Opinion

Cell PRESS

Motivation of extended behaviors by anterior cingulate cortex

Clay B. Holroyd[1] and Nick Yeung[2]

[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]. Longstanding 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 the 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.

Corresponding author: Holroyd, C.B. (holroyd@uvic.ca)

1364-6613/$ – see front matter doi:10.1016/j.tics.2011.12.008 Trends in Cognitive Sciences, February 2012, Vol. 16, No. 2

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,

Box 1. Reinforcement learning: standard and hierarchical approaches

Reinforcement learning (RL) algorithms provide a simple but powerful framework for understanding how agents learn to behave in complex and uncertain environments [27]. Standard RL approaches find an intuitive and influential implementation in actor–critic architectures, which propose a division of labor between two components of the learning system: an actor that selects actions according to their weighted associations with the current state of the world (termed a policy), and a critic that generates an estimate of the long-term reward associated with each world state (termed a value function). The policy of the actor and the value function of the critic are both learned through experience, specifically through computation by the critic of a reward prediction error that indicates whether ongoing events are better or worse than expected. Suppose, for example, that a rat has just discovered food in the left arm of a novel T-maze. This outcome would be coded by the critic as a positive prediction error – events are better than expected – that would lead to an increase in the value of the immediately preceding world state (the choice point of the T-maze) and a strengthening of the association of the immediately preceding action (turning left) with that state. Next time, the rat would tend to turn left when put back in the maze.

RL algorithms are powerful enough to find routes through more complex mazes, forage efficiently for food, and even play international-class backgammon. However, they become increasingly inefficient as the world to be learned about becomes more complex in terms of the number of possible states and available actions. The resulting combinatorial explosion renders standard RL infeasible in even moderately complex tasks. Hierarchical RL algorithms attempt to address this scaling problem by grouping together interrelated states and actions to form higher-level behavioral plans – termed options – that comprise structured sequences of actions directed towards specified subgoals [25]. Options can be learned about as coherent, temporally extended steps towards the overall goal, reducing the complexity of the learning task. Importantly, learning occurs via straightforward extensions of standard RL: options that lead to better-than-expected outcomes are reinforced, whereas successful completion of a chosen option serves as a pseudo-reward that reinforces preceding lower-level actions according to the same RL principles. In effect, the learning task is solved simultaneously at different levels of abstraction, identifying both low-level actions and high-level options that most efficiently achieve their respective goals.

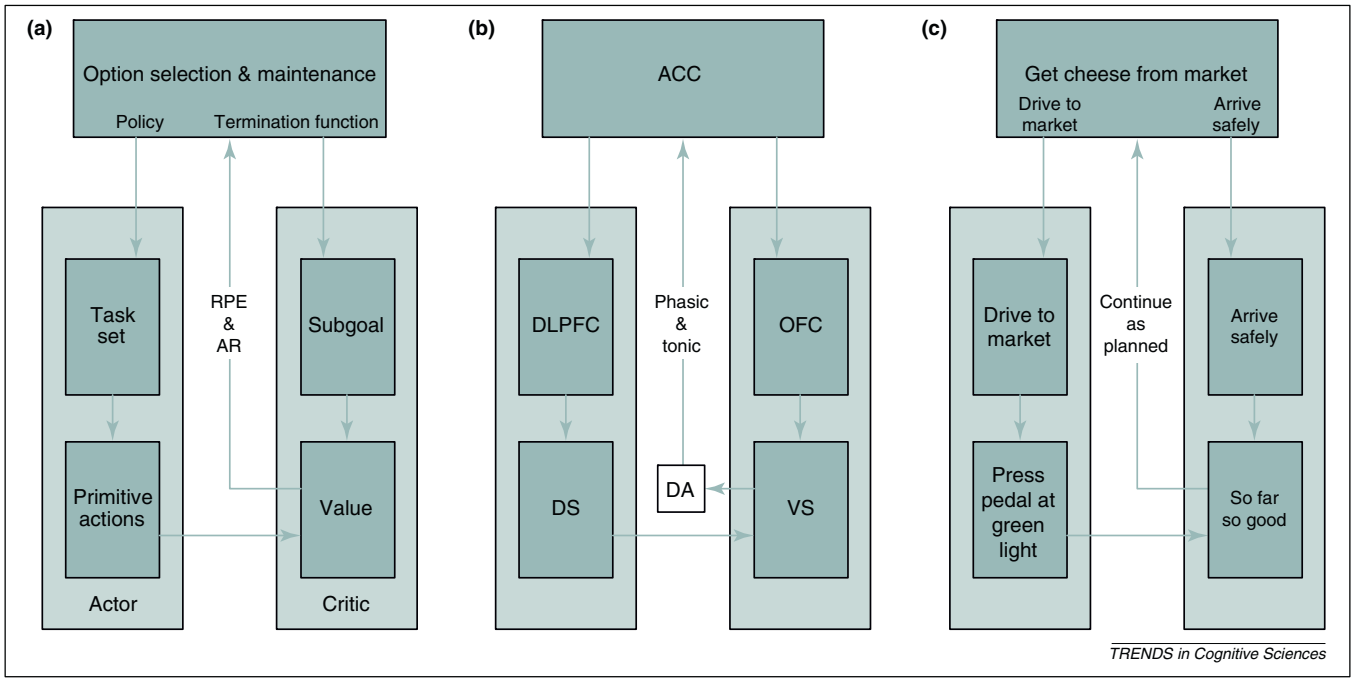

Figure 1. Proposed implementation of the hierarchical reinforcement learning mechanism. **(a)** Abstract function associated with each component. The option selection mechanism sits at the apex of a standard actor–critic architecture for reinforcement learning: it determines the appropriate task to implement given the state of the external environment and specifies the goal-state defining successful task completion. The selected option-specific policy is communicated to the actor, which implements the policy via two interacting modules. A high-level module implements the task set by biasing the activity of a low-level module, which in turn executes behaviors appropriate to the policy given the current state of the environment (not shown). In parallel, a high-level module within the critic associates the termination state of the option with a subgoal, providing pseudo-rewards and contextual information to a low-level critic module that evaluates the progress of the actor towards the option termination state. The critic outputs a slowly changing signal related to average reward and a fast reward prediction error signal indicating when events are better or worse than expected. The option mechanism utilizes these signals, together with information related to experienced costs (not shown), for learning the value of options, for selecting options for execution, and for maintaining the system on-task after an option is selected. Also not shown are additional connections to the actor and critic modules that carry reward-related information from the critic and state-related information from the external environment. **(b)** Proposed neural implementation of the hierarchy. **(c)** An illustrative example. The agent is at home fixing a macaroni and cheese dinner only to find that he missing a key ingredient: cheese. Confronted with this obstacle, ACC selects and coordinates a sequence of options for driving to a nearby market, purchasing the ingredient, and returning home (as opposed to an alternative sequence of options such as ordering the same meal at a local diner). The DLPFC manages the individual policies associated with each option, including driving to the market, by biasing neural activity in the dorsal striatum and in other motor structures that implement the particular steps of the sequence, such as accelerating at green lights. Note that the DLPFC and dorsal striatum work together to execute the policy, but DLPFC input is especially important for tasks that are incompatible with overlearned behaviors, for example, for driving on the wrong side of the road in a foreign country. Meanwhile, the orbitofrontal cortex represents the termination state of the option – arriving safely at the market – as the goal of the action sequence, and the ventral striatum utilizes this contextual information to determine whether or not the individual actions are consistent with the goal. Finally, the dopamine system indicates to ACC whether or not the current state of the task is associated with high predictive value (tonic dopamine) and when events are suddenly better or worse than predicted (phasic dopamine). Thus, if on a long drive the DLPFC momentarily loses control over the desired task set, such that the motor system turns onto the wrong side of the road, the critic can alert ACC via decreased dopamine levels that the action is inconsistent with the goal, which would in turn boost activation of the appropriate task set in DLPFC and correct for the error. ACC, anterior cingulate cortex; AR, average reward; DA, midbrain dopamine system; DLPFC, dorsolateral prefrontal cortex; DS, dorsal striatum; OFC, orbitofrontal cortex; RPE, reward prediction error; VS, 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 to 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].

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 biasing 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.

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 akinetic mutism [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

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].

learning – makes transparent the relationship between this proposal and standard RL theories of ACC function. Yet our position goes beyond current RL models by identifying the unique computational function of ACC: selection and maintenance of high-level plans associated with extended behaviors. Notably, option values used for selection are learned in part by integrating the cumulative reward received across the primitive actions that comprise the option [25]. Thus, the theory explains several findings that are otherwise troubling for standard RL accounts of ACC function, including recent extensions of these models [43], namely, that ACC lesions have little impact on trial-and-error learning tasks in which feedback relates to simple stimulus–action mappings rather than to global options [20] and, conversely, that ACC is activated by rewards and punishments that are not contingent on low-level actions but that nevertheless reinforce extended behaviors, such as the decision to continue with the experiment [44].

Second, HRL theory is in broad agreement with accounts emphasizing the role of ACC in effort and motivation. The theory naturally accommodates neuropsychological evidence that the primary effect of ACC lesions is to reduce spontaneous speech and action, impede effortful behavior, and produce global slowing of responding [5,8,10,20], 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

Box 4. Outstanding questions

- Can the HRL framework be used to identify laboratory tasks for which performance is selectively and dramatically impaired by ACC lesions, in contrast to the subtle and quantitative differences observed in standard conflict and trial-and-error learning tasks that nevertheless reliably elicit ACC activity?
- How do ACC and the basal ganglia interact to select specific actions when there is conflict between short- and long-term goals? When long-term plans are compromised because of short-term expediency (e.g. when we choose to stay home and watch TV rather than exercise), does this reflect a failure of top-down control or an attempt to strike an optimal balance between immediate and future needs?
- Input from the midbrain dopamine system is well characterized as providing information about positive reward and reinforcement (Box 3). From what other neural system(s) might ACC receive information about effort and punishment in support of balanced evaluations of the costs and benefits of potential options?
- It has been proposed that the frontal cortex implements an anterior–posterior gradient in hierarchical control such that the progressively rostral regions mediate progressively abstract representations [6]. Although our account has focused on a two-level hierarchy of control over primitive actions, many real-world problems involve multi-level hierarchical processing with options that call sub-options. Is hierarchical option selection and maintenance organized within the frontal cortex along this anterior–posterior gradient?
- Ongoing debate surrounds whether medial–frontal cortical activity observed in functional neuroimaging experiments, as observed, for example, in conflict [14] and response generation [12] studies, is specific to ACC or to the neighboring pre-supplementary motor area. In addition, human brain lesions that give rise to akinetic mutism include but typically extend well beyond ACC [2]. To what extent is the option-selection mechanism supported specifically by ACC, by the pre-supplementary motor area, and/or by the connections between these regions?

immediately rewarding (or punishing) but is simultaneously predictive of poor (or good) overall outcomes, such as reaching a point in a maze that holds a small reward but represents a dead end. Without a hierarchical structure, standard RL approaches would have much greater difficulty in learning the relationship between present states and distal outcomes and thus would be disproportionately affected by immediate rewards. Box 4

Our computational account focuses on explaining behavioral evidence and delineating the abstract neurocognitive functions of multiple brain areas, rather than providing a detailed neurophysiological-level account of those functions [28]. A key direction for future research thus concerns exactly how options are coded within ACC at the cellular level. Computational simulations of sequential behavior suggest that hierarchical task structure may be encoded by the internal dynamics of collections of neurons, as revealed by the evolution of the system through state space using multi-dimensional scaling [55]. Strikingly, application of similar statistical techniques to neuronal data from rat ACC indicates that neuronal ensembles track the progression of an animal through a task-dependent frame of reference, or task space [50], accompanied by abrupt transitions as the animal learns novel task contingencies [56]. Our proposal predicts that such network-dependent ACC activity will manifest most clearly in hierarchical tasks, that the proportion of task-sensitive (as opposed to response-specific) neurons will increase with the degree of hierarchical organization, and that this activity should be especially sensitive to manipulations of dopaminergic input to ACC. Evidence that fMRI multivariate pattern analysis can identify task-selective activation patterns in medial frontal cortex [57] also raises the tantalizing possibility of identifying option-specific neural signatures in humans, because the HRL framework suggests that distinct options should be associated with distinct patterns of ACC BOLD activity. Specifically, we predict that seemingly task-independent sustained activity previously observed in ACC [49] will, at a finer grain of analysis, reflect evolving patterns of activity that are distinctive for specific tasks.

Conclusion

We began with a rarely acknowledged conundrum: currently, dominant theories of ACC function from the neuroimaging literature struggle to explain the most basic clinical observation about the impact of ACC lesions, a reduction in spontaneous speech and action. We suggest that the missing link is consideration of the hierarchical structure of everyday human behavior. We propose that ACC is responsible for learning and selecting high-level behavioral plans that provide the meaning behind, and thus the motivation for, our moment-to-moment actions.

Acknowledgments

C.B.H. is supported, in part, by funding from the Canada Research Chairs program, a Michael Smith Foundation for Health Research Scholar Award, and a Natural Sciences and Engineering Research Council of Canada Discovery Grant and Discovery Accelerator Supplement (312409-05). N.Y. is supported by a grant from the National Institutes of Health (P50-MH62196).

References

1 Damasio, A.R. and Van Hoesen, G.W. (1983) Emotional disturbances associated with focal lesions of the limbic frontal lobe. In *Neuropsychology of Human Emotion* (Heilman, K.M. and Satz, P., eds), pp. 85–110, Guilford Press
2 Devinsky, O. *et al.* (1995) Contributions of anterior cingulate cortex to behaviour. *Brain* 118, 279–306
3 Németh, G. *et al.* (1988) Akinetic mutism associated with bicingular lesions: clinicopathological and functional anatomical correlates. *Eur. Arch. Psychiatr. Neurol. Sci.* 237, 218–222
4 Vogt, B.A. (2009) Regions and subregions of the cingulate cortex. In *Cingulate Neurobiology and Disease* (Vogt, B.A., ed.), pp. 3–30, Oxford University Press
5 Stuss, D.T. *et al.* (2005) Multiple frontal systems controlling response speed. *Neuropsychologia* 43, 396–417
6 Kouneiher, F. *et al.* (2009) Motivation and cognitive control in the human prefrontal cortex. *Nat. Neurosci.* 12, 939–945
7 Mulert, C. *et al.* (2005) Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *Int. J. Psychophysiol.* 56, 65–80
8 Naccache, L. *et al.* (2005) Effortless control: Executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia* 43, 1318–1328
9 Croxson, P.L. *et al.* (2009) Effort-based cost–benefit valuation and the human brain. *J. Neurosci.* 29, 4531–4541
10 Walton, M.E. *et al.* (2006) Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Netw.* 19, 1302–1314
11 Critchley, H.D. *et al.* (2005) Anterior cingulate activity during error and autonomic response. *NeuroImage* 27, 885–895
12 Mueller, V.A. *et al.* (2007) The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage* 37, 1354–1361
13 Kerns, J.G. *et al.* (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303, 1023–1026

14 Yeung, N. Conflict monitoring and cognitive control. In: *Oxford Handbook of Cognitive Neuroscience* (Ochsner, K. and Kosslyn, S., eds), Oxford University Press (in press)
15 Holroyd, C.B. and Coles, M.G.H. (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709
16 Rushworth, M.F.S. *et al.* (2007) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci.* 11, 168–176
17 Botvinick, M.M. (2007) Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* 7, 356–366
18 Nachev, P. (2011) The blind executive. *NeuroImage* 57, 312–313
19 Amiez, C. *et al.* (2006) Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex* 16, 1040–1055
20 Cohen, R.A. *et al.* (1999) Impairments of attention after cingulotomy. *Neurology* 53, 819–824
21 Kennerley, S.W. *et al.* (2006) Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* 9, 940–947
22 Pears, A. *et al.* (2003) Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. *J. Neurosci.* 23, 11189–11201
23 Rushworth, M.F.S. *et al.* (2003) The effect of cingulate cortex lesions on task switching and working memory. *J. Cogn. Neurosci.* 15, 338–353
24 Holroyd, C.B. and Yeung, N. (2011) An integrative theory of anterior cingulate cortex function: option selection in hierarchical reinforcement learning. In *Neural Basis of Motivational and Cognitive Control* (Mars, R.B. *et al.*, eds), pp. 333–349, MIT Press
25 Botvinick, M.M. *et al.* (2009) Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 113, 262–280
26 Sakai, K. (2008) Task set and prefrontal cortex. *Annu. Rev. Neurosci.* 31, 219–245
27 Niv, Y. (2009) Reinforcement learning in the brain. *J. Math. Psychol.* 53, 139–154
28 Cohen, M.X. and Frank, M.J. (2009) Neurocomputational models of basal ganglia function in learning, memory and choice. *Behav. Brain Res.* 199, 141–156
29 O'Doherty, J. *et al.* (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454
30 Gläscher, J. *et al.* (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb. Cortex* 19, 483–495
31 Tremblay, L. and Schultz, W. (1999) Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704–708
32 Schultz, W. (2007) Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* 30, 259–288
33 McClure, S.M. *et al.* (2003) A computational substrate for incentive salience. *Trends Neurosci.* 26, 423–428
34 Niv, Y. *et al.* (2006) A normative perspective on motivation. *Trends Cogn. Sci.* 10, 375–381
35 Denk, F. *et al.* (2005) Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology* 179, 587–596
36 Schweimer, J. *et al.* (2005) Involvement of catecholamine neurotransmission in the rat anterior cingulate in effort-related decision making. *Behav. Neurosci.* 119, 1687–1692
37 Schweimer, J. and Hauber, W. (2006) Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learn. Mem.* 13, 777–782
38 Assadi, S.M. *et al.* (2009) Dopamine modulates neural networks involved in effort-based decision-making. *Neurosci. Biobehav. Rev.* 33, 383–393
39 Doya, K. (2008) Modulators of decision making. *Nat. Neurosci.* 11, 410–416
40 Niv, Y. (2007) Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation? *Ann. N. Y. Acad. Sci.* 1104, 357–376
41 Durstewitz, D. and Seamans, J.K. (2008) The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-*O*-methyltransferase genotypes and schizophrenia. *Biol. Psychiatry* 64, 739–749
42 O'Reilly, R.C. (2006) Biologically based computational models of high-level cognition. *Science* 314, 91–94
43 Alexander, W.H. and Brown, J.W. (2011) Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344
44 Yeung, N. *et al.* (2005) ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544
45 Coulthard, E.J. *et al.* (2008) Control over conflict during movement preparation: role of posterior parietal cortex. *Neuron* 58, 144–157
46 Forstmann, B.U. *et al.* (2006) Voluntary selection of task sets revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* 18, 388–398
47 Johnston, K. *et al.* (2007) Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron* 53, 453–462
48 Rushworth, M.F.S. *et al.* (2002) Role of the human medial frontal cortex in task switching: a combined fMRI and TMS study. *J. Neurophysiol.* 87, 2577–2592
49 Dosenbach, N.U.F. *et al.* (2008) A dual-networks architecture of top-down control. *Trends Cogn. Sci.* 12, 99–105
50 Balaguer-Ballester, E. *et al.* (2011) Attracting dynamics of frontal cortex ensembles during memory-guided decision-making. *PLoS Comput. Biol.* 7, e1002057
51 Hoshi, E. *et al.* (2005) Neurons in the rostral cingulate motor area monitor multiple phases of visuomotor behavior with modest parametric selectivity. *J. Neurophysiol.* 94, 640–656
52 Shidara, M. and Richmond, B.J. (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* 296, 1709–1711
53 Duncan, J. (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–179
54 Ribas-Fernandes, J.J.F. (2011) A neural signature of hierarchical reinforcement learning. *Neuron* 71, 370–379
55 Botvinick, M.M. (2008) Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* 12, 201–208
56 Durstewitz, D. *et al.* (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448
57 Haynes, J-D. *et al.* (2007) Reading hidden intentions in the human brain. *Curr. Biol.* 17, 323–328
58 Shen, W. *et al.* (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321, 848–851
59 Yin, H.H. *et al.* (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur. J. Neurosci.* 19, 181–189
60 Seo, H. and Lee, D. (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377
61 Jocham, G. and Ullsperger, M. (2009) Neuropharmacology of performance monitoring. *Neurosci. Biobehav. Rev.* 33, 48–60
62 Cohen, J.D. *et al.* (2002) Computational perspectives on dopamine function in prefrontal cortex. *Curr. Opin. Neurobiol.* 12, 223–229
63 Gorelova, N. *et al.* (in press) The glutamatergic component of the mesocortical pathway emanating from different subregions of the ventral midbrain. *Cereb. Cortex* DOI:10.1093/cercor/bhr107
64 Holroyd, C.B. *et al.* (2008) The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology* 45, 688–697
65 Carlson, J.M. *et al.* (2011) Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *NeuroImage* 57, 1608–1616
66 Marco-Pallarés, J. *et al.* (2007) Learning by doing: an fMRI study of feedback-related brain activations. *Neuroreport* 18, 1423–1426
67 Zaghloul, K.A. *et al.* (2009) Human substantia nigra neurons encode unexpected financial rewards. *Science* 323, 1496–1499
68 Overbeek, T.J.M. *et al.* (2005) Dissociable components of error processing: on the functional significance of the Pe vis-à-vis the ERN/Ne. *J. Psychophysiol.* 19, 319–329
69 Emeric, E.E. *et al.* (2008) Performance monitoring local field potentials in the medial frontal cortex of primates: anterior cingulate cortex. *J. Neurophysiol.* 99, 759–772
70 Vezoli, J. and Procyk, E. (2009) Frontal feedback-related potentials in nonhuman primates: modulation during learning and under haloperidol. *J. Neurosci.* 29, 15675–15683