Motivation of extended behaviors by anterior cingulate cortex

Clay B. Holroyd1 and Nick Yeung2

1 Department of Psychology, University of Victoria, P.O. Box 3050 Victoria, BC V8W 3P5, Canada
2 Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

Intense research interest over the past decade has yielded diverse and often discrepant theories about the function of anterior cingulate cortex (ACC). In particular, a dichotomy has emerged between neuropsychological theories suggesting a primary role for ACC in motivating or ‘energizing’ behavior, and neuroimaging-inspired theories emphasizing its contribution to cognitive control and reinforcement learning. To reconcile these views, we propose that ACC supports the selection and maintenance of ‘options’ – extended, context-specific sequences of behavior directed toward particular goals – that are learned through a process of hierarchical reinforcement learning. This theory accounts for ACC activity in relation to learning and control while simultaneously explaining the effects of ACC damage as disrupting the motivational context supporting the production of goal-directed action sequences.

Theories of ACC function

Akinetic mutism (see Glossary) is a striking neurological condition characterized by dramatic reductions in spontaneous speech and movement despite preserved motor ability [1–3]. As one patient reported following remission of the illness:

She didn’t talk because she had ‘nothing to say’. Her mind was ‘empty’. Nothing ‘mattered’. She apparently was able to follow … conversations even during the early period of the illness, but felt no ‘will’ to reply. [1]

The disorder is typically associated with lesions to the anterior portion of midcingulate cortex [4] in a region often referred to as ACC [2,3]. On the basis of decades of such observations, neurologists proposed that ACC serves to motivate goal-directed behaviors [2]. A modern take on this idea holds that ACC provides a global ‘energizing’ factor necessary to support speeded responding [5,6], effortful behavior [7–10] and autonomic arousal [11]. Long-standing neuroimaging evidence of ACC activity during voluntary action selection also appears consistent with this hypothesis [12]. However, although useful as a point of departure, the proposal that ACC motivates goal-directed behavior relies on intuitive but computationally imprecise terms such as ‘effort’ and ‘energy’.

Meanwhile, current dominant theories of ACC function – derived primarily from neuroimaging data – surprisingly point to a role for ACC not in motivating effortful behavior per se but rather in decision making and the deployment of cognitive control. These theories emphasize the ACC contribution to instigating punctate, trial-to-trial changes in behavior such as an increase in the strength of top-down control following experienced response conflict [13,14] or adaptive modification of behavior as a result of action–reinforcement contingencies [15,16]. However, although these theories have enjoyed substantial support from confirmatory neuroimaging studies [13,14,16], they have yet

Glossary

Actor: in reinforcement learning theory, a module that executes the policy of the agent.
Akinetic mutism: neuropsychological disorder characterized by a reduction in or absence of spontaneous behavior in the presence of preserved motor ability.
Anterior cingulate cortex: region of the frontal midline cortex approximately ventral to the cingulate sulcus. This Opinion focuses on the dorsal and caudal portion of this region, which is alternatively termed the anterior midcingulate cortex.
Basal ganglia: collection of subcortical nuclei primarily concerned with motor control.
Conflict monitoring theory: theory that the anterior cingulate cortex is responsible for detecting the simultaneous activation of incompatible response channels to facilitate the deployment of cognitive control.
Critic: in reinforcement learning theory, a module that evaluates the value of the current state and computes reward prediction errors; also termed adaptive critic.
Midbrain dopamine system: collection of subcortical nuclei that project mainly to the basal ganglia and frontal cortex; the system releases the neurotransmitter dopamine, which is strongly implicated in motor activation and reinforcement learning.
Hierarchical reinforcement learning: branch of reinforcement learning theory concerned with hierarchical organization of behavior.
Option: in hierarchical reinforcement learning, a temporally abstract behavior that describes extended and potentially variable sequences of actions.
Policy: in reinforcement learning theory, a mapping of states to actions that determines the agent’s behavior.
Primitive action: in hierarchical reinforcement learning, the smallest unit of behavior that implements simple mappings between stimuli and responses.
Pseudo-reward: in hierarchical reinforcement learning, a quantity that indicates the abstract reward value associated with a subgoal.
Reinforcement learning: process by which rewards and punishments adaptively modify behavior.
Reward positivity: component of the event-related brain potential, also termed the feedback error-related negativity, that indexes a mechanism for reward processing; it is hypothesized to reflect the impact of dopaminergic reward prediction error signals on the anterior cingulate cortex to facilitate adaptive decision-making.
Reward prediction error: in reinforcement learning theory, the instantaneous change in value; positive and negative reward prediction error signals are said to indicate that ongoing events are better or worse than expected, respectively.
Task set: set of mappings between stimuli and responses necessary to effect flexible task-dependent behavior.
Top-down biasing signals: excitatory activity from the prefrontal cortex that facilitates task-dependent processing in other brain areas.
Value: in reinforcement learning theory, a quantity that predicts expected cumulative reward given the system state and policy.
to be satisfactorily reconciled [17]. Of still greater concern, a growing body of lesion and neurophysiological studies have yielded inconsistent or contradictory results for both the conflict [14,18] and reinforcement learning (RL) [19–23] hypotheses, because ACC lesions in humans typically result in global slowing and increased response variability rather than inflexibility of control or an inability to learn from feedback [5].

Why have neuropsychological and neuroimaging approaches not converged on a unified theory of ACC function? We suggest that research to date has focused too narrowly on the types of simple stimulus–response associations that characterize response-conflict and trial-and-error learning tasks, and has therefore neglected the high-level structure that characterizes everyday human behavior. In this Opinion we argue that ACC supports the selection and execution of coherent behaviors over extended periods [24], an idea we formalize in terms of recent advances in RL theory that utilize a hierarchical mechanism for action selection, hierarchical reinforcement learning (HRL) [25]. In this view, ACC is more concerned with the selection and maintenance of the task itself than with the minutiae of task execution. Thus, ACC would be responsible for engaging in a psychology experiment until its completion as opposed to implementing subtle behavioral adjustments along the way. As we argue below, this proposal holds the promise of reconciling diverging theories of ACC function into a formal, unified theoretical framework.

**ACC and option selection**

HRL can provide increased computational efficiency over standard RL approaches for problems involving extended sequences of actions (Box 1). HRL incorporates the concept of options that represent action policies comprising sequences of simple primitive actions. As an everyday example, a set of primitive actions might consist of the individual steps needed to drive a car—releasing the emergency break, turning on the ignition, pressing the gas pedal, etc.—whereas an option might comprise the sequence of primitive actions that bring a driver to a specific supermarket. Crucially, each option comprises not only the given sequence of actions, but the entire set of actions that map various possible initiation states to the goal state, such that a single option could be employed to reach the market from home, work or any arbitrary starting point. Thus, options are defined by their associated goal states (the market) and the set of initiation states that trigger the option (hunger, access to a vehicle, etc.), in addition to the action sequences (the policy) that map the transitions from initiation states to the goal state. The increased computational efficiency of HRL results from the ability to learn and organize sequences of behavior at the option level (drive to the market, get groceries, drive to school, pick up the kids, etc.) rather than at the level of primitive actions (stop at 5th and Main, turn left, accelerate on Main, etc.).

Recently, Botvinick and colleagues have explored how HRL principles might be implemented by the brain [25]. A key insight of their work is that the option idea maps neatly onto the concept in cognitive psychology of the task set, the set of a mappings between stimuli and responses necessary to effect flexible task-dependent behavior, such as stopping the car at red lights and accelerating at green lights. Moreover, they suggested that this function could be implemented by prefrontal cortex, a region widely believed to be involved in supporting task sets [26]. Here we develop their account by proposing that ACC, rather than prefrontal cortex, is responsible for option selection and maintenance.

Our proposal builds on the well-established actor–critic architecture, which separates control into parallel modules for action selection and performance monitoring [25,27], but departs from previous work by placing ACC at the apex of both pathways (Figure 1). Specifically, we propose that ACC selects and maintains options, that dorsolateral prefrontal cortex (DLPFC) and motor structures in the dorsal striatum (which together comprise the actor) execute those options, and that orbitofrontal cortex and the ventral striatum (which together comprise the critic) evaluate progress toward the goal states of the options. This proposal leverages existing concepts about the computational function of these systems with which ACC interacts. First, the dorsal striatum implements the policy of the actor, for example, by stopping the car at red lights and accelerating at green lights [27–29]. Second, the ventral striatum,
which forms the core of the critic, evaluates whether or not ongoing events predict future reward (or punishment), for example, by indicating that accelerating at the red light constitutes maladaptive behavior [27–29]. Third, DLPFC provides top-down biasing signals to the dorsal striatum (and other brain areas) that facilitate execution of the current policy [14]; these signals are most important when the appropriate policy has not been learned by the dorsal striatum or is inconsistent with past behavior, such as when driving on the wrong side of the road in a foreign country. Finally, orbitofrontal cortex provides the ventral striatum with information related to abstract goals, consistent with its role in contextually based action and reward evaluation [30,31]; this information affords the basal ganglia flexibility to learn not only about primary rewards and punishments (such as the pain resulting from a car accident) but also about goal-related outcomes (such as stopping at a red light).

This neurocomputational architecture, which underlies several popular models of cognitive control, RL and decision making [25,27,28], leaves several important issues unaddressed. First, although the model provides a role for the basal ganglia in learning about and executing simple stimulus–response mappings, it falls short of describing how the system can do so efficiently for complex sequences of actions – such as driving the supermarket and returning home with a bag of groceries – because the computational load on the system increases nonlinearly with the number of steps comprising the action sequence [25]. Second, the architecture does not specify what task DLPFC should implement nor what goal orbitofrontal cortex should take as appropriate for the current task context. Third, the framework leaves undetermined the degree of vigor with which the task should be executed.

We argue that ACC, in its role in selecting and maintaining options according to principles of HRL, provides the solution to these problems. By design, the HRL framework alleviates the computational burden on systems such as the basal ganglia that are responsible for learning about primitive actions (Box 2). Second, ACC learns to associate
Box 2. ACC and the basal ganglia

The basal ganglia have an extensively documented role in reinforcement learning (RL). In animals, dopaminergic reward signals facilitate long-term potentiation at cortico-striatal synapses in a manner that supports instrumental conditioning [58], and basal ganglia lesions retard such learning [59]. In humans, neuroimaging studies reveal dissociable correlates of actor and critic components of an RL system in the dorsal and ventral striatum, respectively [29]. Several influential models have converged to document the neurocomputational mechanisms by which the basal ganglia implement associative learning on the basis of dopaminergic input according to RL principles [28].

Several prominent theories have also proposed a role for ACC in associating actions with their outcomes [15,16]. However, these theories have rarely addressed – and never satisfactorily answered – the question of the specific contribution of ACC to this process: If the function ascribed to ACC overlaps so closely with those more typically associated with the basal ganglia [28], then what is the specific ACC contribution to RL? This question is complicated by the fact that ACC damage results in varied task-dependent RL impairments. For example, in humans, ACC lesions promote response slowing and variability [5] but spare the ability to learn from feedback in the Wisconsin Card Sort Test [20]. In non-human primates, ACC lesions spare acquisition of a new response to conditioned reinforcement [22] but impair performance when animals are required to sustain a rewarded response across multiple trials [19,21]. It has also been observed that such lesions increase error rates in a task-switching paradigm because of an apparent failure of subjects to sustain attention to the task [23].

HRL theory offers a clear solution to this conundrum. It proposes that ACC and the basal ganglia operate in parallel and according to common RL principles, but do so at different levels of hierarchical organization: ACC selects, maintains and learns about high-level options, whereas basal ganglia are concerned with lower-level actions. However, in tasks without a hierarchical structure, ACC may come to encode lower-level actions redundantly with the basal ganglia. A key feature of this proposal is that ACC will be most involved in tasks that involve reward integration across multiple trials [60], which ACC utilizes to learn not the value of individual actions but rather the value of the task itself. ACC damage then results in a failure to associate the task with a positive value, resulting in decreased attention to task demands and a concomitant inability to sustain optimal performance [19,21,23].

Dopamine and option maintenance

We propose that option selection and maintenance in ACC are supported by input from the midbrain dopamine system. The dopamine system carries both phasic (brief) and tonic (extended) signals that mediate related but distinct functions [32]. An influential theory holds that the phasic component constitutes a reward prediction error signal as defined within the standard RL framework [27], which we suggest is utilized by ACC to learn option values [15] (Box 3). By contrast, the tonic component is said to be responsible for motivating the pursuit of hedonic rewards by coding for expected future reward [33] or for promoting effortful behavior by coding for the average reward rate [34]. Tellingly, both systematic injection of a dopamine antagonist [35] and direct ACC lesions [10] impair effortful behavior. This process appears to be mediated by the dopamine–ACC interface, because disruption of midbrain dopamine projections to ACC [36] and infusion of a D1 receptor antagonist into ACC [37] also impair effortful behavior. In addition, loss of dopaminergic input to ACC in humans promotes akinesis in movement [3]. Neurocomputational theories of decision making have thus pointed toward the dopamine–ACC interface as a crucial nexus for the selection and execution of effortful behaviors [38–40].

How might dopamine facilitate effortful behavior by ACC? An influential theory of the impact of DA on frontal cortex holds that dopamine levels regulate the balance of D1 and D2 receptor-dependent synaptic activity: D1 receptor activation, promoted by high levels of dopamine, favors stable working memory representations; by contrast, D2 receptor activation, promoted by lower levels of dopamine, favors response flexibility and task switching [41]. These network dynamics may constitute a mechanism for gating and maintaining information in working memory [42]. Understood in the HRL framework, low dopamine levels could facilitate the gating of a high-valued option into working memory (option selection), whereas high dopamine levels could maintain that information in working memory until the option is completed (option maintenance). In this way the system might protect options associated with high-valued subgoals (such as completion of a long, tedious drive for pay) against primitive actions that provide relief against immediate costs but that also impede progress towards the overall goal (such as pausing for libations along the way).

The HRL framework and previous theories of ACC function

The HRL framework incorporates key elements of existing theories of ACC while addressing many of their weaknesses. First, the name itself – hierarchical reinforcement

values with different options and chooses the appropriate option for the current environmental state according to standard RL principles. In this way, ACC decides what task to perform and then directs DLPFC to implement that task, which in turn provides top-down control signals to the dorsal striatum to facilitate execution of the chosen policy. Thus, whereas ACC selects the option-specific policy, DLPFC and the dorsal striatum together execute that policy. Note that a key difference between the role of DLPFC and ACC is that the latter instigates switches between tasks to achieve a higher-level goal (get in the car, drive to the market, get groceries, etc.), whereas the former implements the task at hand (e.g. drive to the market). Furthermore, orbitofrontal cortex associates the termination state of each option with pseudo-reward (Box 1), providing contextually appropriate reward information to the ventral striatum such that, for example, the system is provided with a pseudo-reward on successful completion of the drive-to-market option.

Finally, we propose that ACC not only chooses the option but also determines the level of effort to be applied toward executing the policy, and maintains this signal until the option reaches its termination state. Thus, when the selected option is associated with a high value, ACC directs DLPFC to exert vigorous top-down control. Conversely, when the level of activation provided by ACC to DLPFC is weak because the option is undervalued, then the level of control exerted by DLPFC over the basal ganglia is commensurately weak and subject to decay. As a consequence, responses associated with the desired task set become slow and variable and ultimately dominated by the primitive actions mediated by the basal ganglia: reflexive and characterized by immediate gratification.
Box 3. ACC, dopamine and option-value learning

We propose that reward prediction error signals (RPEs) communicated to anterior cingulate cortex (ACC) from the midbrain dopamine system underlie the development of option-specific values used for option selection. RPEs, which are fundamental to RL algorithms for solving complex sequence learning and control problems [27,42], are believed to be carried by the fast phasic component of dopamine neuron activity: brief increases in the firing rate code for positive RPEs, which indicate that events are better than expected, and brief pauses in the firing rate code for negative RPEs, which indicate that events are worse than expected. The learning process might be mediated directly by the impact of dopamine RPEs on pyramidal cells in ACC [15] (but see [61]) or indirectly by other mechanisms [62,63]. Irrespective of the neural implementation, these RPEs would enable the system to learn that the task itself is valuable rather than the individual components that comprise the task.

We have previously proposed that the impact of these fast phasic dopamine signals on ACC elicits a component of the human event-related brain potential (ERP) [15]: commonly referred to as error-related negativity, this ERP component has recently been termed the correct-related or reward positivity owing to its observed sensitivity to positive rather than negative RPEs [64,65]. Several sources of evidence support the dopamine-ERP link. First, functional magnetic resonance imaging (fMRI) and ERP findings indicate that ACC RPEs are highly correlated with neural activity in the ventral tegmental area (the source of dopamine projection to the cortex) [66] and in the ventral striatum (a primary target of the midbrain dopamine system) [65]. Second, the timing of dopamine RPEs recorded from the human midbrain [67] coincides with that of the reward positivity recorded at the scalp, suggesting a robust functional connection. Third, the reward positivity is highly sensitive to neurological and pharmacological insults of the DA system [68]. And fourth, a homolog of the reward positivity has been identified in the monkey cingulate sulcus [69], the scalp manifestation of which is sensitive to administration of dopamine antagonists [70].

Our proposal finds support from a recent ERP and fMRI study that identified ACC-dependent RPEs associated with progress towards a subgoal as formally defined within the HRL theoretical framework [54]. Taken together, these considerations converge in suggesting that reward-related scalp potentials reflect the impact of dopamine RPE signals on ACC for learning option-specific values [15], while providing a computationally precise formalization of the underlying mechanisms. Thus, because each individual action assumes a small cost, sequential behavior will be associated with a positive value only when taken as a whole. When the whole is disrupted, as occurs following ACC lesions, the costs associated with individual actions become prohibitive and action initiation is suppressed.

Third, the theory holds that option selection and maintenance – not conflict monitoring – are the cardinal ACC functions. According to this view, ACC activation in conflict tasks reflects the broader role of the region in maintaining task sets and sustaining effortful behavior, with other regions such as posterior parietal cortex responsible for resolving interference effects [45]. Instead, just as ACC integrates reward signals across trials to determine the appropriate level of task engagement, ACC integrates information about costs and effort – both correlated with but not reducible to conflict – to the same end [17]. Thus, ACC damage spares conflict adaptation effects [14,18] while causing a more global slowing of responses [5] and, in at least some patients, a lack of awareness of conflict-related costs [8] in these tasks.

Finally, our proposal dovetails with recent neuroimaging and neurophysiological evidence that ACC is active during voluntary task selection [46] and task switching [47,48]. Once a task is selected, ACC appears to provide stable, across-trial maintenance of the task set in support of goal-directed behavior [49] by facilitating DLPFC function towards this end [6]. Hence, ACC lesions result in increased and more variable response times in humans [5] and in impaired attention to task demands and disrupted task-switching in non-human primates [23]. Detailed examination of ACC activity from rats [50] to monkeys [51,52] to people [53] provides converging evidence that the region is responsible for managing transitions between successive task stages, consonant with its organization of multiple options across sequences of behavior. In this way, our theory aligns with recent suggestions that ACC and DLPFC implement high-level cognitive control by organizing behavior according to extended mental programs rather than by selecting individual discrete actions [53].

Future directions

The HRL framework suggests several productive avenues for future research. In particular, the theory points to the development of novel tasks involving a high-level structure to demonstrate the crucial contribution of ACC in motivating extended behaviors. For example, ACC damage should result in abnormal performance in voluntary task-switching studies (i.e. option selection) and in vigilance tasks that require infrequent responding (option maintenance). Along these lines, encouraging evidence from a recent study demonstrated the sensitivity of ACC to HRL reward-prediction error signals in a task in which a subgoal must be completed (collection of a package) before an overall goal is reached (delivery of the package to a set location) [54]. In addition, the hierarchical structure imposed by ACC should be necessary for problems that the basal ganglia would otherwise find computationally intractable. This would be especially important when the outcome of a given action is
立即有奖（或惩罚）但又是同时性预测为差（或好）的总体结果，尤其是那些达到一个峰的障碍，而且在学习过程中，要么是相互关联的，要么是相互依赖的。Box 4

我们的计算模型关注于解释行为性证据和定义判别抽象认知功能的多个大脑区域，而不是提供详细的神经生理学水平的转换过程。[28]。一个关键的方向是未来研究，其焦点是准确地理解选择性编码的特征，以及这些特征如何在ACC中被编码。计算模型通过模拟盲目的执行性行为暗示，表明高级任务结构可能被编码为内部的动态性，以及这些动态性如何在系统层面上，通过多维尺度分析来反映。[55]。严格地说，类似统计技术对神经元数据从大鼠ACC数据表明，神经元组成代表动物行为模式的跟踪。参考文献


活动应特别敏感于多巴胺输入到ACC。证据表明，MRI多元模式分析可以识别任务选择的激活模式在内侧前额叶皮层[57]也提高了识别任务特异性信号在人脑的可能性，因为HRL框架表明，这表明不同的选项可能与特定类型的ACC活动有关。具体而言，我们预测似乎任务独立的持久活动，其在ACC之前已观察到[49]，会以更精细的分析，反映活动模式的差异，这些差异是针对特定任务的。

结论

我们探讨了一个鲜为人知的窘境：当前，主要的ACC理论来自于神经成像文献，它试图解释ACC损伤导致的临床观察，减少自发性语言和行为。我们推测，这个链接是考虑该体系结构的协调性问题，从而推测，ACC应负责学习和选择高级水平行为的计划，这些计划提供了意义的背景，并且是我们行动的动机。

致谢

C.B.H.受加拿大研究委员会、Michael Smith健康研究奖学金奖和自然科学与工程研究理事会的加拿大发现基金资助。N.Y.受国家健康研究所（P50-MH62916）资助。


