (paired t-test, p=0.087). This suggests that self-restricted CA mice do not exhibit increased baseline motivation post-transition and, instead, points to a shift in decision control, possibly one from goal-directed to habitual.

Our analysis thus far required categorization of behavior based on pre-defined criteria (low versus high variability, pre- versus post-transition) and experimenter-defined parameters. We sought to test whether a bottom-up, model-based approach could identify behavioral states in an unbiased, and trial-by-trial, manner. To do this, we applied a generalized linear model that incorporates a hidden Markov process (HMM-GLM)[25] on trial-by-trial behavioral data after animals reached expert discrimination performance (Fig. 2e). The HMM-GLM identified two states that best described the behavior in expert animals, as defined by the lowest cross-validated Bayesian Information Criterion value (BIC) (Fig. 2f-g). Both states were sensitive to the stimulus but with distinct action biases. State 1 (pink) exhibited high engagement, evident by a positive bias and high hit rate (i.e. action rate on target trials), while State 2 (dark green) exhibited strong disengagement, evidenced by a strongly negative bias and low hit rate (Fig. 2g-2j). The HMM-GLM accurately recapitulated the behavioral data (Fig. 2i$_i$ and Supplementary Fig. 3e; solid line) but also allowed us to resolve state-like transitions in behavior that are obscured when using temporally-binned behavioral data (Supplementary Figure 3a-d). Interestingly, before the transition, self-restricted CA mice regularly switched between the two states both within and across sessions. After the transition, however, State 1 (engaged) dominated

behavioral performance (Fig. 2i_{ii}-2l and Supplementary Fig. 3e). We then used the HMM-GLM model to predict the transition in a bottom-up manner which we found to be similar to the behaviorally predicted one while providing greater temporal specificity (Supplementary Fig. 3f) (Wilcoxon rank sum test, p=0.82). The model-defined shifts from engaged to disengaged states before the transition were strikingly abrupt (Fig. 2k). Moreover, the last transition, reflecting the putative transition between goal-directed and habitual behavior, was as abrupt as the earlier ones, occurring in ~3 trials (Fig. 2l; last). Consistent with this interpretation, hit rate declined across a session immediately pre-transition, but remained stable post-transition (Supplementary Fig. 4a), suggesting that post-transition responding is less sensitive to within-session factors such as satiety. Together, both quantification of behavioral data and model-based approaches using the HMM-GLM converge on the abruptness of the identified transition thereby motivating a more direct test of decision control.

Sensory-specific outcome devaluation validates transition to habit

To directly test whether the observed transitions in response stability relate to the underlying psychological constructs of goal-directed and habit, we employed sensory-specific outcome devaluation—a gold standard for differentiating goal-directed and habitual control of behavior. A behavior is considered habitual when animals continue to perform an action even when the outcome is devalued via pre-task satiation[15,16,26,27]. This sensitivity to devaluation should be specific to the task-associated reward so as to differentiate general motivation (for water or food) from outcome-specific value[3,28–30]. We conducted sensory-specific outcome devaluation tests in the same animals at two discrete time points during behavioral training, one pre-transition and the other post-transition (Fig. 3a). Vanilla-sucralose flavored water was used as the valued condition while plain water was used as the devalued condition. Animals received free access to either reward before a given test session and were put into the auditory go/no-go task under extinction.

Pre-transition, animals had lower hit rates (normalized, see Methods) in the devalued condition compared to the valued condition, indicating that behavior is under goal-directed control (Fig. 3b). Post-transition, this devaluation effect was abolished, with animals now having similar hit rates in both the valued and devalued conditions. Moreover, the devaluation index (see Methods) of the animals decreased significantly across the transition (Fig. 3c). These

results demonstrate that the transitions in response reliability in our task directly relate to the nature of decision control, namely a transition from goal-directed to habit.

Bilateral lesions to the DLS prevent transitions to habit

The two decision processes (goal-directed and habitual) are thought to be sub-served by distinct neural circuits in the dorsal striatum[31]. The dorsomedial striatum (DMS) and dorsolateral striatum (DLS) are thought to enable goal-directed and habitual behavior, respectively[32,33]. Using standard outcome devaluation procedures, rodents with lesions to the DLS persist in goal-directed mode (i.e. when they should be sensitive to reward devaluation) even after significant amounts of overtraining[34]. Thus, DLS lesions provide a powerful and orthogonal approach to further test the validity of this paradigm in assessing the underlying decision mode. To do this, we bilaterally lesioned the DLS in self-restricted CA mice (NMDA 20µg/µl, 100nl/site) before head post implantation and behavioral training (Fig. 3d). All DLS-lesioned self-restricted CA mice had visible, localized, and overlapping lesions (Fig. 3e-f, and Supplementary Fig. 4d), while shams did not (Supplementary Fig. 4e-f). DLS-lesioned self-restricted CA mice exhibited high variability in hit rates that persisted for much longer than in sham CA mice and rarely transitioned to low variability (Fig. 3g-3i and Supplementary Fig. 4g) (60% sham vs 20% lesioned). Importantly, self-restricted CA mice with DLS lesions exhibited no significant deficits in the learning of the task contingencies or in lick-related motor behavior (Supplementary Fig. 4h-k), ruling out the possibility that the observed effect was driven by motor refinement or differences in task efficiency, two functions also ascribed to the DLS. These data offer independent evidence, through the manipulation of habit-relevant striatal circuits, that the transitions we observe are indeed genuine transitions from goal-directed to habitual behavior.

Licking microstructures demonstrate motor automaticity post-transition

In skill-based learning, motor patterns of ‘automaticity’ can be used as evidence for the expression of habits[35]. We analyzed lick microstructures in detail to determine the extent to which transitions in hit-rate variability were concomitant with changes in motor automaticity. Before transitions, self-restricted CA mice exhibited highly variable lick microstructures (Fig. 4a; top). Post-transition, however, three aspects of their licking behavior abruptly appear: a uniform lick stereotypy (Fig. 4a; Post), an increase in consummatory licks (Fig. 4b-c), and a reduction in

reaction time (Fig. 4d). These patterns were consistent across trials and sessions after habit transitions demonstrating the simultaneous appearance of motor automaticity.

Pupillometry as biomarker for decision control

Pupillary response dynamics are thought to reflect multiple different features, including sensory cues[36,37], arousal/task engagement[38,39], motor activity[40,41], and motivation[42,43]. The extent to which pupil dynamics can provide a biomarker for decision control, however, remains unexplored. We hypothesized that pupil dynamics may track outcome sensitivity and thus performed a detailed inspection of trial-level, phasic pupillary dynamics across behavioral transitions.

We focused our analysis on hit trials in which the animals consumed the reward (rewarded) and those in which they made the instrumental lick to the S+ correctly but did not consume the reward (no-reward). Pre-transition, no-reward trials elicited lower pupil response than the rewarded trials (Fig. 4e-f; Pre-NR vs Pre-R), suggesting a differential coding of the absence/presence of reward. This difference was significantly reduced post-transition, and more strikingly, the post-transition no-reward trials elicited significantly higher pupillary responses than pre-transition (Fig. 4e-f; Post-NR vs Pre-NR) with noticeable effects occurring within 10 trials of the transition (Fig. 4g).

The observed pupillary dynamics could reflect a loss of outcome sensitivity after transitions to habit or could be driven by transition-related differences in task engagement[33,34], motor activity, or motivation (but not sensory cues, since all analysis is done on S+ trials). First, we excluded all disengaged trials from our analysis, given that pupil responses were lower on disengaged versus engaged trials (see Methods, Supplementary Fig. 5b). Second, we focused our analysis on no-reward trials when licking ceased after the first instrumental lick since animals exhibit higher lick numbers post-transition (Supplementary Fig. 5d and Fig. 4c). In addition, we observed that the pupil response had little correlation to lick number per trial (Supplementary Fig. 5e). Third, within-session reduction in motivation due to satiety led to fewer licks per trial (Supplementary Fig. 5f), but had no effect on the phasic pupillary response, both pre- and post-transition, for both rewarded and no-reward trials (Supplementary Fig. 5f and 5g). Given that these key factors were ruled out, we then sought to further test whether the pupil response we observed was related to outcome sensitivity by exploiting experimenter-triggered reward omission trials (versus no-reward trials when animals occasionally did not consume the reward).

On these trials, animals correctly respond to the S+, continued to lick, but a reward was not delivered. The pupil response on omission trials shared the same feature as no-reward trials: the signal increased across the transition and became closer to that of rewarded trials (Supplementary Fig. 5c). These combined data demonstrate that pupillary dynamics are less outcome sensitive immediately after habit transitions, suggesting that phasic pupillary responses can provide a useful biomarker of decision control.

Abrupt reductions in outcome-related signaling in the dorsal striatum mark the moment of a transition

The behavioral identification of abrupt transitions to habit, and the critical involvement of the DLS, led us to next focus on how neural activity in the dorsal striatum evolves across transitions. In prior work, it has been challenging to observe neural activity changes in real-time as animals initially learn in a goal-directed manner and then transition to habit. Our behavioral findings could be explained by a gradual model – whereby neural activity related to habitual behavior slowly ramps up and takes control of behavior only after reaching a threshold. This could lead to the behavioral appearance of an abrupt shift, despite the gradual change in the underlying circuit activity, and would suggest that habits arise through the gradual strengthening of a habitual controller. Alternatively, neural activity could experience a switch-like change in activity at the moment of the transition. This would suggest that a habit controller is readily available but only comes online when recruited at a precise moment in time. To test between these two plausible models, we performed dual fiber photometry in the DLS and DMS (Fig. 5a) as animals transitioned from goal-directed to habitual behavior.

We first assessed the within-trial dynamics of the DLS and DMS by focusing on one hundred trials around the behaviorally-identified transition (Fig. 5b). To minimize effects driven purely by changes in general motivation in engaged versus disengaged blocks of trials (Supplementary Fig. 6a; see Methods for classification), we selected only engaged trials for appropriate comparison across the transition. As has been described previously[44–46], we observed both early-in-trial (cue-response related) and late-in-trial (outcome-related) activity patterns (Supplementary Fig. 6b-c). When aligning neural data to the transition, we observed a striking reduction in the late-in-trial activity in both the DLS and DMS post-transition (Fig. 5c), with a more pronounced reduction in the DLS (Fig. 5; top, and Supplementary Fig. 6l-m). Accompanying this reduction was a sharpening of the early-in-trial signal in the DLS (Fig. 5e

and 5f). These data suggest that post-transition, the DLS, and to a lesser extent the DMS, shows reduced outcome-related activity and more stereotyped responses to the cue itself, consistent with a stimulus-response mode of control. Importantly, these differences were locked to the transition and were not present when average activity was computed on random trials from the transition day (Supplementary Fig. 6n), nor when activity was averaged around the same within-session transition trial, but 3 days before the true transition day (Supplementary Fig. 6o).

These findings suggest that both the DLS and DMS exhibit stable patterns of activity prior to the behavioral shift, indicating that neural signatures of habitual control may emerge prior to their behavioral manifestation. To assess that possibility, we sought to understand when the DMS and DLS encoded the different trial types (hit, miss, false alarm, correct reject) versus the individual task events (cue, action, outcome). At the outset of learning (i.e. the first 20 trials), we observed no evidence of contingency-specific activity, with activity in the DLS and DMS being dominated by individual task events (e.g. cues, licking vs no licking) (Supplementary Fig. 6b, and Supplementary Fig. 6d; Pre-Acq.). When animals became expert at the discriminative task, but hundreds to thousands of trials before the transition (pre-transition), we observed robust contingency-specific activity that then remained stable until the moment of the transition (Supplementary Fig. 6c-d; Pre-transition, and Supplementary Fig. 6e; Pre-transition early vs. late).

Finally, we sought to determine how precise the within-trial dynamics in the DLS and DMS tracked the behaviorally-identified transition. We performed a trial-by-trial analysis after aligning all animals to the transition. The DLS exhibited an abrupt drop in late-in-trial activity at the moment of a behaviorally-identified transition (Fig. 5g, top, five trials pre- and post-transition and Fig. 5h, bottom, black, 50 trials pre- and post-transition), with a similar drop in DMS late-in-trial activity occurring over the next several trials (Fig. 5g; bottom, and Fig. 5h; bottom, gray). One possibility is that this change could be driven by changes in lick structure (observed post-transition, Fig. 4a-d, and Supplementary Fig. 6f-g), and not a change in decision control. We found no relationship between consummatory licks and late-in-trial signal changes (Supplementary Fig. 6i), highlighting the independence of these signals from motor automaticity linked to habit. Even when restricting our analysis to trials with the same response latency pre- and post-transition, we continued to observe this robust drop in outcome-related activity (Supplementary Fig. 6m). To further test whether these signals reflected outcome-related computations, we compared the amplitude of the late-in-trial signal in no-reward pre-transition

trials (i.e. trials where the animal executes an operant lick but no consummatory licks, Pre-NR, gray), with rewarded post-transition trials (Post-R, red) in both DLS and DMS. Despite the completely distinct lick structures, these two signals were indistinguishable (Supplementary Fig. 6j-k). Moreover, this similarity suggests that the dorsal striatum becomes less sensitive to the presence of the outcome when behavior is habitual. Taken together, these neural data show that transitions to habit are indeed abrupt, occur at trial-level resolution, and are governed by a switch-like change in outcome-related signaling in the dorsal striatum.

Discussion

A fundamental tenet of animal behavior is the existence of multiple controllers that govern decision-making. One prevailing framework is that instrumental decisions come about from two distinct processes: goal-directed and habitual[47]. In rodent-based learning paradigms, goal-directed behaviors are thought to become habitual upon overtraining[48–50]. The goal-directed system dominates early in learning when an outcome is desirable. This requires both a representation of the action-outcome contingency and the recognition of the outcome as an incentive. Goal-directed control is value-based and flexible but also cognitively demanding[51]. With overtraining, the habitual system simplifies decision-making by shifting to a stimulus-response mode of behavior[51], making decisions potentially value-less and inflexible but less cognitively demanding[52]. Over the past 50 years, behavioral, neural, and theoretical support for these two distinct decision processes (but also the complexity of their interaction) has grown largely due to behavioral manipulations, including different reinforcement schedules coupled with outcome devaluation and contingency degradation. These behavioral tools have been invaluable to gain a deeper understanding of the behavioral, neural, and theoretical basis of the multiple systems controlling decision-making. Nevertheless, the extent to which discrete measures of sensitivity to outcome devaluation sufficiently distinguish goal-directed from habitual control is still under scrutiny[53,54] as sensitivity to outcome devaluation can also be triggered by unexpected cues[29] and in situations where habits are expected to form[55]. More broadly, the current methodologies remain fundamentally limited in their temporal resolution and individual specificity, limiting the assessment of nature, timing, and properties of habits[5,53] in individual animals. As a result, an essential question first posed by Clark Hull in 1943 has remained unresolved: ‘Is this transition abrupt, or is it gradual and progressive?’

Here, we lay out an approach that allows real-time assessment of the underlying decision process without the need for discrete testing sessions and without the implementation of specified training schedules to bias decision modes[21,56,57]. We hypothesized that by allowing animals to balance hydration needs against their mild distaste for CA water (self-restriction), they would gain agency on their decision to engage in a task based on the desire for an outcome (in our case, plain water droplets). Here, we show one approach to address this question by giving animals ad libitum access to CA water[23,24] in the home cage, which reduces reward-seeking for plain water provided in an auditory cued go/no-go task (Fig. 1b-d). In this way, when in goal-directed mode, animals occasionally engage and disengage from the task, reflected in the behavior as state-like switching in response rate (see Fig. 2a and Fig. 2d). Importantly, our task embeds a readout of task learning that is independent of response rate. Specifically, we used a discriminative experimental design, i.e. a cued go/no-go, that explicitly separates the learning of the discriminative contingencies (S+ → lick → water; S- → withhold → avoid timeout) from response stability (rate of licking to the S+, hit rate). The hardest aspect to learn for rodents in this task is to withhold licking to the S- and thus mastering the task can be measured, independent of engagement, by the rate of false alarms (incorrect licking to the S-).

One potential challenge is the extent to which this task reflects an instrumental versus Pavlovian learning process. Despite reliance on licking which could also reflect an appetitive Pavlovian response, multiple factors point to the initial lick in this task as instrumental. First, animals must perform a specific action (lick) at a specific time (after cue offset + 100 ms) to trigger reward delivery. This action is not performed during the cue or immediately after, ruling out typical Pavlovian anticipatory responses. Second, animals took longer to suppress licking to the S– cue, which is consistent with the well-known asymmetry in learning to perform versus inhibit actions (especially when suppression is not explicitly reinforced). If the differential response to S+ and S- simply reflects failed Pavlovian conditioning due to lack of stimulus-reward pairing, the animals should not exhibit initial high response rate to S-, let alone the lengthy period that it takes to yield reliable response inhibition. This suggests action selection which is one of the characteristic traits of instrumental learning. Most importantly, the trial-by-trial engagement patterns pre-transition provide a behavioral signature of internal evaluation and outcome sensitivity. This was also confirmed with sensory-specific outcome devaluation. A purely Pavlovian account would predict cue-locked responding that does not vary so

substantially with motivational state. In aggregate, our interpretation of instrumental control emerges from this convergence of evidence—not from any single behavioral metric in isolation.

As behavior becomes habitual, animals shifted to constant engagement, behaviorally observed as an abrupt transition to a constantly high hit rate (Fig. 2a, Fig. 2d, Fig. 2m, and Supplementary Fig. 3e). These transitions occurred thousands of trials after reaching expert discrimination performance (Fig. 2f) but at different time points for individual animals. These data argue that the response variability observed during goal-directed behavior reflects intrinsic reward-seeking dynamics rather than incomplete learning. In contrast, habits emerge with a marked reduction in response variability. We then used orthogonal measurements of motor automaticity, pupillary dynamics, sensory-specific outcome devaluation, lesions of the DLS, and neural recordings in the dorsal striatum to confirm that our observations reflected a transition to habit versus differences in vigilance or discrimination ability. In other words, behavioral automaticity and reduced outcome-specific sensitivity occurs concomitant with habitual transitions in our paradigm. While automaticity alone might be a reductionist perspective on habits[55], the composite picture across behavioral, physiological, and neural approaches in our study points toward the habitual nature of post-transition behavior in intermittently engaged mice. This novel approach adds a powerful en passant tool to study decision control. Thus, promoting and quantifying variability in task engagement—regardless of the specific approach, task, or animal model—can provide a useful proxy for behaviorally dissecting transitions from goal-directed to habitual control.

Pinpointing the precise nature of this transition allows us to explore the psychological and neural basis of habits in real-time. Our findings challenge the notion that habits emerge gradually as we demonstrate that habits come online spontaneously and abruptly in individual animals. This finding suggests that the abrupt shift to habit may reflect a more insight-like cognitive resolution process, whereby animals adopt a stable and efficient stimulus-response strategy to minimize cognitive effort under consistent task conditions. This builds on prior work suggesting that animals may experience nonlinear changes in learning either through hypothesis testing in goal-directed tasks[58–60] or step-like changes in associative strength in Pavlovian tasks[61]. The abrupt habit transition suggests that a separate higher-level process might arbitrate between goal-directed and habitual control. Factors such as cognitive demand or environmental uncertainty are likely to contribute to when the commitment emerges to solve the task in a simple and inflexible manner, activating an otherwise dormant habitual controller. Our neural data

support this interpretation, as we find that the nominal habit controller (the DLS) exhibits stable patterns of contingency-specific activity prior to the habit transition.

The current consensus, though contentious[62], is that habitual and goal-directed behavior are supported by the DLS and DMS, respectively[32], but studies utilizing discrete satiety test sessions or experimenter-defined overtraining periods yield confounding and even conflicting results: some evidence argues that DLS activity changes rapidly before the behavior onset of a habit[63], with others finding that the change is more gradual and closely aligned with a behavioral change[64]. Recent reports even observe an eroding distinction in the control of actions between the DLS and DMS as training progressed[62]. From our data, we observed that a habit is instantiated not when the DLS gradually overtakes the DMS, but when outcome-related signaling is reduced in the DLS and a few trials later in the DMS (Fig. 5h). Consistent with prior reports, the DMS appears to remain active even once habits have been instantiated[62]. Given this, our neural data indirectly points to the idea that animals under habitual control may still internally experience periods of goal-directedness as both the DLS and DMS are active in parallel during goal-directed and habitual phases of behavior. This interpretation would be more in line with the human experience where, for example, the initial reach for a cup of coffee in the morning may be habitual but subsequent sips may be driven by the goal for the coffee itself[65]. Future work will be required to directly test this notion in rodent models of habit.

In addition, given that both the DLS and DMS remain active even when a habit emerges, habits need not be permanent. Interestingly, some animals reverted to goal-directed mode after several sessions in habit mode (Fig. 2d, yellow asterisks and Supplementary Fig. 4g, asterisks), suggesting that transitions to habit may not be intrinsically permanent. A higher-level arbitration process that controls the switch may also help explain why techniques such as outcome devaluation yield conflicting results[66] due to individual variability in when transitions occur.

The spontaneous and abrupt appearance of habits in our paradigm serves as a behavioral marker to identify neural signatures associated with habitual transition[5]. We observed that both DLS and DMS signals exhibited a biphasic pattern, characterized by an initial peak aligned to the cue, likely reflecting cue- and action-related contingencies[44], followed by a second late-in-trial peak linked to outcome-related computations, and possibly related to evaluative signals[45,64,67,68]. Consistent with prior reports, the early peak is more pronounced in the DLS[44,45], while the DMS exhibits relatively larger and longer late-in-trial signals, potentially including a third peak (Fig. 5c and Supplementary Fig. 6j-bottom). Our data revealed that at the moment of

habitual transition, both the DLS and DMS exhibit abrupt reductions in the late-in-trial signal, with this third peak in the DMS being prominently reduced (Fig. 5c; bottom). More broadly, large late-in-trial signals in the DMS may indicate a form of deliberative processing, which has been associated with longer response latencies (Supplementary Fig. 6h)[69,70]. In line with this, post-transition response latencies were shorter and displayed reduced late-in-trial signals. Importantly, the reduction in late-in-trial activity post-transition was not related to changes in response vigor (Supplementary Fig. 6i), as pre-transition no-reward trials (when the animal executes an instrumental but no further consummatory licks) and post-transition rewarded trials (when the animal executes both instrumental and consummatory licks) exhibited similar late-in-trial dynamics (Supplementary Fig. 6j-6k). These data argue that reduced outcome-related signaling reflects a state of reduced reward sensitivity and points to diminished post-decisional evaluative processing.

It should be noted that in this study we used a pan-neuronal promoter and fiber photometry which together provide a bulk activity measure that includes somatic and non-somatic signals[71]. While non-specific in this way, these experiments were designed specifically to test whether the underlying regional activity reflects a gradual strengthening of a habitual circuit that reaches a threshold or a 'switch-like' process in which existing circuitry is rapidly engaged. We show that it is the latter. These findings support the idea that the capacity for habitual behavior is neurally instantiated in advance but lies dormant until recruited. One limitation of our current approach reflects the slow temporal kinetics of calcium sensors coupled with the rapid within-trial structure of events. This combination makes it difficult to fully disentangle neural encoding of the cue, action, outcome, and outcome evaluation period. As a result of this, we use two time periods when evaluating neural signals: cue-related, which may include an action component and outcome-related, which may include both outcome and evaluative signals. Future work can combine higher temporal resolution approaches, such as electrophysiological recordings, or other adjustments to the task design to increase the temporal separation between task events.

Our findings raise the important question of what brain regions and neural processes adjudicate decision control. The potentially higher-level nature points to regions such as premotor or prefrontal cortical areas[3,72], which directly interact with the dorsal striatum[73], and have been shown to actively compute higher-order variables such as environmental uncertainty and cognitive effort[3,72–79]. Alternatively, this arbitration process may arise from striatal dynamics

itself, given the confluence of top-down (cortical) and neuromodulatory (dopamine) input. Much future work is required—through careful interrogation of these neural dynamics in a cell-type and projection-specific manner coupled with variations in task design (including extension to more traditional free-operant behaviors) and causal manipulations of task events—such as optogenetic suppression of DLS outcome-related signaling after task acquisition to prompt animals to transition, to define the precise logic of decision control in the dorsal striatum.

While our task differs from traditional instrumental training paradigms used to explore goal-directed and habitual control of behavior, the main value of our approach lies not in the level of restriction or the existence of behavioral variability, but in the recognition and quantification of this variability in reward-seeking as a proxy for underlying decision control. This provides a complementary approach to existing procedures for probing the temporal dynamics and neural drivers of habit transitions in real-time. More broadly, our finding of an abrupt shift in decision control may inform distinct interventional strategies and new approaches for addressing habit-related disorders in humans, including combining desensitization and cognitive control strategies[80,81]. Additionally, our data suggest that it may be possible to predict when transitions will occur and if such predictions are possible and can be extended from rodents to humans, it could provide a powerful tool to interfere with or manipulate the emergence of maladaptive habits.

Methods

Animals

All mice were housed in standard plastic cages with 1-4 littermates and kept in a 12-h/12-h light/dark cycle (10:30 am / 10:30 pm) with controlled temperature (19.5-22°C) and humidity (35-38%). All the mice used in this study were male C57BL/6J from Jackson Labs (strain# 000664) and were between 11 to 15 weeks of age at the start of training. All the experimental and surgical procedures were approved and performed in accordance with the Johns Hopkins University IACUC protocol (license # MO20A272). Sample sizes were determined based on standard cohort sizes from relevant literature. Mouse allocation to specific groups was randomized but the experimenters were not blinded to group types.

Surgical procedures

Mice were anesthetized with isoflurane (5.0% at induction, 1.5-2.5% during surgery) and placed on a stereotactic apparatus (Kopf). The hair over their skull was removed with hair removal cream and the area disinfected with betadine. The skin over the skull was removed and the area was cleaned of

connective tissue with 3% H_2O_2. A custom-made stainless-steel head-post was fixed onto the exposed skull with C&B Metabond dental cement (Parkell). The animals were given 1-3 days to recover.

Striatal lesions

Mice that underwent bilateral DLS excitotoxic lesions or sham injections, received N-Methyl-D-aspartic acid (NMDA; Sigma Aldrich, 20µg/µl NMDA in PBS1x with 10% glycerol) or vehicle (PBS1x with 10% glycerol) respectively (with a Hamilton syringe and a Harvard Apparatus Pump 11 Elite, 100nL/injection-site at a 70nL/min). The injections were made at AP+1.0mm, DV±2.6mm, ML-2.8mm via burr holes which were sealed with Jet Denture Repair Acrylic (Lang Dental) prior to headpost implantation.

Fiber photometry

The animals were injected with AAV-syn-jGCaMP8m-WPRE, (≥1×10^{13} vg/mL, 1:10 diluted in saline), Addgene Cat #162375-AAV9) in one hemisphere targeting DLS and DMS in the other. The injections were made at AP+1.0mm, DV±2.6mm, ML-2.8mm for DLS and AP+1.0mm, DV±1.5mm, ML-2.8mm for DMS via burr holes, which were sealed with Jet Denture Repair Acrylic (Lang Dental) prior to skin closure with suture (5-0 coated vicryl plus undyed 1X27" RB-2, Ethicon #VCP433H). After 2 weeks of expression, animals underwent fiber implantation surgeries. The surgical site was reopened, and optic fibers (Φ 1.25 mm Ceramic Ferrule, 200um Core, 0.39NA, length = 3mm, RWD # R-FOC-BL200C-39NA) were implanted in the burr holes and secured with dental cement, followed by headpost implantation. Habituation began 7-10 days after the animals regained their original weights.

Habituation and water restriction paradigms

After recovery from surgery, animals were handled and habituated prior to the start of training for at least 10 days based on previous studies[82]. Head-fixed experimental CA mice and their littermate controls (WR85) underwent the same surgical, habituation and testing procedure. Animals were handled by the experimenter/s at increasing times every day, exposed, and habituated to the head fixation station. The different water restriction paradigms started after at least 3 days of handling. The standard water restriction (WR85) protocol prevented the animals from accessing water in their home cage. The mice were weighed daily, and a limited amount of water (~1.0 mL) was given individually to maintain 80-85% of their original weight. For naturalistic water restriction with citric acid (CA), animals had ad libitum access in their home cage to a bottle of tap water with citric acid dissolved following reported protocols[23,24]. The mice were slowly introduced to the taste of CA, increasing its concentration daily from 0.5% CA to 1-3%, and adjusted accordingly within this range to keep the animals at ~95% of their original weights.

Behavioral training

All behavioral training was done using Bpod State Machines (r1 or r2, Sanworks). After habituation, lick training was performed for two days in which animals had to lick in order to receive a small water droplet (un-cued). Immediately after water delivery, licks would not be rewarded until 2 seconds had passed. The lick training session ended either when 1 ml of water was consumed, or the session had reached 30 minutes. On a subsequent session, mice began training on a go/no-go auditory task. Behavioral events (trial structure, stimulus and reward delivery, lick detection) were controlled and stored using a custom-written MATLAB program (2018b, The MathWorks) interfacing with the Bpod, an electrostatic speaker driver (E1, TDT) and an infrared beam for lick detection. In a subset of animals, facial movements and pupil size were measured with either: 1) a Raspberry Pi (3B) and a Raspberry Pi camera module (NoIR v2) coupled with a Bright-Pi infrared LED array (PiSupply); or 2) an Arducam 100fps Global Shutter USB Camera and a custom-made infrared LED array. Mice were head-fixed inside a Plexiglass tube facing a lick-tube. A free field electrostatic speaker (ES1, TDT) was located ~5 cm from the animal's left ear and each sound (either 4757 Hz or 8000 Hz, as target or foil stimuli) was calibrated to an intensity of 60-62 dB (SPL). The pupil camera and IR LED array were positioned ~6 cm away from the animal's face in a 60-degree angle. Everything was enclosed in a custom-made sound-attenuated box. Target and foil tones were pseudo-randomly ordered (equilibrated every 20 trials). Each trial consisted of a pre-stimulus no-lick period (2 s), stimulus presentation (100 ms), delay (100 ms), response period (2 s) and variable inter-trial interval (ITI). Typically, mice were trained for ~300-320 trials per with a short block of 20 non-reinforced trials interleaved in the middle of the session[83,84]. Training lasted for a maximum of 30 days.

Sensory-specific Reward Devaluation

A cohort of 8 animals underwent the same citric acid protocol as previously described. Individual-animal action rates were measured in blocks of 50 trials to obtain hit and false-alarm rates in small blocks. A transition was defined by a criterion of >1600 trials of stable action rate and low action variability. We adapted existing protocols for food-driven tasks[29,85,86] in mice to suit our task structure with thirst as the motivator[87]. Tap water (task reward) was the devalued condition, whereas a vanilla-sucralose solution (2.5% vanilla, 0.4% sucralose) was used as the valued condition. Sucralose does not provide caloric intake but is detectable by the animals as a distinguishable flavor[88]. Vanilla was added to further increase the salience of flavors between the two tastants. Animals were pre-exposed to the vanilla-sucralose solution for 1h during habituation to avoid neophobia. Sensory-specific reward devaluation tests were performed on days 7 (pre-transition) and 23 (post-transition) for 6 animals. One animal reached transition criteria on day 32; thus, it received the post-transition reward devaluation (RD) test on day 33. On valued and devalued testing days, animals received ad libitum access to either solution in each day: water (devalued condition) or vanilla-sucralose solution (valued condition) for 10 minutes prior to a non-rewarded devaluation session (sometimes referred to in the literature as devaluation under extinction).

The two devaluation sessions were performed on consecutive days, and the order in which the animals received valued and devalued conditions was counterbalanced. Access to the solution was provided in a separate cage for each animal, which was immediately transferred to the behavioral setup for testing. The weight difference before and after access to the reward was used to calculate the consumption amount (mL). During devaluation tests, animals received 40 trials in total (20 S+ and 20 S-), and no reward was provided for correct licking to the S+. The raw hit rate was calculated as the percentage of responded S+ trials over total S+ trials in a given devaluation session. The trial structure was otherwise identical to training trials. Animals (n = 1) that did not show a habitual transition by day 32 were excluded from the analysis. All other animals (n = 7) were used for the final analysis.

Fiber photometry recordings

Mice underwent the same citric acid training protocol as previously described. Additionally, during each training session, fluorescence signals were recorded in both DLS and DMS for the same animals, using a commercial fiber photometry setup (Neurophotometrics, FP3002), coupled to the Bpod (Sanworks) and controlled via customized Bonsai workflow[89]. Pairs of animals were recorded simultaneously, using a 4x branching patch cord (length = 3m, NA = 0.37, 200 µm/1.25 mm black ceramic FOC, FC connection, mbf Bioscience #NPM-BPC-4). LEDs delivered two excitation wavelengths (470 nm for GCaMP6s, 415 nm for isosbestic) interleaved at 50 Hz. LED power was calibrated to an intensity of 50 µW at the tip of the patch cord for each channel. Fluorescence emission was collected with a CMOS sensor, with a region of interest defined around the end of each patch cord for every recording session. The patch cord was bleached using maximum intensity of both LEDs, for ~8-12 h before the start of every cohort. Photometry data was acquired with Bonsai and subsequently exported to MATLAB for analysis. All signals were aligned to analog TTL pulses generated by the Bpod, marking behavioral events.

Histology

After every experiment, the brains of all animals were obtained via transcardiac perfusion[90] and stored in 4% paraformaldehyde solution in PBS1x overnight. After further dehydration in 30% sucrose (Sigma-Aldrich), the brains were frozen in OCT gel (Tissue-Tek®) and sliced using a cryostat (Leica) into 50um slices.

Striatal lesions

The slices were mounted on gelatin-coated slides (FD Neuro) and left at room temperature to dry overnight. The following day, the slides were stained using 1% Cresyl Violet acetate (Sigma-Aldrich #C5042) solution and cover glasses were placed and fixated with Permount™ Mounting Medium (Fisher Chemical™ #SP15). The slides were imaged under Brightfield settings in a Zeiss upright microscope (Axio Zoom.V16).

Fiber photometry

The slices were washed with 1XPBS at RT on shaker (30 rpm) for 3 times, 5min each. Permeabilization was done with 0.2% PBS-Triton (TritonTM X-100, Sigma-Aldrich #X100-500ML) at RT on shaker for 1h. Blocking was done with 5% Normal Donkey Serum (Sigma-Aldrich # D9663) in 0.2% PBST at RT on shaker for 1h. The slides were stored in primary antibody (Goat Anti-GFP IgG, 1:1000 dilution in blocking solution, Novus Biologicals #NB100-1770SS) at 4°C overnight. The next day, the slices were washed with 0.2% PBST at RT on shaker for 3 times, 5min each, then stained with secondary antibody (Alexa Fluor® 488 AffiniPure™ Donkey Anti-Goat IgG, 1:500 dilution in 0.2% PBST, Jackson Immuno #705-545-003, ex488nm/em496nm) at RT on shaker for 2h. The slices were washed with 1XPBS at RT on a shaker for 3 times, 5min each, before mounted on microscope slides (Premium Superfrost® Plus Microscope Slides, VWR # 48311-703) with DAPI Fluoromount-G® (SouthernBiotech #0100-20, ex405nm/em465nm), and imaged with Zeiss LSM 700 confocal microscope under 10X magnification.

Statistical analysis

All analyses were performed using custom-written MATLAB code (The MathWorks, 2019b, 2021a, or 2022a) or R environment. All datasets were tested for normality using a one-sample Shapiro-Wilk test; then, parametric or non-parametric statistical tests were applied accordingly. Two-sample (paired or unpaired) t-tests were used for parametric data, and Wilcoxon rank sum tests were used for non-parametric data. Where required, paired comparisons were made. For multiple comparison analyses, (1-way, 2-way, repeated measure) ANOVA or ANCOVA was performed for parametric data, and permutation-based ANOVA for non-parametric data. Post-hoc pairwise comparison was conducted with Bonferroni corrections or LSD test. To build a Receiver Operating Characteristic (ROC curve) (Fig. 3i) we used the transition probability of lesioned and sham animals to obtain the area under the curve (AUC) and generate a shuffled probability distribution to statistically test our experimental animals' distribution difference. Significance was determined as the difference in AUC value between lesioned and sham animals when it fell beyond the 95^{th} percentile confidence interval (Fig. 3j). Except for one-sided paired t-test which uses two-sided 90% confidence interval, all confidence intervals correspond to $\alpha=0.05$. Significance is represented as n.s. $p>0.05$, * $p\leq0.05$, ** $p\leq0.01$, *** $p\leq0.001$, and **** $p\leq0.0001$. Effect size was calculated as ηp^2 for F-based ANOVA, partial R^2 for X^2-based ANOVA, Cohen's d for unpaired t-test and Wilcoxon rank-sum test, Cohen's d_z for paired t-test. In most cases of Wilcoxon signed-rank and rank-sum tests, z values were reported when possible (otherwise U or W was reported).

Behavioral analysis

The first instrumental lick to the target tone to trigger target delivery was used to determine whether the trial is a hit or miss. The subsequent licks after reward delivery were quantified as consummatory licks.

The action rate was calculated as the percentage of hit trials among all target trials. Individual-animal action rates were measured in blocks of 50 trials to obtain hit and false-alarm rates in discrete but small blocks that allowed us to observe behavioral variability. Behavioral discriminability was calculated using the z-scored hit rate minus the z-scored false-alarm rate (d'). To avoid infinite values when rates of 1 or 0 are present, the values were corrected by 1-1/2N or 1/2N respectively, where N corresponds to the number of trials. For all the data presented in this paper, we considered animals to have effectively learned the task by calculating d' during non-reinforced trial blocks, which has previously been demonstrated as an accurate measure of task acquisition[83]. Only mice with a d' > 2 during these non-reinforced trial blocks were included in the analysis (corresponding to 45 out of 48 mice tested, 1 WR85, 1 CA-sham and 1 CA-lesioned mice did not learn the task and thus were excluded).

HMM-GLM model implementation

We fit a GLM-HMM model to trial-by-trial choices of each mouse from 4 days before putative habitual transitions to 4 days after the putative transition (9 days in total). Each state in HMM contains a Bernoulli GLM defined by a weight vector specifying how stimulus inputs and bias are integrated in that state. The model was fit using a previously published expectation-maximization (EM) algorithm[25]. To identify the optimal number of states, we evaluated the cross-validated BIC by fitting choice data from the 5th day before and after habitual transition. A 2-state model was sufficient to explain the choice behavior of six animals, capturing an engaged state and a disengaged state, whereas a 3-state model was needed for two animals, capturing an additional low-discrimination state. For these two animals, we focused only on the engaged and the disengaged state in subsequent analysis. To compute state occupancy, we first inferred the behaviorally dominant state as the state with the highest probability in each trial, and then calculated the percentage of trials that a state is dominant in a 50-trial bin. We inferred the habitual transition by identifying the last trial bin where the occupancy of disengaged state was above a threshold of 30%. The number of trials needed for transitions between engaged and disengaged state is calculated by the number of trials needed for the dominant state to reach 75% probability after the transition. To compare the model-inferred transitions with behavioral data, we quantified the slope of inferred state probability by the GLM (z-scored) at the trial of state transition, compared to the slope of hit rate changes during state transitions (z-scored), quantified using various bin sizes around the transition.

Preprocessing of pupillometry data

Pupillometry videos of 30-40 minutes long (n=8) were taken as the training dataset for a DeepLabCut[91,92] (DLC) pre-trained model (resnet_50). Manual labeling of pupil contour consisted of 8 points (up, down, left, right, up-left, up-right, down-left, down-right) across 180 randomly selected frames. The network was trained for >200,000 iterations until the loss rate plateaued. The final network was used to analyze the

pupillometry videos from the experiment. Custom MATLAB code (The MathWorks, 2021a) was then used to remove blink artifact, reconstruct pupil diameter, and apply a low pass filter (3 Hz) to the data. Individual trials for individual animals were normalized to the first frame of stimulus onset.

For behavioral data, 10-trial windows were taken around each hit trial (5 before + 5 after), and only those with window hit rate higher than 0.75 were selected as engaged trials. Trials in which animals make only 1 lick to the target stimulus with no subsequent consummatory licks were classified as non-consumed hit trials. Non-rewarded trials were introduced in 20 trial blocks (10 targets and 10 foils) on random days, and hits to target tones served as omission trials. All trials are within 3 days pre-and post-transition. Numbers of data points were matched across groups via taking averages of randomly selected trials in the larger groups.

Fiber photometry recording analysis

Fiber photometry signals were corrected for photobleaching by customized function. The baseline was calculated using a moving mean function, and the ΔF/F was computed as: $dff = \frac{raw\ signal}{baseline} - 1$. Both the signal and isosbestic reference are converted into z-scores by subtracting the median and dividing by the standard deviation. The reference was fit to the signal using robust linear regression, and the final ΔF/F was calculated as the difference between the z-scored signal and the aligned reference. The ΔF/F was smoothed using a moving mean with a window of 10 o reduce noise.

After photobleaching correction, signals were cut, taking a 4s window 10 frames before tone onset. Each trial trace was aligned and normalized to tone onset. All data shown is aligned to tone onset (time 0 s). For a trial-aligned transition selection, first, engagement blocks were defined as consecutive periods with a hit rate over 85% after task acquisition. The last transition was selected per animal back-tracing to the last block of Misses before a high-engagement block. Engagement for individual trials was defined around Hit trials, using a window of 5 trials before and 5 trials after each hit, and measuring miss rates. If the miss rate was smaller than 75%, that hit and the block of different trials on that 11-trial window were identified as ‘engaged’. All analysis was restricted to engaged trials unless otherwise stated. Pre-transition epochs comprised from after task acquisition to the trial before the transition for individual animals. All AUC calculations were done using the trapz MATLAB function for selected windows. 50 hits were selected pre- and post-transition for each animal in Figure 5c-f and pooled together. For operant latency distributions, a histogram that contained the average of 50 trials per animal with latencies larger than 20 ms from tone onset were composed in 12 bins. The operant latency jitter was obtained using Levene tests based on absolute deviations from the group mean. ‘Cue’ window comprised a 1 s window after tone onset. ‘Outcome’ window comprised the following time after the cue (~2.8s). Logistic fits were applied to the average of 50 trials per animal for the cue + outcome, cue alone or outcome alone AUC

windows. Logistic fits were obtained using a moving window over the averaged trials with a 5-trial window length.

Task acquisition was defined as the day in which animals show consistent d'>1.5 (trial # at the end of that day). Average acquisition trial was 1200, with std=424 trials, n=7 mice for DLS, n=6 mice for DMS (Supplementary Fig. 6d). The average transition trial was 10292, with std=3318 trials, n=7 mice for DLS, n=6 mice for DMS (Supplementary Fig. 6e). For 'early' and 'late' windows of pre-transition data, 'early' trials comprised the first 1171 trials after acquisition, and the 'late' were the last 1171 trials before the transition, to select a strong comparison. Consummatory licks were counted as any licks happening after initial the operant lick. In Supplementary Fig. 6j, for the DLS, n=7 animals pooling together all trials, with a total of 17,369 pre-transition trials with reward, n=3,553 pre-transition trials without water consumption, n=6,340 trials with reward post-transition, and n=477 post-transition without water consumption. For the DMS, n=6 animals pooling together all trials, with a total of n=13,671 pre-transition trials with water consumption, n=3,383 pre-transition trials without water consumption, n=5,972 trials with water consumption post-transition, and n=417 trials without water consumption post-transition. A sub-selection of pre- and post-transition trials was applied to assess the effect of short latency trials in the late-in-trial DLS activity (Supplementary Fig. 6m). From a total of 350 trials shown in Fig. 5c, only 91 trials pre-transition complied with the rule of belonging to trials with 1st lick latencies >0.328 s. These were compared with 29 trials post-transition under the same rule. As a control, we selected 20 DLS activity trials per animal on the transition day and shuffled trial identity into two groups mimicking pre (black) and post (red) transition (Supplementary Fig. 6n). A second control was made by maintaining the transition structure but selecting 20 trials 3 days before the transition, as a 'pseudo-transition'. We grouped pre (black) and post (red) pseudo-transition trials (Supplementary Fig. 6o).

Sensory-specific reward devaluation test analysis

We determined that the differential consumption of vanilla-sucralose and tap water during the devaluation sessions was a covariate for the hit rate[21,29,93,94].

We performed a two-way repeated-measures ANOVA on consumption amount across conditions (Supplementary Fig. 4b). The ANOVA was implemented through a Linear Mixed-Effect Model that used condition, transition, and their interaction term as the factorial predictors for consumption, while also included the animal identities to account for the within-subject variance. The analysis revealed a significant effect of condition, $X^2(1)=5.56$, $p=0.018$, partial $R^2=0.00$, and transition, $X^2(1)=8.00$, $p=0.0047$, partial $R^2=0.00$. Post hoc comparisons using estimated marginal means with Tukey correction and Kenward-Roger df method indicated that the mean consumption for the pre-transition (M = 0.5, SE = 0.041) and post-transition (M = 0.37, SE = 0.041) valued condition was significantly different, $t(24.5) = 2.78$, $p = 0.010$. The pre-transition valued (M = 0.50, SE = 0.041) and water (M = 0.39, SE = 0.048)

conditions also had significantly different mean consumption, t(24.5) = 2.46, p = 0.021. Therefore, animals generally drink more in valued than devalued condition, and in pre- than post-transition. The absence of a significant group × transition interaction indicates that these effects are additive rather than group-specific responses to the transition. Due to low residual variance in repeated-measure ANOVA, a low R2 for the main effects suggests the differences are consistent but small, yet still warrants further analysis.

To test the effect of liquid consumption amount on hit rate, we performed linear regression with the two variables. Hit rate and consumption amount were negatively correlated (Supplementary Fig. 4c).

To verify that the correlation between consumption and hit rate was uniform across animals, we compared different random slope models to the random intercept model currently employed by the analysis. Including condition, X^2 (2) = 1.93, p = 0.38, transition, X^2 (2) = 1.02, p = 0.60, or both, X^2 (5) = 5.17, p = 0.40, as within-subject effects did not significantly increase the overall model fit. Therefore, we proceeded with the random intercept fixed slope model, which considered only the subject identities as the random effect.

We then proceeded to perform a 2-way repeated-measure ANCOVA to analyze the effect of condition and transition on hit rate, with consumption as a covariate. The slope of the consumption covariate was extracted to calculate the normalized hit rate. The homogeneity of regression slopes assumption was not violated (insignificant interaction effect between the predictors and the covariate, 2-way ANOVA, X^2(3, N = 10) = 2.67, p = .45). Since the ANCOVA was implemented through a Linear Mixed-Effects Model, the slope of the consumption covariate (-0.75) was extracted to calculate the normalized hit rate with the formula below:

$$Hit\ Rate_{normalized} = Hit\ Rate_{raw} - slope_{consumption} \times (Consumption - \overline{Consumption})$$

For the 2-way repeated-measure ANOVA on normalized hit rate. Post hoc comparisons using estimated marginal means with Tukey correction and Kenward-Roger degree of freedom method indicated that for pre-transition, the mean normalized rate of valued condition and that of the devalued condition were significantly different, t(24.5) = 2.36, p = .027. The mean normalized rate of the pre-transition and post-transition devalued condition was significantly different, t(24.5) = -4.42, p = .00020. Effect size was calculated as partial R^2 for the repeated-measure two-way ANOVA, and Cohen's d for the t-test.

The devaluation index for each animal was calculated based on the normalized hit rate with the formula below:

$$Devaluation\ Index = \frac{Hit\ Rate_{valued} - Hit\ Rate_{devalued}}{Hit\ Rate_{valued} + Hit\ Rate_{devalued}}$$

A devaluation index score of 1 suggests perfect goal-directed behavior[85,95]. In contrast, a score of 0 is habitual. Since there existed a clear prediction that the post-transition devaluation index is higher than that of the pre-transition, we performed a one-tailed dependent sample t-test to compare the pre- and post-transition devaluation index[96–98], which yielded significant result, $t(6) = 2.42$, $p = 0.026$, two-sided 90%CI [0.078,0.71], Cohen's $d_z=0.92$.

Data Availability

Source data are provided with this paper. Additional data will be made available upon request to the corresponding author.

Code Availability

No specialized software was developed for this work.

References

1. Bouton ME. Learning and Behavior: A Contemporary Synthesis. 2nd ed. Sunderland, Massachusetts: Sinauer Associates is an imprint of Oxford University Press; 2016.
2. Thorndike EL. Animal intelligence: an experimental study of the associative processes in animals / by Edward L. Thorndike. New York :: Macmillan,; 1898.
3. Lingawi NW, Dezfouli A. The psychological and physiological mechanisms of habit formation. The Wiley handbook on …. 2016;
4. Ostlund SB, Balleine BW. On habits and addiction: An associative analysis of compulsive drug seeking. Drug Discov Today Dis Models. 2008;5(4):235–245.
5. de Wit S, Kindt M, Knot SL, Verhoeven AAC, Robbins TW, Gasull-Camos J, et al. Shifting the balance between goals and habits: Five failures in experimental habit induction. J Exp Psychol Gen. 2018 Jul;147(7):1043–1065.
6. Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LHE, Luijten M, et al. Carrots and sticks fail to change behavior in cocaine addiction. Science. 2016 Jun 17;352(6292):1468–1471.
7. Gillan CM, Robbins TW, Sahakian BJ, van den Heuvel OA, van Wingen G. The role of habit in compulsivity. Eur Neuropsychopharmacol. 2016 May;26(5):828–840.
8. Wood W, Rünger D. Psychology of Habit. Annu Rev Psychol. 2016;67:289–314.
9. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat Rev Neurosci. 2006 Jun;7(6):464–476.
10. Nilsen P, Roback K, Broström A, Ellström P-E. Creatures of habit: accounting for the role of habit in implementation research on clinical behaviour change. Implement Sci. 2012 Jun 9;7:53.
11. van Elzelingen W, Warnaar P, Matos J, Bastet W, Jonkman R, Smulders D, et al. Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. Curr Biol. 2022 Mar 14;32(5):1163–1174.e6.
12. Devan BD, Hong NS, McDonald RJ. Parallel associative processing in the dorsal striatum: segregation of stimulus-response and cognitive control subregions. Neurobiol Learn Mem. 2011 Sep;96(2):95–120.
13. Smith KS, Virkud A, Deisseroth K, Graybiel AM. Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. Proc Natl Acad Sci USA. 2012 Nov 13;109(46):18932–18937.
14. Hernández LF, Redgrave P, Obeso JA. Habitual behavior and dopamine cell vulnerability in Parkinson disease. Front Neuroanat. 2015 Aug 6;9:99.
15. Dickinson A. Actions and habits: the development of behavioural autonomy. Philos Trans R Soc Lond B, Biol Sci. 1985 Feb 13;308(1135):67–78.
16. Adams CD, Dickinson A. Instrumental responding following reinforcer devaluation. The Quarterly Journal of Experimental Psychology Section B. 1981 May;33(2):109–121.
17. Schreiner DC, Renteria R, Gremel CM. Fractionating the all-or-nothing definition of goal-directed and habitual decision-making. J Neurosci Res. 2020 Jun;98(6):998–1006.

18. Vandaele Y, Janak PH. Defining the place of habit in substance use disorders. Prog Neuropsychopharmacol Biol Psychiatry. 2018 Dec 20;87(Pt A):22–32.

19. Holland PC. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. J Exp Psychol Anim Behav Process. 2004 Apr;30(2):104–117.

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

21. Balleine BW, Ostlund SB. Still at the choice-point: action selection and initiation in instrumental conditioning. Ann N Y Acad Sci. 2007 May;1104:147–171.

22. Rossi MA, Yin HH. Methods for studying habitual behavior in mice. Curr Protoc Neurosci. 2012 Jul;Chapter 8:Unit 8.29.

23. Reinagel P. Training rats using water rewards without water restriction. Front Behav Neurosci. 2018 May 3;12:84.

24. Urai AE, Aguillon-Rodriguez V, Laranjeira IC, Cazettes F, International Brain Laboratory, Mainen ZF, et al. Citric Acid Water as an Alternative to Water Restriction for High-Yield Mouse Behavior. eNeuro. 2021 Feb 11;8(1).

25. Ashwood ZC, Roy NA, Stone IR, International Brain Laboratory, Urai AE, Churchland AK, et al. Mice alternate between discrete strategies during perceptual decision-making. Nat Neurosci. 2022 Feb 7;25(2):201–212.

26. Dickinson A. Instrumental Conditioning. Animal learning and cognition. Elsevier; 1994. p. 45–79.

27. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology. 1998;37(4-5):407–419.

28. Balleine BW, Dickinson A. The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. Anim Learn Behav. 1998 Mar;26(1):46–59.

29. Vandaele Y, Pribut HJ, Janak PH. Lever insertion as a salient stimulus promoting insensitivity to outcome devaluation. Front Integr Neurosci. 2017 Sep 27;11:23.

30. Mosberger AC, de Clauser L, Kasper H, Schwab ME. Motivational state, reward value, and Pavlovian cues differentially affect skilled forelimb grasping in rats. Learn Mem. 2016 May 18;23(6):289–302.

31. Lipton DM, Gonzales BJ, Citri A. Dorsal striatal circuits for habits, compulsions and addictions. Front Syst Neurosci. 2019 Jul 18;13:28.

32. Mendelsohn AI. Creatures of habit: the neuroscience of habit and purposeful behavior. Biol Psychiatry. 2019 Jun 1;85(11):e49–e51.

33. Amaya KA, Smith KS. Neurobiology of habit formation. Curr Opin Behav Sci. 2018 Apr;20:145–152.

34. Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci. 2004 Jan;19(1):181–189.

35. Aarts H, Dijksterhuis A. Habits as knowledge structures: automaticity in goal-directed behavior. J Pers Soc Psychol. 2000 Jan;78(1):53–63.

36. Lee CR, Margolis DJ. Pupil Dynamics Reflect Behavioral Choice and Learning in a Go/NoGo Tactile Decision-Making Task in Mice. Front Behav Neurosci. 2016 Nov 1;10:200.

37. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron. 2003 Apr 24;38(2):329–337.

38. Bijleveld E, Custers R, Aarts H. The unconscious eye opener: pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. Psychol Sci. 2009 Nov;20(11):1313–1315.

39. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. Psychon Bull Rev. 2018 Dec;25(6):2005–2015.

40. Yamada K, Toda K. Pupillary dynamics of mice performing a Pavlovian delay conditioning task reflect reward-predictive signals. Front Syst Neurosci. 2022 Dec 8;16:1045764.

41. Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. Neuron. 2014 Oct 22;84(2):355–362.

42. Fröber K, Pittino F, Dreisbach G. How sequential changes in reward expectation modulate cognitive control: Pupillometry as a tool to monitor dynamic changes in reward expectation. Int J Psychophysiol. 2020 Feb;148:35–49.

43. Ganea DA, Bexter A, Günther M, Gardères P-M, Kampa BM, Haiss F. Pupillary dilations of mice performing a vibrotactile discrimination task reflect task engagement and response confidence. Front Behav Neurosci. 2020 Sep 3;14:159.

44. Phillips CD, Hodge AT, Myers CC, Leventhal DK, Burgess CR. Striatal dopamine contributions to skilled motor learning. J Neurosci. 2024 Jun 26;44(26).

45. Bernklau TW, Righetti B, Mehrke LS, Jacob SN. Striatal dopamine signals reflect perceived cue-action-outcome associations in mice. Nat Neurosci. 2024 Apr;27(4):747–757.

46. Zareian B, Lam A, Zagha E. Dorsolateral Striatum is a Bottleneck for Responding to Task-Relevant Stimuli in a Learned Whisker Detection Task in Mice. J Neurosci. 2023 Mar 22;43(12):2126–2139.

47. de Wit S, Dickinson A. Associative theories of goal-directed behaviour: a case for animal-human translational models. Psychol Res. 2009 Jul;73(4):463–476.

48. Smith KS, Graybiel AM. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. Neuron. 2013 Jul 24;79(2):361–374.

49. Adams CD. Variations in the sensitivity of instrumental responding to reinforcer devaluation. The Quarterly Journal of Experimental Psychology Section B. 1982 May;34(2):77–98.

50. Coutureau E, Killcross S. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav Brain Res. 2003 Nov 30;146(1-2):167–174.

51. O'Doherty JP, Cockburn J, Pauli WM. Learning, reward, and decision making. Annu Rev Psychol. 2017 Jan 3;68:73–100.

52. Miller KJ, Shenhav A, Ludvig EA. Habits without values. Psychol Rev. 2019 Mar;126(2):292–311.

53. Watson P, de Wit S. Current limits of experimental research into habits and future directions. Curr Opin Behav Sci. 2018 Apr;20:33–39.

54. Garrett N, Allan S, Daw ND. Model based control can give rise to devaluation insensitive choice. BioRxiv. 2022 Aug 22;

55. Garr E, Delamater AR. Exploring the relationship between actions, habits, and automaticity in an action sequence task. Learn Mem. 2019 Apr;26(4):128–132.

56. Thrailkill EA, Bouton ME. Contextual control of instrumental actions and habits. J Exp Psychol Anim Learn Cogn. 2015 Jan;41(1):69–80.

57. Dickinson A, Nicholas DJ, Adams CD. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. The Quarterly Journal of Experimental Psychology Section B. 1983 Feb;35(1b):35–51.

58. Krechevsky I. Hypotheses" in rats. Psychol Rev. 1932;39(6):516–532.

59. Lashley KS. The Problem of Serial Order in Behavior. Jeffress, L A (Ed), Cerebral Mechanisms in Behavior: The Hixon Symposium. 1951;112–136.

60. Zhu Z, Kuchibhotla KV. Performance errors during rodent learning reflect a dynamic choice strategy. Curr Biol. 2024 May 20;34(10):2107–2117.e5.

61. Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. Proc Natl Acad Sci USA. 2004 Sep 7;101(36):13124–13131.

62. Vandaele Y, Mahajan NR, Ottenheimer DJ, Richard JM, Mysore SP, Janak PH. Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. Elife. 2019 Oct 17;8.

63. Smith KS, Graybiel AM. Using optogenetics to study habits. Brain Res. 2013 May 20;1511:102–114.

64. Smith KS, Graybiel AM. Habit formation coincides with shifts in reinforcement representations in the sensorimotor striatum. J Neurophysiol. 2016 Mar;115(3):1487–1498.

65. Wood W, Neal DT. A new look at habits and the habit-goal interface. Psychol Rev. 2007 Oct;114(4):843–863.

66. Shillinglaw JE, Everitt IK, Robinson DL. Assessing behavioral control across reinforcer solutions on a fixed-ratio schedule of reinforcement in rats. Alcohol. 2014 Jun;48(4):337–344.

67. Ito M, Doya K. Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. J Neurosci. 2015 Feb 25;35(8):3499–3514.

68. Her ES, Huh N, Kim J, Jung MW. Neuronal activity in dorsomedial and dorsolateral striatum under the requirement for temporal credit assignment. Sci Rep. 2016 Jun 1;6:27056.

69. White SR, Preston MW, Swanson K, Laubach M. Learning to Choose: Behavioral Dynamics Underlying the Initial Acquisition of Decision-Making. eNeuro. 2024 May 17;11(5).

70. Vandaele Y, Lenoir M, Vouillac-Mendoza C, Guillem K, Ahmed SH. Probing the decision-making mechanisms underlying choice between drug and nondrug rewards in rats. Elife. 2021 Apr 26;10.

71. Legaria AA, Matikainen-Ankney BA, Yang B, Ahanonu B, Licholai JA, Parker JG, et al. Fiber photometry in striatum reflects primarily nonsomatic changes in calcium. Nat Neurosci. 2022 Sep;25(9):1124–1128.

72. Cruz KG, Leow YN, Le NM, Adam E, Huda R, Sur M. Cortical-subcortical interactions in goal-directed behavior. Physiol Rev. 2023 Jan 1;103(1):347–389.

73. Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology. 2010 Jan;35(1):48–69.

74. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005 Dec;8(12):1704–1711.

75. Bogacz R. Dopamine role in learning and action inference. Elife. 2020 Jul 7;9.

76. Young MK, Conn K-A, Das J, Zou S, Alexander S, Burne THJ, et al. Activity in the dorsomedial striatum underlies serial reversal learning performance under probabilistic uncertainty. Biological Psychiatry Global Open Science. 2022 Aug;

77. Hosking JG, Cocker PJ, Winstanley CA. Prefrontal Cortical Inactivations Decrease Willingness to Expend Cognitive Effort on a Rodent Cost/Benefit Decision-Making Task. Cereb Cortex. 2016 Apr;26(4):1529–1538.

78. McGuire JT, Botvinick MM. Prefrontal cortex, cognitive control, and the registration of decision costs. Proc Natl Acad Sci USA. 2010 Apr 27;107(17):7922–7926.

79. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001;24:167–202.

80. Bouton ME. Why behavior change is difficult to sustain. Prev Med. 2014 Nov;68:29–36.

81. Kober H, Mende-Siedlecki P, Kross EF, Weber J, Mischel W, Hart CL, et al. Prefrontal-striatal pathway underlies cognitive regulation of craving. Proc Natl Acad Sci USA. 2010 Aug 17;107(33):14811–14816.

82. Juczewski K, Koussa JA, Kesner AJ, Lee JO, Lovinger DM. Stress and behavioral correlates in the head-fixed method: stress measurements, habituation dynamics, locomotion, and motor-skill learning in mice. Sci Rep. 2020 Jul 22;10(1):12245.

83. Kuchibhotla KV, Hindmarsh Sten T, Papadoyannis ES, Elnozahy S, Fogelson KA, Kumar R, et al. Dissociating task acquisition from expression during learning reveals latent knowledge. Nat Commun. 2019 May 14;10(1):2151.

84. Moore S, Kuchibhotla KV. Slow or sudden: Re-interpreting the learning curve for modern systems neuroscience. IBRO Neuroscience Reports. 2022 Dec;13:9–14.

85. Renteria R, Baltz ET, Gremel CM. Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. Nat Commun. 2018 Jan 15;9(1):211.

86. Yin HH, Knowlton BJ, Balleine BW. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci. 2005 Jul;22(2):505–512.

87. Sood A, Richard JM. Sex-biased effects of outcome devaluation by sensory-specific satiety on Pavlovian-conditioned behavior. Front Behav Neurosci. 2023 Oct 4;17:1259003.

88. Bachmanov AA, Tordoff MG, Beauchamp GK. Sweetener preference of C57BL/6ByJ and 129P3/J mice. Chem Senses. 2001 Sep;26(7):905–913.

89. Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, et al. Bonsai: an event-based framework for processing and controlling data streams. Front Neuroinformatics. 2015 Apr 8;9:7.

90. Wu J, Cai Y, Wu X, Ying Y, Tai Y, He M. Transcardiac perfusion of the mouse for brain tissue dissection and fixation. Bio Protoc. 2021 Mar 5;11(5):e3988.

91. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nat Protoc. 2019 Jul;14(7):2152–2176.

92. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci. 2018 Sep;21(9):1281–1289.

93. Towner TT, Spear LP. Rats exposed to intermittent ethanol during late adolescence exhibit enhanced habitual behavior following reward devaluation. Alcohol. 2021 Mar;91:11–20.

94. Stapf CA, Keefer SE, McInerney JM, Calu DJ. Sex specific effects of dorsomedial striatal cannabinoid receptor-1 signaling on Pavlovian outcome devaluation. BioRxiv. 2024 May 1;

95. Hadjas LC, Lüscher C, Simmler LD. Aberrant habit formation in the Sapap3-knockout mouse model of obsessive-compulsive disorder. Sci Rep. 2019 Aug 19;9(1):12061.

96. Ruxton GD, Neuhäuser M. When should we use one-tailed hypothesis testing? Methods Ecol Evol. 2010 Mar 1;1(2):114–117.

97. Parab S, Bhalerao S. Choosing statistical test. Int J Ayurveda Res. 2010 Jul;1(3):187–191.

98. Du Prel J-B, Röhrig B, Hommel G, Blettner M. Choosing Statistical Tests: Part 12 of a Series on Evaluation of Scientific Publications. Dtsch Arztebl Int. 2010 May 14;107(19):343–348.

99. Paxinos G, Franklin KBJ. The mouse brain in stereotaxic coordinates. 5th ed. London: Academic Press; 2019.

Acknowledgements

We thank PH. Janak, A. Haith, J. Krakauer, N. Kothari, P. Holland, Y. Cheng, and R. Magnard for helpful comments and discussions on the manuscript. We thank S. Pavuluri for technical support. We thank E. Barker and D. Udzinski for animal care-taking. This work was supported by grants from the NIH R01DC018650 and R00DC015014 to KVK, and by Kavli Neuroscience Discovery Institute at Johns Hopkins University (Baltimore, 21218) fellowships to SM and ZZ.

Author Contribution Statement

SM and KVK designed the project. ZW, SM, YL, and JW performed the experiments. ZZ, RS, and AC performed computational modeling. SM and ZW performed final analysis, figures, and data curation. SM, ZW, and KVK wrote the manuscript. KVK provided funding and supervised the project. All authors participated in results interpretation and manuscript editing.

Competing Interests Statement

The authors declare no competing interests.

Supplementary and figures

Supplementary figures 1 to 6, figure legends, and tables are provided.

Figure Legends/Captions (for main text figures)

Figure 1. Self-restriction reduces reward-seeking for plain water without impacting discrimination learning. a, Protocol outline: after head-post implantation, mice underwent a habituation period and introduction to the water restriction paradigms. 'Control' mice underwent a common (maintenance at 80-85% original body weight) water restriction paradigm (WR85, blue). CA mice had *ad libitum* access to a water bottle with a low percentage of a less palatable hydration source (citric acid laced water) in their home cage (CA, yellow) to which they were progressively introduced (starting with 0.5%, reaching maximum 3%). After a few days, both cohorts underwent two days of lick training (lavender), followed by an auditory Go/No-Go training (black). **b,** CA mice lost significantly less weight compared to WR85. Median diff.= 10.06, n=11/12 mice, two-sided Wilcoxon rank-sum test, z=4.03, p=0.000055, HL=8.78, 95% CI[7.37, 11.05]. **c,** CA mice obtain significantly fewer rewards compared to WR85 in one lick training session. Median diff.= 75.00, n=11/12 mice, two-sided Wilcoxon rank-sum test, z=3.36, p=0.000791, HL=77.50, 95%CI [43.00, 110.00]. **d, (i)** After lick-training, mice underwent auditory cued go/no-go training that consisted of ~300 trials per session. **(ii)** Mice learn to lick after a S+ tone to obtain a plain water reward (3µl) and withhold licking to an S- tone to avoid a time-out. **(iii)** Correct responses are hits (licking to the S+ tone) and correct rejects (withhold licking to the S-), while incorrect responses are false alarms (licking to the S-) and misses (not licking to the S+). **e,** Accuracy comparison between WR85 (blue, n=11) and CA (yellow, n=12) mice on highly engaged trial blocks is similar. Two-way ANOVA revealed a significant effect of Group ($F(1, 319) = 5.45$, $p = 0.020$, $\eta p^2=0.017$), and a significant Trial_block main effect ($F(25, 319)=7.93$, $p=1.47x10^{-21}$, $\eta p^2=0.38$), and no significant Group × Trial_block interaction effect ($F(25, 319)=0.80$, $p=0.741$, $\eta p^2=0.059$). Bonferroni-corrected post hoc pairwise t-test identified: Group=2,Trial_block=2 vs Group=1,Trial_block=9 (mean diff. = -0.33, 95%CI [-0.64,-0.010], p_adj=0.025); Group=2,Trial_block=2 vs Group=1,Trial_block=10 (mean diff.=-0.32, 95%CI [-0.63,-0.020], p_adj=0.015); Group=2,Trial_block=2 vs Group=1,Trial_block=11 (mean diff.=-0.33, 95%CI [-0.65,-0.00], p_adj = 0.047), plus 109 additional significant contrasts; full statistics are provided in Supplementary Table (**'SuppTable PostHoc Bonferroni Accuracy'**). Expert accuracy is defined as 75% correct (gray horizontal line). **f,** No differences were observed in the number of trials to reach expert accuracy between groups. Median diff.= 100.00, n=11/12, two-sided Wilcoxon rank-sum test: z=-0.090, p=0.93, HL=-50.00, 95%CI [-1004.00, 304.00]. Except **a** and **b**, all data are presented as mean values +/- SEM.

Figure 2. Abrupt state-like transitions from goal-directed to habitual behavior appear spontaneously within individual animals. a, Hit (color) and FA (black) action rates for example WR85 (left) and a CA (right) animals. **b,** Hit rate of CA example animal (in A) showing periods of high (gray shadow) and low engagement, followed by a spontaneous transition (red vertical line) to low hit rate variability. **c,** CA mice have significantly less engaged blocks compared to WR85. Median diff.= 13.92, n=11/12 mice, two-sided Wilcoxon rank-sum test: z=2.98, p=0.0028, HL=19.27, 95%CI [8.90,37.32]. **d, (i)** Most CA (8/12, 66.7%) showed hit-rate variability and the presence of a transition. Some CA mice that transitioned to low variability hit rate, seemed to transition to high variability after a while (asterisks). **(ii)** Only a few CA mice showed no hit rate variability (2/12, 16.7%). **(iii)** Few CA mice showed initial hit rate variability, but never transitioned to low variability (2/12, 16.7%). **e,** An HMM-GLM was used to model behavioral data, which allows us to analyze the state-like nature of transitions. **f,** An HMM-GLM with two states provides the best fit for most of the CA animals based on a BIC analysis (n=8 CA mice). **g,** Both states are equally stimulus-driven, but state 2 is characterized by a No-go, or disengaged bias (n=8 CA mice). **h,** State 1 (pink) is highly stimulus selective between target and foil trials with high engagement, while State 2 (green) shows overall task disengagement (n=8 CA mice). **i,** A CA exemplar shows that the two-state HMM-GLM model accurately recapitulates the behavior (top), and the two states govern the pre-transition phase, while only one state becomes explanatory of the post-transition phase. **j, (i)** The GLM states reflect transitions between an engaged state (state 1, pink, hit rate = 0.98) and a disengaged state (state2, olive, hit rate = 0.35). Across all mice, the HMM-GLM model predicted that the probability of staying in engaged or disengaged state (trial-by-trial) is 0.995 and 0.975 respectively, whereas the transition probability between states is 0.005 (engaged to disengaged) and 0.025 (disengaged to engaged). **(ii)** State occupancy pre-transition (black) is approximately divided 50%-50% between State 1 and State 2, while post-transition (red), State 1 dominates (n=8 CA mice). Two-way ANOVA revealed a significant State main effect ($F(1, 28)=62.20$, $p=1.37x10^{-8}$, $\eta p^2=0.69$), and no significant Transition main effect ($F(1, 28)=0.020$, $p=0.90$, $\eta p^2=0.0010$), and a significant State × Transition interaction effect ($F(1, 28)=0.99$, $p=8.7x10^{-5}$, $\eta p^2=0.43$). Bonferroni-corrected post hoc pairwise comparisons across State × Transition cell means identified: State=State1,Transition=Pre vs State=State1,Transition=Post (mean diff.=-0.30, 95%CI [-0.56,-0.040], p_adj=0.015); State=State1,Transition=Pre vs State=State2, Transition=Post (mean diff.=0.50, 95%CI [0.24,0.76],

p_adj=4.38x10^{-5}); State=State2, Transition=Pre vs State=State1, Transition=Post (mean diff. = -0.51, 95% CI [-0.77, -0.26], p_adj = 2.72x10^{-5}), plus 2 additional significant contrasts; full statistics are provided in Supplementary Table (**'SuppTable PostHoc Bonferroni Occupancy'**). **k,** No-go to Go (State 2 → State 1) transitions and Go to No-go (State 1 → State 2) transitions happening in the goal-directed phase, are abrupt (n=8 CA mice). **l,** No differences between the first (blue, belonging to the goal-directed phase), and last (red, belonging to goal-directed to habitual) transitions in abruptness. Both happen within 1-4 trials (n=8 CA mice). Two-sided Wilcoxon rank-sum test: z=NaN, U=30, p=0.89, mean diff.=0.054, median diff.=0.50, 95%CI [-2.00, 2.00], HL=0.00, r_{rb}=0.071. In **g**, **h**, **jii**, **k**, and **l**, all data are presented as mean values +/- SEM.

Figure 3. Sensory-specific devaluation and bilateral DLS lesions confirm a transition to habit. a, Sensory-specific devaluation protocol. **b,** Normalized hit rate across condition and transition. 2-way repeated-measure ANOVA revealed a significant condition x transition interaction effect (X^2(1)=5.61, p =0.018, partial R^2=0.39). Two-sided post hoc pairwise t-test of estimated marginal means with Tukey HSD correction and Kenward-Roger degree of freedom method: transition=Pre, condition=Valued vs transition=Pre, condition=Devalued (mean diff.=0.25, 95%CI [0.032,0.47], p=0.027); condition=Devalued, transition=Pre vs condition=Devalued, transition=Post (mean diff.=-0.47, 95%CI [-0.69,-0.25], p=0.00020). n=7 CA mice. For extended statistical analysis see Methods. **c,** Devaluation index, pre-transition vs. post-transition. Mean diff.=0.39, n=7 mice. one-tailed paired-samples t-test: t(6)=2.42, p=0.026, two-sided 90%CI [0.078,0.71], Cohen's d_z=0.92. Pre-transition: median=0.37, Q1=-0.018, Q3= 0.63, lower bound=-0.99, upper bound=1.60, lower whisker (min)=-0.29, upper whisker (max)=0.77. Post-transition: median=0.027, Q1=-0.30, Q3=0.084, lower bound= -0.87, upper bound= 0.66, lower whisker (min)=-0.41, upper whisker (max)=0.095. H=habitual, GD=goal-directed. **d,** Lesion protocol. **e,** Lesion exemplar. Coronal view of the bilateral lesion sites at the DLS. Structural illustration by the Paxinos and Franklin's Mouse Brain Atlas [99]. **f,** Lesion map of all animals (n=11 mice). **g,** Action rate exemplars for sham (top, gray) and lesioned (bottom, purple) animals. **h,** Cumulative distribution of transition trial for animals that showed behavioral variability (n=10 sham and n=9 lesioned mice). **i,** ROC curve built from transition probabilities in g. **j,** AUC values for shuffled labels in our experimental groups. The difference in AUC between sham and lesioned mice (purple line) falls into the 95th significance percentile (red dotted line), compared to the difference between control animals (gray line). In **b** and **c**, data are presented as mean values +/- SEM.

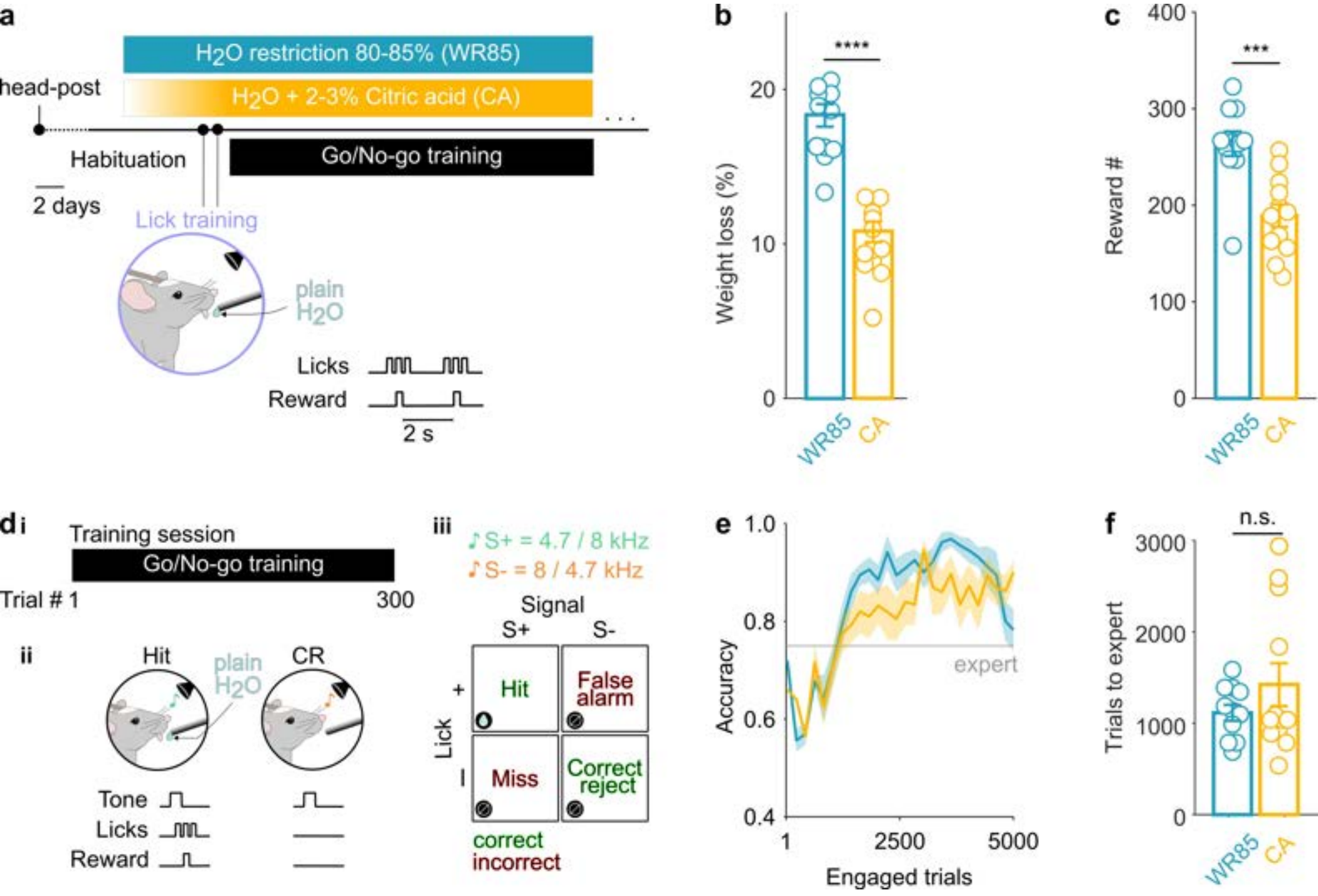
Figure 4. Motor automaticity signatures and pupillary response changes appear concomitant with habit transitions. a, Exemplar lick raster plots of a CA animal, showing individual licks (green lines) to target tones throughout training. Tone onset (0, black note) is followed by a dead period (100ms) and the presence of operant licks (gray rectangle). **b,** CA mice show a strong increase in post-transition number of consummatory licks (bottom). **c,** Average number of licks per trial is significantly higher in CA mice post-transition compared to pre-transition (Mean diff.= -2.60, n=8/8 mice, paired t-test: (t(7.00)=-4.87, p=0.0018), Mean diff.=-2.60, 95%CI[-3.86,-1.34], Cohens dz=-1.72). **d,** A significant reduction in the operant lick latency is observed post-transition (Median diff.= 75.00, n=8/8 mice, two sided paired Wilcoxon signed-rank test (W=28, p=0.016), HL=50.00 (95% CI[50.00, 150.00]). **e,** Evoked pupil dilation separated by trial type (engaged only): black=pre-transition rewarded; red=post-transition rewarded; grey-solid=pre-transition no-reward; pink=post-transition no-reward. **f,** Significant increase in evoked pupil response in no-reward trials. Two-sided post hoc pairwise t-test with LSD test: Trial=Non-consumed,Transition=Pre vs Trial=Non-consumed,Transition=Post (mean diff.= -0.20, 95%CI [-0.65,-1.23], p=0.019) for 3 days pre- and 3 days post-transition. The difference in pupil response between rewarded and no-reward trials decreased across transition, revealed by permutation-based 2-way ANOVA, significant transition x trial interaction effect (F(2, 204)=3.95, p= 0.017, ηp^2=0.037). Data compiled from n=7 CA mice, n=35 trials per group. For extended statistical analysis see **Supplementary Table 1**. **g,** Significant increase in evoked pupil response for 10 engaged, no-reward trials pre- and post-transition. Median diff.= -1.64, n=45/35 trials, two-sided Wilcoxon rank-sum test: z=-2.78, p=0.0054, HL=-1.36, 95%CI [-3.08, -0.34], data compiled from n=7 mice. Except **a**, all data are presented as mean values +/- SEM.

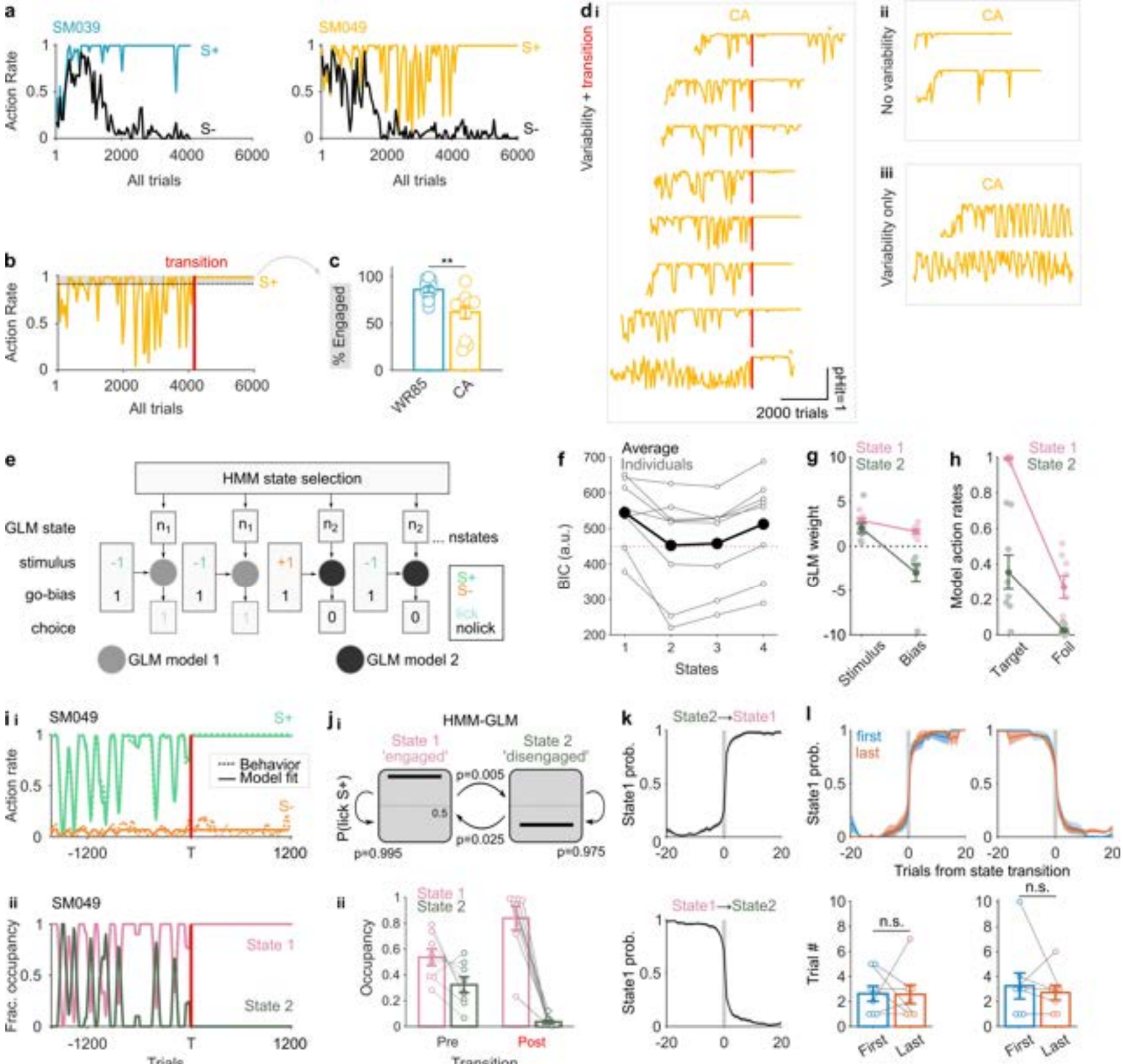
Figure 5. Switch-like reduction in DLS outcome-related signaling at the moment of a habit transition. a, Bilateral simultaneous DLS and DMS fiber photometry recordings in behaving mice across training. Left: Mice were injected with hSyn-GCaMP8m in both structures. Right: Green rectangles show successful fiber placement and viral expression. **b,** Analysis window (±50 trials around the transition). **c,** DLS (top, 50 trials per group from n=7 mice, total n=350 datapoints) and DMS (bottom. 50 trials per group from n=6 mice, total n=300 datapoints) activity pre (black) and post-transition (red). Data are shown as mean ± SEM. **d,** Area under-the-curve (AUC) of the full trace in **c,** for the DLS (top) (Median diff.=70.49, n=350/350 trials from n=7 mice, two-sided paired Wilcoxon signed-

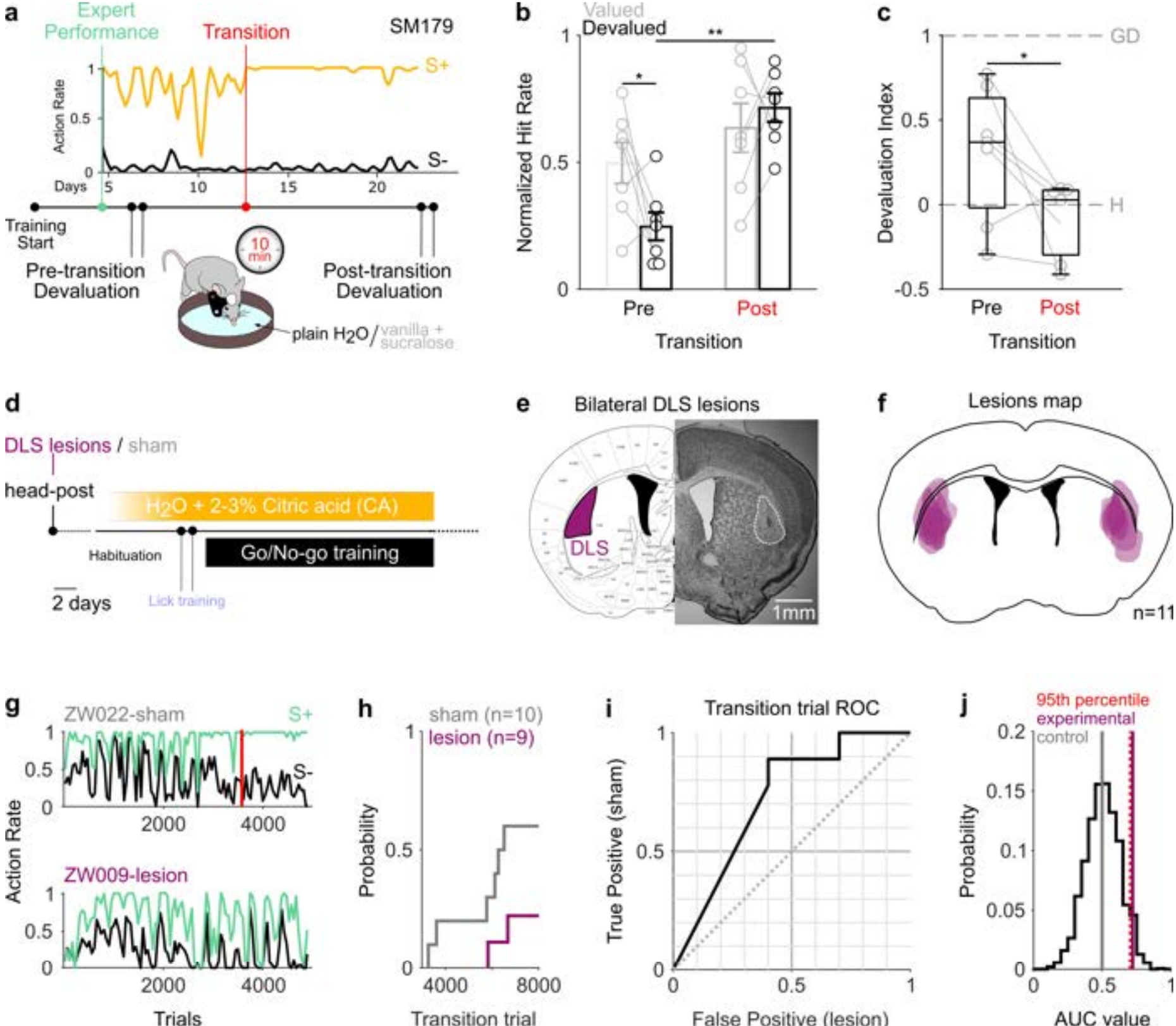
rank test (z=7.53, p=5.04x10^{-14}), HL=63.66 (95% CI[49.01, 77.52])) and DMS (bottom) overall trace (Median diff.= 33.64, n=300/300 trials from n=6mice, two-sided paired Wilcoxon signed-rank test (z=3.37, p=0.00074), HL=31.65 (95% CI[6.58, 54.32])). Data are shown as mean ± SEM. **e,** First peak latency post-transition is significantly different for the DLS (top) (two-sample F-test, (F(349,349)=1.87, p=7.41x10^{-9})) but not the DMS (bottom) (two-sample F-test, F(294,291) = 0.94, p = 0.59). n=350 trials DLS pre; n=330 DLS trials post from n=7 mice; n=295 trials DMS pre; n=292 trials DMS post from n=6 mice. **f,** Jitter quantification of first-peak latencies in DLS (top) and DMS (bottom), showing a significant reduction in DLS jitter post-transition (Levene's test F(1,698)=34.21, p=7.62x10^{-9}), but not in the DMS (Levene's test F(1,585) = 2.94, p = 0.087). n=350 trials DLS pre; n=350 DLS trials post from n=7 mice; n=295 trials DMS pre; n=292 trials DMS post from n=6 mice. Data are shown as mean ± SEM. **g,** Single animal exemplar of 5 trials immediately pre- (black) and post- (red) transition. Individual lines represent individual trials. **h,** Logistic regression fit. Inflection point (in trials): full trace (top) DLS=4.8, DMS=9.4; 'Cue' epoch (middle) DLS none, DMS=9.2; 'Outcome' epoch DLS=3.3, DMS=9.4. Red vertical line represents the transition point. In **d** and **f**, data are presented as mean values +/- SEM.

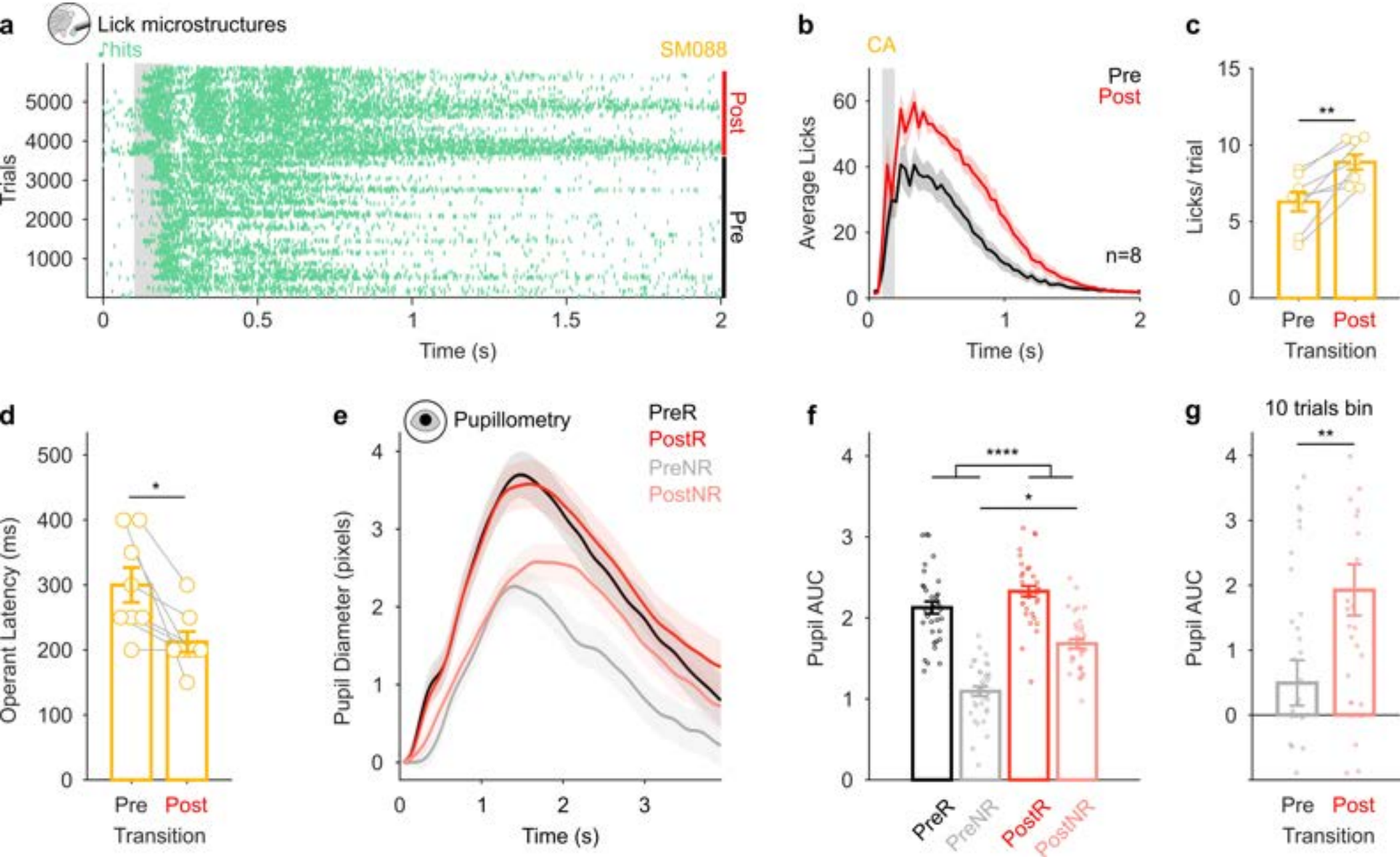
Editorial summary: Whether habits emerge gradually or suddenly remains an open question. Here the authors show that mice abruptly switch from goal-directed to habitual control, marked by a rapid dorsal striatal shift from outcome- to stimulus-driven processing.

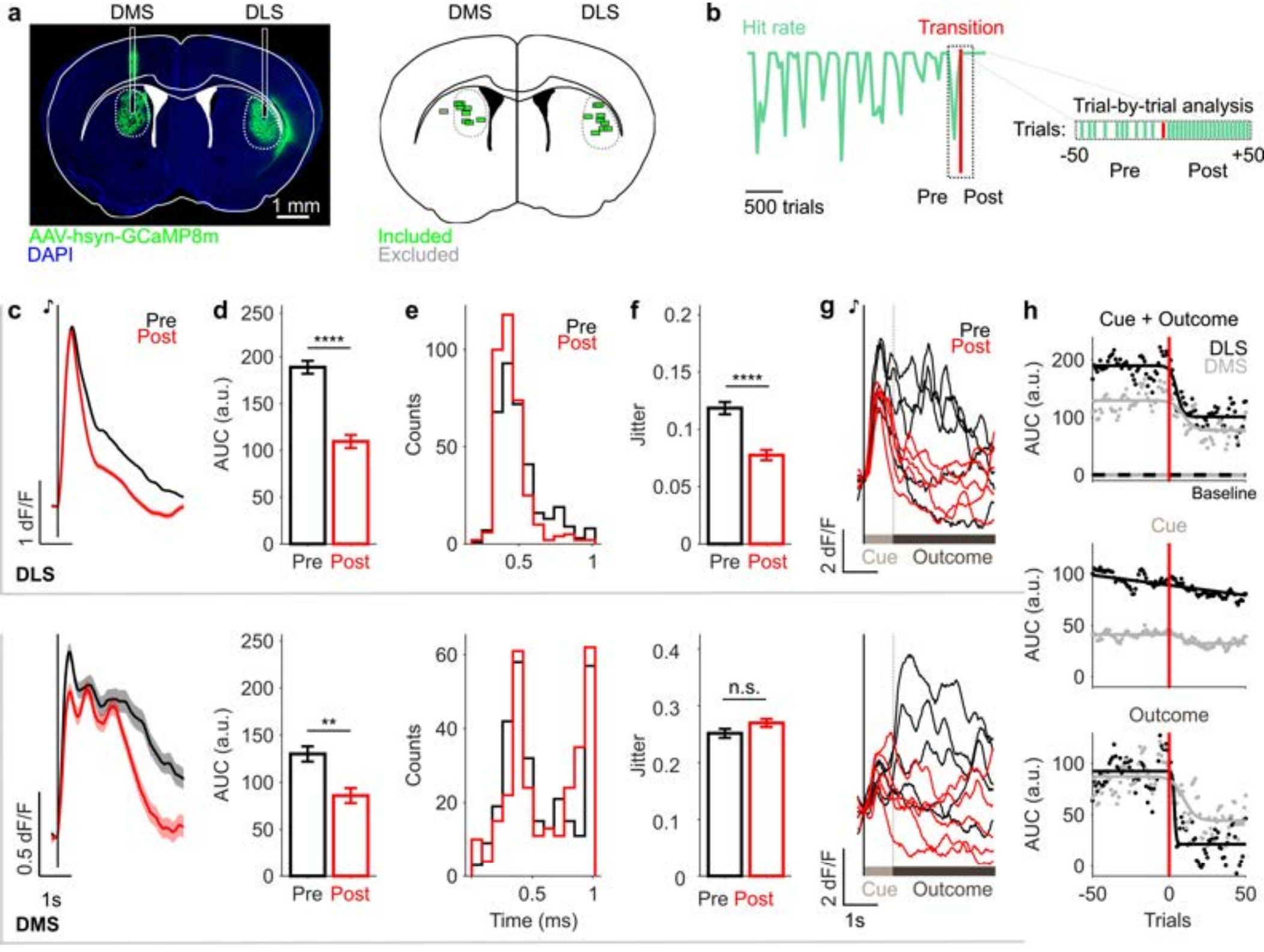
Peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.


a
H2O restriction 80-85% (WR85)
H2O + 2-3% Citric acid (CA)
head-post
Habituation
Go/No-go training
2 days
Lick training
plain H2O
Licks
Reward
2 s
b
Weight loss (%)
****
WR85
CA
c
Reward #
***
WR85
CA
d i
Training session
Go/No-go training
Trial # 1
300
ii
Hit
plain H2O
CR
Tone
Licks
Reward
iii
♪ S+ = 4.7 / 8 kHz
♪ S- = 8 / 4.7 kHz
Signal
S+
S-
Lick
Hit
False alarm
Miss
Correct reject
correct
incorrect
e
Accuracy
expert
Engaged trials
f
n.s.
Trials to expert
WR85
CA


a
SM039
SM049
S+
S-
Action Rate
All trials
b
transition
c
% Engaged
WR85
CA
d i
Variability + transition
2000 trials
ii
No variability
iii
Variability only
e
HMM state selection
GLM state
stimulus
go-bias
choice
nstates
GLM model 1
GLM model 2
f
Average
Individuals
BIC (a.u.)
States
g
GLM weight
Stimulus
Bias
h
Model action rates
Target
Foil
i i
Behavior
Model fit
Frac. occupancy
Trials
j i
HMM-GLM
State 1 'engaged'
State 2 'disengaged'
P(lick S+)
p=0.005
p=0.025
p=0.995
p=0.975
ii
Occupancy
Pre
Post
Transition
k
State2→State1
State1→State2
State1 prob.
l
first
last
Trials from state transition
Trial #
n.s.


a
Expert Performance
Transition
SM179
S+
S-
Action Rate
Days
Training Start
Pre-transition Devaluation
Post-transition Devaluation
10 min
plain H2O / vanilla + sucralose
b
Valued
Devalued
Normalized Hit Rate
Pre
Post
Transition
c
GD
H
Devaluation Index
Pre
Post
Transition
d
DLS lesions / sham
head-post
H2O + 2-3% Citric acid (CA)
Habituation
Go/No-go training
2 days
Lick training
e
Bilateral DLS lesions
DLS
1mm
f
Lesions map
n=11
g
ZW022-sham
S+
S-
ZW009-lesion
Action Rate
Trials
h
sham (n=10)
lesion (n=9)
Probability
Transition trial
i
Transition trial ROC
True Positive (sham)
False Positive (lesion)
j
95th percentile
experimental
control
Probability
AUC value


a
Lick microstructures
hits
SM088
Post
Pre
Trials
5000
4000
3000
2000
1000
0
0.5
1
1.5
2
Time (s)
b
CA
Pre
Post
Average Licks
60
40
20
0
n=8
0
1
2
Time (s)
c
Licks/ trial
15
10
5
0
**
Pre
Post
Transition
d
Operant Latency (ms)
500
400
300
200
100
0
*
Pre
Post
Transition
e
Pupillometry
PreR
PostR
PreNR
PostNR
Pupil Diameter (pixels)
4
3
2
1
0
0
1
2
3
Time (s)
f
Pupil AUC
4
3
2
1
0
****
*
PreR
PreNR
PostR
PostNR
g
10 trials bin
Pupil AUC
4
3
2
1
0
**
Pre
Post
Transition


a
DMS
DLS
1 mm
AAV-hsyn-GCaMP8m
DAPI
DMS
DLS
Included
Excluded
b
Hit rate
Transition
500 trials
Pre
Post
Trial-by-trial analysis
Trials:
-50
Pre
Post
+50
c
Pre
Post
1 dF/F
DLS
d
AUC (a.u.)
****
Pre
Post
e
Counts
Pre
Post
Time (ms)
f
Jitter
****
Pre
Post
g
Pre
Post
2 dF/F
Cue
Outcome
h
Cue + Outcome
DLS
DMS
AUC (a.u.)
Baseline
Cue
Outcome
Trials
0.5 dF/F
1s
**
n.s.
2 dF/F
1s