

LETTERS

Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories

Terra D. Barnes^{1*}, Yasuo Kubota^{1*}, Dan Hu¹, Dezhe Z. Jin^{1,2} & Ann M. Graybiel¹

Learning to perform a behavioural procedure as a well-ingrained habit requires extensive repetition of the behavioural sequence, and learning not to perform such behaviours is notoriously difficult. Yet regaining a habit can occur quickly, with even one or a few exposures to cues previously triggering the behaviour^{1–3}. To identify neural mechanisms that might underlie such learning dynamics, we made long-term recordings from multiple neurons in the sensorimotor striatum, a basal ganglia structure implicated in habit formation^{4–8}, in rats successively trained on a reward-based procedural task, given extinction training and then given reacquisition training. The spike activity of striatal output neurons, nodal points in cortico-basal ganglia circuits, changed markedly across multiple dimensions during each of these phases of learning. First, new patterns of task-related ensemble firing successively formed, reversed and then re-emerged. Second, task-irrelevant firing was suppressed, then rebounded, and then was suppressed again. These changing spike activity patterns were highly correlated with changes in behavioural performance. We propose that these changes in task representation in cortico-basal ganglia circuits represent neural equivalents of the explore–exploit behaviour characteristic of habit learning.

The ability to establish habits, procedures and stereotyped behaviours brings great biological advantages to active organisms, and much evidence indicates that cortico-basal ganglia loops are critical for such learning^{4–10}. If this view were correct, changes in the activity of basal ganglia neurons should accompany changes in behaviour not only as habits and procedures are initially acquired, but also as they are changed in response to altered behavioural contexts. To test for such restructuring of basal ganglia activity, we recorded chronically with multiple tetrodes for up to 63 sessions from the sensorimotor striatum of rats undergoing consecutive acquisition, over-training, extinction and reacquisition training on a conditional T-maze task (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). The rats navigated the T-maze and turned right or left in response to auditory cues indicating whether a chocolate reward was at the left or right choice-arm of the maze (Fig. 1c). This task requires trial-and-error learning, in which initial ‘exploration’ of the environment over successive trials leads, with successful learning, to ‘exploitation’, in which correct choices are made consistently¹¹. Performance accuracy increased during acquisition and was at or near asymptote during over-training (Fig. 1d). Accuracy then steadily deteriorated during extinction training, when reward was reduced ($n = 4$) or withheld entirely ($n = 3$), but recovered rapidly during retraining after extinction. Running times similarly fell, rose and fell (Fig. 1e, f).

As these behavioural changes occurred, the spiking of striatal neurons became redistributed across task time (Fig. 2, Supplementary Figs 2 and 3). We focused on the spike activity of neurons classified as striatal projection neurons, which directly participate in

cortico-basal ganglia loop processing¹² (Fig. 1a, Supplementary Fig. 1 and Supplementary Methods). At the start of acquisition training, the spike responses of the task-responsive projection neurons, as a group, occurred throughout the maze runs (Fig. 2a). By the time that the learning criterion had been met, however, the strongest per-unit firing occurred near the start and end of the runs. This progressive concentration of spike activity continued during the over-training period, even though behavioural performance had reached near-asymptotic values. In addition, early activity advanced from the time of locomotion onset towards the waiting period after the warning cue, and late activity shifted from around goal-reaching to around the end of turning (Fig. 2a, c, and Supplementary Figs 2 and 3).

These acquired spiking patterns were largely reversed during the extinction period (Fig. 2a). Mid-trial firing increased, and the temporal shifts, particularly for the early activity, reversed. When

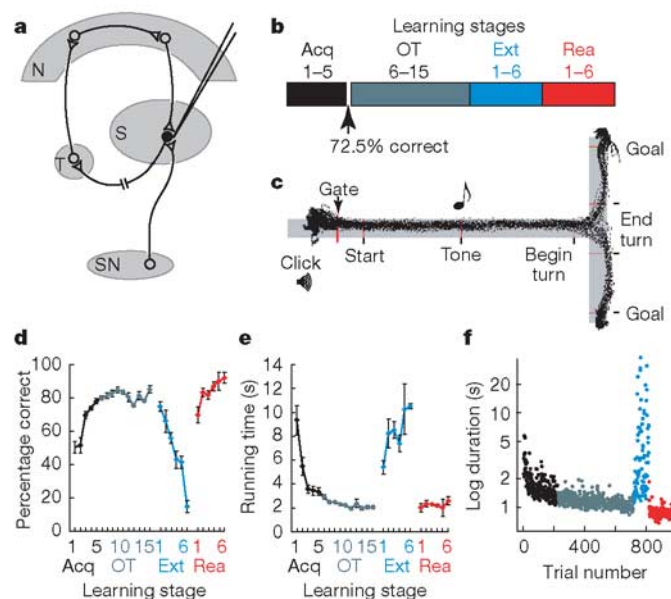


Figure 1 | T-maze task and behavioural learning. **a**, Simplified cortico-basal ganglia circuit, indicating recording of striatal projection neurons. N, neocortex; S, striatum; T, thalamus; SN, substantia nigra. **b**, Training stages (acquisition (Acq, black), 1–5; over-training (OT, grey), 6–15; extinction (Ext, blue), 1–6; reacquisition (Rea, red), 1–6; described in Supplementary Methods). **c**, Run trajectories for one over-training session. **d**, **e**, Average percentages of correct responses (**d**) and average per-trial running times (**e**). Error bars represent s.e.m. **f**, Trial-by-trial running times for the interval between tone onset and turn onset by a rat during successive training. Each dot represents performance in one trial.

¹Department of Brain and Cognitive Sciences and the McGovern Institute for Brain Research, Massachusetts Institute of Technology, 43 Vassar Street, 46-6133, Cambridge, Massachusetts 02139, USA. ²Department of Physics, Pennsylvania State University, 0104 Davey Laboratory, University Park, Pennsylvania 16802, USA.

*These authors contributed equally to this work.

the reacquisition period then was initiated by returning the reward at the end of each correct run, there was another sudden shift in the spike patterns, producing reduced mid-trial firing and a temporal advance of start activity resembling that seen during initial acquisition (Fig. 2a, and Supplementary Figs 2–5).

To estimate the randomness of the population spike activity across the entire trial time, we calculated the entropy of the average per-neuron firing across learning stages (Supplementary Methods). The

entropy fell sharply during acquisition, rebounded during extinction, and fell again during reacquisition (Fig. 2e), in the absence of significant changes in average per-trial firing rates (Supplementary Fig. 6). The changes in the spike patterns were highly correlated with the changes in behavioural accuracy (Fig. 2f, g).

Remarkably, we found equally striking lability in the spiking patterns of the striatal projection neurons that lacked detectable phasic peri-event activity during the task (Fig. 2b). Some of these

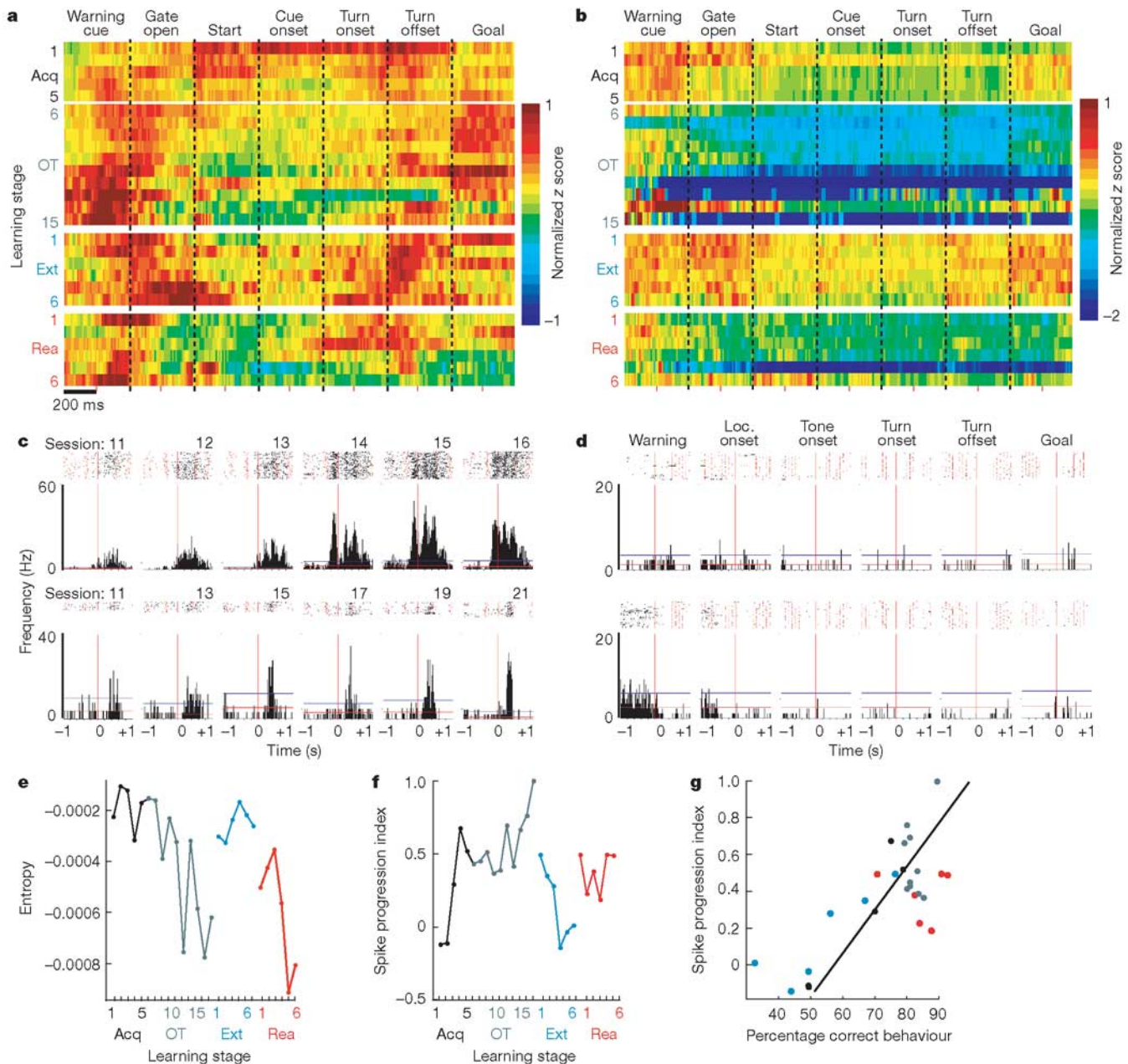


Figure 2 | Plasticity in spike activity patterns of striatal projection neurons. **a, b,** Average activity of units classified as task-responsive (**a**) and non-task-responsive (**b**) neurons plotted in 10-ms bins as z scores normalized for each neuron relative to that neuron's baseline activity, according to pseudo-colour scales shown at the right, with one row per training session. Plots show ± 200 -ms time windows around task events, abutted in the order of occurrence within a trial. **c,** Peri-event time histograms (PETHs, ± 1 -s window) for units recorded on consecutive days at single sites (putative single units), illustrating strengthening and time-shift in responses around locomotion onset over 6 consecutive sessions (top) and sharpening of phasic responses at turn onset over 13 sessions (bottom). Horizontal lines indicate mean pre-trial baseline firing rates (red) and two

standard deviations above the mean (blue, threshold for task-related activity). **d,** Typical PETHs for neurons lacking in-trial phasic peri-event activity ('non-task-responsive' units). **e,** Entropy of per-trial spike activity of task-responsive units calculated for each training stage. **f,** Spike progression index (SPI) illustrating correlation between per-trial spike activity of task-responsive projection neurons at each training stage and the neural activity at the last stage of over-training. **g,** Significant correlation ($r = 0.74$, $P < 0.0001$) between SPI and progressive changes in percentage correct behaviour during training. Black and grey, acquisition and over-training, $r = 0.82$, $P = 0.0002$; blue, extinction, $r = 0.87$, $P = 0.02$; red, reacquisition, $r = 0.09$, $P = 0.87$.

non-task-responsive neurons fired at low rates both in task and out of task, and some fired more out of task than in task (Fig. 2d). The in-task activity of these neurons dwindled during acquisition and then nearly ceased. Yet, on the first day of extinction, the average per-neuron firing of these neurons returned to pre-training levels and remained elevated. Then, when reacquisition training began, their activity declined sharply. These abrupt shifts were evident whether the activity of the neurons during the task was classified with respect to pre-trial baseline firing (Fig. 2b) or relative to in-trial activity (Supplementary Fig. 2)

To determine whether the tuning of task-related responses changed during learning, we measured multiple parameters (for example, height, width, peri-event peak timing) of the phasic spike responses to specific task events detected by a slope threshold (Supplementary Methods). None of these was altered during learning. By contrast, we found large-scale changes in the proportion of spikes per entire trial run that occurred within phasic responses (Fig. 3a, b). This proportion tripled during acquisition, fell abruptly during extinction and abruptly rose again during reacquisition (Fig. 3b). The number of phasic responses also changed successively (Fig. 3c). Reinforcing these redistributions of spike activity, the proportions of task-responsive projection neurons responding to different task events also progressively emphasized⁷, de-emphasized and re-emphasized the beginning and end of the task (Fig. 3d, and Supplementary Fig. 6). Notably, the sharpening of phasic responses during acquisition held not only for those occurring in the early and late parts of the task runs in which overall spiking increased, but also for responses occurring in the middle parts of the runs in which spike activity decreased (Supplementary Fig. 6). This result indicates that even when fewer neurons responded, some 'expert' responders with sharpened responses developed in the striatum as the task was acquired. This property, too, was subject to reversal and reappearance during subsequent extinction and reacquisition training.

Both the increase in concentration of spikes within phasic peaks during acquisition and the redistribution of spikes across run times had the effect of reducing the spread of spiking across trial performance time as the rats learned the task. We looked for, but did not

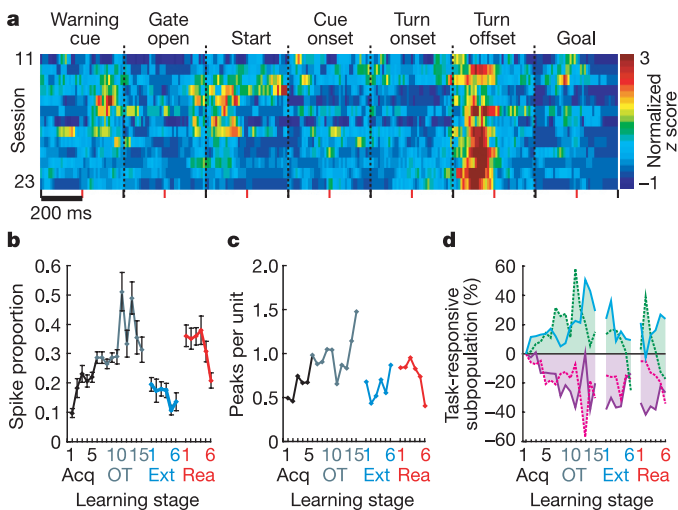


Figure 3 | Multiple changes in projection neuron activity in the sensorimotor striatum during acquisition, extinction and reacquisition training. **a**, Unit activity at a single site recorded over 13 sessions. Each row represents a session. **b**, Proportions of spikes concentrated in phasic responses in each trial, averaged for each session. Error bars indicate s.e.m. **c**, Average numbers of phasic responses per unit. **d**, Percentages of task-responsive projection neurons with responses at warning cue (solid blue line), goal reaching (dotted green line), locomotion onset (solid purple line) or turning (dotted magenta line). Values are plotted relative to the first training stage.

observe, significant changes in the variability of firing rates within peri-event or phasic-response windows across learning. However, we found major changes in the entropy (Fig. 2e) and in the variance (Supplementary Fig. 6) of spiking activity across the entire maze runs. Changes in spike distribution within the time frame of the entire procedural performance thus represented the key modulation of spike variability that we detected during learning.

Taken together, our findings show that per-trial spike distributions, response tuning and task selectivity were dynamically reconfigured as the procedural behaviour was acquired, extinguished and reacquired. Composite neural activity scores based on these factors were highly correlated with both behavioural accuracy and running times, especially during acquisition and extinction (Fig. 4, Supplementary Fig. 7 and Supplementary Methods). Restructuring of the day-by-day neural activity patterns in the 'fast learners' ($n = 5$) but not in the 'slow learners' ($n = 2$) early during acquisition (Supplementary Fig. 8) favoured a primary correspondence between the evolution of the neural restructuring over time and associative learning. The acquired patterns were detectable in both correct and incorrect trials (Supplementary Fig. 9), however, so that the ensemble patterns were not tied to performance in individual trials.

It has been proposed that the basal ganglia promote variability in behaviour during trial-and-error learning (exploration) and serve to evaluate behavioural changes to promote the acquisition of optimal behaviour (exploitation)^{9,10,13,14}. Our findings suggest that there might be a direct neural analogue to this explore-exploit behaviour in the firing patterns of projection neurons in the sensorimotor striatum. We demonstrate two fundamental changes in the spike activity of striatal projection neurons during procedural learning.

First, there was a global modulation of the firing of projection neurons. Early in training, the spike activity of the task-responsive population was spread throughout task time, as though all task events were salient (neural exploration). Even neurons without detectable phasic task-responsive activity fired at low rates during the task. Then, with continued training, this widespread spiking of the task-responsive population diminished, and their spike activity became focused (neural exploitation). At the same time, the non-task-responsive population fell silent, further reducing the task-irrelevant firing of the total projection neuron population. These changes in firing thus altered the distribution of striatal output neuron firing across the actual time-frame of the behaviour to be learned (the entire task run time). The reversal of the acquired task-related patterning during extinction and its reappearance in reacquisition fits the idea of increased neural exploration in the new extinction context and then a return to neural exploitation in the reinstated original context during reacquisition^{15–17}. The vivid

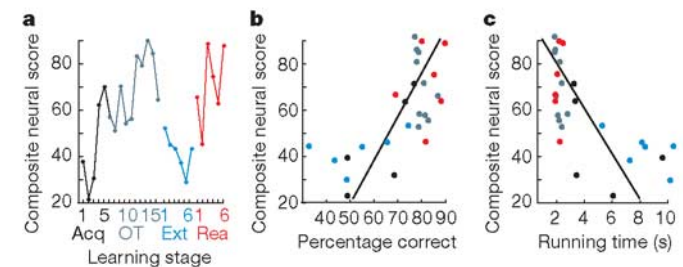


Figure 4 | Striatal neural activity predictive of behavioural performance. **a**, Composite neural scores based on weighted neural measures at trial start (normalized per-neuron firing rates during the ± 200 -ms interval around the warning cue, proportions of warning cue-responsive neurons, and proportions of spikes within phasic warning cue responses). **b**, Significant correlation between the composite scores (shown in **a**) and actual behavioural accuracy for each training stage (colour-coded as in **a**; $r = 0.69$, $P < 0.001$). **c**, Plot as in **b**, showing significant correlation between the composite neural scores and actual running times for each training stage ($r = -0.69$, $P < 0.001$).

modulation of the spiking of striatal projection neurons without detectable task-related activity also accords with this interpretation.

Second, in the exploitation phases of learning, ensemble firing at the start and end of the learned procedure strengthened, and sharply tuned responses of 'expert neurons' appeared. These changes indicate that early in training many candidate neurons fired, but that, with training and presumably competitive selection^{18–20}, neurons with sharply tuned responses appeared and, as a population, were tuned preferentially to respond near the start and end of the entire procedure performed. Our experiments leave open the question of where within the sensorimotor cortico-basal ganglia loop such changes were initiated. Because we recorded from striatal projection neurons, however, our findings show that such learning-related changes in neuronal firing occur as part of cortico-basal ganglia loop processing. The learning-related reduction in firing during the middle of the task time could indicate that striatal activity during this time was no longer needed for task execution, but could reflect the marking of behavioural boundaries in the process of chunking of the entire task performance⁵. These changing patterns could, in turn, reflect ongoing reorganization of cortico-basal ganglia activity^{20–23}. If so, the patterns could reflect neural representations related to the ready release of the learned behaviours by the appropriate context⁵.

Cortico-basal ganglia circuits probably act in determining, through reinforcement-based evaluation, which actions to enhance or diminish as learning proceeds^{4–6,9,10,19,20,24–30}. Viewed in the context of such selection functions, our findings indicate that dynamic neural representations in the striatum could adjust the encoded salience of task events and behavioural responses as habits are formed, lost and regained.

METHODS

The spike activity of neurons in the sensorimotor striatum was recorded chronically during behavioural training on a conditional reward-based T-maze task for 24–63 daily sessions from seven rats in which seven tetrode headstage assemblies had been implanted. Recordings began on the first day that the rats received training (about 40 trials per day) on the task, and were continued through successive acquisition training (stages 1–5), over-training (stages 6–15), extinction (stages 1–6) and reacquisition training (stages 1–6; Fig. 1b, Supplementary Table 2 and Supplementary Methods). In this task, rats learned to run down the maze and to turn right or left as instructed by auditory cues in order to receive reward. Behavioural data were acquired by means of photo-beams and a CCD camera. Neural data (32 kHz sampling) were collected by means of a Cheetah Data Acquisition System (Neuralynx Inc.). Well-isolated units accepted after cluster cutting were classified as striatal projection neurons or interneurons (Supplementary Fig. 1b–d). Behavioural and neural data were aligned by time stamps and were analysed by in-house software. The properties of both task-responsive and non-task-responsive projection neurons were analysed. Task-related responses of putative projection neurons were identified with respect to activity during a pre-trial 500-ms baseline period (threshold: 2 s.d. above baseline mean) and used to define task-responsive and non-task-responsive populations (Supplementary Methods). Unit data were analysed per neuron and per neuronal population across task events (Fig. 1c). To analyse population activity, normalized firing rates were averaged for each learning stage, and indices of spike firing patterns across learning stages were computed. The proportions of neurons with different task-related response types, the proportions of spikes that occurred within peri-event phasic responses per session, and trial-to-trial spike variability were also calculated, along with composite neural scores and measures of the entropy of neural firing. Changes in these measures were compared to changes in per cent correct performance and running times of the rats across stages of training.

Received 2 June; accepted 21 July 2005.

1. James, W. *The Principles of Psychology* 104–127 (Dover, New York, 1890).
2. Dickinson, A. Actions and habits: the development of behavioural autonomy. *Phil. Trans. R. Soc. Lond. B* **308**, 67–78 (1985).

3. Pavlov, I. P. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (ed. and transl. Anrep, G. V.) (Oxford Univ. Press, London, 1927).
4. Packard, M. G. & Knowlton, B. J. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* **25**, 563–593 (2002).
5. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
6. Poldrack, R. A. *et al.* Interactive memory systems in the human brain. *Nature* **414**, 546–550 (2001).
7. Jog, M., Kubota, Y., Connolly, C. I., Hillegaart, V. & Graybiel, A. M. Building neural representations of habits. *Science* **286**, 1745–1749 (1999).
8. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* **19**, 181–189 (2004).
9. Olveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3**, e153 (2005).
10. Kao, M. H., Doupe, A. J. & Brainard, M. S. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* **433**, 638–643 (2005).
11. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
12. Wilson, C. J. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 361–413 (Oxford Univ. Press, New York, 2004).
13. Doya, K. & Sejnowski, T. J. in *Advances in Neural Information Processing Systems* Vol. 7 (eds Tesauro, G., Touretzky, D. S. & Leen, T. K.) 101–108 (MIT Press, Cambridge, Massachusetts, 1995).
14. Doya, K. & Sejnowski, T. J. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 469–482 (MIT Press, Cambridge, Massachusetts, 2000).
15. Bouton, M. E. Context and behavioural processes in extinction. *Learn. Mem.* **11**, 485–494 (2004).
16. Routtenberg, A. & Kim, H.-J. in *Cholinergic–Monoaminergic Interactions in the Brain* (ed. Butcher, L. L.) 305–331 (Academic, New York, 1978).
17. Myers, K. M. & Davis, M. Behavioral and neural analysis of extinction. *Neuron* **36**, 567–584 (2002).
18. Gurney, K., Prescott, T. J. & Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* **84**, 401–410 (2001).
19. Graybiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
20. Djurfeldt, M., Ekeberg, Ö. & Graybiel, A. M. Cortex-basal ganglia interaction and attractor states. *Neurocomputing* **38–40**, 573–579 (2001).
21. Frank, M. J., Loughry, B. & O'Reilly, R. C. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* **1**, 137–160 (2001).
22. Houk, J. C. & Wise, S. P. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb. Cortex* **5**, 95–110 (1995).
23. Doya, K. Metalearning and neuromodulation. *Neural Netw.* **15**, 495–506 (2002).
24. Montague, P. R., Hyman, S. E. & Cohen, J. D. Computational roles for dopamine in behavioural control. *Nature* **431**, 760–767 (2004).
25. Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7**, 887–893 (2004).
26. Reynolds, J. N. J., Hyland, B. I. & Wickens, J. R. A cellular mechanism of reward-related learning. *Nature* **413**, 67–70 (2001).
27. Mink, J. W. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog. Neurobiol.* **50**, 381–425 (1996).
28. McClure, S. M., Berns, G. S. & Montague, P. R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339–346 (2003).
29. Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J., Davis, J. & Beiser, D.) 215–232 (MIT Press, Cambridge, Massachusetts, 1995).
30. Dickinson, A. & Balleine, B. W. Motivational control of goal-directed action. *Anim. Learn. Behav.* **22**, 1–18 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. F. Hall, P. A. Harlan and C. Thorn for their help. This work was funded by the National Institutes of Health and the Office of Naval Research.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M.G. (graybiel@mit.edu).