Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

Neuron
Article

A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics

Kyle S. Smith[1,*] and Ann M. Graybiel[1,*]
[1]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: kyle.s.smith@dartmouth.edu (K.S.S.), graybiel@mit.edu (A.M.G.)
http://dx.doi.org/10.1016/j.neuron.2013.05.038

SUMMARY

Habits are notoriously difficult to break and, if broken, are usually replaced by new routines. To examine the neural basis of these characteristics, we recorded spike activity in cortical and striatal habit sites as rats learned maze tasks. Overtraining induced a shift from purposeful to habitual behavior. This shift coincided with the activation of neuronal ensembles in the infralimbic neocortex and the sensorimotor striatum, which became engaged simultaneously but developed changes in spike activity with distinct time courses and stability. The striatum rapidly acquired an action-bracketing activity pattern insensitive to reward devaluation but sensitive to running automaticity. A similar pattern developed in the upper layers of the infralimbic cortex, but it formed only late during overtraining and closely tracked habit states. Selective optogenetic disruption of infralimbic activity during overtraining prevented habit formation. We suggest that learning-related spiking dynamics of both striatum and neocortex are necessary, as dual operators, for habit crystallization.

INTRODUCTION

Across the animal kingdom, and across the range from normal to dysfunctional states in humans, the balance between flexible and repetitive behaviors is critical for optimal performance of tasks (Aston-Jones and Cohen, 2005; Balleine et al., 2009; Brainard and Doupe, 2002; Daw et al., 2005; Graybiel, 2008; Hikosaka and Isoda, 2010; Yin and Knowlton, 2006). Flexible goal seeking is advantageous in many situations, but a narrowing of behavioral focus is necessary to reach specific goals. Conversely, fixed routines are advantageous in freeing up attention and decision-making resources, but habits can be harmful and difficult to break (Everitt and Robbins, 2005; Graybiel, 2008; Hyman et al., 2006; Kalivas and Volkow, 2005; Redish et al., 2008).

Classic experimental studies based on lesion and chemical inactivation methods have identified two major brain regions as being essential for performing habits in animal studies. One, the sensorimotor striatum (called the dorsolateral striatum, DLS, in rodents), is embedded in sensorimotor basal ganglia circuitry (McGeorge and Faull, 1989). This striatal region is thought to store action plans for habit learning based on its anatomical position, its neural activity related to behavioral responses, and evidence that damage to it disrupts the stability of well-honed behaviors (Aldridge et al., 2004; Balleine et al., 2009; Carelli et al., 1997; Graybiel, 2008; Kimchi et al., 2009; Packard, 2009; Tang et al., 2007; Tricomi et al., 2009; Yin and Knowlton, 2006). This site has repeatedly been shown to develop a pattern of neuronal activity that brackets the beginning and end actions of a well-learned behavior sequence (Barnes et al., 2005; Jin and Costa, 2010; Jog et al., 1999; Thorn et al., 2010).

Less is known about the neural activity patterns related to habit formation in the other key habit-promoting site, the infralimbic (IL) cortex. This medial prefrontal cortical region lacks direct connections with the DLS but must also be intact in order for habits to be expressed (Coutureau and Killcross, 2003; Hitchcott et al., 2007; Killcross and Coutureau, 2003). This control is exerted online during habit performance (Smith et al., 2012). Based on its connections with prefrontal-limbic networks, the IL cortex has been proposed as exerting an executive-level control in the selection of habits (Daw et al., 2005; Hitchcott et al., 2007; Killcross and Coutureau, 2003), whereas representations of the habit itself would reside in sensorimotor networks. However, such findings raise the possibility that the IL cortex and DLS might need to operate coordinately in order for habits to form, both being responsible for building a habit, probably along with a distributed network of other regions (Balleine et al., 2009; Coutureau and Killcross, 2003; Daw et al., 2005; Graybiel, 2008; Yin and Knowlton, 2006).

To test this possibility, we simultaneously monitored neural activity in the IL cortex and the DLS with chronic tetrode recordings over months as animals learned a maze habit through training and overtraining, then as the habit was lost after reward devaluation, and finally as it was replaced by a new habit. We found strikingly different dynamics of ensemble spike activity in the two regions as habits formed, yet we found that the IL cortex eventually joins the DLS in forming a consensus task-bracketing activity pattern as the habits become crystallized. We then used optogenetic methods to perturb the IL cortex online during this critical crystallization period and found that daily online IL inhibition prevented the habit formation. These findings suggest that the crystallization of habits does not simply result from the storing of fixed values in the sensorimotor system but, instead, represents the consensus operation of both sensorimotor and limbic circuits.

Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

RESULTS

T-Maze Overtraining Induces a Habit

We designed a task for rat subjects allowing us to determine the time during learning at which the animals switched from flexible, goal-directed behavior to habitual, repetitive routines. We adapted a classic devaluation protocol to determine whether a behavior qualifies as a habit (Dickinson, 1985). The test involves training animals on a task that is rewarded and then determining whether the reward still drives the behavior after it has been made aversive or nonrewarding, a procedure called devaluation. If subjects continue to perform the task to obtain the newly devalued reward, that behavior is considered to be outcome independent and habitual. If, however, the subjects quit performing the task, the behavior is considered to be goal directed, as though the subjects were keeping the specific outcome in mind. We used this approach by having rats perform a T-maze task in which they could receive different reward (chocolate milk or sucrose solution) at the two end-arms of the maze (Figure 1A). This strategy allowed us to devalue one reward and then to test for habitual running to the end-arm baited with the now-devalued reward, as compared to running to the other end-arm as a control (Smith et al., 2012).

We tracked the learning curves of multiple sets of rat subjects (Figure 1B). Over 8 to 16 weeks of training, for ca. 40 or more trials per daily session, the rats were required to initiate maze runs in response to a warning cue and gate opening, run down the maze, and turn right or left, depending on an auditory instruction cue, in order to receive reward. Each reward type was assigned to one arm for each rat. Entry into an incorrect arm resulted in no reward. One set of rats (CT group) was trained just until they reached a criterion of statistically significant performance accuracy (at least 72.5% correct for 2 days, stage 6; Figure 1B). A second set of rats (OT group) was trained past learning criterion during an overtraining period for ten or more additional sessions. Both groups of rats learned the task, reaching about 90% correct (Figure 1B).

Each set of rats was then exposed to the devaluation protocol, in which we exposed the rats to home-cage pairings of one reward with a nauseogenic dose of lithium chloride to induce devaluation (Adams, 1982; Holland and Straub, 1979). After establishing that this procedure produced an aversion to the paired reward, as measured by reduced home-cage intake (Figure 1C), we tested the rats in the maze in a probe session. Reward was not given in this probe test in order to estimate whether running was outcome-guided behavior and sensitive to the change in reward value, or whether instead running was habitual. The results of this probe test were clear cut: the rats trained only to criterion immediately reduced by nearly 50% their running to the end-arm that would have been baited with the devalued reward (Figure 1D). The overtrained rats, however, kept running to the devalued reward (Figure 1D). All of the rats ran correctly when they were cued to go to the nondevalued end-arm (Figure 1E). These results suggest that T-maze overtraining had induced an outcome-insensitive running habit, confirming our previous finding (Smith et al., 2012), but that the full habit had not yet been induced in the animals trained only to the criterion level for behavioral acquisition.

A Replacement Habit Forms with Postdevaluation Training

We next tested the behavior of the rats when we again rewarded correct performance during 6 or more days of maze training. In accord with the powerful effect of conditioned taste aversion on reward pursuit (Adams, 1982; Garcia and Ervin, 1968; Holland and Straub, 1979), even the overtrained animals reduced their running to the end-arm with the devalued reward after tasting that reward again on the maze. Their runs to the devalued side, when so instructed, fell to the same 50% level that control rats had reached during the probe session (Figures 1F and 1G). Moreover, the rats drank the devalued reward on average fewer than half the times when they did run to it (Figure 1H). Instead, they ran the "wrong way" to the nondevalued goal in response to the instruction cues directing them to the devalued side (Figure 1I). Despite remaining unrewarded, the wrong-way runs increased in frequency over days (Figure 1I) and grew equivalent in speed to correct runs to the same goal and to predevaluation behavior, suggesting that they became insensitive to outcome value and became habitual (Smith et al., 2012).

The occurrence of deliberative head movements also suggested that these wrong-way runs represented a new habit. The head movements, in which the rats looked to the nonchosen run side before running the other way at the choice point (Figure 1J), decreased in frequency as performance improved during training and overtraining (Figure 1K). This result is in accord with previous suggestions that they reflect purposefulness in decision making (Muenzinger, 1938; Redish et al., 2008; Tolman, 1948). In the sessions after devaluation, the deliberative movements during wrong-way runs were initially high, but then they fell again (see Figure 3B). Run speeds similarly rose during overtraining and, after devaluation, were eventually higher for both wrong-way runs and correct runs to the nondevalued goal, and lower for runs to the devalued goal (Figures 1L and 1M).

Contrasting Cortical and Striatal Activity Dynamics Track Habit Formation

Based on these behavioral indices of habit formation, blockade, and replacement, we analyzed the spike activity patterns of IL and DLS neurons relative to the rats' performance across both the early training and overtraining periods and also the postdevaluation period. We recorded activity in the IL cortex and DLS simultaneously for up to 4 months with chronically implanted multiple-tetrode assemblies as rats learned the tasks (n = 7, OT rats in Figure 1). Tetrodes were not moved or were lowered only in small (ca. 40 μm) steps to maintain the quality of recordings. For the DLS recordings, we focused on putative striatal projection neurons (n = 1,479 total and n = 858 task-related units; Supplemental Experimental Procedures available online). For the IL cortical recordings, we analyzed 1,694 units, of which 1,013 were task-related units. Because of the near-vertical orientation of the medially situated IL cortex, we were able to monitor activity recorded from tetrodes placed in relatively more superficial (ILs) or deep (ILd) depths of the neocortex (Figures 2A and S1).

We found a marked contrast between the changes in ensemble activity in the DLS and IL cortex that occurred as learning proceeded. During initial training, ensemble activity in the DLS was at first heightened throughout the maze runs.

Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

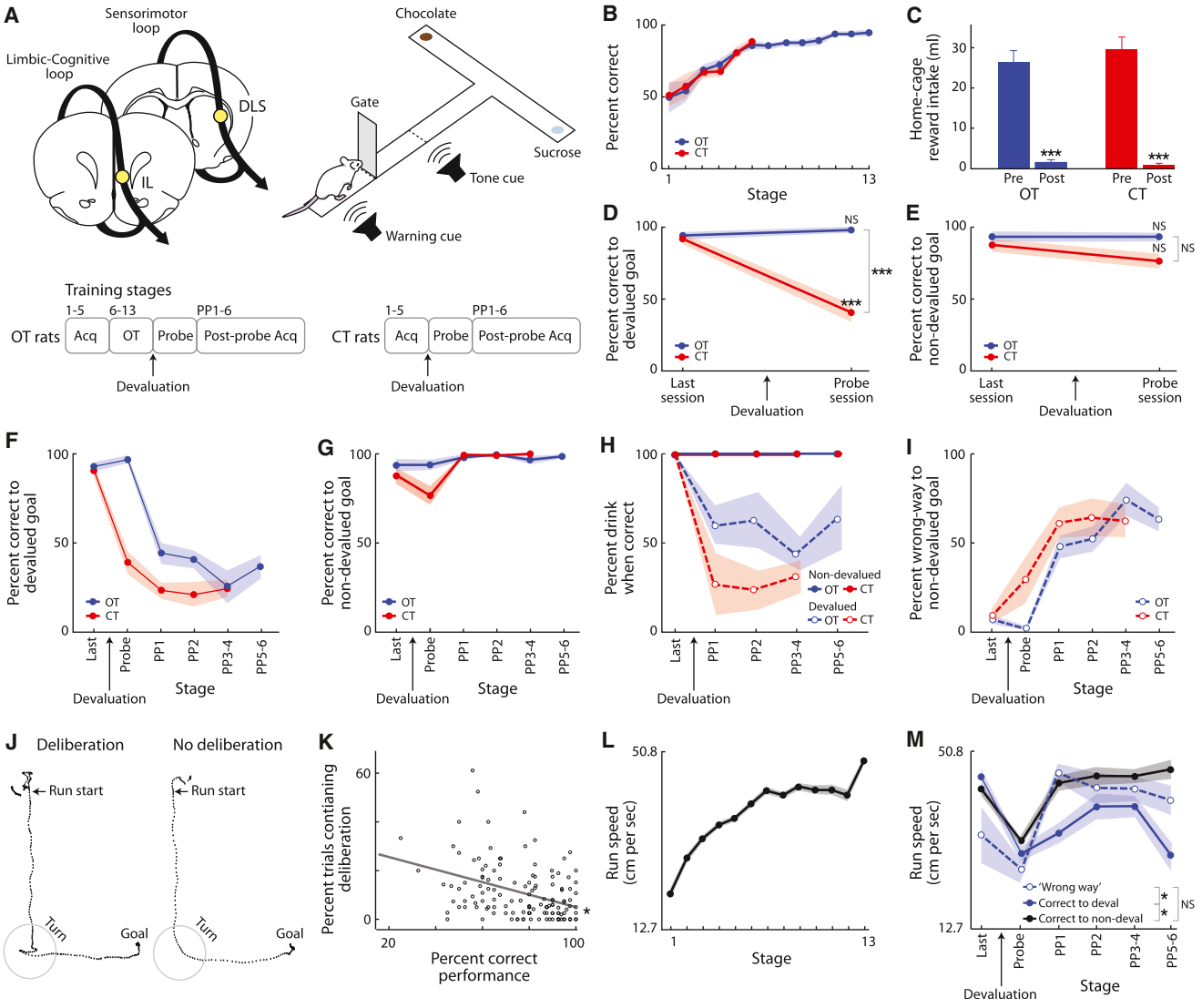

Figure 1. Experimental Design and Behavioral Performance

(A) Recording locations and T-maze task. Below, protocols for overtrained rats (OT, n = 7) and criterion-trained rats (CT, n = 5). Acq, task acquisition; Probe, unrewarded session after devaluation; PP, postprobe rewarded acquisition sessions.

(B) Performance accuracy for OT (blue) and CT (red) rats.

(C) Home cage reward intake pre- and postdevaluation. ***$p < 0.001$.

(D and E) Performance on runs cued to devalued (D) and nondevalued (E) goals on days before devaluation and after (unrewarded probe day). ***$p < 0.001$; NS, not significantly different.

(F and G) Correct cued runs to devalued (F) and nondevalued (G) goals in PP sessions.

(H) Percent of correctly performed trials with reward intake, during runs to the devalued (dashed) and nondevalued (solid) goals. Drinking of devalued reward was low after devaluation (e.g., 50% drinking from 25% correct runs = 2.5 drinks or ~0.75 ml).

(I) "Wrong way" runs to nondevalued goal before and after devaluation.

(J) Representative videotracker traces of maze runs with (left) and without (right) deliberation at the choice point.

(K) Scatter plot and regression fit for performance accuracy and deliberation occurrence in OT rats (dot = session), showing fewer deliberation trials with greater performance accuracy (Pearson's R = −0.37; *$p < 0.001$).

(L) Run speed for OT group during training and overtraining. Apparent increase at stage 13 due to lack of stage 13 data for three slower rats.

(M) Speed of OT rats on runs to devalued (blue) and nondevalued (black) goals, and wrong-way runs to nondevalued goal (blue dashed) on days before and after devaluation. *$p < 0.05$.

Data are presented as mean ± SEM.

Around the time the learning criterion was reached, this pattern gave way to one in which the activity decreased at midrun and became high early and late during the maze runs, and at the turns (Figures 2B–2E and S2), consistent with previous findings (Barnes et al., 2005; Thorn et al., 2010). By contrast, during the entire initial training period, ensemble activity in the IL cortex

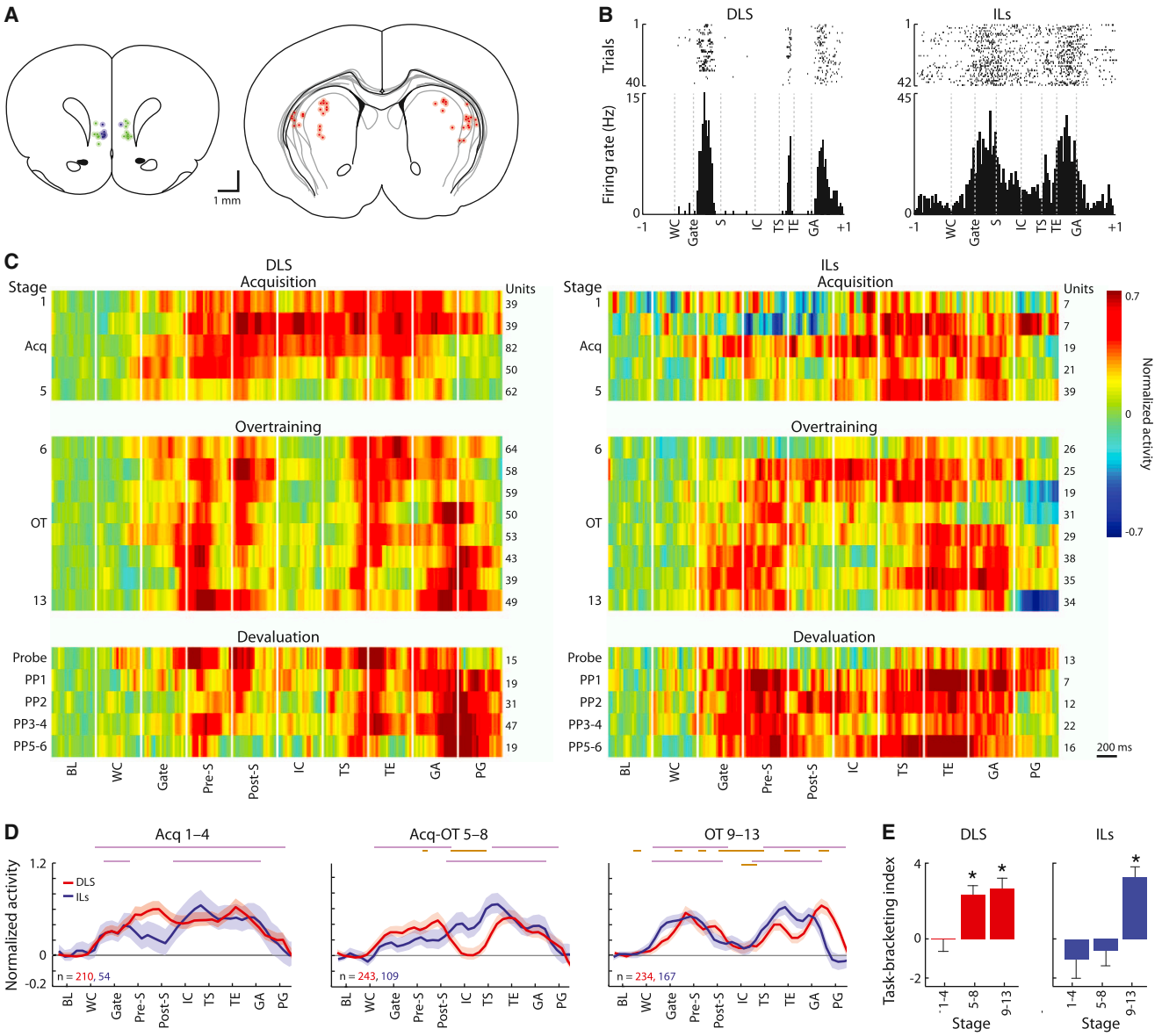

Figure 2. Formation of Task-Bracketing Activity in DLS and ILs

(A) Schematic sections of tetrode recording locations (circles) in IL cortex (left) and DLS (right). IL recordings split by mediolateral position into "superficial" (blue) and "deep" (green) placements. Circle sizes indicate estimated recording coverage (inner circle: 0.05 mm radius of peak spike recording; outer halo: 0.14 mm radius of maximal recording; from Henze et al. (2000). See also Figure S5C.

(B) Spike raster plots (top) and histograms (bottom) of sample DLS (left) and ILs (right) units recorded during overtraining (50 ms bins, ±1 s before and after run). Perievent windows display middle half of median perievent time between the prior and next events, averaged across trials. WC, warning cue; Gate, gate opening; S, run start; IC, instruction cue; TS, turn start; TE, turn end; and GA, goal arrival.

(C) Normalized (baseline-subtracted *Z* scores) activity of DLS (left) and ILs (right) task-related units for seven rats, constructed from abutted ±200 ms perievent periods (20 ms bins) during acquisition (stages 1–5), overtraining (6–13), and postdevaluation probe and rewarded (PP 1–6) sessions. Number of units and color scale are shown on the right. BL, baseline; Pre-S, 200 ms before run start; Post-S, 200 ms after run start; PG, 0.5 s after goal arrival.

(D) Activity in ±200 ms perievent windows (100 ms bins) for DLS (red) and ILs (blue) for successive training stages (Acq, 1–4; Acq-early OT, 5–8; late OT, 9–13). Number of task-related units is shown on the bottom left. Purple bars, bins with activity significantly different from prerun baseline; orange bars, significant difference from activity in same time bins in Acq 1–4 ($p < 0.05$).

(E) Index of task-bracketing ensemble pattern strength (mean activity in start and end periods minus mean midrun activity) across training stages and recording locations. *$p < 0.05$ from zero.

Data are presented as mean ± SEM. See also Figure S1.

scarcely changed, despite the fact that the animals were learning (Figures 2C–2E, S1, and S2). Then, nearly halfway through the overtraining period, the IL ensembles acquired a run-bracketing pattern quite similar to the pattern that had developed much earlier in the DLS recordings (Figures 2B–2E). This change occurred during the time period in which behavior shifted from goal directed to habitual. Thus, by the time overtraining was completed, the ensemble activities in both DLS and ILs exhibited task-bracketing patterns with low activity midrun and highest activity early and late during the runs. However, this patterning was reached in the two regions at different times during training, as confirmed by analysis of task-bracketing index scores for the ensembles, defined as ([mean activity during run start and end periods] − [mean activity around the instruction cue]) (Figure 2E).

Contrasting Cortical and Striatal Activity Dynamics Track the Suppression of an Acquired Habit and the Emergence of a Second Habit

The similarity in the task-bracketing patterns that formed early in DLS and late in ILs raised the possibility that, in order for the habit to become established, both the DLS and the ILs had to form a beginning-and-end pattern. We therefore assessed whether these patterns also changed after the reward devaluation protocol (Figures 2, 3, 4, and 5). Surprisingly, the task-bracketing pattern of ensemble activity in the DLS remained almost completely stable after devaluation (Figures 2C and 5A), despite the major changes in behavior and outcome occurring during this time (Figures 1F, 1H, 1I, and 1M). By contrast, ILs activity changed sharply. The magnitude of ensemble activity during runs rose immediately after devaluation on the first training day, postprobe day 1 (PP1) (Figures 2C and 5B), so that midrun activity became as strong as it had been at the task boundaries before devaluation. The trial-to-trial variability of ILs spiking during runs also increased markedly on this PP1 day (Figures 5C and 5D). The task-bracketing pattern remained evident but became obscured by generalized higher activity by the second postdevaluation training day (Figures 2C, 3D, and 5A). These results suggested that the task-bracketing ensemble pattern in the striatum, viewed across sessions, was insensitive to the devaluation but that activity in the medial prefrontal cortex was sensitive to exposure to the devalued goal during task performance.

We next tracked the session-by-session ensemble activity in the ILs and in the DLS in relation to the behavioral measure of deliberative head movements at the choice point of the maze. We calculated the task-bracketing index for the neural activity for each unit recorded per session (Figure 2E) and then compared the index scores to the percentage of trials in which deliberative head movements occurred during these same sessions. As the deliberations fell during the initial acquisition and overtraining periods, the ILs task-bracketing pattern gradually emerged (Figures 3A and 3C). After devaluation, the session-wide level of deliberative head movements again was correlated inversely with the ILs task-bracketing pattern. Deliberations were somewhat low on PP1 when the pattern mostly remained, then rose on subsequent days as the pattern decayed, and finally fell again at the end of testing when the pattern re-emerged (Figures 3B, 3D, and 5A). These changes in total deliberations were driven chiefly by the number of deliberations during trials in which the rats ran the wrong way when instructed to the devalued goal (Figure 3B). Deliberations during correct running to the same, nondevalued side were almost nil throughout postdevaluation training (Figure 3B).

When viewed across all training stages, the session-by-session changes in deliberative head movements were significantly anticorrelated with the strength of the task-bracketing patterning index score calculated for each recorded ILs unit (Figure 3F). The total numbers of recorded ILs units with significant responses to the start and/or end of the runs tended to follow a similar inverse relationship with deliberations (Figure 3E). We further divided the ILs units into those with positive index scores (task-bracketing activity) or negative scores (higher midrun activity) and assessed the population activity changes of these two subgroups relative to learning stages and deliberations. During initial training and early overtraining, there were more units with negative index scores than with positive scores. Then, during the late overtraining phase, the balance shifted: more of the recorded ILs units exhibited a positive task-bracketing pattern, resulting in a significant interaction of the index score with learning stage (Figure 3G). It was the units with positive task-bracketing scores that accounted for the significant correlation with deliberative movements; units with negative task-bracketing scores were not significantly correlated with deliberations (Figure 3H). This result suggested that as the habit emerged during late overtraining, there was a concomitant increase in the number of ILs units with task-bracketing activity, a decrease in those with opposite patterning, and an increase in the strength of task-bracketing in the ILs ensemble.

DLS activity did not covary with the number of deliberations occurring in a given session, whether analyzed as total ensemble activity (Figure 3F) or after division of the units into subgroups based on positive and negative task-bracketing scores. The session-averaged DLS task-bracketing pattern remained relatively stable across overtraining and postdevaluation test days (Figures 3C–3E), even though the net number of deliberations fluctuated.

When we assessed the DLS spike activity trial by trial, however, we found a nearly opposite result. In the DLS, there was a clear trial-level modulation of the bracketing pattern in relation to the occurrence of deliberative movements. The bracketing index was higher on single runs lacking a deliberation at the choice point (Figure 4A), most prominently during learning and late overtraining (Figure 4B). This modulation involved weaker levels of DLS spike activity at the start of the single runs in which a subsequent deliberation occurred (Figure 4C). Activity during the deliberation and turn itself was only moderately and nonsignificantly lower during such trials and thus did not solely account for the effect. By contrast, in the ILs, spike activity during individual trials was similar whether the runs contained or lacked a deliberation (Figures 4A and 4C), and whether units were considered as an ensemble or were divided based on positive or negative task-bracketing scores.

This contrast suggests that the task-bracketing pattern that forms in ILs ensembles covaried over sessions with states of habitual behavior in which the majority of runs were nondeliberative, whereas the relatively similar ensemble pattern in the DLS

Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

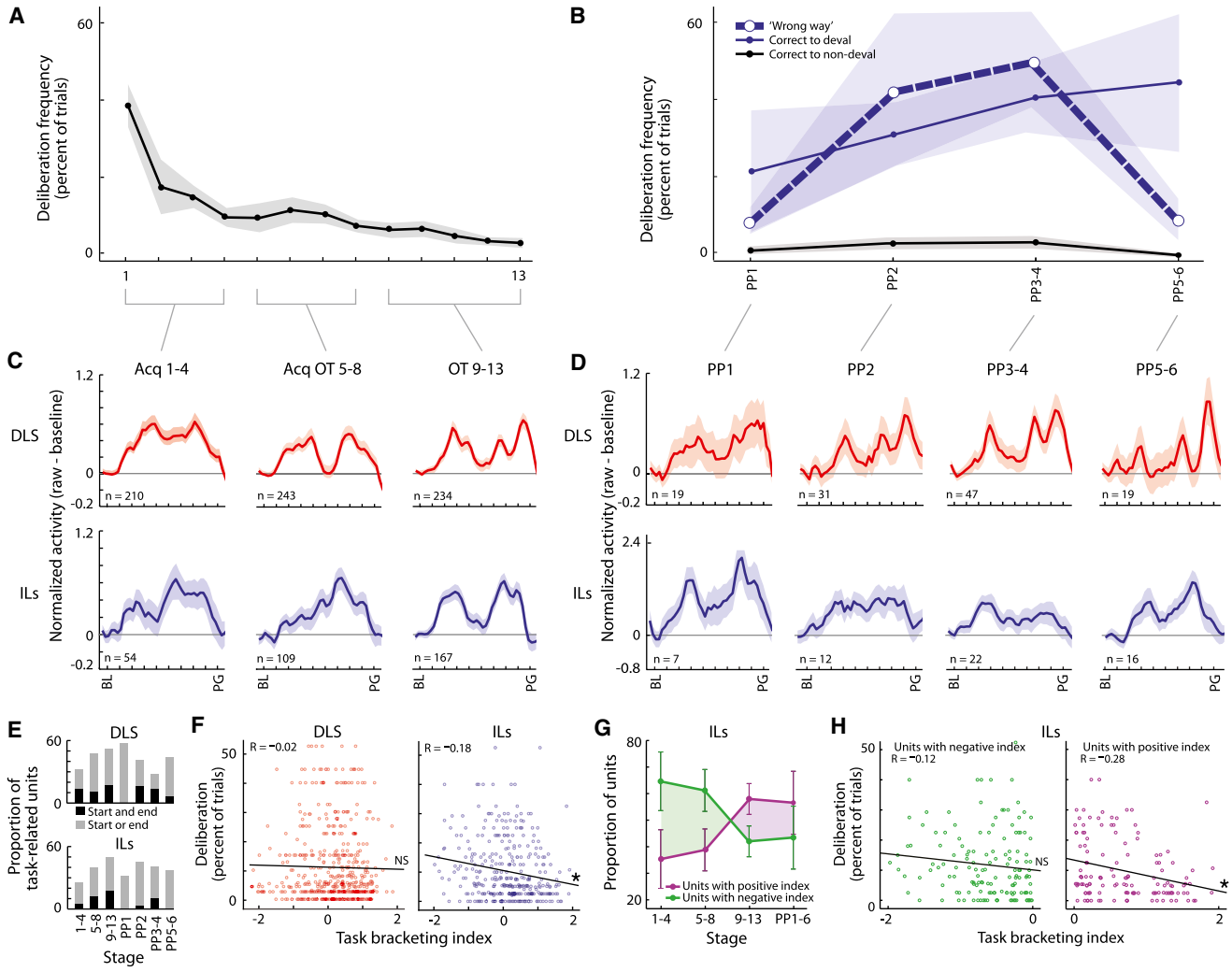

Figure 3. Comodulation of ILs Ensemble Activity and Deliberative Behaviors at the Session Level

(A and B) Percent of trials containing a deliberation across training (A) and during PP days (B) for cued runs to the nondevalued (black) and devalued (blue) goals and wrong-way runs (dashed blue).

(C and D) Normalized ensemble activity (baseline-subtracted spiking) during acquisition (Acq) and overtraining (OT, C) and postdevaluation stages (D) for the DLS (top) and ILs (bottom). Note expanded y axis for ILs in (D). Plotting is as in Figure 2D.

(E) Proportions of task-related DLS (top) and ILs (bottom) units that contribute to task-bracketing activity, including those with activity at run start and end (black, task-bracketing units) or activity specific to start or end (gray).

(F) Scatter plots and regression fit of DLS (left) and ILs (right) task-bracketing index per unit and percent of trials containing deliberation during the session the unit was recorded. *R = −0.18; regression, t = −3.56, p < 0.001.

(G) Proportion of all ILs units with a positive task-bracketing index (purple, index above zero) or a negative task-bracketing index (green, index below zero). The relative number of positive task-bracketing units increased sharply at late OT (interaction of training time and proportion of units with positive task bracketing: F = 3.6, p = 0.017), just as the task-bracketing pattern emerged in ensemble activity.

(H) Split regression on ILs units with a negative (left) or positive (right) index score and percent of trials containing deliberation per session (positive: *t = −3.30, p = 0.001; negative: t = −1.42, p = 0.16). Thus, correlation in (F) was driven by units with positive task-bracketing activity.

Data are presented as mean ± SEM. See also Figure S2.

appeared stable over the time span of sessions but was modulated trial to trial, especially at run start (Figure 3E). The DLS task-bracketing activity was also influenced by the stage of behavioral training that the rats had reached, however, as the pattern emerged after initial learning, suggesting that the presence of the DLS ensemble pattern was a function of learning or experience as well as the automaticity in individual runs.

Distinct Pattern of Activity in Deep IL Cortex Related to Habitual Maze Runs

Units recorded from tetrodes placed in the deeper layers of the IL cortex responded differently from those in the upper layers (Figures 5 and 6). ILd units did not form a pattern marking particular phases of the task but, rather, showed a general increase in activity as ensembles in the superficial layers formed a

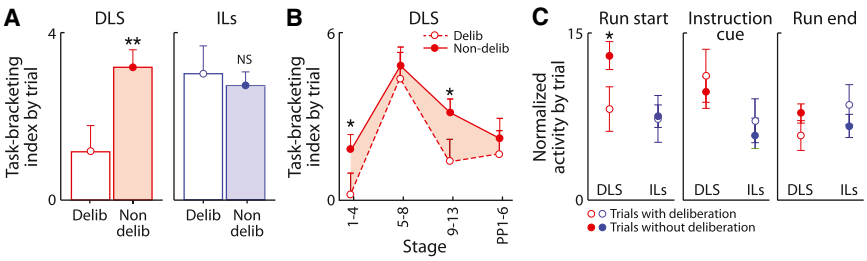

Figure 4. Trial-Level Modulation of DLS Spiking by Deliberations

(A) Task-bracketing index averaged over stages on trials with (empty bars) or without (solid bars) deliberation, for DLS (left) and ILs (right) units. $^{**}p < 0.01$.

(B) DLS task-bracketing index for trials with (dotted) and without (solid) deliberation across stage blocks. $^{*}p < 0.05$.

(C) Normalized activity (baseline-subtracted spiking) around run start, instruction cue, and run end for each trial type and site. $^{*}p < 0.05$.

Data are presented as mean ± SEM. See also Figure S5.

task-bracketing pattern (Figures 6, S1, and S2). We evaluated these superficial and deep ensembles across the cortical depth in small sliding spatial windows starting from the white matter and moving to more superficially situated levels, with the windows adjusted to include an average of at least five units per session (ca. 0.1 mm steps) (Figure S1). Ensembles sampled from tetrodes placed within about 0.5–0.6 mm of the midline exhibited a task-bracketing activity. As the samples shifted farther lateral (deeper, >0.6 mm), this pattern gave way during overtraining to one in which activity was pronounced through most of the run period.

Despite the strikingly different forms of ensemble patterning in the ILs and ILd, the changes in their activity patterns followed similar time courses. Both patterns emerged only during overtraining, and activity at both sites changed rapidly after devaluation (Figures 5, 6E, and 6F). ILd activity increased during the midrun decision period as accuracy increased, as opposite activity modulations occurred in the ILs (and in the DLS) (Figures 6C and 6D). Moreover, in the ILd, the panrun activity became suppressed during sessions after devaluation, just as the ILs activity increased (Figures 5 and 6). The activity in ILd did not change across postdevaluation days, remaining consistently as low as it had been during initial acquisition (Figure 5B and 6F). This activity did not correlate with deliberative behavior at either session or trial levels. These results demonstrate that ensembles sampled from superficial and deep depth levels of IL cortex exhibit highly contrasting patterns of activity during procedural learning, even though the time courses of their plasticity were similar.

Other parameters of activity that we assessed in the IL sites, as well as in the DLS, mostly did not change or changed only subtly across learning stages, including the magnitudes of spike activity averaged over the full run period, spiking variability, and the proportions of task-related units and single-event-related subpopulations (Figure S3). One exception was the selectivity of units to single task events (Figure S3H). The number of DLS and ILs units with selective responses to single events increased with training, perhaps contributing to more structured task representations (Barnes et al., 2005), whereas in the ILd, units became less selective.

Outcome, Goal Value, Goal Location, and Turn Direction Variables Do Not Account for Habit-Related Activity Patterns

For each recording site, we also assessed the activity of each unit in relation to other trial variables within sessions: correct versus incorrect runs, right versus left turn, right versus left goal location, and run outcome after devaluation (for runs to devalued goal, runs to nondevalued goal, or wrong-way runs). These variables did not appear to account for the changes in ensemble activity patterns that occurred across learning and habit expression (Figure S3). Even the average firing frequencies of subsets of units that responded differentially to turn direction (percent of turn-related units; DLS = 49%, ILs = 56%, ILd = 54%) or goal location (percent of goal-related units; DLS = 64%, ILs = 66%, ILd = 68%) were similar and were stable across learning stages. These findings suggest that changes in activity during training reflected the relative levels of purposeful as opposed to semiautomatic behavior, as indicated by the level of deliberative behavior expressed by the animals and their outcome sensitivity, rather than these particular performance parameters.

Double Devaluation Leads to Loss of the DLS Task-Bracketing Pattern

The strategy after devaluation of nearly always running to the nondevalued side suggested that the stable DLS pattern might reflect stability of running a familiar and valued route. To test this possibility, we asked whether the stable DLS pattern would be lost after a second devaluation procedure, which would render all outcomes aversive. In these double-devaluation conditions, the rats eventually learned to quit completing the maze runs, stopping at the instruction cue on over a quarter of the trials (Figure S5A). During the maze runs that were completed, the DLS ensemble activity no longer accentuated run start and end. Instead, activity was variably distributed throughout the run as the activity had been early in task learning (Figure S5B). This result suggests a correspondence between the DLS task-bracketing pattern and conditions under which thoroughly learned and valued runs are completed, but little correspondence with the specific outcome value of a given run.

Neuronal Activity in Prelimbic Cortex Declines during Habit Formation

To assess the selectivity of the IL response patterns, we recorded in the overlying prelimbic/cingulate (PL) cortex, a cortical region thought to promote flexibility and to oppose habit formation (Balleine and Dickinson, 1998; Killcross and Coutureau, 2003). Recordings were made during the overtraining period, the time during which the habits became stabilized and IL units developed task-bracketing or panrun patterns (n = 399 total

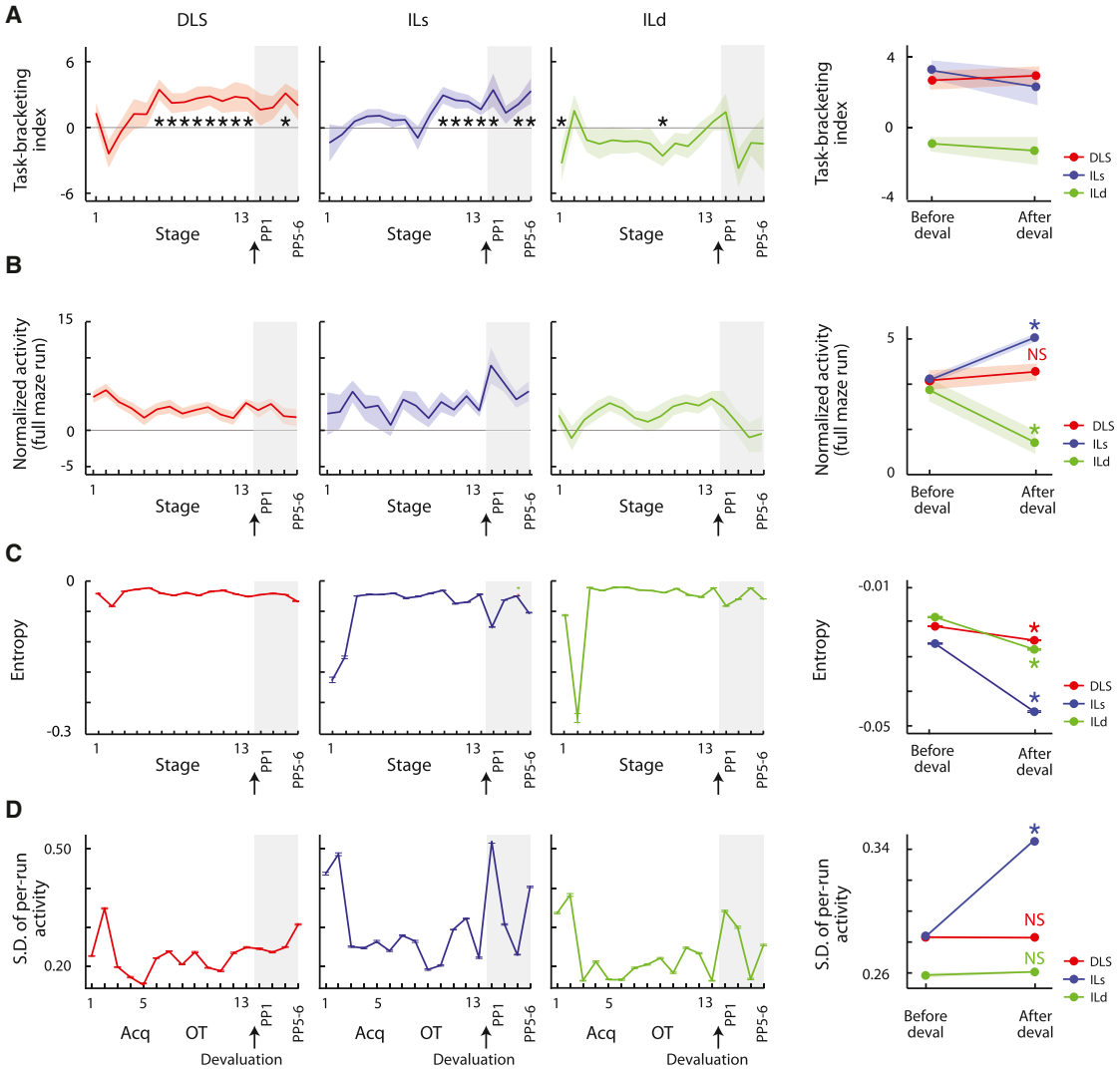

Figure 5. Fluctuations in Firing Strength and Variability Related to Learning and Devaluation

Task-bracketing index (A), baseline-subtracted raw firing during the full run (from start to goal, B), entropy of ensemble spike activity during the full run across trials within a session (SEM of 1,000 bootstrapped units, C), and SD of ensemble spike activity during full maze runs (SEM of 1,000 bootstrapped units, D), calculated for ensemble activity by recording site and training stage. Right: averages over five stages before and after devaluation. *$p < 0.05$ compared to no index (zero, left) or to before devaluation (right).

Data are presented as mean ± SEM. See also Figure S3.

and n = 184 task-related units). In contrast to activity in the adjoining IL cortex, ensemble activity in the PL cortex, both in superficial and deep depth levels, gradually declined from early to late overtraining as the runs grew outcome insensitive and habitual (Figure 7). We found no evidence for a task-bracketing ensemble pattern.

Online IL Perturbation during Overtraining Prevents Habit Formation

The fact that marked plasticity of ensemble plasticity appeared in both depth levels of IL only during the critical overtraining period in which habits became crystallized suggested an unexpected role of IL in the formation of habits, not only in their expression. To test this hypothesis, we perturbed the activity of IL cortex during this overtraining period to determine whether this might prevent the formation of the maze habit. We leveraged the high spatiotemporal resolution and repeatability of optical neuromodulation to disrupt IL activity just during the runs performed during overtraining (Figure 8A). Separate animals received bilateral IL injections of an eNpHR3.0 (halorhodopsin) viral construct (n = 6) or a control construct lacking the opsin gene (n = 4) and bilateral optical fibers aimed at IL cortex to permit light delivery. After training, rats received 10 days of overtraining during which 593.5 nm light was delivered on each trial from run start to goal arrival. This protocol results in time-locked perturbation of IL spiking over many repetitions (Smith et al.,

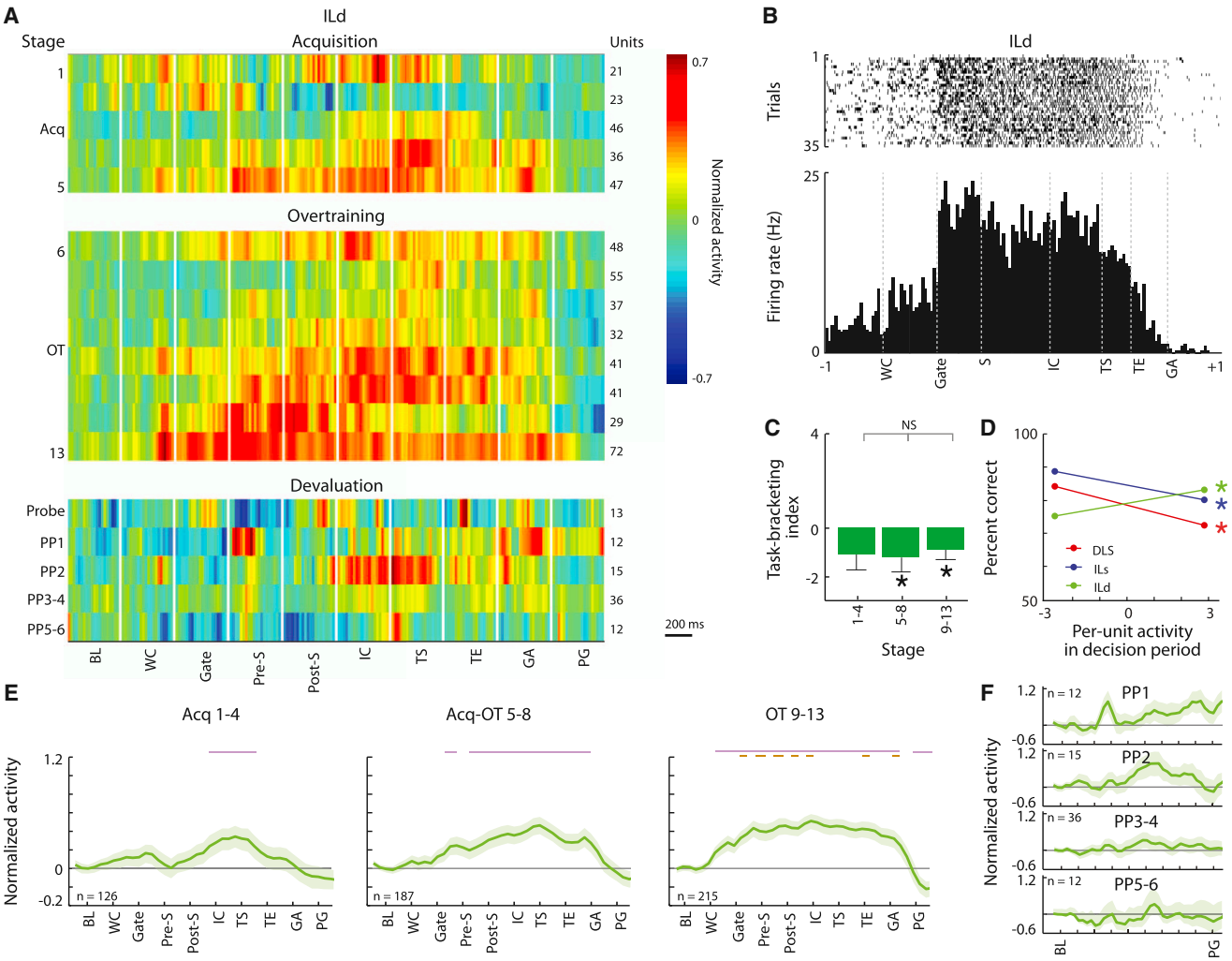

Figure 6. Distinct Pattern of Activity in ILd during Habit Learning

(A) Ensemble activity of ILd units in individual training stages, as in Figure 2C.

(B) Raster plot and histogram of single ILd unit activity during an overtraining session, as in Figure 2B.

(C) ILd task-bracketing activity index, as in Figure 2E. *$p < 0.05$.

(D) Opposite changes in decision-period activity in ILd compared to ILs and DLS. Regression line between normalized activity of each task-related unit during the decision period (from cue onset to turn start, 0 = baseline) and performance accuracy, for training stages 1–13. *$p < 0.05$.

(E and F) Normalized activity of ILd units across learning stages (E) and across PP stages (F), as in Figures 2D and 3D.

Data are presented as mean ± SEM. See also Figure S4.

2012) and did not affect running or accuracy during the perturbation time (Figure 8B). Then, without further IL illumination, the rats underwent reward devaluation, probe testing, and 2 PP test days to determine whether they had developed an outcome-insensitive habit. On the probe day, the control rats ran habitually to both devalued and nondevalued goals (Figures 8C and 8D), as had normal overtrained rats (Figure 1). By contrast, rats with IL perturbation did not exhibit a full habit: they avoided the devalued goal on ca. 50% of trials instructed there and ran accurately to the nondevalued goal (Figures 8C and 8D). Their behavior was thus similar to that of normal rats trained only up to the initial criterion for acquisition (Figure 1). On subsequent PP rewarded days, all rats learned to avoid the devalued goal with tasting experience (Figures 8C and 8D). Thus, targeted disruption of IL activity during the overtraining period selectively prevented habit acquisition.

DISCUSSION

Our findings demonstrate that both DLS-associated sensorimotor circuits and IL-associated limbic circuits register habits by heightened representations of action boundaries with diminished spike activity during decision-making periods. As the structure of these bracketing patterns increased with habit formation in both regions, variability in spike timing declined and single-event selectivity of individual units increased, suggesting a cross-circuit shift from neural exploration to exploitation as behavior became automatized into a habit (Barnes

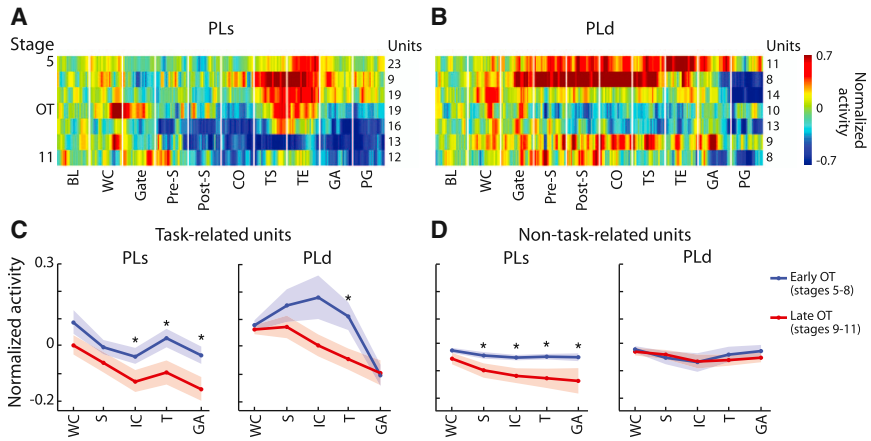

Figure 7. Activity of Prelimbic/Cingulate Neurons during Overtraining

(A and B) Normalized ensemble activity of PL units in superficial (PLs, A) and deep (PLd, B) layers from early to late overtraining (stages 5–11).

(C and D) Baseline-subtracted raw firing activity of task-related (C) and non-task-related (D) units, separated by early overtraining (blue, stages 5–8) and late overtraining (red, stages 9–11). Turn-related activity in superficial layers, and panrun activity in deep layers, declined as overtraining progressed. *$p < 0.05$.

Data are presented as mean ± SEM.

et al., 2005). Despite these similarities, the IL cortex and the DLS expressed spiking changes with strikingly different temporal dynamics during learning and with different relations to the behavioral parameters being acquired. Even within the IL cortex, different depth levels acquired different patterns. The perturbation of IL activity that we applied by optogenetic neuromodulation during overtraining established that IL activity during this habit crystallization period is necessary for full habit acquisition. We suggest an extension of current habit learning models to incorporate dynamic neural operators in both IL cortex and DLS. By this dual-operator account, habits are composites of multiple core neural components working simultaneously, and the mark of a fully formed habit could include the alignment of task-bracketing activity patterns in both limbic and sensorimotor circuits.

DLS and IL Cortex Dynamics: Dual Operators for Habit Control

In accord with experimental evidence, associative learning models have suggested that the brain has goal-directed, action-outcome (A-O) systems comprising model-based (e.g., tree-search) planning systems and that these compete for behavioral control with habit systems viewed as stimulus-response (S-R) or model-free systems (Balleine et al., 2009; Daw et al., 2005; Dickinson, 1985; Killcross and Coutureau, 2003). In these frameworks, the DLS is considered to represent the core S-R association or cached model-free predictions of a habit that can be acquired early and can control behavior when selected, whereas the IL cortex serves as an executive controller or arbiter favoring habit systems (Balleine et al., 2009; Daw et al., 2005; Dickinson, 1985; Killcross and Coutureau, 2003). The dynamics of neural activity that we observed are consistent with some predictions of these models, but there are also inconsistencies that encourage extensions of these views.

At a behavioral level, we found that deliberations did not covary perfectly with outcome value expectations. Nor did outcome insensitivity covary perfectly with the lack of deliberations. These observations suggest a distinction between goal directedness and deliberation scales for understanding an action sequence as a habit. At a mechanistic level, we found aspects of DLS activity that accord with it storing cached values, in that the task-bracketing activity formed early and was maintained across changes in outcome value as though ready to influence behavior whenever selected. However, surprisingly, DLS activity was most clearly related to the amount of deliberation rather than to other variables. Its task-bracketing activity not only remained fixed when values and behavior first changed after devaluation but even after new values had been incorporated into a putative second habit. The dominant task-bracketing ensemble spike activity pattern in the DLS might therefore not relate to specific S-R associations, which would probably have changed as the second habit overtook the first one. Some units might still retain such S-R associations but might be in the minority, in accord with observations in related work (Berke et al., 2009; de Wit et al., 2011; Root et al., 2010; Thorn et al., 2010). Our findings, instead, link the DLS bracketing pattern to the automatic execution of a familiar course of action, almost irrespective of actual outcome value or route-related details once the pattern is acquired. One interesting possibility is that this pattern represents a value bound to the learned behavior that has been bracketed, as though through the reinforcement history the behavior itself had grown to be an incentive (Glickman and Schiff, 1967). Other open alternatives include that the pattern reflected a stored S-R value of initially learned runs only, that S-R representations occurred in features of activity not assessed here, or that sensory stimuli in the maze environment guided behavior apart from instrumental processes despite the shift from outcome-sensitive to outcome-insensitive performance.

For the IL cortex, the close relationship between task-bracketing activity and the expression of outcome-insensitive behavior is consistent with its participation in an executive control process that selects habits. We found, however, that this relationship did not hold uniformly at the level of individual instances of execution of the behavior. If the IL cortex were an arbiter, it might be expected to "choose" the habitual or nonhabitual mode on any given trial (Wunderlich et al., 2012), but its activity did not suggest this. IL activity instead appeared to result in a general state permissive of habitual behaviors; it tracked, in general, the goal directedness of the behavior but not the detailed S-R type of behavior usually considered as a habit. These results suggest that IL activity could reflect a state function in promoting the emergence of habitual behavior, analogous to stressful states

Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

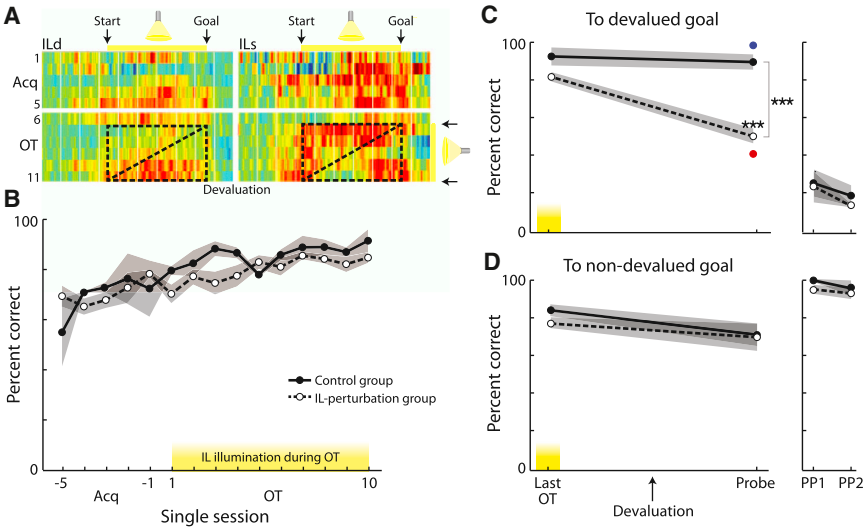

Figure 8. Optogenetic Perturbation of IL Cortex Blocks Habit Formation

(A) Light delivery, related to IL activity, from run start to stop, for 10 OT days only (stages 7–11). Box demarcates time of IL illumination.

(B) Performance accuracy during last five Acq sessions and ten OT sessions. No effect of light on performance: session ($F = 4.80$, $p < 0.001$; group, $F = 2.82$, $p = 0.10$; interaction, $F = 0.63$, $p = 0.84$).

(C) Correct turns to devalued goal on last OT day with IL light and on postdevaluation probe day without light. Group, $F = 44.80$, $p < 0.001$; session, $F = 21.12$, $p < 0.001$; interaction, $F = 14.44$, $p < 0.01$. Interaction of goal value and group on probe day: $F = 18.46$, $p < 0.001$. ***$p < 0.001$ post hoc. All other comparisons $p > 0.05$. Red dot, normal CT devaluation-sensitive behavior from Figure 1; blue dot, normal OT devaluation-insensitive behavior. Right: behavior during PP days.

(D) Correct runs to nondevalued goal, which did not change (group, $F = 0.52$, $p = 0.48$; session, $F = 3.51$, $p = 0.078$; interaction, $F = 0.23$, $p = 0.64$). Data are presented as mean ± SEM.

promoting the occurrence of repetitive behaviors without dictating the behavioral details (for example, cribbing versus pacing in horses).

The IL cortex is part of visceromotor/autonomic circuits that could influence behavior in this way, as similarly suggested by the involvement of IL cortex (or its presumed human homolog) in affective states (Holtzheimer and Mayberg, 2011; Quirk and Beer, 2006). Based on a reinforcement learning perspective, the IL cortex could categorize situation-action associations into discrete state-based habits (Redish et al., 2007; Sutton and Barto, 1998). Within IL, the task-bracketing pattern in the ILs supports a direct role for IL cortex in the crystallization or "chunking" of behavior (Graybiel, 1998), and the panrun pattern in ILd could relate to the tracking or invigoration of the full behavior that occurred during the critical overtraining phase. The results of our optogenetic experiments support this possibility: disrupting IL activity across depth levels during overtraining prevented the maze habit from forming. These findings suggest that the IL cortex participates in the actual formation of a habit, along with the DLS. The ebb and flow of the ILs task-bracketing pattern could potentially determine when limbic and sensorimotor circuits are aligned temporally to allow a learned habit to be fully expressed, thus providing habit "permission."

These findings suggest the working hypothesis that the DLS and the IL cortex conjointly influence, as dual operators, both the formation and the maintenance of habits. Habits, understood as devaluation-insensitive and nondeliberative behaviors, could have multiple core building blocks rather than involving a single component (e.g., an S-R association or set of associations). Such multicircuit modulation of habitual behavior is consistent with evidence that even simple reflexes underpinned by central pattern generators can be dynamically modulated (Graybiel, 2008; Marder, 2011). This conjunctive organization also raises the possibility that habits can be "incomplete" if composed of only some of several building blocks (as opposed to behaviors that oscillate between habitual and nonhabitual). Incomplete habits could have occurred in the experiments documented here when deliberations and outcome sensitivity did not go together, or when the ILs and DLS patterns were not both present.

IL Cortex as an Online Operator to Build and Permit Habitual Behavior

The IL cortex has been found to be important for maintaining new task strategies and conditioned responses, especially when they compete with alternate ones (Ghazizadeh et al., 2012; Peters et al., 2009; Rhodes and Killcross, 2004; Rich and Shapiro, 2009; Smith et al., 2012). Our findings help to characterize the activity of IL neurons in the context of organizing action sequences as habits. We demonstrate a close correspondence between ILs task-bracketing activity and the learning period at which behavior becomes automatic, but at the same time we failed to find such a close correspondence at the level of single trials as we found for the DLS. A session-wide inverse relationship between spiking activity and automatic running thus is an important and distinct feature of ILs activity. We emphasize that we recorded from only small numbers of IL units, and we used behavioral measures that only indirectly accessed underlying performance strategies; other features of IL activity that track behavior trial-to-trial, directly or through its interactions with other regions, may have been covertly present. It is nonetheless striking that a strong correlation did hold between the dominant IL ensemble activity pattern and habitual features of behavior measured at the level of sessions, which were at particular levels of learning and behavioral plasticity.

Notably, the times at which the task-bracketing activity pattern was observed in IL cortex were nearly identical to the times at which optogenetic IL perturbation (of all layers) could disrupt the maze habits: during overtraining, as shown here, as well as after overtraining and after postdevaluation training when a second habit had become established (Smith et al., 2012). These times, in turn, were highly correlated with the periods in which the numbers of deliberative head movements declined. Together, these results suggest that the task-bracketing pattern

Please cite this article in press as: Smith and Graybiel, A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics, Neuron (2013), http://dx.doi.org/10.1016/j.neuron.2013.05.038

in the IL cortex could reflect the training-related development of a potent and active IL influence over the sculpting of habits as well as an influence over their execution. The lack of trial-level correlation with behavior suggests a contribution to habits at the level of states that bias behavior toward outcome insensitivity (or low deliberation). This view might help account, for example, for the fact that the ILs bracketing pattern remained on PP day 1, when we had previously reported that IL perturbation does not affect behavior (Smith et al., 2012); the pattern, although present, was joined by marked increases in spiking variability and magnitude reflecting perhaps a mixed habit/non-habit state.

If the IL cortex were to have such a state-level influence, how would it interact with the DLS to promote habits, given that direct connections between them have not been detected? Potential indirect connectivity could include fiber projections via the ventral striatum or the amygdala and the substantia nigra or by way of projections to other cortical areas and then to the DLS (Hurley et al., 1991). However, as favored here, the IL cortex and the DLS might work partly in parallel, promoting habits through distinct circuit mechanisms, with the IL cortex providing, by way of its many limbic connections, routes by which it could disrupt flexibility and mnemonic processes or invigorate learned behavior.

Layer-Specific Patterning of Activity in IL Cortex Suggests Simultaneous Operation of Transcortical and Cortical-Subcortical Circuits

An unexpected finding of this study is that the task-bracketing pattern that did form in the IL cortex was evident only in the superficial layers. Superficial cortical layers are especially important for transcortical processing, and deeper layers for cortical projections to subcortical regions including the striatum (Anderson et al., 2010; Douglas and Martin, 2004). The activity in the ILd was reminiscent of that found in the dorsomedial striatum in previous maze experiments, in which midrun activity increased during habit learning but then faded as the fully acquired habit settles (Thorn et al., 2010). The IL cortex and dorsomedial striatum could interact through direct projections from IL cortex to parts of the medial striatum (Hurley et al., 1991). Fiber projections to the amygdala, thought to be related to suppression of conditioned responses, as well as to habits, could also be important (Lingawi and Balleine, 2012; Peters et al., 2009), as could projections to the nucleus accumbens, intralaminar thalamus, and other sites. The emergence of some habits might involve plasticity in layer-selective associative-limbic networks that occurs alongside established sensorimotor representations. From our findings, this plasticity occurs in the IL cortex and does not generalize to activity in the adjoining PL cortex; PL activity instead grew weak as the habit emerged. It would be of great interest to apply layer- and pathway-specific manipulations to these cortical regions.

DLS as an Operator Favoring Nondeliberative Behavior

In the DLS, the sharp accentuation of spike activity at action start and termination phases of behavior has been seen in prior studies on rodents, monkeys, and birds (Barnes et al., 2005; Fujii and Graybiel, 2003; Fujimoto et al., 2011; Jin and Costa, 2010; Jog et al., 1999; Kubota et al., 2009; Thorn et al., 2010). Here, by imposing a reward devaluation protocol, we could evaluate the relationship between this pattern of activity and levels of habitual performance. We confirmed that this DLS task-bracketing pattern is a function of learning stage, and we demonstrated that the pattern is independent of outcome value but sensitive to the automaticity of single maze runs as measured by deliberative head movements. These findings suggest a potential link between DLS task-bracketing activity and the antagonism of purposeful decision making that results in the sequencing together of reinforced actions for fluid expression (Balleine et al., 2009; Graybiel, 1998, 2008; Hikosaka and Isoda, 2010; Packard, 2009; Yin and Knowlton, 2006).

The early time course of DLS spiking plasticity could reflect a mechanism by which sensorimotor elements and action boundaries of a habit could be acquired and stored rapidly, while requiring additional processes for selection and translation into a fully habitual behavior (Balleine et al., 2009; Barnes et al., 2005; Coutureau and Killcross, 2003; Daw et al., 2005; Kimchi et al., 2009; Thorn et al., 2010). This theme resonates across the larger framework of action learning in the brain (Brainard and Doupe, 2002; Graybiel, 2008; Hikosaka and Isoda, 2010), in which studies have demonstrated latent learning of skilled behaviors in rodents and songbirds if basal ganglia regions for execution are blocked (Atallah et al., 2007; Charlesworth et al., 2012), as well as habit expression very early during learning when regions for behavioral flexibility are shut down (Killcross and Coutureau, 2003; Yin and Knowlton, 2006). The early plasticity and subsequent stability of DLS activity during automatic runs could reflect such early action learning.

It was only after the second devaluation procedure was imposed that the stability of the task-bracketing pattern was broken along with extinction of running. This finding is in accord with prior evidence that the DLS pattern, once formed, is insensitive to an instruction cue change requiring new learning (Kubota et al., 2009) but decays when reward is omitted altogether (Barnes et al., 2005). Under conditions of at least partial reinforcement, the acquired DLS pattern remains intact. It is within these conditions that well-learned behaviors can be maintained under some habitual control. Our findings suggest, however, that it is the balance of this sensorimotor striatal activity with value-sensitive limbic IL activity that may ultimately determine the extent of habitual performance. Such dynamics could, in disease or addictive states, provide a route by which behaviors become overly repetitive.

EXPERIMENTAL PROCEDURES

Rats (n = 22) were trained on a T-maze task requiring them to respond to auditory instruction cues by turning into maze end-arms to receive reward (chocolate milk or sucrose, each paired with a distinct cue). Training proceeded over daily sessions through task acquisition (72.5% accuracy for 2 days) and overtraining (10+ more days). For reward devaluation, rats received three pairings of home-cage intake with lithium chloride injection and were returned to the task for an unrewarded probe session and subsequent rewarded sessions. Task events were controlled by computer software (MED-PC or MATLAB). Behavior was monitored by in-maze photobeams and an overhead charge-coupled device camera recording at 30 Hz. Neuronal activity was recorded from 12–24 independently drivable tetrodes using a Cheetah

acquisition system (Neuralynx). Single units were isolated using Offline Sorter (Plexon) and, for DLS recordings, sorted into neuronal subtypes. Task-related spike activity exceeded 2 SD above a baseline period for three 30 ms bins within ±200 ms of a task event. Analysis were conducted on behavior- and learning-related changes in task-related population sizes, spike magnitude, spiking variability, and task-bracketing activity scores (spiking around the cue period subtracted from mean spiking around run start and run stop). Optogenetic perturbation during 10 overtraining days, from run start to stop, was accomplished using bilateral IL injection of AAV5-CaMKIIα-eNpHR3.0-EYFP (halorhodopsin) or AAV5-CaMKIIα-EYFP (control), duel-ferrule fiber implants (Doric Lenses), laser light (2.5–4 mW/side; 593.5 nm; OEM Laser Systems), and a pulse generator (AMPI). ANOVA, linear regression, and neuronal spike distribution statistics assessed behavioral and neuronal activity changes, with significance set at $p < 0.05$. Immunostaining and Nissl-staining procedures were used to label tetrode and fiber tracks, and neurons expressing EYFP. See also Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2013.05.038.

ACKNOWLEDGMENTS

We thank Christine Keller-McGandy, Alex McWhinnie, Dr. Daniel J. Gibson, and Henry F. Hall; Dr. Marshall Shuler, Dr. Catherine Thorn and Dr. Yasuo Kubota; and Karen Sittig, Arti Virkud, and Dordaneh Sugano for their help and advice. This work was supported by NIH grants R01 MH060379 (A.M.G.) and F32 MH085454 (K.S.S.), by Office of Naval Research grant N00014-04-1-0208 (A.M.G.), by the Stanley H. and Sheila G. Sydney Fund (A.M.G.), and by funding from Mr. R. Pourian and Julia Madadi (A.M.G.).

Accepted: May 30, 2013
Published: June 27, 2013

REFERENCES

Adams, C.D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. Q. J. Exp. Psychol. B *34*, 77–98.

Aldridge, J.W., Berridge, K.C., and Rosen, A.R. (2004). Basal ganglia neural mechanisms of natural movement sequences. Can. J. Physiol. Pharmacol. *82*, 732–739.

Anderson, C.T., Sheets, P.L., Kiritani, T., and Shepherd, G.M. (2010). Sublayer-specific microcircuits of corticospinal and corticostriatal neurons in motor cortex. Nat. Neurosci. *13*, 739–744.

Aston-Jones, G., and Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu. Rev. Neurosci. *28*, 403–450.

Atallah, H.E., Lopez-Paniagua, D., Rudy, J.W., and O'Reilly, R.C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nat. Neurosci. *10*, 126–131.

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

Balleine, B.W., Liljeholm, M., and Ostlund, S.B. (2009). The integrative function of the basal ganglia in instrumental conditioning. Behav. Brain Res. *199*, 43–52.

Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158–1161.

Berke, J.D., Breck, J.T., and Eichenbaum, H. (2009). Striatal versus hippocampal representations during win-stay maze performance. J. Neurophysiol. *101*, 1575–1587.

Brainard, M.S., and Doupe, A.J. (2002). What songbirds teach us about learning. Nature *417*, 351–358.

Carelli, R.M., Wolske, M., and West, M.O. (1997). Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. J. Neurosci. *17*, 1804–1814.

Charlesworth, J.D., Warren, T.L., and Brainard, M.S. (2012). Covert skill learning in a cortical-basal ganglia circuit. Nature *486*, 251–255.

Coutureau, E., and Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav. Brain Res. *146*, 167–174.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Actions, policies, values, and the basal ganglia. In Recent Breakthroughs in Basal Ganglia Research, E. Bezard, ed. (Hauppauge: Nova Science Publishers), pp. 91–106.

de Wit, S., Barker, R.A., Dickinson, A.D., and Cools, R. (2011). Habitual versus goal-directed action control in Parkinson disease. J. Cogn. Neurosci. *23*, 1218–1229.

Dickinson, A. (1985). Actions and habits: the development of behavioral autonomy. Philos. Trans. R. Soc. Lond. B Biol. Sci. *308*, 67–78.

Douglas, R.J., and Martin, K.A. (2004). Neuronal circuits of the neocortex. Annu. Rev. Neurosci. *27*, 419–451.

Everitt, B.J., and Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat. Neurosci. *8*, 1481–1489.

Fujii, N., and Graybiel, A.M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. Science *301*, 1246–1249.

Fujimoto, H., Hasegawa, T., and Watanabe, D. (2011). Neural coding of syntactic structure in learned vocalizations in the songbird. J. Neurosci. *31*, 10023–10033.

Garcia, J., and Ervin, F.R. (1968). Appetites, aversions, and addictions: a model for visceral memory. Recent Adv. Biol. Psychiatry *10*, 284–293.

Ghazizadeh, A., Ambroggi, F., Odean, N., and Fields, H.L. (2012). Prefrontal cortex mediates extinction of responding by two distinct neural mechanisms in accumbens shell. J. Neurosci. *32*, 726–737.

Glickman, S.E., and Schiff, B.B. (1967). A biological theory of reinforcement. Psychol. Rev. *74*, 81–109.

Graybiel, A.M. (1998). The basal ganglia and chunking of action repertoires. Neurobiol. Learn. Mem. *70*, 119–136.

Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. Annu. Rev. Neurosci. *31*, 359–387.

Henze, D.A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K.D., and Buzsáki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. J. Neurophysiol. *84*, 390–400.

Hikosaka, O., and Isoda, M. (2010). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. Trends Cogn. Sci. *14*, 154–161.

Hitchcott, P.K., Quinn, J.J., and Taylor, J.R. (2007). Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. Cereb. Cortex *17*, 2820–2827.

Holland, P.C., and Straub, J.J. (1979). Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. J. Exp. Psychol. Anim. Behav. Process. *5*, 65–78.

Holtzheimer, P.E., and Mayberg, H.S. (2011). Deep brain stimulation for psychiatric disorders. Annu. Rev. Neurosci. *34*, 289–307.

Hurley, K.M., Herbert, H., Moga, M.M., and Saper, C.B. (1991). Efferent projections of the infralimbic cortex of the rat. J. Comp. Neurol. *308*, 249–276.

Hyman, S.E., Malenka, R.C., and Nestler, E.J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. Annu. Rev. Neurosci. *29*, 565–598.

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

Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V., and Graybiel, A.M. (1999). Building neural representations of habits. Science *286*, 1745–1749.

Kalivas, P.W., and Volkow, N.D. (2005). The neural basis of addiction: a pathology of motivation and choice. Am. J. Psychiatry *162*, 1403–1413.

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb. Cortex *13*, 400–408.

Kimchi, E.Y., Torregrossa, M.M., Taylor, J.R., and Laubach, M. (2009). Neuronal correlates of instrumental learning in the dorsal striatum. J. Neurophysiol. *102*, 475–489.

Kubota, Y., Liu, J., Hu, D., DeCoteau, W.E., Eden, U.T., Smith, A.C., and Graybiel, A.M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. J. Neurophysiol. *102*, 2142–2160.

Lingawi, N.W., and Balleine, B.W. (2012). Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. J. Neurosci. *32*, 1073–1081.

Marder, E. (2011). Variability, compensation, and modulation in neurons and circuits. Proc. Natl. Acad. Sci. USA *108*(*Suppl 3*), 15542–15548.

McGeorge, A.J., and Faull, R.L. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience *29*, 503–537.

Muenzinger, K.F. (1938). Vicarious trial and error at a point of choice: I. A general survey of its relation to learning efficiency. J. Genet. Psychol. *53*, 75–86.

Packard, M.G. (2009). Exhumed from thought: basal ganglia and response learning in the plus-maze. Behav. Brain Res. *199*, 24–31.

Peters, J., Kalivas, P.W., and Quirk, G.J. (2009). Extinction circuits for fear and addiction overlap in prefrontal cortex. Learn. Mem. *16*, 279–288.

Quirk, G.J., and Beer, J.S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. Curr. Opin. Neurobiol. *16*, 723–727.

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

Redish, A.D., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process. Behav. Brain Sci. *31*, 415–437, discussion 437–487.

Rhodes, S.E., and Killcross, S. (2004). Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. Learn. Mem. *11*, 611–616.

Rich, E.L., and Shapiro, M. (2009). Rat prefrontal cortical neurons selectively code strategy switches. J. Neurosci. *29*, 7208–7219.

Root, D.H., Tang, C.C., Ma, S., Pawlak, A.P., and West, M.O. (2010). Absence of cue-evoked firing in rat dorsolateral striatum neurons. Behav. Brain Res. *211*, 23–32.

Smith, K.S., Virkud, A., Deisseroth, K., and Graybiel, A.M. (2012). Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. Proc. Natl. Acad. Sci. USA *109*, 18932–18937.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction (Cambridge: MIT Press).

Tang, C., Pawlak, A.P., Prokopenko, V., and West, M.O. (2007). Changes in activity of the striatum during formation of a motor habit. Eur. J. Neurosci. *25*, 1212–1227.

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

Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189–208.

Tricomi, E., Balleine, B.W., and O'Doherty, J.P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. Eur. J. Neurosci. *29*, 2225–2232.

Wunderlich, K., Dayan, P., and Dolan, R.J. (2012). Mapping value based planning and extensively trained choice in the human brain. Nat. Neurosci. *15*, 786–791.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. *7*, 464–476.

Neuron, Volume *79*

Supplemental Information

A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics

Kyle S. Smith and Ann M. Graybiel

SUPPLEMENTAL DATA

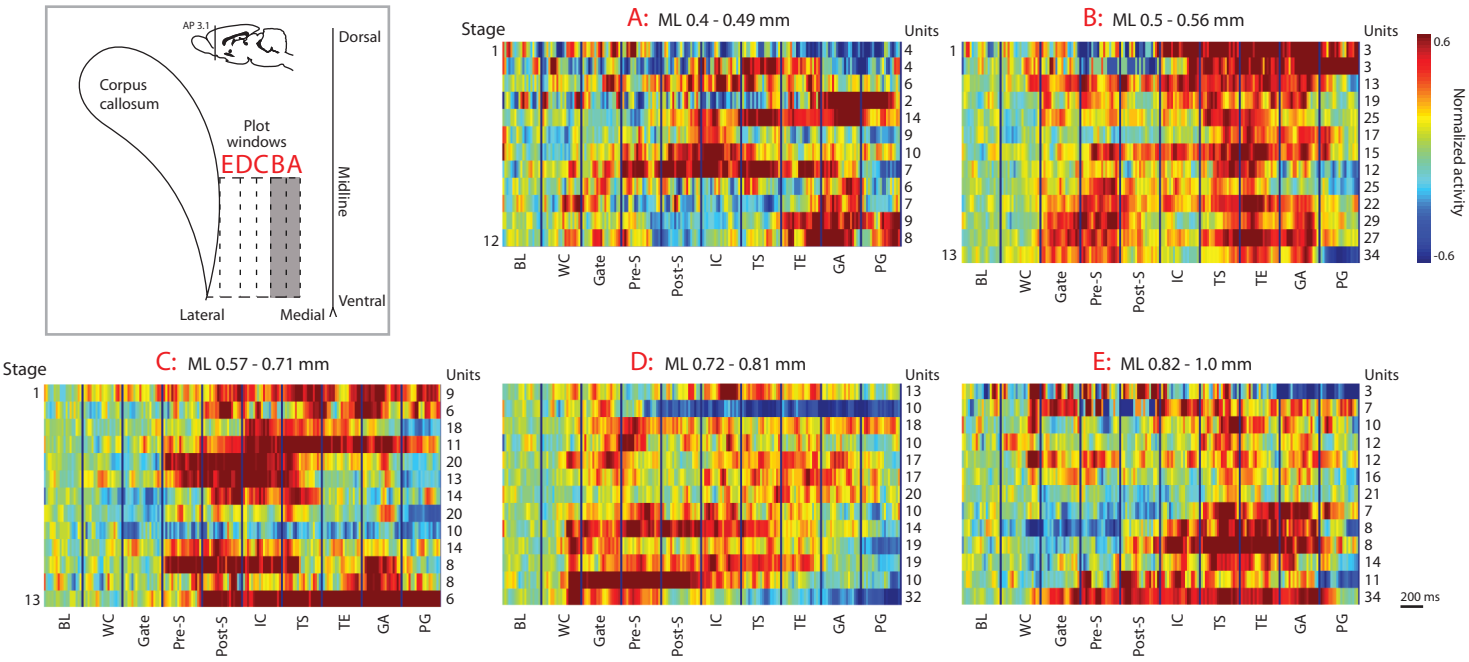

Figure S1. Fine-Scale Division of Medial-Lateral Ensembles of IL Activity, Related to Figure 2

Normalized activity (baseline subtracted Z-score) of IL ensembles in moving windows, as depicted schematically in upper left (from A to E: ensembles recorded more superficial to deeper tetrode placements). Gray shading in upper left denotes the regions in which we recorded ensemble activity designated as “superficial” for analysis. Plots constructed as in Figure 2C.

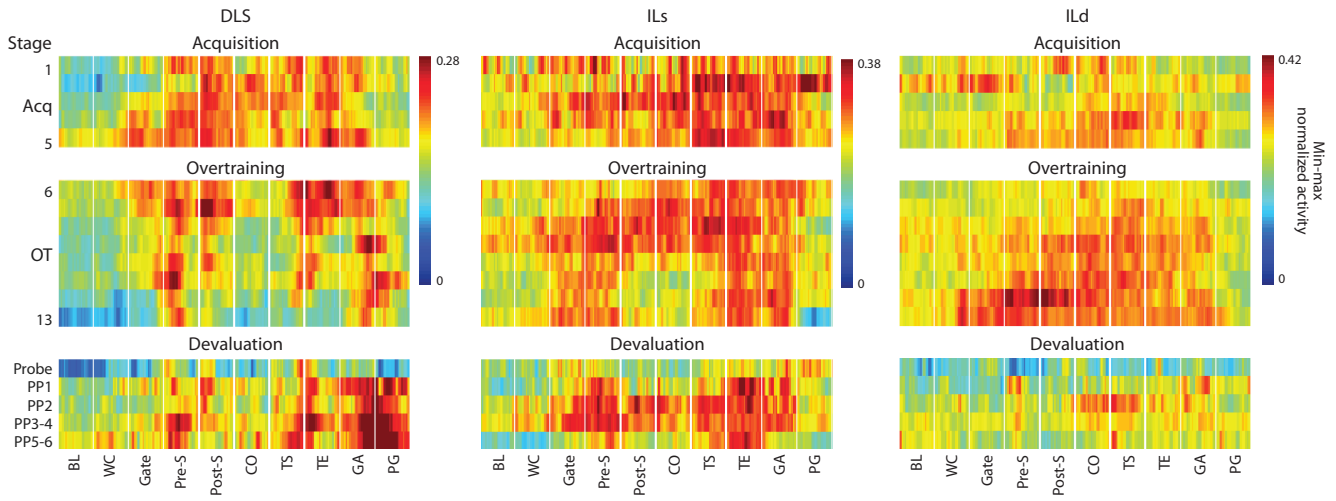

Figure S2. DLS, ILs and ILd Ensemble Activity Following Min-Max Normalization, Related to Figure 3

Min-max normalization of activity as an alternative to normalization of Z-scores relative to pre-cue baseline, for DLS (left), ILs (middle), and ILd (right). Plots constructed as in Figure 2C.

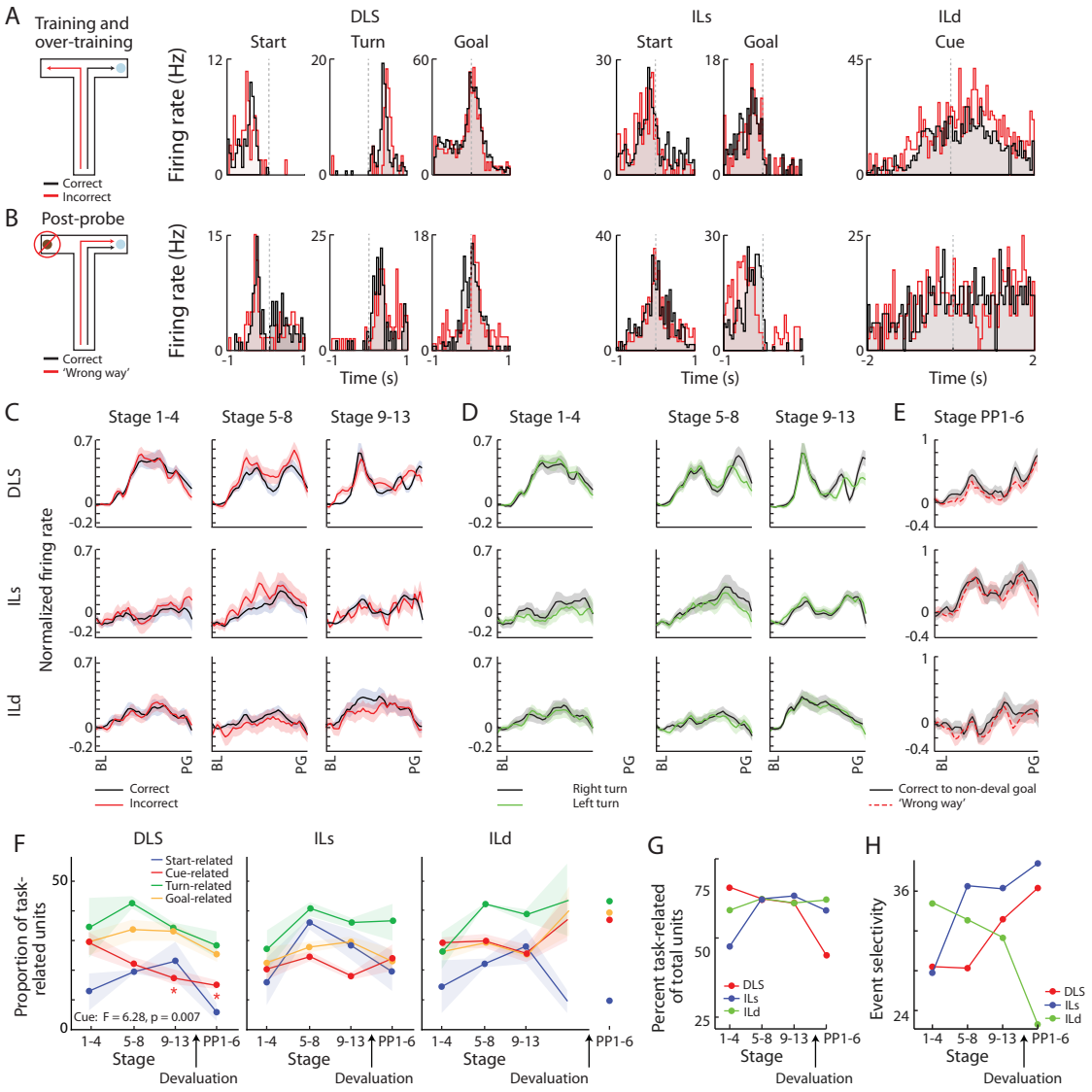

Figure S3. Similarity of Single Unit and Ensemble Activity across Behavioral Variables, Related to Figure 5

(A and B) Spike histograms showing similar task-related activity of sample single units for correct (black) and incorrect (red) trials during training (A), and for correct (black) and wrong-way (red) runs to the non-devalued goal during PP sessions (B).

(C) Normalized firing rate (baseline-subtracted Z-score, ±200 ms around each event) during maze runs on runs performed correctly (black) or incorrectly (red). No main comparisons or interactions with learning significant. Plotting as in Figure 2D.

(D) Normalized firing on runs to right (black) versus left (green) goal-arms. No significant effects of performance type or interaction. Though some units were found in each site that responded preferentially to turn direction, the averaged spike activity for each turn was comparable across learning.

(E) Normalized firing averaged across post-devaluation sessions for correct (black) and 'wrong-way' (red dashed) runs to non-devalued goal. No significant difference found if post-devaluation days were grouped as here or analyzed individually.

(F) Percent, out of total recorded units, of start-related (blue), cue-related (red), turn-related (green), and goal arrival-related (orange) units across sites and stages (stages 1-4, 5-8, 9-13, PP1-6). All main effects and interactions were not significant, except for cue activity of DLS (*$p < 0.05$ post-hoc comparison to stage 1-4).
(G) Percentage of total recorded units with a significant task response across learning phases.
(H) Percentage of recorded units with significant responses to only one task event. Higher percentage denotes more units with single event responses, while lower percentage denotes more units with multiple event responses.

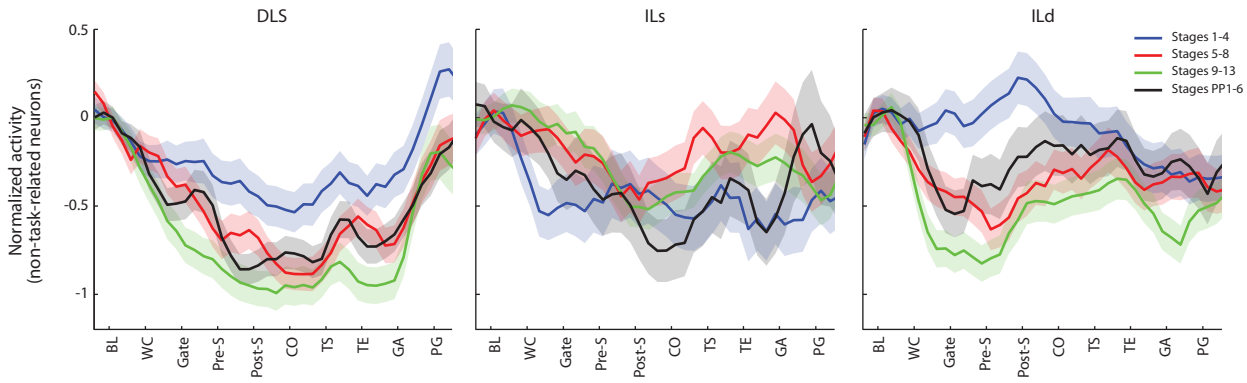

Figure S4. Activity of Non-Task-Related Units, Related to Figure 6
Normalized activity (baseline-subtracted Z-score) of units lacking phasic responses to task events, separated by training phases for each site.

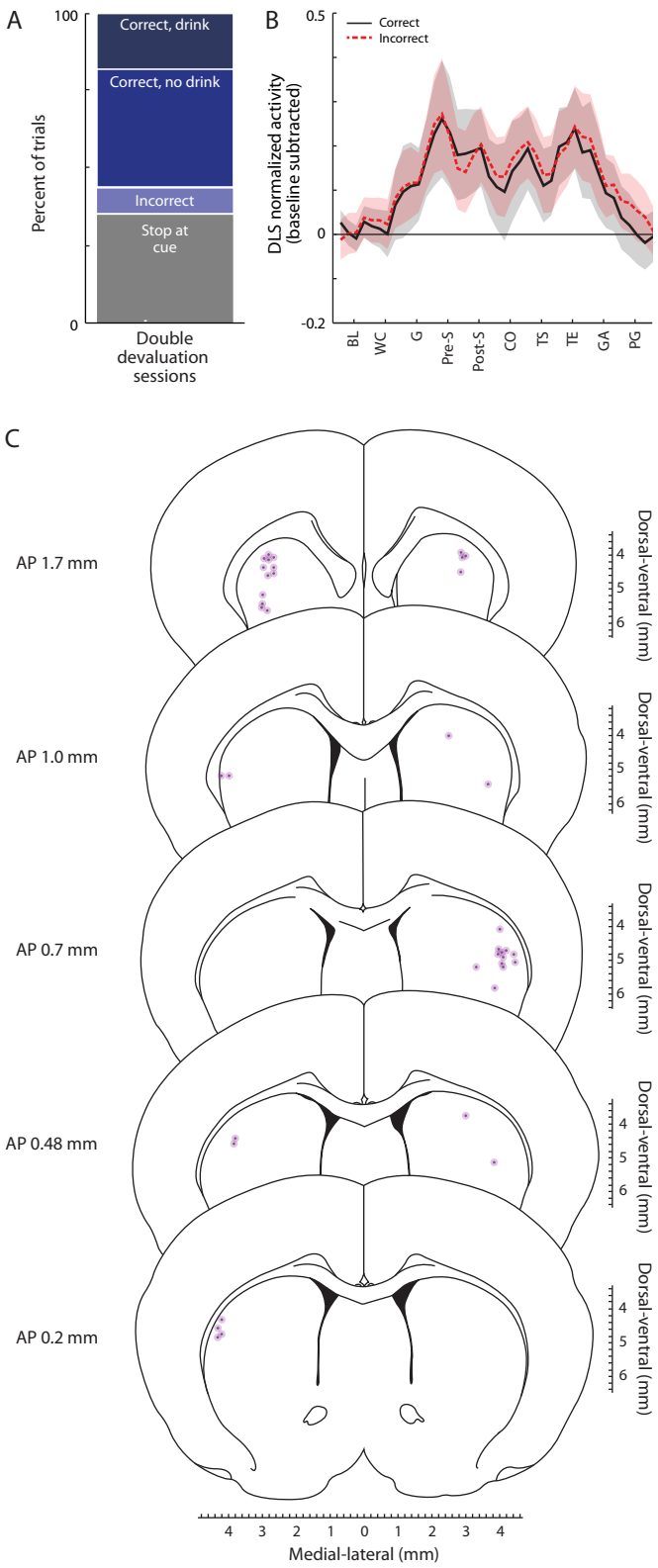
A
100
0
Percent of trials
Correct, drink
Correct, no drink
Incorrect
Stop at cue
Double devaluation sessions
B
0.5
0
-0.2
DLS normalized activity (baseline subtracted)
Correct
Incorrect
BL
WC
G
Pre-S
Post-S
CO
TS
TE
GA
PG
C
AP 1.7 mm
AP 1.0 mm
AP 0.7 mm
AP 0.48 mm
AP 0.2 mm
Dorsal-ventral (mm)
4
5
6
4 3 2 1 0 1 2 3 4
Medial-lateral (mm)

Figure S5. Performance and DLS Ensemble Activity after Devaluation of the Second Reward, Related to Figure 4

(A) Breakdown of performance in the post-double-devaluation test sessions by the percent of trials in which the rats performed correctly and consumed reward, performed correctly and did not consume reward, performed incorrectly by entering the wrong end-arm, or stopped running at the cue (rats always initiated running from the start).

(B) Activity of DLS units during test sessions after this double-devaluation, only for trials in which the task run was completed (baseline-subtracted raw rate in ±200 ms around each event). In this period, the task-bracketing pattern decayed and activity instead resembled that observed during initial learning (see Figure 2D, red). Colors separate correct (black) vs. incorrect (red) runs. DLS activity was again independent of these performance or outcome details.

(C) Cartoon of DLS recording sites separated across multiple anterior-posterior levels, plotted as in Figure 2A.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Subjects and Surgery

Individually housed male Sprague-Dawley rats (n = 22) maintained on a reverse light-dark cycle and within 85% of pre-surgical weight were run in experiments during their dark (active) cycle, with procedures approved by the M.I.T. Committee on Animal Care. For electrophysiology, headstages carrying 12-24 independently movable tetrode drives (Neuralynx, Bozeman, MT) were implanted as described (Barnes et al., 2005; Barnes et al., 2011; Kubota et al., 2009; Thorn et al., 2010). Three of the criterion-trained rats were implanted with non-functional headstages. Tetrodes were placed in the ILs (range: AP 2.3-3.9 mm; ML 0.4-0.55 mm; DV 4.5-5.4 mm from skull), ILd (AP 2.3-3.9 mm; ML 0.60-1.0 mm; DV 4.2-6.0 mm), DLS (AP −0.1-2.0 mm; ML 2.5-4.6 mm; DV 3.8-5.8 mm), and PL (AP 2.3-3.9 mm; ML 0.4-0.7 mm; DV 2.8-4.0 mm). For optogenetics, separate animals were given bilateral injections of a halorhodopsin virus construct (n = 6; AAV5-CaMKIIα-eNpHR3.0-EYFP) or control construct (n = 4; AAV5-CaMKIIα-EYFP), at 10-20 min per 0.2-0.5 µl injection. Injections targeted IL at AP 3.1 mm, ML ±0.6 mm, and DV -5.2 mm. Implanted bilateral dual-ferrule optical fibers (200 µm; Doric Lenses) terminated at DV -5.0 mm. Fibers were shielded with a modified centrifuge tube.

T-maze Apparatus and Training

Two mazes were used for the experiment, one identical and the other comparable to one described previously (Barnes et al., 2005; Barnes et al., 2011; Thorn et al., 2010). Reward was manually delivered via tubing to troughs at the end-arm goal sites. Rats were habituated to the maze and rewards (30% sucrose solution and chocolate flavored whole milk) over several days of free exposure. Training then proceeded in daily ca. 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 an instruction cue (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 for

correct performance (ca. 1 min inter-trial interval). Each reward was assigned to only one arm per rat; tone and reward assignments were pseudorandom across rats. Training continued through acquisition (72.5% accuracy criterion, X^2, $p < 0.01$ compared to chance). Criterion-trained rats (CT group) received one further criterion-session to confirm learning. Over-trained rats (OT group; IL silencing group; control group) ran 10+ additional sessions at or above criterion.

Reward Devaluation

Rats were given 45 min access to one maze reward (e.g., chocolate milk) in their home-cage followed by an injection of lithium chloride (0.6 M 5 ml/kg or 0.3 M 10 ml/kg, i.p.) to induce nausea. Three devaluation procedures at 48 hr intervals were given in multiple laboratory rooms, though never in the maze room, and efficacy was confirmed by reduced home-cage intake. Devalued reward identity was pseudorandomly assigned across rats. Rats were then given a probe session without rewards given, followed by normal post-probe 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 and neural plasticity occurring after encounter with the devalued reward in the maze task. A subset of rats (n = 5) later underwent another identical procedure to devalue the second reward (double-devaluation).

Session Staging

Training sessions were staged: Stages 1-2 (first two sessions), Stage 3-4 (pairs of sessions ≥ 60% correct), Stage 5 (first pair of sessions ≥ 72.5%), Stages 6-13 (subsequent pairs of sessions ≥ 72.5%). Stages after devaluations were: Probe (unrewarded probe session), Stage PP1-2 (first two post-probe rewarded sessions), Stage PP3-6 (subsequent pairs of rewarded sessions). Analysis stopped when < 3 rats or < 5 units were contributing data.

Electrophysiological Data Acquisition

Tetrodes lowered to recording targets over 7 post-surgical days were left in place or moved in <0.04 mm steps. Electrical signals were amplified at 100-10000, sampled at 32 kHz, band-pass filtered for 600-6000 Hz, and recorded by a Cheetah data acquisition system (Neuralynx) as described (Barnes et al., 2005; Barnes et al., 2011; Kubota et al., 2009; Thorn et al., 2010). An overhead CCD camera tracked LEDs on the head-stage preamplifiers (30 Hz sampling rate), and photobeams were placed every ca. 17.5 cm on one maze. Task-control was provided by a MED-PC program (Med Associates, Inc., St. Albans, VT) or MATLAB (Mathworks, Natick, MA).

Unit Sorting and Classification

Single units were identified as isolated waveform clusters using Offline Sorter (Plexon, Inc., Dallas, TX). Striatal units classified as putative medium-spiny neurons (Barnes et al., 2005; Barnes et al., 2011; Berke et al., 2004; Kubota et al., 2009; Schmitzer-Torbert and Redish, 2004; Thorn et al., 2010) were analyzed. Cortical units were classified by cortical (mediolateral) depth. Units were assigned as task-related or non-task-related units based on presence or absence of response to task events (Barnes et al., 2005; Barnes et al., 2011; Kubota et al., 2009; Thorn et al., 2010). Units were designated as task-responsive if spiking exceeded 2 s.d. above a baseline period for three 30 ms bins within ±200 ms of a task event. Post-goal activity was measured during 0.2-2.7 sec after goal arrival.

Optogenetic intervention

Animals were trained until reaching the criterion accuracy for 2 days. 593.5-nm light was then delivered bilaterally to the IL through the implanted fibers on each of 10 days of over-training (stages 7-11) using a laser source (OEM Laser Systems), fiber patch cords, a rotary joint, and a beam splitter (Thorlabs, Newton, NJ; Doric Lenses, Quebec City, Canada). Light delivery was

gated by a Master 8 pulse generator (A.M.P.I.), and power output ranged from 2.5-4.0 mW per hemisphere. Light was delivered from just after gate opening, as the run started, to goal arrival (ca. 2 sec). We have shown that this relatively moderate protocol of illumination of eNpHR3.0-expressing IL neurons results in repeatable, time-locked perturbation of spike activity, with robust light effects on behavior detectible over at least six 40-trial days of illumination (Smith et al., 2012). After over-training, animals underwent devaluation of one maze reward, a probe unrewarded test, and two post-probe rewarded sessions, all without any additional IL illumination.

Analysis

Performance accuracy (percent correct, incorrect and incomplete trials), run speed, deliberative head movements, and reward consumption were analyzed by ANOVA ($p < 0.05$) to compare across learning stages, trial subtypes (e.g., devalued and non-devalued trials) and rat groups (e.g., OT and CT rats; IL-halo and control rats). Tukey-corrected post hoc comparisons were made when significance was obtained for the main effect of variables and/or interaction between variables. Deliberative movements were identified through visual inspection of video-tracker data from the head-mounted LEDs. To count as a deliberation, movement had to slow at the turn, divert toward one end arm, and then proceed down the other arm to the goal location. Video-tracker quality was sufficient to analyze deliberations in 170 sessions (81% of total) from 6 of the 7 over-trained rats.

Per-unit firing was usually normalized by a baseline-subtracted Z-score computation (Thorn et al., 2010). For subpopulations of units (e.g., task-responsive DLS units), a mean and SEM of these normalized Z-scores was calculated for each session or stage and smoothed with a 3-point averaging filter. We also assessed activity using baseline-subtracted raw firing rate (e.g., for comparing conditions with unequal trial numbers), and min-max normalized activity (Kubota et al., 2009). ANOVA was used to detect significant firing changes in 20 ms or 100 ms time-bins

±200 ms around task events compared to pre-trial baseline within sessions, or to activity in the same time-bins during task acquisition. The strength of patterned activity was measured by a proportion index for each unit: [(activity from gate opening to post-start and from turn start to goal arrival) / 2] – [activity around the instruction cue]. We also computed single-unit regressions between event-related activity of units and behaviors occurring during the session in which each unit was recorded. Activity before vs. after devaluation was compared for overall firing rates across task events and the squared mean firing difference across each task event across the 5 sessions before and after devaluation for each rat. Variability was assessed by comparing standard deviations or entropy of firing across task events using 1,000 bootstraps from the neuronal population (Thorn et al., 2010).

Histology

At the end of the recording experiment, small electrical lesions (25 μA, 10 sec) were made at tetrode tips under anesthesia (sodium pentobarbital, 40-50 mg/kg, i.p.). For histological assessments, rats were anesthetized with a lethal dose of sodium pentobarbital (100-145 mg/kg) and transcardially perfused with 0.9% saline followed by 4% paraformaldehyde in 0.1M KNaPO4 buffer. Brains were post-fixed in paraformaldehyde followed by cryoprotectant solution (1:3 glycerol in 0.1 M phosphate buffer with sodium-azide), and sectioned at 30 μm. Sections for tetrode localization were stained for combinations of Nissl substance with cresyletcht violet. Sections for virus and optical fiber localization were stained with cresyletcht violet and GFP antibodies to label EYFP.

SUPPLEMENTAL REFERENCES

Barnes, T.D., Mao, J.B., Hu, D., Kubota, Y., Dreyer, A.A., Stamoulis, C., Brown, E.N., and Graybiel, A.M. (2011). Advance cueing produces enhanced action-boundary patterns of spike activity in the sensorimotor striatum. J Neurophysiol *105*, 1861-1878.

Berke, J.D., Okatan, M., Skurski, J., and Eichenbaum, H.B. (2004). Oscillatory entrainment of striatal neurons in freely moving rats. Neuron *43*, 883-896.

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