The Role of Executive Function in Shaping Reinforcement Learning
Milena Rmus1, Samuel D. McDougle2, Anne G. E. Collins1,3

1Department of Psychology, University of California, Berkeley
2Department of Psychology, Yale University
3Helen 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 some aspects of learning. However, the neural computations that are explained by RL algorithms fall short of explaining many sophisticated aspects of human decision making, including the generalization of learned information, one-shot learning, and the synthesis of task information in complex environments. Instead, these aspects of instrumental behavior are assumed to be supported by the brain’s executive functions (EF). We review recent findings that highlight the importance of EF in 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 the 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 RL computations. In particular, cortico-striatal loops enable state-dependent value-based choice selection [9]. Furthermore, dopaminergic signaling of reward-prediction errors (RPEs) in the midbrain and striatum induces 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 different cognitive systems.

Executive functions (EF) have been identified as psychological faculties that could 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. Another 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]. This suggests that RL operates with EFs like working memory for 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’s contributions to learning are not limited to providing a separate learning substrate: Rather, EF may also directly contribute 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 motivated behavioral experiments and computational modeling approaches aimed at identifying EF-RL interactions [20, 21, 4, 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 considers EF contributions.
The ingredients of RL computations

Past work suggests that a specific brain network (primarily cortico-striatal loops) supports RL computations, like 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 or contexts an agent may be in), 2) the action space (reflecting the possible choices to be made), and 3) the reward function (defining valuable outcomes). The specification of these variables can dramatically impact the behavior of a decision-making agent. How these three variables are input to the brain’s RL network is poorly understood - we explore here a role for EF in contributing this information.

Figure 1. Schematic of EF contributions (WM, attention) to 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).

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 crucially 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 difficulty.

Simplifying the state space is a function sometimes attributed to attentional filters, which can specify important features of the environment [4]. In this framework, attention tags the task features that RL computes over [32, 33, 14; 34]. 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) use it as the relevant state space for RL. That way, an agent can eschew computing values over a larger state space of all possible features [35, 36]. 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 the 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). Thus, defining a low-dimensional representation which can be applied to multiple tasks in service of goal pursuit makes learning both more flexible and robust [37].

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 [38]. Recent work in animals suggests that RL computations in the striatum are likely performed over these latent belief states [39, 40]. For example, by showing markedly different dopamine dynamics if an expected reward is sure to arrive (e.g., 100% chance) versus almost sure to arrive (e.g., 90% chance) [41]. In this example, an inference about the latent state, which indicates whether 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, further supporting a link between EF and RL [42].

**Action space**

Above we reviewed a role of EF, 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 play 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 basic 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 may be 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 as well as to task-irrelevant motor actions (i.e., reinforcing a right-finger button press regardless of the stimulus that was present) [43]. 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 can exert an influence on both choices and reaction times [44], particularly when cognitive load is high. One hypothetical consequence of 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 bias could be linked to maladaptive forms of habitual behavior, such as addiction [45].

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 [46, 47]. These results suggest that simple cognitive inferences about the cause of errors (e.g., choice errors versus action errors) affect RL computations.
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 [48, 49], 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, selecting a task-set can itself be seen as a choice made in a high-level context. Learning to make this abstract choice can also occur via RL, such that RL computations happen over two different state-action spaces in parallel -- an abstract context and task-set space, and a more concrete stimulus-action space [50, 51]. There is recent computational, behavioral, and neural evidence that hierarchies of RL computations happen in parallel over more and more abstract types of states and choices, facilitating complex learning abilities [52, 53, 50]. Such learning may be supported by hierarchies of representations in prefrontal context [54, 55]. This again highlights a role for EF in setting the stage for RL computations to support complex learning.

**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 are primary or secondary rewards, such as food, money, or points. However, the path to an RPE is not always 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 [56, 57].

**Figure 3.** Contribution of executive functions in action selection.
These findings suggest that the currency of RL computations goes beyond primary and secondary rewards. There is early evidence that EF could be implicated in setting this currency.

One such example relates to the value of information. Humans are motivated to reduce uncertainty about their environment [58]. 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 [59]. 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 [60], even when it is not valenced [58]. The prefrontal cortex also appears to track information and uncertainty [41], which can be held in working memory to influence decision making [61] (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 [62]. 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. These subgoals generate pseudo-rewards in hierarchical RL models, and appear to drive activity in the canonical reward-processing regions in the brain, even though these rewards are 1) not inherently rewarding, and 2) are clearly distinguished from terminal rewards [63, 64]. The processing of pseudo-rewards is additionally assumed to be driven by the prefrontal cortex, suggesting a link to EF [65].

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 by the RL network. For example, work by Collins and colleagues [66] 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 RL RPEs \[20,21\]. 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 \[67\], and addiction \[68, 3\] 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


68. 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](https://doi.org/10.1038/s41467-017-02615-9)