Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Current Opinion in Neurobiology

Prefrontal executive function and adaptive behavior in complex environments

Etienne Koechlin

The prefrontal cortex (PFC) subserves higher cognitive abilities such as planning, reasoning and creativity. Here we review recent findings from both empirical and theoretical studies providing new insights about these cognitive abilities and their neural underpinnings in the PFC as overcoming key adaptive limitations in reinforcement learning. We outline a unified theoretical framework describing the PFC function as implementing an algorithmic solution approximating statistically optimal, but computationally intractable, adaptive processes. The resulting PFC functional architecture combines learning, planning, reasoning and creativity processes for balancing exploitation and exploration behaviors and optimizing behavioral adaptations in uncertain, variable and open-ended environments.

Address
INSERM, Département d'Etudes Cognitives, Ecole Normale Supérieure,
29 rue d'Ulm, 75005 Paris, France

Corresponding author: Koechlin, Etienne (etienne.koechlin@upmc.fr)

Current Opinion in Neurobiology 2016, **37**:1–6

This review comes from a themed issue on **Neurobiology of cognitive behavior**

Edited by **Alla Karpova** and **Roozbeh Kiani**

For a complete overview see the Issue and the Editorial

Available online 11th December 2015

http://dx.doi.org/10.1016/j.conb.2015.11.004

0959-4388/Published by Elsevier Ltd.

Introduction

Adaptive behavior is critical for organisms to survive in real-world situations that are often changing. Basal ganglia in vertebrates are subcortical nuclei including the striatum that are thought to implement basic adaptive processes akin to what is usually referred to as (temporal-difference) Reinforcement Learning [1–4]. RL consists of adjusting online stimulus–action associations to the rewarding/punishing values of action outcomes. Importantly, RL is both a very simple and robust process endowing the animal with the ability to learn optimal behavioral strategies even in complex and uncertain situations [5••]. In mammals, basal ganglia further form loop circuits with the prefrontal cortex (PFC) [6] to further the flexibility and complexity of the behavioral repertoire, in essence overcoming the critical limitations of the RL processes. Here, we review recent findings from both empirical and computational studies and outline a general theoretical framework describing the PFC function as implementing adaptive processes devoted to overcoming key RL adaptive limitations.

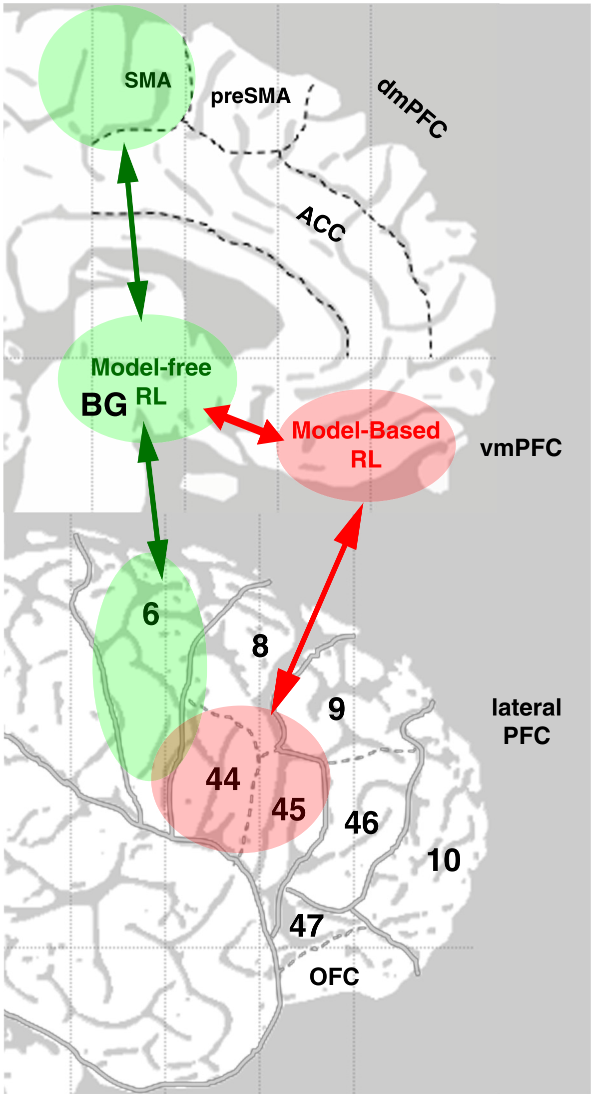
From reinforcement learning to adaptive planning

A first critical limitation in basic RL (also named *model-free* RL) is that behavior cannot adjust to internal changes in subjective values of action outcomes [7,8]. Consider, for instance, action A in a given situation leads to water and action B leads to food. If you are thirsty but replete, RL will reinforce action A relative to B in this situation. When the situation reoccurs, you will then select action A rather than B. If you are now hungry rather than thirsty, however, this is certainly a maladaptive behavior. The problem arises because basic RL make no distinctions between rewarding values of action outcomes and action outcomes *per se*.

Overcoming this limitation requires learning an internal model that specifies the outcomes resulting from actions, regardless of rewarding values. Learning this model is simply based on outcome likelihoods given actions and current states. This *predictive* model is thus learned besides the stimulus–action associations learned through RL (collectively named the *selective* model here). The predictive model especially enables to internally emulate RL without physically acting [5••]: This model predicts the outcomes of actions derived from the selective model, so that their rewarding values may be internally experienced according to the current motivational state of the agent (e.g. thirsty or hungry). Stimulus–action associations are then adjusted accordingly through standard RL algorithms. This emulation is commonly referred to as *model-based* RL [5••]. Behavior is thus adjusted to the agent's motivational state before acting and reflects internal planning. Model-based RL also enables to generally adapt faster than RL to external changes in action-outcome contingencies and/or outcome values [5••].

Empirical studies confirm that human behaviors cannot be fully explained by model-free RL, but instead have a model-based component. [9,10••,11]. Neuroimaging studies show that both the inferior parietal cortex and lateral PFC are involved in learning predictive models [12], with the former possibly encoding these models [13] and the latter, in association with the hippocampus, retrieving these models for emulating model-based RL [9]. Furthermore, empirical evidence argues that the orbitofrontal cortex (ventromedial PFC in humans) in association with the striatum encode action outcomes from predictive models and their actual rewarding values [14–18,19•]. Together, these studies also suggest that the ventromedial PFC may directly learn and encode simple predictive models directly mapping stimulus–action pairs onto expected valued outcomes [20•], while the inferior parietal cortex and lateral PFC may be involved in implementing more complex predictive models as multi-step state–action–state maps (Figure 1).

Figure 1



Current Opinion in Neurobiology

Reinforcement learning in the human frontal lobes. Schematic diagram showing main subcortical and cortical structures involved in reinforcement learning (RL). Green: brain regions involved in model-free RL. BG: basal ganglia. SMA: supplementary motor area. Numbers indicate broadmann's area. BA 6: premotor cortex. Red: Brain regions involved in model-based RL. vmPFC: ventromedial PFC. dmPFC: dorsomedial PFC. OFC: orbitofrontal cortex. ACC: anterior cingulate cortex. Arrows indicate critical interregional connectivity presumably underpinning RL.

Some authors have proposed that in the brain, model-free and model-based RL form two concurrent instrumental controllers. In this view, their relative contribution to action selection is a function of the relative uncertainty and/or reliability about reward and outcome expectations derived from selective and predictive models, respectively [21,22]. Others have proposed that model-free and model-based RL form two cooperative systems with model-free RL driving online behavior and model-based RL working off-line in the background to continuously adjust model-free RL [5••,10••,23]. Recent behavioral results support the second view [10••]. As shown below, this view is also more consistent with the present theoretical framework.

From adaptive planning to Bayesian inference

A second critical limitation of RL systems described above is that adapting and learning new external contingencies gradually erases previously learned ones. This again leads to maladaptive behavior in environments exhibiting periodically recurring external contingencies (i.e. recurrent situations): RL systems have no memory and need to entirely relearn previously encountered situations. In uncertain and open-ended environments where new situations may always arise (i.e. the environment corresponds to an infinite-multidimensional space), overcoming this limitation requires solving a nonparametric probabilistic inference problem [24•] for constantly arbitrating between continuing to adjust ongoing behavior through (model-free and/or model-based) RL, switching to previously learned behaviors and even creating/learning new behaviors. Previously learned behaviors along with the ongoing behavior thus form a collection of discrete entities stored in long-term memory and referred to as *task sets* [25]. Task sets are abstract instantiations of the situations the agent inferred to have encountered so far and comprises the selective and predictive model learned when the task set guided behavior [26]. Task sets further comprise an additional internal model — the *contextual* model of the likelihood of any external cues — learned when the task set guided behavior in the past [27,28], and likely encoded in lateral PFC regions [29,30]. The aforementioned arbitration problem has optimal statistical solutions based on Dirichlet process mixtures [31,32] which in practical cases, are actually computationally intractable and consequently, biologically implausible.

Recent studies, however, show that a biologically plausible, online algorithm approximating Dirichlet process mixtures can account for human behavior in both recurrent and open-ended environments [24•,33•,34••]. This