Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Computational approaches to habits in a model-free world

Wolfgang M Pauli, Jeffrey Cockburn, Eva R Pool, Omar D Pérez and John P O'Doherty

Model-free (MF) reinforcement learning (RL) algorithms account for a wealth of neuroscientific and behavioral data pertinent to habits; however, conspicuous disparities between model-predicted response patterns and experimental data have exposed the inadequacy of MF-RL to fully capture the domain of habitual behavior. We review several extensions to generic MF-RL algorithms that could narrow the gap between theory and empirical data. We discuss insights gained from extending RL algorithms to operate in complex environments with multidimensional continuous state spaces. We also review recent advances in hierarchical RL and their potential relevance to habits. Neurobiological evidence suggests that similar mechanisms for habitual learning and control may apply across diverse psychological domains.

Address
Computation and Neural Systems Program, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, United States

Corresponding authors: Pauli, Wolfgang M (pauli@caltech.edu), O'Doherty, John P (jdoherty@caltech.edu)

Current Opinion in Behavioral Sciences 2018, **20**:104–109

This review comes from a themed issue on **Habits and skills**

Edited by **Barbara Knowlton** and **Jörn Diedrichsen**

For a complete overview see the Issue and the Editorial

Available online 20th December 2017

https://doi.org/10.1016/j.cobeha.2017.12.001

Introduction

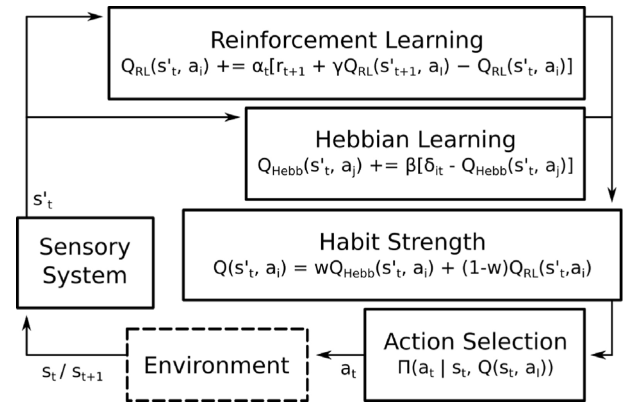
The advantages of habits have been recognized since the founding days of experimental psychology, when William and Harter summarized the results of their seminal experimental studies on habit learning in telegraphers, noting that their participants had 'no useful freedom for higher language units [sentences] which [they have] not earned by making the lower ones automatic' [1]. Their characterization of habits has influenced scientific inquiry to this day. In general, the execution of a single goal (preparing a favorite dish) might involve assembly of several frequently performed subtasks (e.g. turning the stove element on, or salting the boiling water) that are habitual in nature. Requiring minimal cognitive effort, relying on habits releases cognitive resources that can be applied to more demanding tasks. But there's no free lunch; the computational benefits of habits come at the cost of relative inflexibility, occasionally rendering behavior maladaptive if ingrained habits are difficult to overcome. Thus, adaptive behavior is generally argued to require a balanced mixture of habitual efficiency and goal-directed flexibility.

Current computational models of habit learning can be categorized according to their emphasis on three distinct aspects of habit learning. One category of models aims to capture the mechanisms of improving the accuracy and efficiency of motor movements. Challenged with noisy or delayed feedback, error-based learning mechanisms improve forward models, which make predictions about the outcome of motor movements, taking into account that both the body and its surrounding environment may have moved between the initiation of a motor command and its completion [2] (for a review, see Shadmehr *et al.* [3]). A second category of models focuses on use-dependent learning [4]. These models predict that habitual behaviors evolve merely from extended context-dependent repetition of a behavior [5,6].

Reinforcement learning (RL) algorithms represent a third category of computational models. In this context, habitual behavior occupies the middle ground between learned reflexes and goal-directed behavior. Learned reflexes are stereotyped such that sensory stimuli have innate activating tendencies, such as quickly withdrawing one's hand after noticing its placement on a stove-top before realizing that the stove top is cold. In contrast to reflexes, both habitual and goal-directed learning produces behavior not previously associated with a stimulus [7]. Goal-directed behavior is produced because it is expected to lead to a desirable outcome [8]. In contrast, habitual behavior is not produced because of an expectation of a particular outcome, but because its execution in a particular context has been consistently reinforced, resulting in the acquisition of stimulus–response (S–R) associations, as proposed by Thorndike's law of effect [9], or Hull's later drive reduction theory [10].

Two alternative algorithmic accounts have attempted to parsimoniously approximate habitual and goal-directed behavior. They have assumed that goal-directed behavior is the result of the belief in a causal association based upon the rate of responding and the rate of reward [11•], or the result of careful deliberation involving a cognitive model of environmental contingencies (model-based RL, MB-RL) [12,13]. Both accounts posit that habitual behavior can be approximated with model-free RL (MF-RL). MF-RL algorithms come in different flavors [14], but share the common principle that an action's value (representing habit strength) is determined by its reinforcement history whereby appetitive and aversive outcomes strengthen and weaken a habit respectively [15,16]. In the following, we review how the MF-RL account of habitual behavior handles three key experimental manipulations: its ability to approximate persistent responding after a previously desired outcome of an action has been devalued, or action-outcome contingencies have changed, as well as rapid reinstatement of behavior when rewards are reintroduced after extended periods without reinforcement. We then extend the developed framework to return to hierarchies of habits in other, more complex task domains.

Figure 1



Habit strength as the combined result of Hebbian learning (e.g. [6]) and RL (e.g. sarsa [16]). The sensory system interprets the state *s* of the environment at time *t* as s'_t, according to internal needs, goals, or believes about hidden states of the environment. The rate α_t of RL is reduced by the experience of stable contingencies (e.g. [26]). According to Hebbian learning, whether an action a_i is strengthened or weakened at rate β, depends on whether a_i was selected at time *t* ($\delta_{it} = 1$, if $a_i = a_t$, else $\delta_{it} = 0$). Which action is executed (a_t) is the result of weighting alternative actions a_l (e.g. pulling a chain or pressing a lever). *w* weights the respective contribution of Hebbian learning and RL mechanisms to behavior. λ is a temporal discounting factor. In an alternative implementation, learning may itself be the combined effect of Hebbian learning and RL [30••].

Model-free reinforcement learning

The benchmark for habits has traditionally been their insensitivity to outcome devaluation [17–20] and action-outcome contingency degradation [21–23]. MF-RL successfully captures the insensitivity to outcome devaluation (induced by for example pairing the outcome with illness), which is tested under extinction; as long as the devalued outcome is not re-experienced as a reinforcer of the learned action, subjects reduce the rate of responding merely because of the lack of reinforcement, but without additionally taking into account the devalued outcome [13].

However, MF-RL has difficulty accounting for the insensitivity of habitual behavior to changes in action-outcome contingencies during two distinct types of behavioral procedures, action-outcome contingency degradation and omission training. During action-outcome contingency degradation, subjects experience an increase in non-contingent reward delivery. MF-RL incorrectly predicts that animals are sensitive to non-contingent reward delivery, as long as the experimental protocol allows for alternative behaviors (e.g. grooming, rearing) to be reinforced. At the same time, MF-RL also has difficulty explaining the resilience of behavior during omission training, when they experience an increase in non-reinforced behavior. Here, MF-RL incorrectly predicts that the ensuing negative reward prediction errors (RPEs) lead to a rapid reduction in habitual response rates if behavior is no longer reinforced contiguously.

Two modifications to generic MF-RL enable it to account for this insensitivity to action-contingency degradation (see Figure 1). First, many MF-RL implementations assume that the amount of learning is constantly proportional to RPE magnitude, independent of whether RPEs are experienced early or late in training. Existing approaches to this limitation either suggest faster learning rates for acquisition than for unlearning [24], or a decrease in learning rates with extended experience of stable contingencies [25,26]. The latter proposal of experience-dependent variations in associability has successfully explained behavioral effects of backward blocking [27] and attenuated learning after forward blocking [28].

A second modification offers an opportunity for a unification of use-dependent learning [4,6] and MF-RL models of habits. Assuming that synaptic plasticity is the result of Hebbian learning mechanisms ubiquitous throughout the brain [29], RPEs are thought to be modulating the rate of synaptic plasticity [30••,31]. The contiguous expression of habitual behavior would therefore further ingrain a habit via Hebbian processes after the learned value of an action matches the value of the reinforcer and RPEs are no longer experienced [32]. At the same time, this implies that Hebbian learning may lead to the acquisition of distinct stimulus–response associations, encoded in synapses that are less susceptible to RPE driven plasticity [33]. Hebbian learning is thus a primary candidate mechanism for explaining insensitivity to outcome contingency degradation. Furthermore, in addition to accounting for the insensitivity of habits to action-outcome contingencies, Hebbian learning mechanisms also predict that the contiguous expression of goal-directed behavior may eventually render behavior habitual.

Other computational models of habit learning have focused on characterizing the behavioral context of a habit. Instead of simulating behavior in discrete state space environments, with individual states and a set of actions to transition among states, these models operate in RL environments consisting of continuous state spaces with a temporally evolving behavioral context that has multiple dimensions and attributes, and a set of actions that can affect particular aspects of states. These models posit that internal goals, and beliefs about hidden states of the environment, influence the interpretation of the behavioral context [34,35]. In doing so, these models can account for the rapid reinstatement of behavior after extended periods without reinforcement during an extinction test: rather than unlearning, agents instead assume that the context has changed, preserving existing stimulus–response associations, but temporarily rendering them irrelevant in the extinction context [34]. Enabling MF-RL algorithms to learn about the relevant dimensions of the behavioral context has significant implications for computational psychiatry [34], but also leads to qualitative performance increases in artificial cognitive architectures [36,37].

Hierarchical integration of behavior

A separate but related question is whether and how animals assemble habits hierarchically to efficiently solve familiar tasks with minimal oversight [1]. In continuous state space models with multidimensional behavioral contexts, RL and Hebbian learning mechanisms independently contribute to the aggregation of individual responses, which become associated with overlapping context characteristics. Model-free hierarchical RL (MF-HRL) [38,5] provides a formal account of how agents may learn to aggregate actions into reusable subroutines and skills, and how agents can identify the potential relevance for action routines to be applied to a wide range of future problems. Similar to MF-HRL, hierarchical dual system models have focused on how the acquisition of action sequences leads to saltatory behavioral control, where actions within each sequence are no longer evaluated individually [39•]. Reversion to goal-directed control occurs when action sequences no longer lead to the desired goals, prompting the sequence to be decomposed into its constituent actions for reevaluation. A complementary account posits that goals may be selected according to their model-free values, but that goal-directed planning is deployed to attain desired outcomes [40•]. Thus, these hierarchical models differ in their formalization of the trade-off between flexible goal-directed actions and computationally efficient habits when it comes to goal selection, deliberation, and monitoring.

By assuming a hierarchical integration of habitual and goal-directed systems, these models go beyond existing proposals for arbitration mechanisms that determine the contribution of either system to behavior. Arbitration models differ in their assumption about the criteria by which an arbitrator weights the contribution of each system (e.g. the respective uncertainty or expected inaccuracy [13], or reliability of the two systems [41], based on cost–benefit analyses [42,43], or based on deviations of the reward rate from the expected reward rate [6]). Because of the significant computational effort associated with evaluating the performance and predictions of constituent models during arbitration, these models do not speak to the benefits of hierarchical integration of behavior.

Biological substrates of habitual behavior

Although MF-RL approximates habit learning on the algorithmic level of analysis [44], significant progress has also been made in characterizing analogous neurocomputational mechanisms across mammalian species (for a review, see [45]). Briefly, convergent inputs to the midbrain dopamine system [46,47••] drive phasic activity of dopamine (DA) neurons that resembles a RPE learning signal [48–51]. These dopaminergic signals modulate synaptic plasticity in the striatum [52•,53,54], leading to the acquisition of stimulus–response associations in the dorsolateral striatum (DLS) [23,55,56].[1] Other evidence suggests that habits eventually become independent of striatal and dopaminergic mechanisms, relying instead on cortical areas [33]. Still more evidence suggests that the rodent infralimbic cortex [60], or human subgenual cortex [20], represents values of actions, in terms of their reinforcement history, to mediate the incremental ability of habits to out-compete goal-directed behavior after over-training.

In sum, the available evidence converges on the idea that the striatum learns to act as a gate-keeper for tentative motor plan representations in posterior frontal cortex through RL and Hebbian mechanisms. Decisions as to whether to execute a motor plan rely on highly convergent input to the striatum, including action value representations in ventromedial frontal cortex. Interestingly, several lines of research indicate a regional specialization within the striatum for diverse psychological functions [61•]. Critically, anatomical studies in primates [62] and rodents [63] describe topographically organized circuit architectures analogue to those supporting stimulus–response behavior. Topographic connections among the functionally organized frontal cortex [64] with distinct striatal regions may provide some of the neural architecture required to support hierarchical integration between

[1] DA also plays a role in model-based RL [57], working memory [58], and synaptic plasticity in motor cortex during acquisition of motor skills [59].

goal-directed and habitual behavior. At the same time, this topography may support habits in diverse psychological domains [65•], potentially including psycho-linguistic habits [66].

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by NIDA-NIH R01 grant (1R01DA040011-01A1). The authors would like to thank all members of the O'Doherty Human Reward and Decision Making laboratory for intriguing discussions.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. William LB, Harter N: **Studies on the telegraphic language: the acquisition of a hierarchy of habits**. *Psychol Rev* 1899, **6**:345-375.

2. Kawato M: **Internal models for motor control and trajectory planning**. *Curr Opin Neurobiol* 1999, **9**:718-727 http://dx.doi.org/10.1016/S0959-4388(99)00028-8.

3. Shadmehr R, Smith MA, Krakauer JW: **Error correction, sensory prediction, and adaptation in motor control**. *Annu Rev Neurosci* 2010, **33**:89-108 http://dx.doi.org/10.1146/annurev-neuro-060909-153135.

4. Orban de Xivry J-J, Criscimagna-Hemminger SE, Shadmehr R: **Contributions of the motor cortex to adaptive control of reaching depend on the perturbation schedule**. *Cereb Cortex* 2011, **21**:1475-1484 http://dx.doi.org/10.1093/cercor/bhq192.

5. Botvinick MM, Niv Y, Barto AC: **Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective**. *Cognition* 2009, **113**:262-280 http://dx.doi.org/10.1016/j.cognition.2008.08.011.

6. Miller K, Shenhav A, Ludvig E: **Habits without values**. *bioRxiv* 2016:067603 http://dx.doi.org/10.1101/067603.

7. Dickinson A: **Actions and habits: the development of behavioural autonomy**. *Philos Trans R Soc Lond B: Biol Sci* 1985, **308**:67-78 http://dx.doi.org/10.1098/rstb.1985.0010.

8. Tolman EC: **Cognitive maps in rats and men**. *Psychol Rev* 1948, **55**:189-208.

9. Thorndike EL: **Animal intelligence: an experimental study of the associative processes in animals**. *Psycholl Rev: Monogr Suppl* 1898, **2**:1125-1127 http://dx.doi.org/10.1037/h0092987.

10. Hull C: *Principles of Behavior*. Appleton-Century-Crofts; 1943.

11. • Dickinson A, Perez OD: **Actions and habits: psychological issues in dual-system theory**. *Understanding Goal-Directed Decision Making: Computations and Circuits*. Elsevier; 2017.

This chapter explains how a rate-correlation account for goal-directed actions may interact cooperatively with habits to explain performance and behavioural control under different reward schedules and extensions of training.

12. Doya K, Samejima K, Katagiri K-i, Kawato M: **Multiple model-based reinforcement learning**. *Neural Comput* 2002, **14**:1347-1369 http://dx.doi.org/10.1162/089976602753712972.

13. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nat Neurosci* 2005, **8**:1704-1711 http://dx.doi.org/10.1038/nn1560.

14. Sutton RS: **Learning to predict by the methods of temporal differences**. *Mach Learn* 1988, **3**:9-44 http://dx.doi.org/10.1007/BF00115009.

15. Bellman R: *Dynamic Programming*. Courier Corporation; 2013.

16. Sutton R, Barto A: *Reinforcement Learning*. Cambridge, MA: MIT Press; 1998.

17. Adams CD: **Variations in the sensitivity of instrumental responding to reinforcer devaluation**. *Q J Exp Psychol Sect B* 1982, **34**:77-98 http://dx.doi.org/10.1080/14640748208400878.

18. Yin HH, Knowlton BJ, Balleine BW: **Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning**. *Eur J Neurosci* 2004, **19**:181-189 http://dx.doi.org/10.1111/j.1460-9568.2004.03095.x.

19. Tricomi E, Balleine BW, O'Doherty JP: **A specific role for posterior dorsolateral striatum in human habit learning**. *Eur J Neurosci* 2009, **29**:2225-2232 http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x.

20. Liljeholm M, Dunne S, O'Doherty JP: **Differentiating neural systems mediating the acquisition vs. expression of goal-directed and habitual behavioral control**. *Eur J Neurosci* 2015, **41**:1358-1371 http://dx.doi.org/10.1111/ejn.12897.

21. Dickinson A, Charnock DJ: **Contingency effects with maintained instrumental reinforcement**. *Q J Exp Psychol Sect B* 1985, **37**:397-416 http://dx.doi.org/10.1080/14640748508401177 URL https://doi.org/10.1080/14640748508401177.

22. Dickinson A: **Omission learning after instrumental pretraining**. *Q J Exp Psychol Sect B* 1998, **51**:271-286 http://dx.doi.org/10.1080/713932679.

23. Yin HH, Knowlton BJ, Balleine BW: **Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning**. *Behav Brain Res* 2006, **166**:189-196 http://dx.doi.org/10.1016/j.bbr.2005.07.012.

24. Bush RR, Mosteller F: **A mathematical model for simple learning**. *Psychol Rev* 1951, **58**:313.

25. Mackintosh NJ: **A theory of attention: variations in the associability of stimuli with reinforcement**. *Psychol Rev* 1975, **82**:276.

26. Pearce JM, Hall G: **A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli**. *Psychol Rev* 1980, **87**:532-552 http://dx.doi.org/10.1037/0033-295X.87.6.532.

27. Dayan P, Kakade S: **Explaining away in weight space**. *Advances in Neural Information Processing Systems*. 2001:451-457.

28. Pauli WM, O'Reilly RC: **Attentional control of associative learning — a possible role of the central cholinergic system**. *Brain Res* 2008, **1202**:43-53 http://dx.doi.org/10.1016/j.brainres.2007.06.097.

29. Urakubo H, Honda M, Froemke RC, Kuroda S: **Requirement of an allosteric kinetics of NMDA receptors for spike timing-dependent plasticity**. *J Neurosci* 2008, **28**:3310-3323 http://dx.doi.org/10.1523/JNEUROSCI.0303-08.2008.

30. •• Rao RPN, Sejnowski TJ: **Spike-timing-dependent Hebbian plasticity as temporal difference learning**. *Neural Comput* 2001, **13**:2221-2237 http://dx.doi.org/10.1162/089976601750541787.

The authors developed a biophysical model of spike-timing-dependent plasticity that influenced later work on the integration of Hebbian mechanisms with temporal difference (TD) learning.

31. Pan W-X, Schmidt R, Wickens JR, Hyland BI: **Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network**. *J Neurosci* 2005, **25**:6235-6242 http://dx.doi.org/10.1523/JNEUROSCI.1478-05.2005.

32. Pauli WM, Hazy TE, O'Reilly RC: **Expectancy, ambiguity, and behavioral flexibility: separable and complementary roles of the orbital frontal cortex and amygdala in processing reward expectancies**. *J Cogn Neurosci* 2012, **24**:351-366 http://dx.doi.org/10.1162/jocn_a_00155.

33. Ashby FG, Turner BO, Horvitz JC: **Cortical and basal ganglia contributions to habit learning and automaticity**. *Trends Cogn Sci* 2010, **14**:208-215 http://dx.doi.org/10.1016/j.tics.2010.02.001.

34. Redish AD, Jensen S, Johnson A, Kurth-Nelson Z: **Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling**. *Psychol Rev* 2007, **114**:784-805 http://dx.doi.org/10.1037/0033-295X.114.3.784.

35. Gershman SJ, Blei DM, Niv Y: **Context, learning, and extinction**. *Psychol Rev* 2010, **117**:197-209 http://dx.doi.org/10.1037/a0017808.

36. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D: **Human-level control through deep reinforcement learning**. *Nature* 2015, **518**:529-533 http://dx.doi.org/10.1038/nature14236.

37. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D: **Mastering the game of Go with deep neural networks and tree search**. *Nature* 2016, **529**:484-489 http://dx.doi.org/10.1038/nature16961.

38. Sutton R, Precup D, Singh S: **Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning**. *Artif Intell* 1999, **1–2**:181-211.

39. • Dezfouli A, Lingawi NW, Balleine BW: **Habits as action sequences: hierarchical action control and changes in outcome value**. *Philos Trans R Soc B* 2014, **369**:20130482 http://dx.doi.org/10.1098/rstb.2013.0482.

The authors provide an integrative account of goal-directed and habitual behavior. They posit that complex tasks engage goal-directed deliberation over alternative strategies, followed by habitual execution of previously acquired action sequences. Outcomes of individual actions within a sequence are not evaluated, unless persistent lack of success necessitates a decomposition of action sequences.

40. • Cushman F, Morris A: **Habitual control of goal selection in humans**. *Proc Natl Acad Sci U S A* 2015, **112**:13817-13822 http://dx.doi.org/10.1073/pnas.1506367112.

The authors provide an integrative account of goal-directed and habitual behavior. They posit that solving complex tasks involves habitual selection of goals, while their attainment involves goal-directed planning.

41. Lee SW, Shimojo S, O'Doherty JP: **Neural computations underlying arbitration between model-based and model-free learning**. *Neuron* 2014, **81**:687-699 http://dx.doi.org/10.1016/j.neuron.2013.11.028.

42. Pezzulo G, Rigoli F, Chersi F: **The mixed instrumental controller: using value of information to combine habitual choice and mental simulation**. *Front Psychol* 2013, **4**:1-15 http://dx.doi.org/10.3389/fpsyg.2013.00092.

43. Kool W, Gershman SJ, Cushman FA: **Cost–benefit arbitration between multiple reinforcement-learning systems**. *Psychol Sci* 2017, **28**:1321-1333.

44. Marr DC, Poggio T: **From understanding computation to understanding neural circuitry**. *AI Memo* 1976, **357**:1-22.

45. O'Doherty JP, Cockburn J, Pauli WM: **Learning, reward, and decision making**. *Annu Rev Psychol* 2017, **68**:73-100 http://dx.doi.org/10.1146/annurev-psych-010416-044216.

46. Hazy TE, Frank MJ, O'Reilly RC: **Neural mechanisms of acquired phasic dopamine responses in learning**. *Neurosci Biobehav Rev* 2010, **34**:701-720 http://dx.doi.org/10.1016/j.neubiorev.2009.11.019.

47. •• Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N: **Arithmetic and local circuitry underlying dopamine prediction errors**. *Nature* 2015, **525**:243-246 http://dx.doi.org/10.1038/nature14855.

The authors combined optogenetic manipulations with extracellular recordings in the ventral tegmental area of the midbrain dopamine system to deconstruct the neuronal computations underlying reward prediction errors.

48. Montague PR, Dayan P, Sejnowski TJ: **A framework for mesencephalic dopamine systems based on predictive Hebbian learning**. *J Neurosci* 1996, **16**:1936-1947.

49. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599 http://dx.doi.org/10.1126/science.275.5306.1593.

50. O'Doherty JP, Buchanan TW, Seymour B, Dolan RJ: **Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum**. *Neuron* 2006, **49**:157-166 http://dx.doi.org/10.1016/j.neuron.2005.11.014.

51. Pauli WM, Larsen T, Collette S, Tyszka JM, Seymour B, O'Doherty JP: **Distinct contributions of ventromedial and dorsolateral subregions of the human substantia nigra to appetitive and aversive learning**. *J Neurosci* 2015, **35**:14220-14233.

52. • Reynolds JNJ, Hyland BI, Wickens JR: **A cellular mechanism of reward-related learning**. *Nature* 2001, **413**:67-70 http://dx.doi.org/10.1038/35092560.

Using intracranial self-stimulation of the substantia nigra, the authors provided causal evidence for the role of phasic dopamine release in the vicinity of cortico-striatal synapses in inducing behaviorally-relevant synaptic potentiation.

53. Shen W, Flajolet M, Greengard P, Surmeier DJ: **Dichotomous dopaminergic control of striatal synaptic plasticity**. *Science* 2008, **321**:848-851 http://dx.doi.org/10.1126/science.1160575.

54. Tsai H-C, Zhang F, Adamantidis A, Stuber GD, Bonci A, Lecea Ld, Deisseroth K: **Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning**. *Science* 2009, **324**:1080-1084 http://dx.doi.org/10.1126/science.1168878.

55. Pauli WM, Clark AD, Guenther HJ, O'Reilly RC, Rudy JW: **Inhibiting PKMzeta reveals dorsal lateral and dorsal medial striatum store the different memories needed to support adaptive behavior**. *Learn Mem* 2012, **19**:307-314 http://dx.doi.org/10.1101/lm.025148.111.

56. McNamee D, Liljeholm M, Zika O, O'Doherty JP: **Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fMRI study**. *J Neurosci* 2015, **35**:3764-3771 http://dx.doi.org/10.1523/JNEUROSCI.4677-14.2015.

57. Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G: **Dopamine transients are sufficient and necessary for acquisition of model-based associations**. *Nat Neurosci* 2017, **20**:735-742.

58. Williams GV, Goldman-Rakic PS: **Modulation of memory fields by dopamine D1 receptors in prefrontal cortex**. *Nature* 1995, **376**:572-575 http://dx.doi.org/10.1038/376572a0.

59. Hosp JA, Pekanovic A, Rioult-Pedotti MS, Luft AR: **Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning**. *J Neurosci* 2011, **31**:2481-2487 http://dx.doi.org/10.1523/JNEUROSCI.5411-10.2011.

60. Killcross S, Coutureau E: **Coordination of actions and habits in the medial prefrontal cortex of rats**. *Cereb Cortex* 2003, **13**:400-408 http://dx.doi.org/10.1093/cercor/13.4.400.

61. • Pauli WM, O'Reilly RC, Yarkoni T, Wager TD: **Regional specialization within the human striatum for diverse psychological functions**. *Proc Natl Acad Sci U S A* 2016, **113**:1907-1912 http://dx.doi.org/10.1073/pnas.1507610113.

Using an unbiased, data-driven approach, the authors analyzed large-scale coactivation data from human imaging studies. They identified distinct striatal zones that exhibited discrete patterns of coactivation with cortical brain regions across distinct psychological processes, and identified the different psychological processes associated with each zone.

62. Alexander GE, DeLong MR, Strick PL: **Parallel organization of functionally segregated circuits linking basal ganglia and cortex**. *Annu Rev Neurosci* 1986, **9**:357-381.

63. McGeorge A, Faull R: **The organization of the projection from the cerebral cortex to the striatum in the rat**. *Neuroscience* 1989, **29**:503-537 http://dx.doi.org/10.1016/0306-4522(89)90128-0.

64. Badre D: **Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes**. *Trends Cogn Sci* 2008, **12**:193-200 http://dx.doi.org/10.1016/j.tics.2008.02.004.

65. Hazy TE, Frank MJ, O'Reilly RC: **Banishing the homunculus:**
• **making working memory work**. *Neuroscience* 2006, **139**:105-118 http://dx.doi.org/10.1016/j.neuroscience.2005.04.067.

The authors provide a biologically inspired mechanistic account of pre-frontal cortical–basal ganglia interactions subserving executive functions.

66. Cohen JD, Dunbar K, McClelland JL: **On the control of automatic processes: a parallel distributed processing account of the Stroop effect**. *Psychol Rev* 1990, **97**:332-361.