The Role of Executive Function in Shaping Reinforcement Learning

Milena Rmus[1], Samuel D. McDougle[2], Anne G. E. Collins[1,3]

[1]Department of Psychology, University of California, Berkeley

[2]Department of Psychology, Yale University

[3]Helen Wills Neuroscience Institute, University of California, Berkeley

Abstract

Reinforcement learning (RL) models have advanced our understanding of how animals learn and make decisions, and how the brain supports learning. However, the neural computations that are explained by RL algorithms fall short of explaining many sophisticated aspects of human learning and decision making, including the generalization of behavior to novel contexts, one-shot learning, and the synthesis of task information in complex environments. Instead, these aspects of behavior are assumed to be supported by the brain's executive functions (EF). We review recent findings that highlight the importance of EF in instrumental learning. Specifically, we advance the theory that EF sets the stage for canonical RL computations in the brain, providing inputs that broaden their flexibility and applicability. Our theory has important implications for how to interpret RL computations in both brain and behavior.

Introduction

Our ability to learn rewarding actions lies at the core of goal-directed decision-making. Reward-driven choice processes have been extensively modeled using reinforcement learning (RL) algorithms [1]. This formalized account of learning and decision making has contributed significantly to expanding the frontiers of artificial intelligence research [2], our understanding of clinical pathologies [3, 4], and research on developmental changes in learning [5, 6].

A key reason for the success of the RL framework is its ability to capture learning not only at the behavioral level, but also at the neural level. The neural foundations of reward-dependent learning [7], and its various successors [8], have established a well-defined brain network that performs key RL computations. In particular, cortico-striatal loops enable state-dependent value-based choice [9]. Furthermore, dopaminergic signaling of reward-prediction errors (RPEs) in the midbrain and striatum induces neural plasticity consistent with RL algorithms, incrementally increasing/decreasing the value of actions that yield better/worse than expected outcomes.

Despite its tremendous success, there are well known limitations of canonical RL algorithms [10]. Historically, many insights provided by RL research have been demonstrated in relatively simplistic learning tasks, casting doubt on how useful classic RL models are in explaining how

humans learn and make choices in everyday life. To solve this problem, recent research often augments RL algorithms with learning and memory mechanisms from other cognitive systems.

Executive functions (EF) have been identified as a key set of psychological faculties that appear to interact with RL computations. For instance, working memory (WM), as a short-term cache which allows us to retain and manipulate task-relevant information over brief periods [11, 12, 13], occupies a central position in our ability to organize goal-directed behavior. A related core EF, attention, also contributes to behavioral efficiency through selective processing of subsets of environmental features relevant for learning [14, 15, 16]. Research on WM and attention points to the prefrontal cortex (PFC) as the primary site of these processes [17, 18], suggesting that this network shapes information processing in the RL system during learning.

Several straightforward experimental manipulations have revealed that an isolated RL system fails to effectively capture human instrumental learning behavior. For example, while online maintenance of representations in WM is capacity-limited [19], standard RL models have no explicit capacity constraints. This property of RL suggests that if individuals rely on RL alone, learning should not be affected by the number of rewarding stimulus-response associations they are required to learn in a given task. However, humans learn much less efficiently when the number of associations to be learned in parallel exceeds WM capacity [20, 6, 21], suggesting that RL operates side by side with working memory during learning. Other work has similarly shown that EF-dependent planning contributes to choice alongside core RL computations implemented in the brain [22, 23].

However, there is also evidence that EF does not only contribute as a distinct learning system operating independently of the brain's RL network: Additionally, EF may interact with RL by directly contributing to RL computations in the brain. Models of PFC-striatal loops [24, 25], which posit that brain regions associated with EF and RL interact directly, has inspired behavioral experiments and computational models aimed at identifying EF-RL interactions [20, 21, 14, 5]. The advent of these modeling tools has shown that an interaction of multiple neurocognitive domains (e.g., RL, WM, attention) may provide a more robust account of goal-directed behavior, one that still maintains the centrality of canonical RL computations in instrumental learning [26, 27].

In this paper, we review recent work that provides converging evidence for direct, functionally coherent contributions of EF to RL computations. More specifically, we review how EF (WM and attention in particular) might set the stage for RL computations in the brain by defining the relevant state space, action space, and reward function (Figure 1). The ideas reviewed here can help inform future computational modeling efforts and experimental designs in the study of goal-directed behavior. Furthermore, it may shift our interpretations of past and future findings focused on isolated RL computations towards a broader framework that also considers EF contributions.

The ingredients of RL computations

Past work suggests that a specific brain network (primarily cortico-striatal loops) supports RL computations, such as temporal difference learning [28, 29] and actor-critic learning [30, 31]. These learning algorithms update estimates of values via reward prediction errors (RPEs). In machine learning, such algorithms are defined not only by how they estimate value, but also by (at least) three fundamental components: 1) the *state space* (reflecting the possible states *s* or contexts an agent may be in), 2) the *action space* (reflecting the possible choices *a* to be made in a given state), and 3) the *reward function R(s,a)*(signaling reinforcing outcomes). The specification of these variables can dramatically impact the behavior of a decision-making agent, but how these three variables are supplied to the brain's RL network is poorly understood.

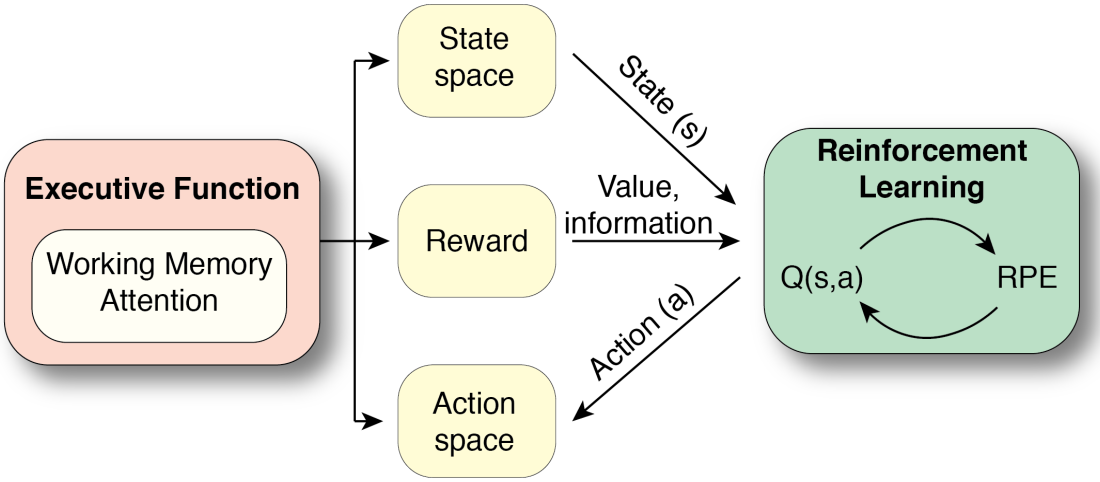

Figure 1. Schematic of EF contributions (WM, attention) to the brain's RL computations. EF can optimize RL computations in the domain of 3 relevant RL-components: state space, reward functions, and action space. Q(s,a) reflects the estimated value of a state and action. RPE is the reward prediction error used to update Q(s,a). Additional RL-independent contributions of EF to learning are not shown.

State space

The RL framework defines a state space over which learning occurs. A state can be a location in the environment, a sensory feature of the environment (e.g., the presence of a stimulus, such as a light), or a more abstract, internally represented context (such as a point in time). At each state, a decision-making agent enacts a choice in pursuit of rewards [1]. The specification of the state space significantly impacts the behavior of artificial RL agents. For example, in a large state space, RL performance is limited by what is known as the curse of dimensionality [1,10]: Learning a vast number of state-action values quickly becomes computationally intractable. Defining a smaller state space limited to only task-relevant states is one path toward overcoming this challenge.

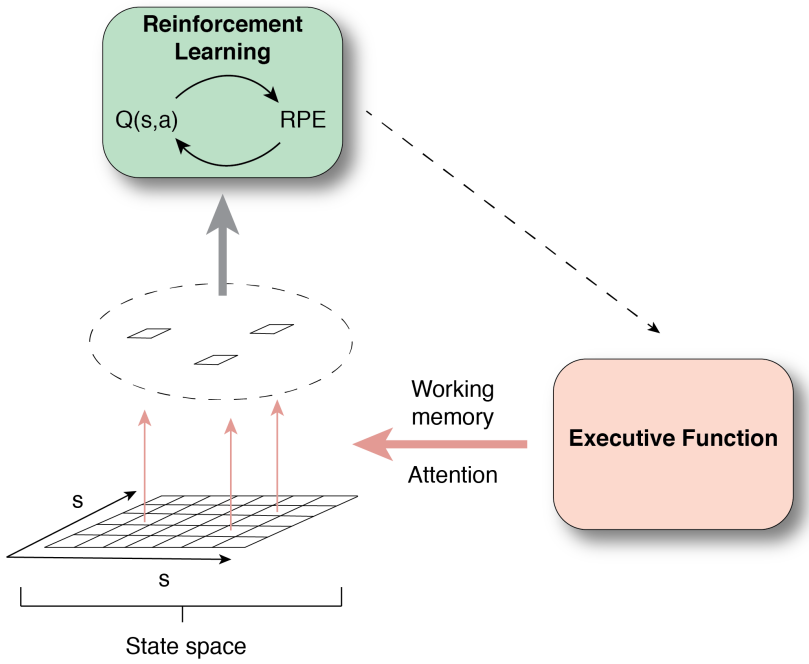

Figure 2. EF specifies the relevant state space, allowing the RL system to efficiently operate over a subset of task-relevant states. See Figure 1 for notations.

Simplifying the state space is a function sometimes attributed to attentional filters, which can specify important features of the environment [14]. In this framework, attention tags the features of the environment that RL variables are computed over [32, 33, 34, 14]. This is accomplished by attention differentially weighing environmental features, assigning a higher weight to task-relevant ones [33] (Figure 2). For instance, if an agent is attempting to earn reward from various stimuli that differ along several dimensions (e.g. color, shape), with only one dimension predicting reward, an optimized learning agent would 1) identify that dimension, and 2) specify it as the relevant state space for RL. That way, an agent can avoid computing values over a larger state space containing all possible features [35]. Computationally, this can be achieved by implementing Bayesian inference to discover relevant task features that RL operates over [14]. In addition to attention affording the reduction of task complexity, attentional mechanisms serve another purpose: Many tasks share overlapping/competing state spaces, leading to potential interference in correct action selection (e.g., the Stroop Task). Here again, defining a low-dimensional representation that can be applied to multiple tasks in the service of goal attainment makes learning simultaneously more flexible and more robust [36].

Importantly, the relevant state space is not always signaled by explicit sensory cues. Thus, an agent often has to make an inference about their current state [37]. Recent work in animals suggests that RL computations in the striatum are likely performed over these latent belief states [38, 39]. For example, markedly different dopamine dynamics are observed if an expected reward is sure to arrive (e.g., 100% chance) versus almost sure to arrive (e.g., 90% chance) [40]. In this example, an inference about the latent state, which indicates the probability that a reward will arrive or not, dramatically alters RL computations. It is hypothesized that RL computations over these belief states may be mediated by input from frontal cortices involved in the discovery

and representation of state spaces (e.g., orbitofrontal cortex), further supporting a link between EF and RL [41].

Action space

Above we reviewed a role for EFs, such as working memory and attention, in attending to and carving out the appropriate state space for RL. A complementary idea is that EF also plays a role in specifying (or simplifying) the action space for the RL system (Figure 3). The action space in the RL formalism is defined as the set of choices an agent can make. The choice can take the form of a simple motor action (e.g., a key press), a complex movement (e.g., walking to the door), or an abstract choice not defined by specific motor actions (e.g., choosing soup vs. salad). Defining the relevant action space is arguably as essential for learning as defining the relevant state space.

Recent studies indicate that the action space is a separable dimension for RL. First, behavioral evidence suggests that reward outcomes can simultaneously be assigned to task-relevant choices in addition to task-irrelevant motor actions (i.e., reinforcing a right-finger button press regardless of the stimulus that was present) [42]. Moreover, this process appears to be negatively related to the use of goal-directed planning strategies, suggesting that EF enables RL to focus in on the task-relevant action space. Similarly, recent modeling work suggests that a stateless form of action values – that is, action values computed independently of any specific context – can exert an influence on both choices and reaction times [43], particularly when cognitive load is high. One hypothetical consequence of independently learning over the action dimension is that when executive functions are disrupted or taxed, and thus cannot properly conjoin states and actions, action values may be learned in a vacuum. Speculatively, this could lead to actions being performed perseveratively even when they are maladaptive in certain states, which could be further linked to pathological forms of habitual behavior, such as addiction [44].

Because actions link predicted choice values with observed outcomes, one natural question beyond the selection of actions is how the RL system differentiates choice errors (e.g., which is the best object?) from choice execution errors (e.g., did I grasp the desired object?). In RL tasks that require reaching movements, behavioral data and fMRI responses in the striatum suggest that perceived action errors influence RPEs. That is, if the credit for a negative outcome is assigned to the motor system, the RL system appears to eschew updating the value of the choice that was made [45, 46]. These results suggest that simple cognitive inferences about the cause of errors (e.g., choice errors versus action execution errors) are incorporated into RL computations.

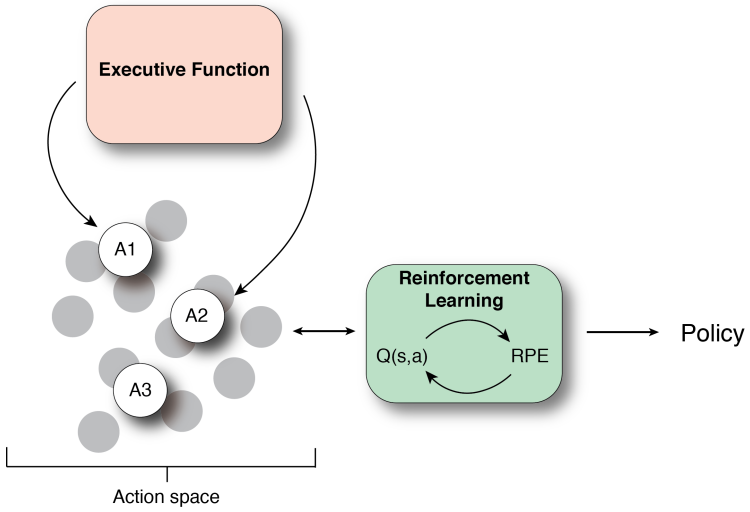

Figure 3. Contribution of executive functions in action selection.

In more complex situations with a large action space, EF can aid the learning process by attempting to reduce the size of this space. That is, the brain can create "task-sets", or selective groupings of state-action associations and use contextual cues to retrieve the appropriate task set. To illustrate, if one learns the motor commands for copying text on both a PC and a Mac, to avoid interference it is beneficial to associate the specific motor sequences (ctrl-c versus command-c) with their respective contexts (typing on a PC keyboard versus a Mac keyboard). Indeed, humans appear to cluster subsets of actions with associated sensory contexts during instrumental learning [47, 48], and they do so in a manner which suggests that high-level inferences about task structure shape low-level reinforcement learning computations over actions. Moreover, such behaviors echo the important role of affordances [49], which describe the link between specific environmental states and the actions they afford. This concept has recently been proposed as a novel method for making RL more efficient in complex state-spaces [50].

Selecting a task-set can itself be seen as a choice made in an abstract, high-level context. Learning to make this abstract choice may also involve RL, such that RL computations occur over two different state-action spaces in parallel – an abstract context and task-set space, and a more concrete stimulus-action space [51, 52]. There is recent computational, behavioral, and neural evidence that stacked hierarchies of RL computations happen in parallel over more and more abstract types of states and choices, facilitating complex learning abilities [53, 54, 51]. Such learning may be supported by hierarchies of representations in prefrontal cortex [55, 56]. This again highlights a role for EF in setting the stage for RL computations to solve complex learning problems.

Rewards & expectations

Goal-directed behavior is dependent on making correct predictions about the outcome of our choices. RPEs, which serve as a teaching signal, occupy a central position in the RL framework, linking midbrain dopaminergic activity with RL computations [7]. Most RL research since has focused on simple forms of learning from outcomes that act as primary or secondary rewards, such as food, money, or numeric points in a game. However, the path to an RPE is not always so

straightforward: For instance, recent work departs from the role of dopaminergic signaling in standard RPEs based on scalar rewards, extending the domain of RL to learning from indirect experiences (e.g., secondary conditioning) and more abstract learning of associations based on sensory features [57, 58]. These findings suggest that RL value computations integrate information beyond primary and secondary rewards. There is early evidence that EF could be implicated in signaling what information is treated as a reinforcer by the brain's RL network.

One such example relates to the value of information. Humans are motivated to reduce uncertainty about their environment [59]. Thus, acquisition of novel information should in itself function as reinforcement. Most information-seeking mechanisms, however, are not accounted for in the traditional RL framework. By contrast, recent work has shown that uncertainty reduction and information gain are indeed reflected in neural RL computations [60]. Evidence from fMRI studies suggests that corticostriatal circuits incorporate the utility of information in reward computations, such that information is conceptualized as a reward that reinforces choices [61], even when it is not valenced [59]. The prefrontal cortex also appears to track information and uncertainty [40], which can be held in working memory to influence decision making [62] (Figure 4).

The theoretical framework of hierarchical RL also dissociates the role of exploiting information about the environment from the role of primary/secondary rewards, while emphasizing that both act as a teaching signal [63]. In particular, when learning a multi-step policy that ultimately leads to a rewarding goal, agents identify and use subgoals en route to terminal rewards. In the hierarchical RL framework, reaching these subgoals generates pseudo-rewards, and appears to drive activity in canonical reward-processing regions in the brain, even though these rewards are 1) not inherently rewarding, and 2) are clearly distinguished from terminal rewards [64, 65]. The processing of pseudo-rewards is additionally assumed to be driven by the prefrontal cortex, suggesting a link to EF [66].

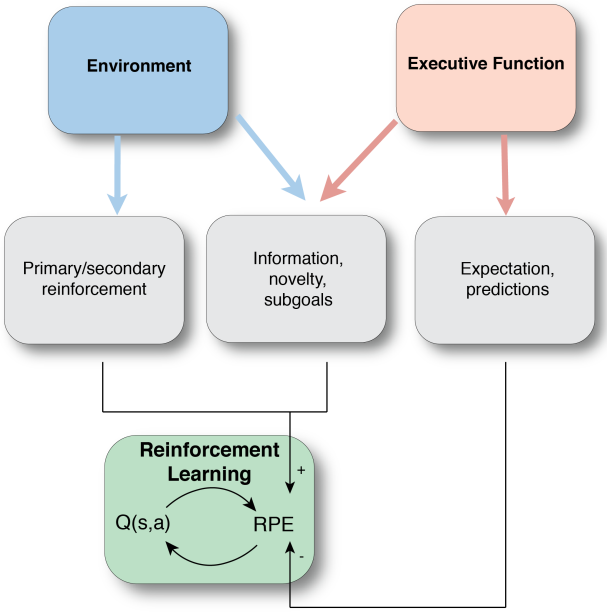

Figure 4. Traditional model of putative neural mechanisms involved in reward learning suggests that the RPE, are primarily driven by the primary and/or secondary reinforcement. More recent work posits that RPEs are also influenced by the

Beyond expanding the space of rewarding outcomes, there is also evidence that EF may affect RPEs in an alternative way: namely, by inputting reward expectations that have not yet been learned via the RL network. For example, work by [67] has shown that the magnitude of RPEs in the striatum is affected by cognitive load such that learning a small number of stimulus-response associations leads to attenuated striatal RPEs. This result is explained by "top-down" input of predictions from working memory: Information held in working memory in simple learning environments creates expectations of reward that are learned faster than in the RL system, and thus weaken RPEs [21, 20]. Similar results are observed in planning tasks, where an EF-dependent planned expectation of reward modulates the classic representation of RPEs in the striatum [22]. Taken together, these results demonstrate a key role for EF in defining the reward function for the RL system, and in contributing to the value estimation process.

Conclusions & discussion

We have reviewed and summarized computational, behavioral and neural evidence which collectively suggest that (1) executive function shapes reinforcement learning computations in the brain, and (2) neural and cognitive models of this interaction provide useful accounts of goal-directed behavior. We discussed the EF-RL interaction vis-a-vis the specification of the state space, action space, and reward function that RL operates over.

This new framework has important implications for applying both neural and cognitive computational models to study individual differences in learning. Although it is tempting to study individual differences with simple RL models, it is essential that we carefully consider the role of alternative neurocognitive systems in learning. Evidence of individual learning differences captured by an RL model might not reflect differences in the brain's RL process, but rather in upstream EF that shapes RL. Indeed, recent work on development [5, 34], schizophrenia [68], and addiction [69, 3] has shown that individual variability in learning might be driven by

both EF and RL, and/or the interaction of the two. Thus, building improved models of the interplay between different neurocognitive systems should help us better understand individual differences across the lifespan and in clinical disorders. This expansion of the RL theoretical framework can deepen our understanding of how learning is supported in the brain and inform future interventions and treatments.

References

1. Sutton, R.S., & Barto, A.G. (2018). Reinforcement learning: An introduction. MIT Press.

2. Botvinick, M., Ritter, S., Wang, J. X., Kurth-Nelson, Z., Blundell, C., & Hassabis, D. (2019). Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*, *23*(5), 408–422. https://doi.org/10.1016/j.tics.2019.02.006

3. Wyckmans, F., Otto, A. R., Sebold, M., Daw, N., Bechara, A., Saeremans, M., Kornreich, C., Chatard, A., Jaafari, N., & Noël, X. (2019). Reduced model-based decision-making in gambling disorder. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-56161-z

4. Radulescu, A., & Niv, Y. (2019). State representation in mental illness. *Current Opinion in Neurobiology*, *55*, 160–166. https://doi.org/10.1016/j.conb.2019.03.011

5. Segers, E., Beckers, T., Geurts, H., Claes, L., Danckaerts, M., & van der Oord, S. (2018). Working Memory and Reinforcement Schedule Jointly Determine Reinforcement Learning in Children: Potential Implications for Behavioral Parent Training. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00394

6. Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. E. (2020). Disentangling the systems contributing to changes in learning during adolescence. *Developmental Cognitive Neuroscience*, *41*, 100732. https://doi.org/10.1016/j.dcn.2019.100732

7. Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

8. Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*, *577*(7792), 671–675. https://doi.org/10.1038/s41586-019-1924-6

9. Frank, M. J. (2011). Computational models of motivated action selection in corticostriatal circuits. *Current Opinion in Neurobiology*, *21*(3), 381–386. https://doi.org/10.1016/j.conb.2011.02.013

10. Vong, W. K., Hendrickson, A. T., Navarro, D. J., & Perfors, A. (2019). Do Additional Features Help or Hurt Category Learning? The Curse of Dimensionality in Human Learners. *Cognitive Science*, *43*(3), e12724. https://doi.org/10.1111/cogs.12724

11. Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, *100*(2), 463–475. https://doi.org/10.1016/j.neuron.2018.09.023

12. Lundqvist, M., Herman, P., & Miller, E. K. (2018). Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *The Journal of Neuroscience*, *38*(32), 7013–7019. https://doi.org/10.1523/JNEUROSCI.2485-17.2018

13. Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological Review*, *125*(4), 486–511. https://doi.org/10.1037/rev0000101

14. Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic Reinforcement Learning: The Role of Structure and Attention. *Trends in Cognitive Sciences*, *23*(4), 278–292. https://doi.org/10.1016/j.tics.2019.01.010

15. Norman D.A., Shallice T. (1986) Attention to Action. In: Davidson R.J., Schwartz G.E., Shapiro D. (eds) Consciousness and Self-Regulation. Springer, Boston, MA

16. Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (p. 631–682). The MIT Press.

17. Badre. D. (2020). Brain networks for cognitive control: Four unresolved questions. In P. W. Kalivas and M. P. Paulus (Eds.), Intrusive Thinking across Neuropsychiatric Disorders: From Molecules to Free Will. Strüngmann Forum Reports, vol. 30, J. R. Lupp, series editor. Cambridge, MA: MIT Press, in press.

18. Badre, D., & Desrochers, T. M. (2019). Chapter 9—Hierarchical cognitive control and the frontal lobes. In M. D'Esposito & J. H. Grafman (Eds.), *Handbook of Clinical Neurology* (Vol. 163, pp. 165–177). Elsevier. https://doi.org/10.1016/B978-0-12-804281-6.00009-4

19. Baddeley, A. (2012). Working memory: theories, models, and controversies. Annual Review of Psychology, 63, 1–29. http://doi.org/10.1146/annurev-psych-120710- 100422

20. Collins, A. G. E. (2018). The Tortoise and the Hare: Interactions between Reinforcement Learning and Working Memory. *Journal of Cognitive Neuroscience*, *30*(10), 1422–1432. https://doi.org/10.1162/jocn_a_01238

21. Collins, A.G.E. & Frank, M.J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences, 115,* 2502-2507.

22. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

23. Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free

mechanisms. *PLOS Computational Biology*, *13*(9), e1005768. https://doi.org/10.1371/journal.pcbi.1005768

24. Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1601–1613. https://doi.org/10.1098/rstb.2007.2055

25. Zhao, F., Zeng, Y., Wang, G., Bai, J., & Xu, B. (2018). A Brain-Inspired Decision Making Model Based on Top-Down Biasing of Prefrontal Cortex to Basal Ganglia and Its Application in Autonomous UAV Explorations. *Cognitive Computation*, *10*(2), 296–306. https://doi.org/10.1007/s12559-017-9511-3

26. Hernaus, D., Xu, Z., Brown, E. C., Ruiz, R., Frank, M. J., Gold, J. M., & Waltz, J. A. (2018). Motivational deficits in schizophrenia relate to abnormalities in cortical learning rate signals. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(6), 1338–1351. https://doi.org/10.3758/s13415-018-0643-z

27. Quaedflieg, C. W. E. M., Stoffregen, H., Sebalo, I., & Smeets, T. (2019). Stress-induced impairment in goal-directed instrumental behaviour is moderated by baseline working memory. *Neurobiology of Learning and Memory*, *158*, 42–49. https://doi.org/10.1016/j.nlm.2019.01.010

28. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron*, *38*(2), 329–337. https://doi.org/10.1016/S0896-6273(03)00169-7

29. Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

30. Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15*(4), 535–547. https://doi.org/10.1016/S0893-6080(02)00047-3

31. Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., & Guillot, A. (2005). Actor–Critic Models of Reinforcement Learning in the Basal Ganglia: From Natural to Artificial Rats. *Adaptive Behavior*, *13*(2), 131–148. https://doi.org/10.1177/105971230501300205

32. Zhang, Z., Cheng, Z., Lin, Z., Nie, C., & Yang, T. (2018). A neural network model for the orbitofrontal cortex and task space acquisition during reinforcement learning. *PLOS Computational Biology*, *14*(1), e1005925. https://doi.org/10.1371/journal.pcbi.1005925

33. Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

34. Daniel, R., Radulescu, A., & Niv, Y. (2020). Intact Reinforcement Learning But Impaired Attentional Control During Multidimensional Probabilistic Learning in Older Adults. *Journal of Neuroscience*, *40*(5), 1084–1096. https://doi.org/10.1523/JNEUROSCI.0254-19.20

35. Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, *8*(1), 1768. https://doi.org/10.1038/s41467-017-01874-w

36. Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLOS Computational Biology*, *14*(4), e1006043. https://doi.org/10.1371/journal.pcbi.1006043

37. Gershman, S.J., Jones, C.E., Norman, K.A., Monfils, M.H., Niv, Y. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*. 7: 164

38. Babayan, B.M., Uchida, N. & Gershman, S.J. Belief state representation in the dopamine system. *Nat Communications* 9, 1891 (2018). https://doi.org/10.1038/s41467-018-04397-0

39. Samejima, K., & Doya, K. (2007). Multiple representations of belief states and action values in corticobasal ganglia loops. *Annals of the New York Academy of Sciences*, *1104*, 213–228. https://doi.org/10.1196/annals.1390.024

40. Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. *Nature neuroscience*, *20*(4), 581–589. https://doi.org/10.1038/nn.4520

41. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

42. Shahar, N., Moran, R., Hauser, T. U., Kievit, R. A., McNamee, D., Moutoussis, M., Consortium, N., & Dolan, R. J. (2019). Credit assignment to state-independent task representations and its relationship with model-based decision making. *Proceedings of the National Academy of Sciences*, *116*(32), 15871–15876. https://doi.org/10.1073/pnas.1821647116

43. McDougle, S., & Collins, A. (2020). Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. *Psychonomic Bulletin & Review* [In Press]. https://doi.org/10.3758/s13423-020-01774-z

44. Everitt, B. J., & Robbins, T. W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annual Review of Psychology*, *67*, 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

45. McDougle, S. D., Boggess, M. J., Crossley, M. J., Parvin, D., Ivry, R. B., & Taylor, J. A. (2016). Credit assignment in movement-dependent reinforcement learning. *Proceedings of the National Academy of Sciences*, *113*(24), 6797–6802. https://doi.org/10.1073/pnas.152366911

46. McDougle, S. D., Butcher, P. A., Parvin, D. E., Mushtaq, F., Niv, Y., Ivry, R. B., & Taylor, J. A. (2019). Neural Signatures of Prediction Errors in a Decision-Making Task Are Modulated by Action Execution Failures. *Current Biology*, *29*(10), 1606-1613.e5. https://doi.org/10.1016/j.cub.2019.04.011

47. Collins, A.G.E. & Frank, M.J. (2013). Cognitive control over learning: Creating, clustering and generalizing task-set structure. *Psychological Review, 120,* 190-229.

48. Franklin, N.T. & Frank, M.J. (2018). Compositional clustering in task structure learning. *PLOS Computational Biology, 14(4): e1006116,*.

49. Gibson, J., J. (1977). "The theory of affordances". In Perceiving acting, and knowing, Edited by: Shaw R. Bransford J. 67-82. Hillsdale, NJ: Lawrence Eribaum

50. Khetarpal, K., Ahmed, Z., Comanici, G., Abel, D., & Precup, D. (2020). What can I do here? A Theory of Affordances in Reinforcement Learning. *ArXiv:2006.15085 [Cs, Stat]*. http://arxiv.org/abs/2006.15085

51. Eckstein, M. K., & Collins, A. G. (in press). Computational evidence for hierarchically-structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences.*

52. Ballard, I., Miller, E. M., Piantadosi, S. T., Goodman, N. D., & McClure, S. M. (2018). Beyond Reward Prediction Errors: Human Striatum Updates Rule Values During Learning. *Cerebral Cortex*, *28*(11), 3965–3975. https://doi.org/10.1093/cercor/bhx259

53. Badre, D. & Frank, M.J. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 2: Evidence from fMRI. *Cerebral Cortex, 22,* 527-536.

54. Frank, M.J. & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex, 22,* 509-526.

55. Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235. https://doi.org/10.1016/j.tics.2007.04.005

56. Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*, 659-669.

57. Langdon, A. J., Sharpe, M. J., Schoenbaum, G., & Niv, Y. (2018). Model-based predictions for dopamine. *Current Opinion in Neurobiology*, *49*, 1–7. https://doi.org/10.1016/j.conb.2017.10.006

58. Sharpe, M. J., Batchelor, H. M., Mueller, L. E., Yun Chang, C., Maes, E. J. P., Niv, Y., & Schoenbaum, G. (2020). Dopamine transients do not act as model-free prediction errors during associative learning. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-019-13953-1

59. White, J. K., Bromberg-Martin, E. S., Heilbronner, S. R., Zhang, K., Pai, J., Haber, S. N., & Monosov, I. E. (2019). A neural network for information seeking. *Nature Communications*, *10*(1), 1–19. https://doi.org/10.1038/s41467-019-13135-z

60. Mikhael, J. G., Kim, H. R., Uchida, N., & Gershman, S. J. (2019). *Ramping and State Uncertainty in the Dopamine Signal* [Preprint]. Neuroscience. https://doi.org/10.1101/805366

61. Charpentier, C. J., Bromberg-Martin, E. S., & Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7255–E7264. https://doi.org/10.1073/pnas.1800547115

62. Honig, M., Ma, W. J., & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences*, *117*(15), 8391–8397. https://doi.org/10.1073/pnas.1918143117

63. Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280. https://doi.org/10.1016/j.cognition.2008.08.011

64. Mas-Herrero, E., Sescousse, G., Cools, R., & Marco-Pallarés, J. (2019). The contribution of striatal pseudo-reward prediction errors to value-based decision-making. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.02.052

65. Diuk, C., Tsai, K., Wallis, J., Botvinick, M., & Niv, Y. (2013). Hierarchical Learning Induces Two Simultaneous, But Separable, Prediction Errors in Human Basal Ganglia. *The Journal of Neuroscience*, *33*(13), 5797–5805. https://doi.org/10.1523/JNEUROSCI.5445-12.2013

66. Ribas-Fernandes, J. J. F., Shahnazian, D., Holroyd, C. B., & Botvinick, M. M. (2019). Subgoal- and goal-related reward prediction errors in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *31*(1), 8–23. https://doi.org/10.1162/jocn_a_01341

67. Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working Memory Load Strengthens Reward Prediction Errors. *Journal of Neuroscience*, *37*(16), 4332–4342. https://doi.org/10.1523/JNEUROSCI.2700-16.2017

68. Collins, A. G. E., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working Memory Contributions to Reinforcement Learning Impairments in Schizophrenia. *Journal of Neuroscience*, *34*(41), 13747–13756. https://doi.org/10.1523/JNEUROSCI.0989-14.2014

69. Renteria, R., Baltz, E. T., & Gremel, C. M. (2018). Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nature Communications*, *9*(1), 1–11. https://doi.org/10.1038/s41467-017-02615-9