

Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex

Kyle S. Smith^{a,1}, Arti Virkud^a, Karl Deisseroth^b, and Ann M. Graybiel^{a,1}

^aMcGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^bDepartment of Bioengineering, Stanford University School of Medicine, Stanford, CA 94305

Contributed by Ann M. Graybiel, September 18, 2012 (sent for review August 20, 2012)

Habits tend to form slowly but, once formed, can have great stability. We probed these temporal characteristics of habitual behaviors by intervening optogenetically in forebrain habit circuits as rats performed well-ingrained habitual runs in a T-maze. We trained rats to perform a maze habit, confirmed the habitual behavior by devaluation tests, and then, during the maze runs (ca. 3 s), we disrupted population activity in a small region in the medial prefrontal cortex, the infralimbic cortex. In accordance with evidence that this region is necessary for the expression of habits, we found that this cortical disruption blocked habitual behavior. Notably, however, this blockade of habitual performance occurred on line, within an average of three trials (ca. 9 s of inhibition), and as soon as during the first trial (<3 s). During subsequent weeks of training, the rats acquired a new behavioral pattern. When we again imposed the same cortical perturbation, the rats regained the suppressed maze-running that typified the original habit, and, simultaneously, the more recently acquired habit was blocked. These online changes occurred within an average of two trials (ca. 6 s of infralimbic inhibition). Measured changes in generalized performance ability and motivation to consume reward were unaffected. This immediate toggling between breaking old habits and returning to them demonstrates that even semiautomatic behaviors are under cortical control and that this control occurs online, second by second. These temporal characteristics define a framework for uncovering cellular transitions between fixed and flexible behaviors, and corresponding disturbances in pathologies.

learning | limbic | time | reinforcement | neuroplasticity

Habits are among the most stable and powerful behaviors that we have. Forming such strongly ingrained behaviors requires that a durable representation of the movement repertoire be acquired. Much evidence suggests that this process involves a gradual transition from flexible and goal-directed behavior to a more fixed, habitual behavioral strategy (1–7). How these properties of habits map onto neural circuitry has been the focus of classic lesion and chemical inactivation studies, which have identified regions of the striatum as essential for the expression of habits (2–4). In addition, such studies have established that a region in the medial prefrontal cortex also must be intact for habits to be expressed (7–9). This medial prefrontal region [called infralimbic (IL) cortex in rodents] is linked to emotion-related limbic circuitry and projects to sites that promote behavioral flexibility at the expense of habits (e.g., prelimbic cortex and medial striatum) (3, 4, 7, 10). Based on this anatomy, the IL cortex is thought to be at an executive level in the control of habits and behavioral strategies (1, 8, 9, 11).

This sketch of the circuitry for habits and skilled habit-like behaviors opens key questions about how habitual behaviors are controlled on a moment-by-moment basis. The slow emergence of habits favors a gradual biasing of these behaviors toward automaticity, but the stability of habits favors their being performed without moment-to-moment biasing or executive cortical control. We tested these dynamics of habit control by the prefrontal IL region by targeting precisely timed optogenetic inhibition to the IL cortex with light to drive virally introduced halorhodopsin (eNpHR3.0) (12). This tool allowed us to examine not only the

online contribution of cortical activity to the control of habitual behavior, but also, because of its unique repeatability, to track the effects of such brief, seconds-long manipulations over weeks of subsequent performance. To our surprise, we found that despite the automaticity of habitual behavior, it is subject to online control by the prefrontal cortex.

Results

Overtrained T-Maze Running Is Habitual. We used a navigational T-maze task (Fig. 1A) similar to one developed to examine neural activity in the striatum during habit formation (13–15). We tracked the learning curves of multiple sets of rat subjects ($n = 10$). In daily sessions of ca. 40 trials, the rats were required to initiate maze runs in response to a warning cue, run down the maze, and turn right or left, depending on an auditory instruction cue, to receive one of two rewards (chocolate milk or sucrose solution). For each rat, each reward type was assigned to one end arm, and entry into an incorrect arm resulted in no reward. Rats were trained to a criterion of statistically significant performance accuracy (72.5% correct; criterion training stage 5; Fig. 1B) and then were overtrained for 10 or more additional sessions. Peak performance accuracy was high at ~90% correct (Fig. 1B).

We applied a reward-devaluation protocol as a quantitative test for habit formation after this extensive training (16). We devalued one of the two maze rewards in home-cage sessions and then tested in the maze experiments for habitual running to the end arm baited with the now-devalued reward, relative to runs to the normally valued reward in the other end arm. To induce devaluation, we paired one reward with a nauseogenic dose of lithium chloride in home-cage sessions (17, 18). In home-cage tests, we confirmed that this method produced an aversion to the devalued reward (Fig. 1C). After devaluation, the rats were placed in the maze to run the task again, but the end arms were not baited. This procedure allowed us to determine whether the animals would still run when instructed to the side with the now-devalued reward, suggesting that this running was habit-driven.

The behavior of the control rats ($n = 4$) established that overtrained T-maze running was normally habitual. In the probe test after reward devaluation, these rats continued running to the devalued goal when so instructed, just as they had in the session before devaluation (Fig. 1D). Their runs to the nondevalued goal were unchanged and accurate as well (Fig. 1D). This insensitivity to reward devaluation confirmed that the training procedures produced ingrained, habitual maze runs.

To test for the effects of IL cortical activity on this habitual behavior, we used the strategy of perturbing the IL cortex during the unrewarded probe session. We applied the intervention exclusively during the performance of the runs to ensure that it was

Author contributions: K.S.S. and A.M.G. designed research; K.S.S. and A.V. performed research; K.D. contributed new reagents/analytic tools; K.S.S. and A.M.G. analyzed data; and K.S.S., K.D., and A.M.G. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. E-mail: kyless@mit.edu or graybiel@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216264109/-DCSupplemental.

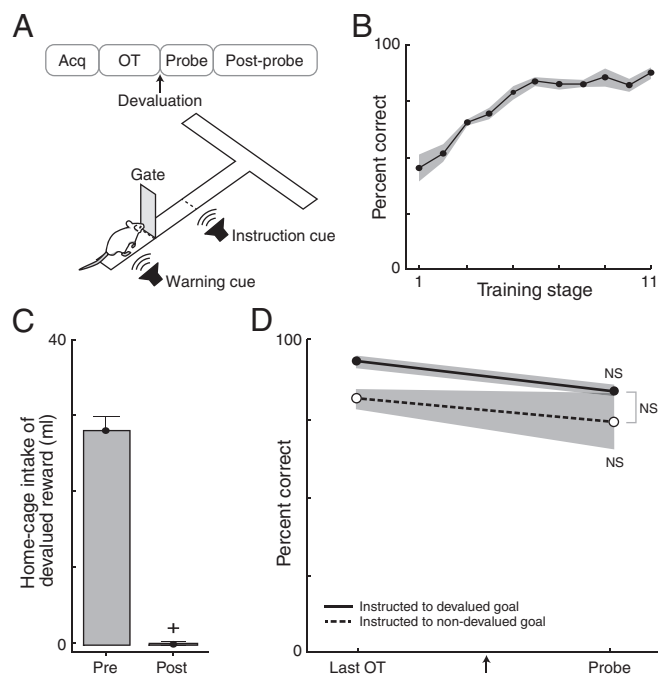


Fig. 1. T-maze training and habitual behavior of control rats. (A) Protocol for task training. (B) Performance across training stages for control rats. (C) Home-cage reward drinking before (Pre) and after (Post) devaluation. (D) Performance during the last session before devaluation (Left) and then during probe test after (Right), for control rats (solid, cued to devalued goal; dashed, cued to nondevalued goal). NS, not significantly different. * $P < 0.05$. Data are presented as means \pm SEM throughout.

online. We did not include either the prerun periods, the periods during reward delivery, or intertrial intervals. In the experimental rats ($n = 6$), an AAV-5 viral vector encoding eNpHR3.0 fused to EYFP, and targeted preferentially to glutamatergic pyramidal neurons through the calcium/calmodulin-dependent protein kinase II α promoter (AAV5-CaMKII α -eNpHR3.0-EYFP) (12), was injected bilaterally into the IL cortex 1–3 mo before behavioral training (Fig. 2A). Anterograde viral labeling was later observed in the medial striatum but not in the sensorimotor striatum, confirming a lack of a direct corticostriatal projection from the IL cortex to the dorsolateral region of the striatum known to also be essential for habit expression (Fig. 2B) (3, 10). Two of the control rats were given laser exposure following injections of control virus lacking halorhodopsin (AAV5-CaMKII α -EYFP). The other two were injected with halorhodopsin-containing virus, but the laser was inactive.

Optical Stimulation of Halorhodopsin-Expressing IL Neurons Inhibits Spike Activity. IL manipulation was accomplished by delivery of 593.5-nm-wavelength laser light through optical fibers implanted bilaterally so as to extend to the dorsal edge of the targeted IL cortex (Fig. 2A). Partial inhibition of IL activity was verified in two behaving rats fitted with a head stage carrying tetrodes and optical fibers by applying equivalent illumination parameters to those that would be used in the maze task (Fig. 2C–G). Neuronal spike activity was suppressed by 56% on average during illumination in 11 units, and the suppression was consistent over 40+ trials. This protocol resulted in a more temperate, as well as temporally specific, suppression of spike activity than the full blockade of activity produced, presumably, by lesions or drug infusions. In six units, we observed increased firing rates during light delivery, and these units were often recorded on the same tetrode as others showing inhibition (Fig. 2F). Thus, the optogenetic inhibition targeted to pyramidal cells influenced local

microcircuitry (19, 20), yielding a time-locked disruption of population spike activity (inhibition:excitation ratio, 1.8:1).

IL Perturbation Blocks Habits Online. We began by disrupting the IL cortex online while the rats performed the postdevaluation probe test (Fig. 3). Light-on occurred immediately after gate-opening, and light-off occurred immediately after goal-reaching (ca. 3 s of light per trial; <2 min total per day). This within-run treatment produced a dramatic effect: the rats acted as though they had not acquired the habit of running to the devalued goal. As shown in Fig. 3D, the rats with IL perturbation sharply reduced their runs to the end arm that would have had the devalued reward—runs characteristically made by the overtrained control rats (Fig. 3E). Instead, during the intervention, the rats had a propensity to run to the nondevalued end arm (Fig. 4A and B). Such avoidance behavior was seldom observed in the control group (Fig. 4B).

The devaluation sensitivity induced by IL perturbation was evident almost immediately (Fig. 4C). It took on average of three light trials for rats to begin avoiding the devalued goal when instructed there. In one animal, the avoidance was present during the first trial. This onset of habit blockade corresponded to an average within-run illumination time of ca. 7–9 s, with the most immediate effect within 3 s in the single rat. The online IL intervention did not have a generalized effect on the performance ability of the rats, however, because they ran accurately and seemingly automatically with IL illumination when they were instructed to run to the nondevalued reward side (Fig. 3D). These experiments suggested that IL cortical activity was required online, during the actual maze runs, in order for the expression of running behavior as a habit.

Replacement Habit Forms with Postdevaluation Training. We next analyzed the behavior of the rats when they again received reward for correct performance in subsequent days of maze training (Figs. 4 and 5). In accordance with classic findings (3, 17, 18, 21), all of the rats, including the control animals, avoided the devalued end arm after being reexposed to the reward that had been devalued (Figs. 4A and B and 5A and B), and they almost never consumed it (Fig. 5C). Thus, the control rats required actual contact with the devalued reward on the maze to trigger a loss of habitual behavior that rats with a disrupted IL cortex already exhibited in the unrewarded probe session.

Under IL perturbation, when the rats avoided the nondevalued goal, they almost never stopped the task. Instead, they continued to run to the nondevalued end arm (Figs. 4B and 5B). These “wrong-way runs” increased in frequency over days, despite the fact that the rats had not been instructed to go to that end arm and did not ever receive reward for these runs. The rats showed no overt anticipatory behavior such as licking or signs of distress at the lack of reward. In control rats, the high frequency of wrong-way runs lasted for as long as we tested (>3 wk). Moreover, these runs also appeared immune to the modest loss of in-maze aversion to the devalued outcome that appeared to occur during this time, as indicated by a small recovery of in-maze drinking of the devalued reward when the rats did run to it (Fig. 5C). This pattern of behavior suggested that the rats developed a new habit in these postdevaluation days, namely of always running to the nondevalued end arm.

IL Perturbation Blocks Replacement Habit and Returns Original Behavior. The nearly immediate loss of habitual behavior during online perturbation of IL activity suggested that switching off IL cortex might be like flipping an off-switch for habitual behavior, consistent with prior work (8, 9).

To evaluate this possibility further, we extended the laser experiments after the original probe test, gradually increasing the time for up to a month. The IL cortex was first disrupted during one to two rewarded sessions immediately after the initial probe test intervention [postprobe stage (PP)1 (first laser-on days after probe); Figs. 4 and 5]. This intervention produced no detectable

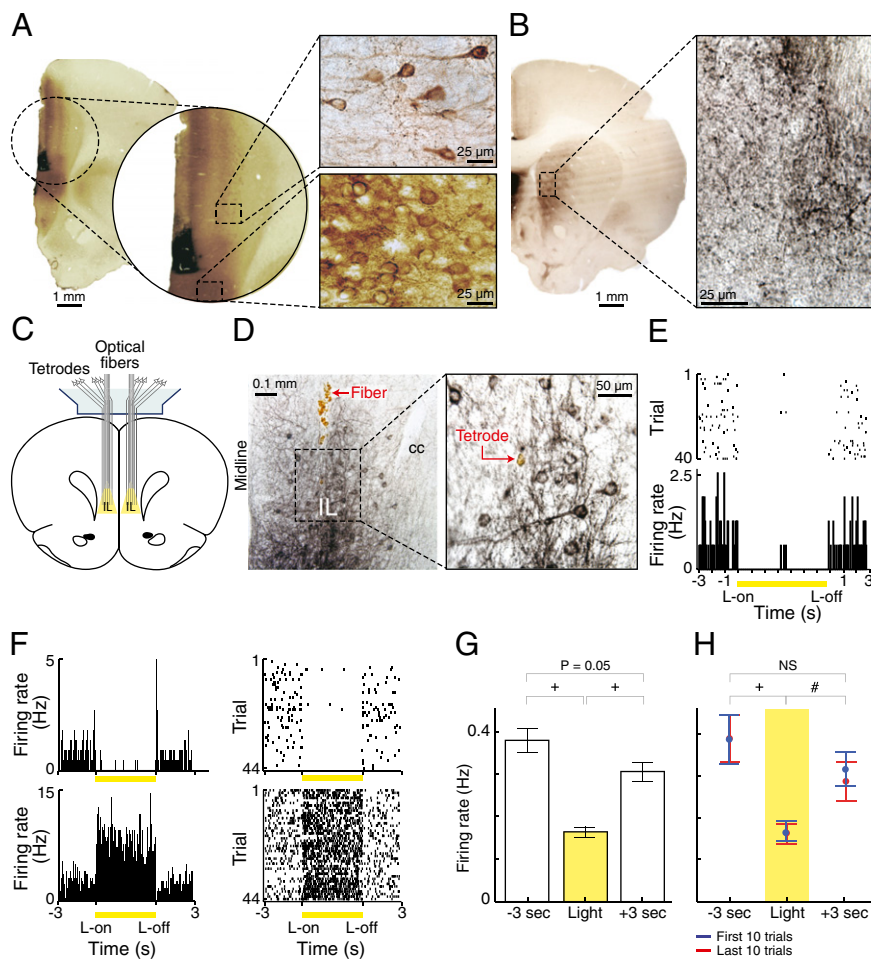


Fig. 2. Optogenetic modulation of IL neuronal activity. (A) Photomicrograph of optical fiber track (black, activated microglia stain) and virally infected neuropil (brown, EYFP stain for eNpHR3.0-EYFP). Images show areas of weak (*Upper*) and dense (*Lower*) expression. (B) Photomicrograph of virally infected IL axons and terminals in the medial, but not lateral, striatum (black, EYFP stain). The high-magnification image shows dense terminal field. (C) Arrangement of tetrodes and optical fibers in IL cortex. (D) Photomicrographs of virally infected neurons in IL cortex and track of optical fiber implant (*Left*) and tetrode (*Right*). cc, corpus callosum. (E) Raster and histogram plots of spike activity for an IL unit (50-ms bins), recorded ± 3 s around light delivery (yellow) during each of 40 trials. (F) Opposite modulation by light of the activity of two IL units recorded simultaneously with the same tetrode. One is inhibited during light delivery (with single spike rebound at light offset), whereas the other shows excitation (with momentary inhibition at light offset). (G) Average per-session spike activity of inhibited units ($n = 11$, from 2 rats) 3 s before light onset, during light delivery, and 3 s after light offset. (H) Average spiking during first 10 illumination trials (blue) and last 10 trials (red). # $P < 0.01$; * $P < 0.001$.

effect on behavior; both groups of rats equally avoided the devalued end arm on most trials. Nor did further IL perturbation introduced at up to 6 d after the initial procedure change the behavior of the rats ($n = 2$), which remained stable for up to 14 d [stages PP2–PP5 (laser-off days); Figs. 4–6]. This result suggested that once the habitual behavior had been blocked, and the IL-disrupted rats were outcome-guided in behavior like the control rats, further IL disruption failed to affect maze runs.

We obtained sharply different results when we extended the time before imposing the further perturbation (Figs. 4–6). When we again disrupted the IL cortex online 2–3 wk after the initial disruption, on days 13 ($n = 4$), 15 ($n = 1$), or 21 ($n = 1$) (stage PP6), the effect was immediate and surprising: the rats now readily ran to the devalued side when so instructed (Figs. 4A and 5A). They displayed the original behavior that they had had before any IL manipulation and did so within fewer than two laser trials on average (and on the first trial for three of the rats) (Fig. 4D). Moreover, they drank the devalued reward every time they ran to it (Fig. 5C). Simultaneously, the rats nearly stopped performing wrong-way runs (Fig. 4B). Thus, their behavior became similar for runs to the devalued and nondevalued sides (Fig. 5A–D).

This change in behavior was abrupt. These rats, on the day before the late IL perturbation, ran to the devalued reward when so cued 6 times on average and drank it 2.17 times on average (~ 0.7 mL), at most twice consecutively (Figs. 4D and 6A and B). Despite having sampled the devalued reward on this prior day, as well as on days before that (average = 1.88 drinks/day), these rats did not return to their original habitual behavior of running to the devalued reward when cued to do so (Figs. 4D and 6A and

B). By contrast, on the following day with IL illumination, the rats ran to the devalued reward and drank it earlier in the session, quickly reached an equivalent number as on the prior day, surpassed it by the fifth such trial on average, and then kept going: they ran there and drank 14.5 times on average (~ 4.4 mL), often in long repeated sequences of consecutive runs and drinks (Fig. 6B). Simultaneously, this late IL perturbation blocked the new wrong-way running habit that had developed. Both effects occurred within seconds, and within few trials, as had the original habit blockade. Control rats rarely ran to the devalued reward on the equivalent test trials (mean runs and drinks, 1.25 times) and, instead, continued to mainly avoid the devalued reward (Figs. 4–6, black lines).

This apparent return of the original habitual behavior after the late IL disruption remained during subsequent laser-off days (stages PP7–PP8, up to 4 test days; Figs. 4 and 5). When we then administered another laser session (stage PP9), this treatment fully returned the original habitual behavior, which remained for as long as we tested (up to 20 d after the third silencing, stage PP10) (Figs. 4 and 5). Over the same length of time, control rats continued to avoid the devalued goal and to not consume it (Figs. 4 and 5).

Day-by-day analysis of the running patterns suggested that a tipping point for return of the original learned behavior occurred between 6 and 13 d after the initial probe test, following which the effect of IL perturbation changed from blocking the initial habit to seemingly promoting it (Fig. 6C). This reversal of the effect of IL perturbation corresponded to the time when the new wrong-way runs had been repeated over a number of sessions. The coordinate effects on the original and new behavior of

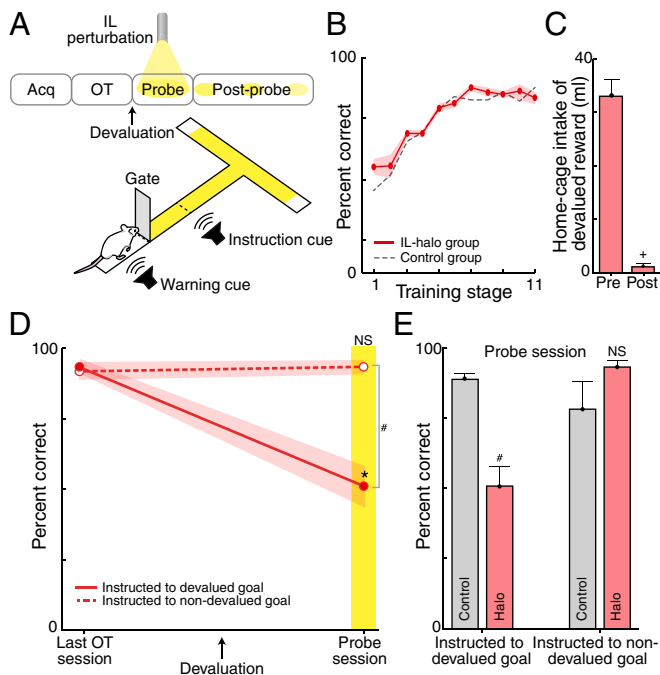


Fig. 3. Optogenetic perturbation of IL cortex blocks habits. (A) IL-illumination protocol. (B) Equivalent performance across training for halorhodopsin-treated (red) and control (dashed gray) rats before devaluation and silencing. (C) Home-cage drinking pre- and postdevaluation for IL-halorhodopsin (IL-halo) rats. (D) Performance during last session before devaluation (*Left*) and probe test after (*Right*), for rats with IL-halo (solid, cued to devalued goal; dashed, cued to nondevalued goal). (E) Performance during the probe session for control and IL-halo rat groups. Interaction of goal value and rat group on probe: $P = 0.001$; main comparisons shown: * $P < 0.05$; # $P < 0.01$, + $P < 0.001$. NS, not significantly different.

the rats suggested that IL perturbation might have turned off both the initial habit and the new habit and that blockade of the second habit might, thus, have uncovered the original habitual behavior.

IL Perturbation Does Not Affect Generalized Motivation to Consume the Devalued Reward or Taste-Aversion Memory. We tested the alternative possibility that the IL manipulation was changing a generalized motivation to drink or the associated taste-aversion memory irrespective of maze habits by examining home-cage drinking of the devalued reward on test days following the late IL perturbation (around stage PP9). The cages were in the maze room to provide context similarity (Fig. 5E). On sequential test days, light was delivered or not delivered while the reward was freely available to the rats in their cages ($n = 6$). This IL manipulation had no effect on drinking of the devalued reward (Fig. 5E, “Late”), which was also similar to home-cage drinking assessed at an equivalent time in control rats (Fig. 5E). We likewise tested the effect of IL perturbation on in-cage drinking just after devaluation and again found no effect (Fig. 5E, “Early”). This lack of effect sharply contrasted to the major effect of IL intervention on drinking after correct performance in the maze experiment proper, both at the first IL intervention and at the late interventions (Figs. 5C and 6B). Thus, if IL perturbation was affecting appetitive motivation, it was doing so only in the maze.

The in-cage test showed that both the control and IL-manipulated rats drank more at the late time than they did right after devaluation, with control rats reaching 33% (*ca.* 10 mL) of predevaluation drinking. This partial recovery also contrasted sharply with the consistently low levels of drinking in the maze during the same time period: in the maze, the rats still rejected drinking over

half of the times that they ran to it (Figs. 5C and 6A). Thus, even though the incentive value of the reward at home was partly recovering over time, it had remained weak in the maze normally. These disconnects between in-cage and in-maze drinking accord with the strong context-dependency of habits.

Discussion

Our findings demonstrate that well-ingrained habits can be controlled online by optogenetic manipulation of a specific site in the prefrontal cortex, the IL cortex. Strikingly, this control was effective even when exerted only during performance of the behavior. Moreover, although optogenetic inhibition of the IL cortex could block expression of an ingrained habit, further IL inhibition could return this habitual behavior if enough time had elapsed to allow the formation of a new habitual behavior. These effects occurred within seconds of online performance time.

These findings carry implications for the temporal dynamics and online scope of behavioral control exerted by the neocortex over habitual behavior. First, despite the seemingly automatic behavior typified by habits, their behavioral automaticity requires ongoing permissive or supervisory activity in the medial prefrontal cortex. Second, this prefrontal control appears to favor new habits if there is a competition between old and new habits; IL perturbation can block an old habit abruptly but can also abruptly bring back an old habitual behavior by blocking the newer one. Third, online manipulation of the medial prefrontal cortex affects the expression of habitual behavior almost immediately (within a performance cycle or two, totaling only

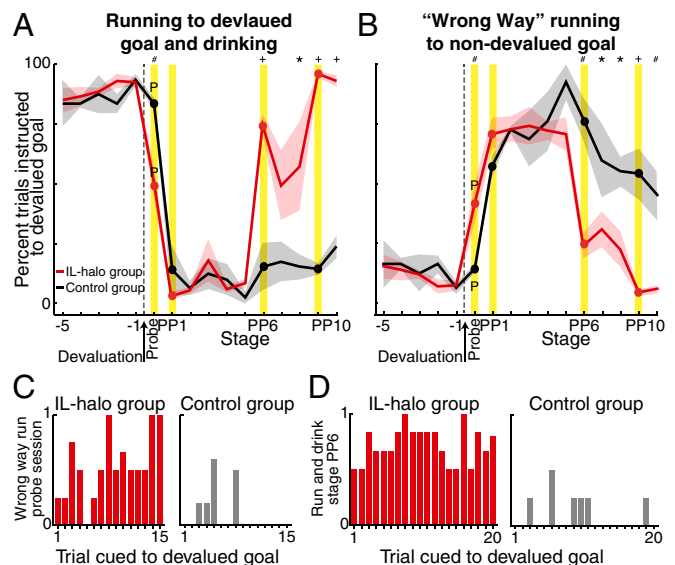


Fig. 4. IL perturbation reinstates original habit and simultaneously breaks replacement habit. (A) Percentage of trials in which rats instructed to run to the devalued goal went there correctly and also drank the reward (IL-halo group, red; control group, black; see Fig. 5 for runs and drinks separately). Shown are five sessions leading up to reward devaluation, the unrewarded probe session (only correct runs shown), and the postdevaluation stages [PP1 initial light-on session(s) or first postprobe session for controls; PP2–PP5 and PP7–PP8: sessions without light; PP6: light-on session or equivalent for control; PP9: light-on session(s); PP10: final session(s) without light]. * $P < 0.05$; # $P < 0.01$, + $P < 0.001$ between groups within stage; differences lacking symbols indicate not significantly different. (B) Percentage of wrong-way runs to nondevalued goal. (C) Trial-by-trial plot for the IL-halo group and control group, in which rats ran the wrong way during probe session (i.e., habit blockade). For each trial, wrong-way run was scored “1,” and correct run was scored “0,” and then scores were averaged (e.g., “0.5”: half of rats ran the wrong way and half ran correctly; “1”: all rats ran the wrong way). (D) Trial plot for IL-halo and control groups, in which rats ran to the devalued reward when instructed and drank it, during stage PP6 (i.e., habit reinstatement).

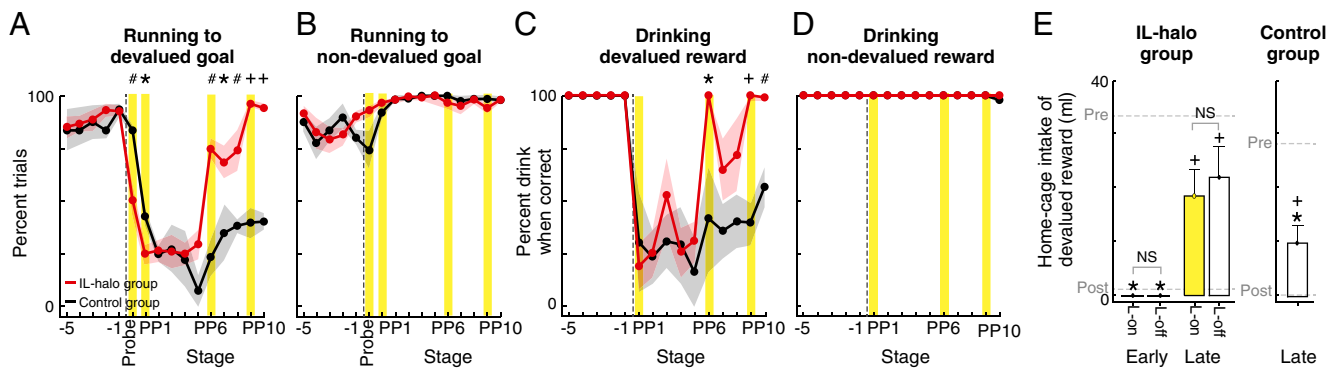


Fig. 5. Maze running and reward drinking during IL perturbation. (A and B) Percentage of correct performance on trials with instruction to devalued goal (A) or to nondevalued goal (B). * $P < 0.05$; # $P < 0.01$; + $P < 0.001$; differences lacking symbols indicate not significantly different. (C and D) Percentage of drinking of devalued reward (C) and nondevalued reward (D) when the trial was run correctly. The inverse is runs to the devalued goal without drinking. (E) Home-cage drinking of devalued reward for IL-halo group in two light-on and light-off sessions after devaluation (“Early” test, conducted around PP1–PP2), and in light-on and light-off sessions after habit reinstatement (“Late” test, conducted around PP9). * $P < 0.05$ compared with predevaluation intake (“Pre,” dashed gray line); + $P < 0.05$ compared with intake just after devaluation (“Post”). (Right) Home-cage drinking of devalued reward for control group (around PP9).

seconds of disrupted neuronal activity) when exerted only during performance, not during pretask anticipation and planning or during postperformance reinforcement or consolidation times.

These results support the findings of classic work on the habit system suggesting that in rodents, the IL cortex is necessary for habit expression (7–9) and point to the startling extent and potency of control exerted by this small cortical region. Our experiments also generated the unanticipated finding that IL perturbation can have the opposite effect when applied later to the same animals, when they had acquired a new habit: it can result in the expression of the apparently same habitual behavior that earlier IL perturbation originally blocked. Runs to the nondevalued goal, and home-cage drinking, were unaffected by the intervention, ruling out generalized effects on performance ability, motivation to drink, or devaluation memory.

One interpretation of these findings is that after devaluation, the original habit lost access to behavior, but its representation was maintained in the brain, and the late IL perturbation unmasked it. This conclusion is in good accordance with evidence, dating from the time of Pavlov (22), that when a habit is broken, it is not forgotten; rather, a new one replaces it. By this view, the IL cortex might serve, in part, as an online executive controller that favors newly acquired habits over old strategies (8, 11). Certainly, the IL cortex did not appear to act as a simple bidirectional on/off switch for habitual behavior: when we performed the second manipulation before the 1- to 2-wk period

after which the new habit was well developed, the perturbation did not reinstate the original habit (nor did it block the emerging new behavior). We take this result to suggest that the reinstatement was locked to the blockade of the second habit. The view that the IL cortex supervises newly established habits over old strategies, even habits, is consistent with evidence that the IL cortex helps maintain current response tendencies when they compete with prior ones (7, 11, 23).

An alternative view is that the late perturbation could have returned the rats to a state of value-driven behavior. There is usually a close coupling of reward-proximal stimuli and actions to current value (7, 21, 24, 25); the fact that the rats drank the devalued reward after this late manipulation of IL suggests that the reward might no longer have been aversive. However, this interpretation must face three issues. First, in-maze drinking was consistently low throughout testing in control rats and was low in the IL-halorhodopsin rats up to the putative “reinstatement.” Even when these rats ran to the devalued reward, they drank it <50% of the time; they had the opportunity to drink more on the maze but did not. Thus, the devalued reward was consistently aversive on the maze up to the point of IL-perturbation. However, then, during the late IL intervention, pursuit of this devalued reward jumped far above this level. Second, the IL-halorhodopsin rats had experience with the devalued reward in sessions before IL perturbation on the few trials they drank it, and yet this experience failed to evoke a return of the original

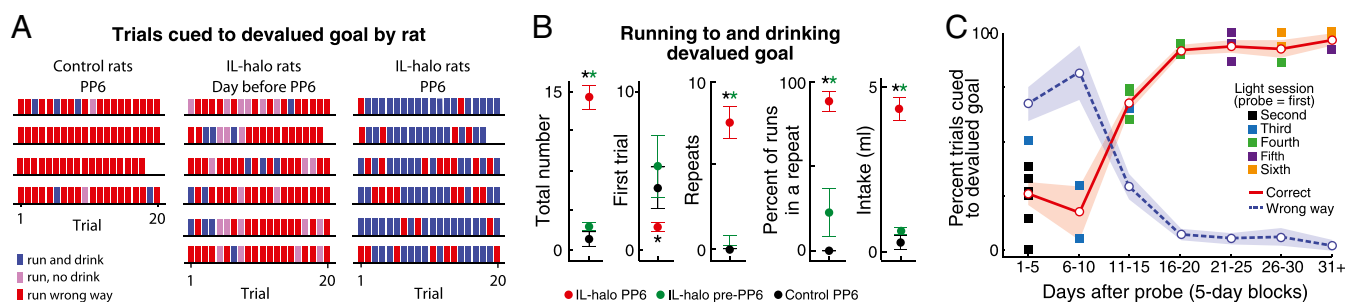


Fig. 6. Timeline of late IL perturbation effects. (A) Behavior when cued to devalued goal for each rat (top to bottom) and each trial, showing control rats on PP6, and IL-halo rats the day before PP6 and on PP6. (B) Habit reinstatement on PP6 in IL-halo rats (red), compared with the prior day (green) and to control rats on PP6 (black). Measures of running and drinking devalued goal: first trial of occurrence in the session (if ≥ 1 occurrence), number of repeats (two consecutive), percentage occurring in a repeat, and resulting intake volume. * $P < 0.05$ IL-halo rats on PP6 compared with the prior day (green) or control rats on PP6 (black). (C) Effects over postprobe days in 5-d blocks on performance during trials instructed to the devalued goal (solid, correct runs; dashed, wrong-way runs). Dots show individual data points for correct runs during light sessions, color-coded by order of light delivery after the initial probe session (e.g., black, second time overall that the rat received IL light).

behavior. Only during the IL intervention, and only in the later period of testing (PP6), did rats continue to pursue the devalued goal beyond a few samples. Third, the return of pursuing the devalued goal was nearly immediate, as soon as the very first laser trial, suggesting that the runs were based on a stored value rather than rats changing their performance within the session after contact with the devalued reward.

The maze behavior we analyzed here constituted a complex habit, with components related to the two different rewards, only one of which was devalued. Our findings suggest, as a favored interpretation, that the return of the original behavior of running to the devalued goal was a readjustment of a component of the larger behavioral repertoire learned by the rats and may have influenced the coupling of the components and subcomponents of running and drinking as well.

This evidence, collectively, places the temporal control of habit expression in a paradoxical context: despite the apparent automatic performance of habits, classically considered as outcome-independent and noncognitive, the prefrontal cortex still monitors ongoing contingencies time step by time step, and does so with the capacity to reverse the semiautomatic behavioral expression. Our findings further raise the possibility that scripts for alternate habitual behaviors are somehow stored when not expressed and that they can be unmasked if IL activity is disturbed, suggesting coordination with other brain regions. Our experiments do not settle how this coordination is accomplished or whether it is accomplished exclusively online. Still, our anatomical evidence from using the viral vector to trace the connections from the IL-injection sites confirmed that this region did not have detectable direct connections with the sensorimotor striatum but, rather, with regions (10) that should give the IL cortex direct access to circuits involved in flexibility and reinforcement as well as addiction (e.g., via projections to prelimbic neocortex and to the medial and ventral striatum) (3, 4, 6, 7, 15, 23, 25, 26), and also to habit-promoting circuits through the central amygdala (27). Each could be important for the IL habit-toggling function.

These observations raise the question of whether the IL cortex, or its human brain homolog, could similarly control addictions

or states in which behavioral flexibility and behavioral fixity are out of balance, as seen in major neurologic and neuropsychiatric disorders (5, 6, 23, 26, 28–30). Evoking fast, robust, and enduring behavioral change by targeting the IL cortex or its homolog could be of substantial value in treating disease states in a range of clinical settings, in addition to serving as a powerful approach to studying the real-time making and breaking of normal habitual behaviors.

Materials and Methods

Rats were trained on a T-maze task requiring them to respond to auditory instruction cues by turning into maze end arms to receive reward (~0.3 mL chocolate milk or sucrose, each paired with a distinct cue). Training proceeded over daily sessions through task acquisition (72.5% accuracy) and through 10+ additional overtraining sessions. For reward devaluation, rats received three pairings of free home-cage intake with lithium chloride injection and were later returned to the task for an unrewarded probe session and rewarded sessions. Task events were controlled by computer software (MED-PC; Med Associates). Behavior was monitored by in-maze photobeams and an overhead video camera. For optogenetic manipulation, injections of AAV5-eNpHR3.0-CaMKII α -EYFP or AAV5-CaMKII α -EYFP were made bilaterally into the IL cortex, and bilateral dual-ferrule optical fibers were implanted to terminate at the dorsal aspect of the IL cortex. To perturb neurons during maze runs, yellow light (2.5–5 mW) was delivered from the warning cue to goal arrival (ca. 3 s). In tests to analyze spiking dynamics, neuronal activity was recorded from 12 to 24 tetrodes while light was delivered (3-s-on/10-s-off pulses). Immunostaining and Nissl-staining procedures were used to label tetrode tracks, fiber optic cannulae tracks, and YFP-positive neurons. ANOVA and neuronal spike distribution statistics were used to assess behavioral and neuronal activity changes, with significance set at $P < 0.05$. Also see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Christine Keller-McGandy, Alex McWhinnie, Dr. Daniel J. Gibson, Henry F. Hall, Dr. Dan Hu, Dr. Yasuo Kubota, and Dordaneh Sugano for their technical help. This work was supported by National Institutes of Health (NIH) Grants R01 MH060379 (to A.M.G.) and F32 MH085454 (to K.S.S.); by the Stanley H. and Sheila G. Sydney Fund (A.M.G.); funding from Mr. R. Pourian and Julia Madadi (A.M.G.); and grants from the NIH, Defense Advanced Research Projects Agency, and the Gatsby Foundation (to K.D.).

- Daw ND, Niv Y, Dayan P (2005) Actions, policies, values, and the basal ganglia. *Recent Breakthroughs in Basal Ganglia Research*, ed Bezdard E (Nova Science Publishers, Hauppauge, NY), pp 91–106.
- Packard MG (2009) Exhumed from thought: Basal ganglia and response learning in the plus-maze. *Behav Brain Res* 199(1):24–31.
- Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. *Nat Rev Neurosci* 7(6):464–476.
- Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35(1):48–69.
- Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387.
- Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat Neurosci* 8(11):1481–1489.
- Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex* 13(4):400–408.
- Coutureau E, Killcross S (2003) Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav Brain Res* 146(1–2):167–174.
- Hitchcock PK, Quinn JJ, Taylor JR (2007) Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cereb Cortex* 17(12):2820–2827.
- Hurley KM, Herbert H, Moga MM, Saper CB (1991) Efferent projections of the infralimbic cortex of the rat. *J Comp Neurol* 308(2):249–276.
- Rich EL, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. *J Neurosci* 29(22):7208–7219.
- Gradinaru V, et al. (2010) Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* 141(1):154–165.
- Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. *Science* 286(5445):1745–1749.
- Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437(7062):1158–1161.
- Thorn CA, Atallah H, Howe M, Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66(5):781–795.
- Dickinson A (1985) Actions and habits: The development of behavioral autonomy. *Philos Trans R Soc Lond B Biol Sci* 308:67–78.
- Adams CD (1982) Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q J Exp Psychol B* 34:77–98.
- Holland PC, Straub JJ (1979) Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. *J Exp Psychol Anim Behav Process* 5(1):65–78.
- Anikeeva P, et al. (2012) Optetrode: A multichannel readout for optogenetic control in freely moving mice. *Nat Neurosci* 15(1):163–170.
- Han X, et al. (2009) Millisecond-timescale optical control of neural dynamics in the nonhuman primate brain. *Neuron* 62(2):191–198.
- Holland PC, Wheeler DS (2009) Representation-mediated food aversions. *Conditioned Taste Aversion: Behavioral and Neural Processes*, eds Reilly S, Schachtman T (Oxford Univ Press, Oxford), pp 196–225.
- Pavlov I (1927) *Conditioned Reflexes* (Dover Publications, Mineola, NY), 448 pp.
- Peters J, Kalivas PW, Quirk GJ (2009) Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn Mem* 16(5):279–288.
- Balleine BW, Garner C, Gonzalez F, Dickinson A (1995) Motivational control of heterogeneous instrumental chains. *J Exp Psychol Anim Behav Process* 21:203–217.
- Smith KS, Berridge KC, Aldridge JW (2011) Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proc Natl Acad Sci USA* 108(27):E255–E264.
- Pascoli V, Turiault M, Lüscher C (2012) Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. *Nature* 481(7379):71–75.
- Lingawi NW, Balleine BW (2012) Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *J Neurosci* 32(3):1073–1081.
- Hyman SE, Malenka RC, Nestler EJ (2006) Neural mechanisms of addiction: The role of reward-related learning and memory. *Annu Rev Neurosci* 29:565–598.
- Gillan CM, et al. (2011) Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry* 168(7):718–726.
- Leckman JF, Riddle MA (2000) Tourette's syndrome: When habit-forming systems form habits of their own? *Neuron* 28(2):349–354.

Supporting Information

Smith et al. 10.1073/pnas.1216264109

SI Materials and Methods

Subjects and Surgery. Individually housed male Sprague–Dawley rats ($n = 12$) maintained on a reverse light–dark cycle and within 85% of presurgical weight were run in experiments during their dark (active) cycle, with procedures approved by the Massachusetts Institute of Technology Committee on Animal Care. For optogenetics, AAV5–CaMKII α -eNpHR3.0-EYFP (halorhodopsin) or AAV5–CaMKII α -EYFP (control) was injected bilaterally into the IL cortex [0.2–0.7 μ L/20 min per injection; anterior–posterior (AP), 3.1 mm; medial–lateral (ML), \pm 0.6 mm; dorsal–ventral (DV), 5.3 mm]. A dual-ferrule optical fiber (diameter, 200 μ m; Doric Lenses) was implanted bilaterally in the dorsal IL cortex, anchored, and shielded. For recording light-induced firing changes, rats were implanted with recording head stages carrying 12–24 tetrodes and two optical fibers aimed at the IL cortex. Recording head stages were implanted as described (1–4).

T-Maze Apparatus and Training. A T-maze was used for the experiment and was identical to one described previously (1–3). Reward was manually delivered via tubing to troughs at the end-arm goal sites. Rats were habituated to task conditions and rewards (30% sucrose solution and chocolate-flavored whole milk) as described (1–4). Training proceeded in daily 40-trial sessions consisting of the following: the rat waited on a platform, a warning click sounded, the start gate was lowered, the rats traversed the maze, and instruction cues (1 or 8 kHz) sounded as the rat approached the decision point and remained on until a goal was reached, where the rat was rewarded (about 0.3 mL) for correct performance (*ca.* 1-min intertrial interval). Each reward was assigned to only one arm per rat; turn, tone, and reward assignments were pseudorandom across rats. Training continued through acquisition (72.5% accuracy criterion, χ^2 ; $P < 0.01$ compared with chance) and overtraining (10+ additional sessions at or above criterion). Photobeams placed every *ca.* 17.5 cm on the maze-tracked behavior. Task control was by a MED-PC program (Med Associates). Nearly all trials terminated with goal-reaching. Very rare trials in which rats stopped at the tone were accepted but did not contribute to tallies of correct or wrong-way runs. Some rats eventually chose not to run during the unrewarded probe session (*i.e.*, not all rats reached 40 probe trials); the session was ended at these instances.

Reward Devaluation. Rats were given 45-min access to one maze reward (*e.g.*, chocolate milk) in their home cage and then received an injection of lithium chloride (0.6 M 5 mL/kg or 0.3 M 10 mL/kg, *i.p.*). Three devaluation procedures at 48-h intervals were given in multiple laboratory rooms, although never in the maze room, and efficacy was confirmed by reduced intake. Devalued reward identity was pseudorandomly assigned across rats. Rats were then given a probe session without rewards given, followed by normal rewarded sessions. The purpose of these rewarded sessions was to confirm that the taste aversion developed in the home-cage environment generalized to the task environment, as well as to assess behavioral plasticity occurring after encounter with the devalued reward in the maze task.

Session Staging. Training sessions were staged as follows: stages 1–2 (first two sessions); stages 3–4 (pairs of sessions $\geq 60\%$ correct); stage 5 (first pair of sessions $\geq 72.5\%$); and stages 6+ (subsequent pairs of sessions $\geq 72.5\%$). Stages for comparing post-devaluation days of IL light delivery were as follows: probe (IL light delivery, unrewarded); stage PP1 (first one to two rewarded sessions with IL light delivery); stages PP2–PP5 (subsequent sessions without IL light delivery); stage PP6 (session with light delivery); stages PP7–PP8 (sessions without IL light delivery); stage PP9 (final IL light delivery sessions); and stage PP10 (final sessions without light delivery).

Optogenetic Light Delivery. Light was delivered to the IL cortex from a laser (593.5-nm diode-pumped solid-state laser; OEM Laser Systems) from warning cue to goal arrival (*ca.* 3-s duration; 2.5–5.0 mW). Fibers and equipment were connected during light-off training and test days. Light effects on reward consumption were measured in two, pseudorandomly ordered 40-min tests with rats placed in their home cage on a table in the maze room. The devalued reward was delivered via tubing into a maze trough placed in the cage while illumination was given (5 mW per fiber; 3-s-on/10-s-off pulses) or not given. The same illumination parameters were used for tests of firing rate changes evoked by light delivery, during which rats were allowed to freely explore the maze for 40+ illumination trials while IL activity was recorded.

Electrophysiological Data Acquisition. Tetrodes lowered to recording targets over 7 postsurgical days were left in place or moved in <0.04 -mm steps. Electrical signals were amplified at 100–10,000, sampled at 32 kHz, filtered at 600–6,000 Hz, and recorded by a Cheetah data acquisition system (Neuralynx) as described (1–4). Single units were identified as isolated waveform clusters using Offline Sorter (Plexon).

Analysis. Performance (percentage of correct, incorrect, and incomplete trials) and reward consumption were analyzed by ANOVA ($P < 0.05$) to compare across learning stages, trial subtypes (*e.g.*, devalued and nondevalued trials), and conditions of IL light delivery. Within sessions, we also assessed the time scale of runs to the devalued goal and drinks: first trial in which it occurred, number of two consecutive trials (repeat doublets), average number of repeated runs and drinks, and volume of drinks (*ca.* 0.3 mL per reward). Bonferroni-corrected post hoc comparisons were made when significance was obtained for the main effect of variables and/or interaction between variables. For neuronal recordings, per-unit firing 3 s before, during, and 3 s after light delivery was compared using ANOVA on time epoch and firing rate and separate ANOVAs for comparing first and last trials blocks within a session. Responsive units were those with a significant change in firing rate during the light-on period compared with 3 s prior.

Histology. Tetrode and fiber cannula tracks were identified histologically (1–4). For double immunostaining, sections were immunostained for YFP (GFP antibody) and activated microglia (CD11b/c antibody).

1. Barnes TD, Kubota Y, Hu D, Jin DZ, Graybiel AM (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* 437(7062): 1158–1161.
2. Barnes TD, et al. (2011) Advance cueing produces enhanced action-boundary patterns of spike activity in the sensorimotor striatum. *J Neurophysiol* 105(4):1861–1878.

3. Thorn CA, Atallah H, Howe M, Graybiel AM (2010) Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* 66(5): 781–795.
4. Kubota Y, et al. (2009) Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. *J Neurophysiol* 102(4):2142–2160.