15

Models of Anterior Cingulate Cortex Function in Cognitive Control

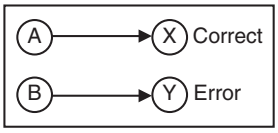
Joshua W. Brown

The anterior cingulate cortex (ACC) shows a variety of effects and has been found to be active in over 20% of all fMRI studies (Poldrack, 2011). Among the earliest neuropsychological findings was a pronounced akinetic mutism that resulted from focal lesions of the ACC (Barris & Schumann, 1953; Nemeth, Hegedus, & Molnar, 1988). Neuroscience research on the ACC began in the late 1970s, when it was shown that individual monkey cingulate neurons are more active when errors are committed (Niki & Watanabe, 1979). This finding was supported by subsequent human ERP studies in the early 1990s showing that human medial frontal activity is greater when errors are committed (Falkenstein, Hohnsbein, Hoorman, & Blanke, 1991; Gehring, Coles, Meyer, & Donchin, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993). In turn, this led to some speculation about what mechanisms could account for the newly discovered error effects. Some argued that the error signal in the ACC reflected a comparison of the actual versus intended response (Bernstein, Scheffers, & Coles, 1995), whereas others suggested that the error may reflect a comparison of predicted versus actual consequences (Ito, Stuphorn, Brown, & Schall, 2003).

While some were envisioning the error signal as a kind of comparison to detect discrepancies, another series of studies argued that the ACC signalled a co-activation of mutually incompatible (conflicting) signals. An initial study showed that events and responses with a low prior probability led to greater ACC activation, presumably as a result of the more likely but incorrect response representation competing with the less likely but correct response representation (Carter et al., 1998). This key finding showed that the ACC could be activated even when an error did not occur. A follow-up study showed that the ACC activation reflected a transient monitoring function, but it did not remain active as a driver of greater cognitive control during subsequent task performance. This indicated that the ACC might serve more as a performance monitor than a controller (Botvinick, Nystrom, Fissel, Carter, & Cohen, 1999). Later studies revealed a dissociation in which the ACC provided a monitoring signal, while the lateral prefrontal cortex (lPFC) provided a control signal over ongoing behaviour (MacDonald, Cohen, Stenger, & Carter, 2000). The following year, a formal model of conflict detection was proposed. Table 15.1 summarises a number of models of error detection in the ACC (see also Chapter 17 by Ullsperger in this volume).

The Wiley Handbook of Cognitive Control, First Edition. Edited by Tobias Egner.

Table 15.1 Summary of ACC Error Monitoring Theories.



Model	*Method*	*Error detection pseudocode*	*References*
Discrepancy	Subtractive (reactive)	Y-X+1	Scheffers & Coles (2000)
Response selection	Learned conjunction (reactive, proactive)	$V_r(A, Y) - V_{r-1}(A)$	Holroyd et al. (2005)
Correction detection	Sequence detection (reactive)	X after Y?	Steinhauser et al. (2008)
Conflict	Multiplicative (reactive, proactive)	A * B	Botvinick et al. (2001)
Error likelihood	Learned probability (proactive)	P(Y | A,B)	Brown & Braver (2005)
PRO	Learned prediction, comparison (reactive, proactive)	abs(Y-P(Y|A))	Alexander & Brown (2011)
Double integrator	Integration of task cue integrator	$\dot{x}_1 = A + B + X + Y - const$ $\dot{x}_2 = \beta x_1$	Bekolay et al. (2014)

Note. Consider two representations of cued responses with continuous activities A and B in [0, 1] that lead to binary 0 or 1 responses X and/or Y, representing correct and error responses, respectively. For concreteness, A may correspond to a valid cue in the Stroop task and drive correct response X, whereas B may represent an incongruent distractor that drives incorrect response Y. Source: Adapted from Alexander 2010. Reproduced with permission of Wiley.

The Conflict Model

According to the conflict computational model (Botvinick, Braver, Barch, Carter, & Cohen, 2001), the ACC activity reflects a simple multiplication of the activity representing two mutually incompatible response processes (Figure 15.1). If 'A' represents the activity of response A, and 'B' represents the activity of a competing response B, then the ACC activation will simply be the product A × B. The simplicity of this computational model, along with its ability to account for a considerable array of effects found in the ACC, led to it rapidly becoming a leading computational account of ACC function, to the point that ACC activation was nearly synonymous with conflict. The conflict model was applied to various problems and subdivided into various kinds of conflict detection, for example, conflict between task sets and conflict across sequences of changing responses (Brown, Reynolds, & Braver, 2007).

The first question to be addressed by the conflict model is that if the ACC computes conflict, what is the source of the error signals? According to a subsequent analysis of the conflict model, error signals may result simply from 'post-error conflict'. That is, once an error response is committed, there is still motor-related activity representing the planned correct response. The co-activation of the motor activities representing the planned error and the planned (but not executed) correct response leads to a state of conflict. This was proposed as the origin of the error signal (Yeung, Botvinick, & Cohen, 2004).

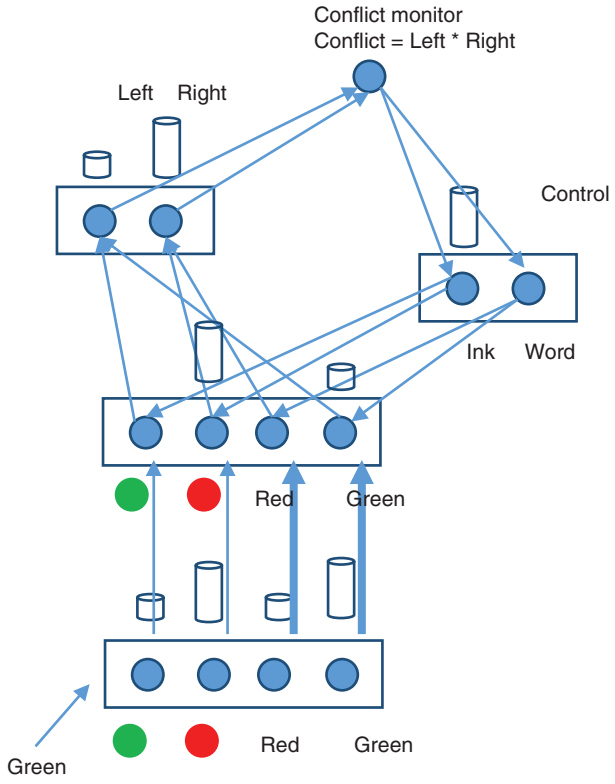


Figure 15.1 Conflict model of the Stroop task. The Stroop task requires subjects to name the ink color of a word and ignore the semantic meaning. When an incongruent Stroop stimulus is presented, both the left and right response representations are simultaneously activated. The conflict model suggests that the ACC detects the simultaneous co-activation of the mutually incompatible responses, that is, Conflict = LEFT * RIGHT response activation levels (Botvinick et al., 2001). The conflict detector in turn increases the gain of the control signals, which bias the network to produce the correct response according to the task rules.

Challenges to the Conflict Model

In the wake of the conflict model proposal, a number of challenges arose. One key challenge to the conflict model was a finding that the ACC activation (as indexed by the error-related negativity, or ERN) could be elicited by feedback that an error has occurred, and not simply by the commission of an error per se. This is an essential distinction. Once a response has been made, the response-related activity presumably decays rapidly. Nevertheless, Holroyd and Coles found that the ERN could be elicited by feedback that an error occurred, even when feedback was given well after the responses were made (Holroyd & Coles, 2002). This was termed the 'feedback ERN', or fERN. It is unclear how the conflict model could account for this effect, because the response-related activations would have decayed well before the outcome was revealed (Holroyd, Yeung, Coles, & Cohen, 2005).

Aside from the conflict model's difficulty in accounting for the fERN, a direct test for conflict cells in the macaque monkey ACC failed to turn up any cells that signalled response conflict (Ito et al., 2003), although conflict-related activity was apparently found more dorsally in the supplementary eye fields (Stuphorn, Taylor, & Schall, 2000). By 2009, some

had suggested that perhaps the monkey and human ACC regions were simply different in this regard, and that the monkey ACC did not compute conflict, whereas the human ACC did (Cole, Yeung, Freiwald, & Botvinick, 2009). The lack of findings of conflict in monkey ACC cells has persisted until very recently (Ebitz & Platt, 2015). Unfortunately, the recent claim of monkey conflict cells actually found what turned out to be 'task conflict' cells, and no actual response conflict cells were found (Ebitz & Platt, 2015). The absence of single units that apparently compute response conflict in the monkey ACC remains.

Likewise, several papers challenged the conflict model on the basis of ERP or fMRI results. One paper found that in a Stroop task, ACC activity was greater for neutral than for congruent stimuli, in which neither condition entailed response conflict (Roelofs, van Turennout, & Coles, 2006). An ERP study showed that as conflicting response representations overlapped more in time, the ERN amplitude actually decreased rather than increased, contrary to the predictions of the conflict model (Burle, Roger, Allain, Vidal, & Hasbroucq, 2008). Another set of papers proposed that the ACC activity simply correlates with response time (Carp, Kim, Taylor, Fitzgerald, & Weissman, 2010; Grinband et al., 2011), and that apparent response conflict effects are an epiphenomenal reflection of this effect, because incongruent trials generally entail longer response times.

Reinforcement Theories of the ERN

Shortly after the conflict model was put forward, an alternative model of ACC function was proposed, based on reinforcement learning. This new model suggested how the ACC might compute a signal that indicates an error has occurred on the basis of the fERN, rather than on the basis of internal motor signals. According to the RL-ERN reinforcement model (Holroyd & Coles, 2002), the ACC is activated by pauses in dopamine cell activity (Figure 15.2a), which occur when the expected reward is not received (Schultz, 1998). Thus, an error results in reward omission, which pauses dopamine cell firing, which disinhibits the ACC. The RL-ERN model further posits that the role of the ACC is to select which other frontal areas will be permitted to drive action generation, although the same function has elsewhere been ascribed to the basal ganglia (Brown, Bullock, & Grossberg, 2004). In a follow-up ACC model (Figure 15.2b), the ACC is cast as learning to respond to reduced value, which again may originate as reduced dopamine signalling from the basal ganglia (Holroyd et al., 2005). The later model simulates how the ACC may learn to respond to conjunctions of events that collectively indicate that an error has occurred. In this way, the fERN can be accounted for as being activated when the events indicate an error.

The Error Likelihood Model

Around the same time, an alternative computational model of the ACC was proposed, in which the ACC was cast as learning to predict the likelihood of an error (Brown & Braver, 2005). According to this model, the ACC learns to represent conjunctions of events for which the probability of an error was high (Figure 15.2c). Building on the RL-ERN model (Holroyd & Coles, 2002), the model proposed that dopamine pauses not only activated the ACC, but also trained it to respond in anticipation of errors in situations where they were likely to occur, on the basis of the modulatory influence of dopamine on cortical plasticity (Bao, Chan, & Merzenich, 2001). The main evidence for this model came from an fMRI study in which task conditions associated with a higher likelihood of errors led to greater ACC activity. Critically, this contrast held even when both error commission and

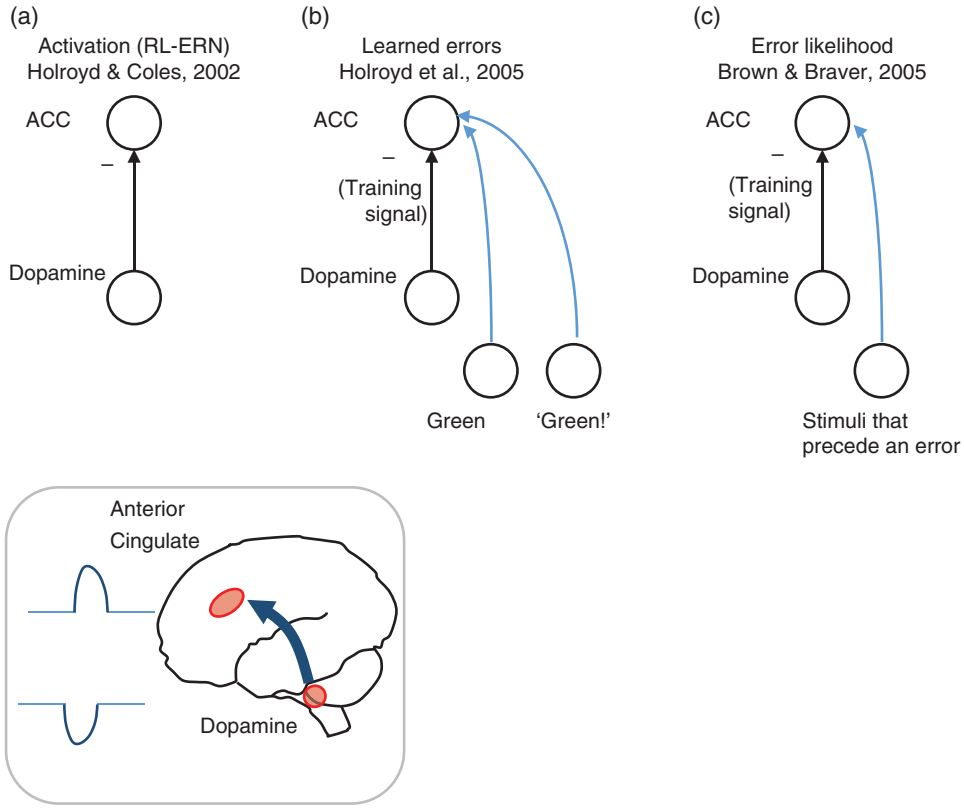


Figure 15.2 Reinforcement learning models of ACC. Left: Dopaminergic (DA) signals from the midbrain disinhibit the anterior cingulate cortex. (a) The RL-ERN model (Holroyd & Coles, 2002) simulates the ACC error activity as resulting from a pause of DA activity, which disinhibits the ACC. (b) The Holroyd et al. (2005) model, in which the DA pause trains the ACC to recognize conjunctions of events that indicate an error has *already* occurred. (c) The error likelihood model of Brown & Braver (2005), in which a DA pause trains the ACC to be activated by cues that reliably *precede* an error.

conflict were controlled for (Brown & Braver, 2005). According to the error likelihood model, error signals could be accounted for in that when an error occurred, the probability of an error was essentially P(error) = 1, so the ACC would signal that an error was maximally likely (as it had, in fact, occurred).

The error likelihood model was quickly challenged, and the results highlighted the importance of individual differences. First, a nearly direct replication study of the ACC error likelihood effect failed (Yeung & Nieuwenhuis, 2009). Subsequent studies revealed that the ACC error likelihood effect was strongest in those who are most risk averse, but virtually absent in those who are risk seeking (Brown & Braver, 2007). Thus, a population sample that is more risk tolerant overall may fail to show an error likelihood effect in the ACC. Subsequent computational neural models formalised this idea—as the ability of the model ACC to learn from dopamine cell activity pauses decreased, the model ACC cells showed progressively weaker error likelihood effects (Brown & Braver, 2008). Overall this suggested that greater ACC activity led to greater risk or loss avoidance, a finding that has been replicated in several fMRI studies (Fukunaga, Bogg, Finn, & Brown, 2013; Fukunaga, Brown, & Bogg, 2012; Krawitz, Fukunaga, & Brown, 2010).

More Constraints on ACC Models

While the conflict, RL-ERN, and error likelihood models gained traction, new studies of the ACC reported additional effects that challenged the existing models. One study showed that simply manipulating whether subjects were required to detect or correct their own errors, by a certain deadline, led to measurable effects on responding; this in turn motivated a model of ACC error detection in which the ACC was proposed to detect not the errors, nor the likelihood of errors, but rather a correction response that follows an error (Steinhauser, Maier, & Hübner, 2008). In particular, the model posited ACC activity when a correction response followed an initial error response.

Other new effects followed. Several studies showed that ACC activity has a strong positive correlation with response time (RT), so that when this correlation is accounted for, no response conflict effects remain. This was construed to mean that perhaps the ACC does not compute response conflict but rather signals a measure of the RT (Carp et al., 2010; Grinband et al., 2011). Other earlier studies found similar effects, although they were construed as either simply inconsistent with the conflict model (Burle et al., 2008) or not relevant to the conflict model (Yeung & Nieuwenhuis, 2009).

Still another influential study showed that ACC activity correlated with the degree of volatility (non-stationarity) in the environment, which in turn correlated with the effective learning rates seen in task performance (Behrens, Woolrich, Walton, & Rushworth, 2007). It is unclear how any of the existing models could have accounted for this. Relatedly, a number of studies began to show effects of prediction errors in the medial prefrontal cortex and ACC in particular. One study in monkeys showed that some ACC cells signal unsigned prediction errors (Hayden, Heilbronner, et al., 2011). A human study showed that error effects in the ACC could actually be inverted, so that there was more activation for correct responses than errors when errors were very likely (Jessup, Busemeyer, & Brown, 2010). This challenged the prevailing view that the ACC responded primarily to negatively valenced events such as errors, and instead, it suggested that the ACC responds to surprising events, whether affectively positive or negative, a finding that has been confirmed by other fMRI and ERP studies (Ferdinand, Mecklinger, Kray, & Gehring, 2012; Garofalo, Maier, & di Pellegrino, 2014; Nee, Kastner, & Brown, 2011; Oliveira, McDonald, & Goodman, 2007).

The Prediction of Responses and Outcomes (PRO) Model

With the mounting empirical evidence, it became clear that the ACC yielded signals consistent with both predictions of outcomes and surprise. Two groups (including the author) began working independently to develop computational neural models of the ACC as predicting outcomes and then comparing the outcomes against the predictions (Alexander & Brown, 2011; Silvetti, Alexander, Verguts, & Brown, 2014; Silvetti, Seurinck, & Verguts, 2011). The two models are remarkably similar (Silvetti et al., 2014). Each builds on the temporal difference model. The Reward Value and Prediction Model (RVPM; Silvetti et al., 2011) casts the ACC in the role of learning to predict reward, and then separately detecting better versus worse than expected reward. The PRO model (Alexander & Brown, 2011) casts the ACC as learning to predict the variety of possible outcomes (good, bad, or neutral), each with some probability. Both models predict the timing of when the reward (RVPM) or various possible outcomes (PRO) might occur.

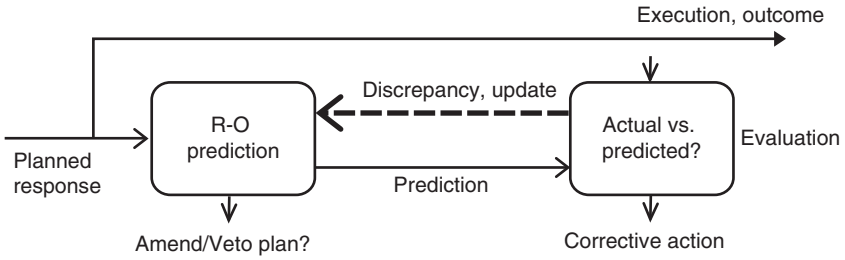


Figure 15.3 PRO model. Planned responses activate learned response–outcome predictions. These predicted outcome signals can feed back to amend or veto a planned action. Once an action is generated, the actual outcome (the movement itself or the feedback from the environment) is compared against the expected outcome, and any discrepancy leads to an update of the learned response–outcome predictions. Source: Adapted from Alexander 2010. Reproduced with permission of Wiley.

Similarly, both models compare the actual outcomes against the prior predictions. The RVPM model has cells that specifically signal when an outcome was more rewarding than expected, versus other cells that signal specifically when a reward was less rewarding than expected. These are simply the positively and negatively rectified signals from the temporal difference signal in the model.

The PRO model is shown in more detail in Figure 15.3. Specifically, the R-O prediction layer has cells that represent specific possible conjunctions of responses and outcomes, whether good or bad. Greater activity signals a greater likelihood of the outcome. When the timing of the expected outcome approaches, the cells increase their activities and peak around the time of the expected outcome. These prediction signals in turn activate an evaluation layer, which compares the actual versus expected outcomes. A large number of empirical effects can be simulated when the evaluation units are excited by the prediction, and inhibited by the actual outcome (Alexander & Brown, 2011). This is referred to as *negative surprise*, that is, the omission of an expected outcome whether subjectively good or bad. In this case, the evaluation cells are most active when an outcome is predicted, but it does not occur. This is how the PRO model generates greater activity when errors occur: It signals not an error per se, but rather an expected correct outcome that failed to occur. Conversely, if errors are more common than correct trials, then correct trials will elicit greater activity when the expected error fails to occur. Such effects have been found empirically (Ferdinand et al., 2012; Garofalo et al., 2014; Jessup et al., 2010; Nee et al., 2011; Oliveira et al., 2007).

The PRO model has been able to simulate and account for a large body of data. The original PRO model publication (Alexander & Brown, 2011) showed that it could simulate human ERP and fMRI effects of error, conflict, greater error likelihood, greater error unexpectedness (Brown & Braver, 2005), volatility (Behrens et al., 2007), positive correlations with RT (Carp et al., 2010; Grinband et al., 2011), and individual difference effects (Brown & Braver, 2008, 2007), and also monkey single-unit effects of reward, unexpected reward, and prediction error (Amador et al., 2000; Hayden, Heilbronner, et al., 2011; Ito et al., 2003).

A subsequent set of simulations (Alexander & Brown, 2014) showed that the PRO model could account for general predictions of subsequent events, including stimuli as 'outcomes'. This generalisation of the model allowed it to simulate ACC effects of trial type frequency, item-specific versus global control (Blais & Bunge, 2010), the mismatch negativity (Crottaz-Herbette & Menon, 2006), informative versus uninformative cues

(Aarts, Roelofs, & van Turennout, 2008), Bayesian surprise (Ide, Shenoy, Yu, & Li, 2013), and additional monkey neurophysiology studies showing surprising gain and loss effects (Sallet et al., 2007).

One of the distinctive features of the PRO model is that it generates timed predictions of expected outcomes. In other words, the prediction units signal not only *what* outcomes are likely to occur and with what probability, but also *when* those outcomes are likely to occur. The R-O prediction unit activities peak at the time when the expected outcome is most likely to occur. This is consistent with empirical results, which show that an expected outcome can still generate ACC activity if it is received at an unexpected time (Forster & Brown, 2011). The PRO model uses a tapped delay line, which is a mechanism in which an array of cells are activated at specific delay intervals following a stimulus. Thus, the timing of an expected outcome (relative to an earlier stimulus) can be learned by training a specific delay interval cell to activate a prediction unit. The PRO model is agnostic about where these delay intervals may be represented, but one possibility is the cerebellum via the thalamus, as the cerebellum is known to both represent adaptive timing (Fiala, Grossberg, & Bullock, 1996) and play an essential role in cognitive as well as motor functions (Middleton & Strick, 1994). In this scenario, the adaptive timing functions of the ACC would be derived from an *external* basis in the cerebellum. An alternative approach has recently been suggested, in which the prediction timing is computed by mechanisms *intrinsic* to the ACC. In this double integrator model (Bekolay, Laubach, & Eliasmith, 2014), the 'ramping up' signals within the ACC are generated by two integrators. The first integrates a signal that the trial is under way, and thus the magnitude of the integrated signal correlates with the elapsed trial time. Likewise, the second signal integrates the first integrator, thus providing an approximate measure of how much time has elapsed since the first integrator changed. The timing of events in correct versus error trials is different, leading to divergent states of the two integrators. The authors showed that the two integrators yielded signals that appeared qualitatively similar to the principal components found in a population of units from the rat prelimbic cortex, which may be the rat equivalent of the ACC (Bekolay et al., 2014).

Value Models

So far, the models discussed have mainly focused on cognitive constructs such as conflict detection, error detection, and error likelihood. Another class of models constructs the ACC as representing different kinds of affective and motivational values in a way that influences decision making, providing a kind of motivational signal to exert cognitive control over decision making. Cognitive control essentially biases decisions away from the immediate demands of strong stimuli and towards the longer-term, higher-level goals of the agent. There is some evidence that ACC activity in humans provides a kind of motivational value to exert effort towards a goal—humans receiving direct electrical stimulation of the dorsal ACC report an impetus to 'push harder to try and get through this' (Parvizi, Rangarajan, Shirer, Desai, & Greicius, 2013).

In some ways, the models above reflect aspects of such values. The error likelihood model could be thought of as representing an aversive signal of impending errors, which in turn inhibits responding (Brown & Braver, 2005, 2007, 2008). Similarly, other models can be viewed as learning states associated with a negative reward value (Holroyd et al., 2005).

A recent review argues that the ACC computes the expected value of control, which is defined as the expected rewards of exerting control over decision making, less the 'costs' of exerting such control. This has been called the 'expected value of control' (EVC) model (Shenhav, Botvinick, & Cohen, 2013; see also Chapter 10 by Kool, Shenhav, & Botvinick in

this volume). The costs in this case are internal, subjective, and therefore somewhat difficult to quantify, although there is some qualitative evidence to support the general notion (Kool, McGuire, Wang, & Botvinick, 2013), and the model has not been simulated or used to generate predictions thus far.

A recent paper has proposed that the ACC represents the value of foraging (Kolling, Behrens, Mars, & Rushworth, 2012). According to this view, animals must generally choose whether to engage (exploit) a known but diminishing resource or else forage (explore) for a better resource, with no guarantee that such a better resource will be found. Humans performing this kind of foraging task generally have a bias to engage, even when the expected value is lower relative to foraging. To overcome this bias, the ACC was proposed to generate a control signal whose activity correlates positively with the relative value of foraging instead of engaging. When the activity is higher, it will bias the animal to forage for a better resource. This idea is consistent with monkey single-unit findings, showing that animals forage when ACC cell activity exceeds a specific threshold (Hayden, Pearson, et al., 2011). A recent paper has shown that ACC activity correlates positively with the relative value of foraging (Kolling et al., 2012). A subsequent paper has challenged this interpretation (Shenhav, Straccia, Cohen, & Botvinick, 2014). In it, the authors show that as the relative value of foraging increases, a region in the dorsal ACC shows activity that peaks at a certain relative foraging value and then decreases as the relative value of foraging increases further. The authors argued that the ACC therefore may not represent relative foraging value but instead may represent a more cognitive signal of choice difficulty. Specifically, ACC activity peaks when the choice probabilities of foraging versus engaging are near the indifference point, that is, each 50%, which is the point at which the decision to forage or engage is most difficult. The authors did not specify a computational model of how choice difficulty may be computed, but choice difficulty is not the only possible explanation of this phenomenon. In the PRO model, activity in the evaluation layer (Figure 15.3) might be expected to peak at the indifference point. This is due to the fact that at the indifference point, whichever action is not chosen nevertheless was 50% likely, and therefore the fact that it was not chosen is somewhat surprising. In contrast, when the relative value of foraging is either very high or very low, unlikely actions are rarely chosen (by definition, because they are unlikely), so there is on average less prediction error and therefore less overall activity.

Oscillatory Models

EEG studies of the ACC show oscillatory activity especially in the theta band (approximately 4–8 Hz). The prominent theta signals have been found in EEG signals with a likely ACC source (see also Chapter 14 by M.X. Cohen in this volume). These include the N2 effects of novelty and conflict, and the ERN and feedback ERN effects (Cavanagh & Frank, 2014). A number of models of the ACC simulate neural activity as a rate code rather than by simulating individual spiking neurons. This simplifies the computation, but it also renders such models unable to simulate the moment-by-moment spectrum of oscillatory power. A recent theoretical model (Cohen, 2014) proposes that conflict detection is implemented as coincidence detection in cortical layer 5 of the ACC, whereas the EEG conflict-related theta oscillations may be generated in layers 2 and 3. The idea is that layer 5 cells detect the coincident activation that represents conflict, and these in turn activate layers 2 and 3 cells. The input from layer 5 has the effect of amplifying the processing in layers 2 and 3. The layer 2 and 3 signals may be the source of the observed theta rhythms, but it is unclear whether the theta rhythm plays an important functional role or is merely an epiphenomenal reflection of afferent signals.

Control Signals: How Does the ACC Influence Other Brain Regions?

So far this chapter has considered mainly what signals are generated within the ACC and how they are generated. The next question is exactly what effects those signals have on other parts of the brain and behaviour. Here again, a variety of effects have been proposed. Before considering those, it must be noted that animals and humans generally show surprisingly little effects of ACC lesions. In the most severe cases, massive lesions result in akinetic mutism (Barris & Schumann, 1953; Nemeth et al., 1988). More recent studies show that human ACC lesions do not impair cognitive control in general (Fellows & Farah, 2005) and only moderately slow error correction (Modirrousta & Fellows, 2008). Animal studies show that cingulate lesions impair the ability to take the recent reward history into account (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006) and an unwillingness to exert extra effort to obtain a larger reward (Walton, Bannerman, & Rushworth, 2002). Any proposed model of ACC effects on other brain regions and behaviour must take these findings into account.

The proposed influences of the ACC on behaviour can generally be categorised into proactive and reactive control signals (Braver, Gray, & Burgess, 2007; see also Chapter 9 by Chiew & Braver in this volume), and this chapter will view such effects through this lens of proactive versus reactive control. Proactive control signals refer to those that are activated prior to a cognitively demanding event, typically in response to cues that precede and correlate with demanding events. Reactive control signals, on the other hand, are activated in response to errors or unexpected events. The ACC may generate both kinds of signals. Error likelihood signals would be an example of a signal that is sufficient to provide proactive control (Brown & Braver, 2005), especially as a means of inducing risk avoidance (Brown & Braver, 2007). Error signals (Gehring et al., 1990; Hohnsbein, Falkenstein, & Hoorman, 1989) would be an example of signals sufficient to drive reactive control.

One proposed way of understanding the ACC is that it does not so much learn specific behaviours as motivate or energise those behaviours, especially to achieve larger task goals (Holroyd & Yeung, 2012). This idea of proactive control was recently fleshed out in a computational model (Holroyd & Mcclure, 2015). According to the model, the ACC represents the values and costs of various higher-level options and then exerts a control signal specifically over the striatum. This proposal is similar to an earlier proposal, which, however, did not focus on a role of the striatum and did not include computational model simulations (Shenhav et al., 2013). The model of Holroyd and Yeung (2012) casts the ACC in the role of an actor rather than a performance monitor, and the effects of such ACC activity would amount to a proactive control signal. This control signal in the model ensures that higher-level goals appropriately guide lower-level action selection, so that behaviour remains directed towards higher-level goals. Simulations of rat prelimbic cortex lesions show that, as with rat empirical data, lesions of the computational model led to a higher number of perseverations and a longer time to criterion in a maze task with regularly switching rules.

Another recent model argues that the ACC reacts to strategy failures by forcing the brain to discard current behavioural strategies in favour of newly synthesised strategies (Donoso, Collins, & Koechlin, 2014). This bears a resemblance to the proposed role of the ACC in driving foraging (Kolling et al., 2012), except that here the foraging would be done at the cognitive level in terms of strategies, rather than at the behavioural level in terms of afforded actions. In this model, the brain creates approximately three possible strategies for the current behavioural task at any given time. One strategy is used to guide behaviour, while the other two strategies are held in reserve and evaluated to see if they would have provided a more successful match to the current task environment (see also Chapter 12 by Duverne and Koechlin in this volume). The frontopolar cortex activity correlated with the reliabilities of

the alternative strategies, whereas the left lateral PFC was activated when the current strategy was rejected in favour of an alternative strategy. It is noteworthy that this model does not cast the ACC as driving a strategy switch in every case. Instead, the ACC is active only when both the current *and the alternative* strategies are all found wanting, so that a completely new set of strategies must be implemented.

The PRO model as discussed above (Alexander & Brown, 2011) focuses much on how the ACC generates the empirically observed activity patterns, but it is relatively silent on what functional role those activity patterns have on other brain regions. The only proposed control effect is on learning. Specifically, when the environmental contingencies are relatively stable, the environment can be described as statistically more stationary or less volatile. When the environment is more stationary, the effective learning rate that can be recovered from observed behaviour is relatively lower, and this correlates with reduced ACC activity (Behrens et al., 2007). Conversely, in more non-stationary environments, the effective learning rates are larger. The PRO model captures these phenomena by modelling surprise signals (i.e., prediction errors) as reducing the overall learning rate. Although this may seem counterintuitive, it turns out that in the PRO model, non-stationary environments cause the predicted outcomes of actions to be less certain, and therefore any particular outcome is less surprising. Thus, in non-stationary environments, surprise has a weaker effect on reducing the learning rate, which allows the learning rate to be effectively higher.

The ACC may also provide a proactive control signal that drives risk avoidance. Several studies have shown that ACC activity is greater when subjects choose to engage in a riskier option (Fukui, Murai, Fukuyama, Hayashi, & Hanakawa, 2005; Fukunaga et al., 2012; Krawitz et al., 2010). The error likelihood model simulates how the ACC may learn to represent conditions when errors are more likely, including incongruent stimuli that may entail response conflict (Brown & Braver, 2005). Still, the initial error likelihood proposal did not investigate the potential role of the ACC in risk avoidance. To study this, subjects were asked to self-report their risk-taking preferences. Those who were more risk averse showed stronger error likelihood effects, whereas the more risk-seeking subjects showed essentially no error likelihood effect (Brown & Braver, 2007). This individual difference effect was modelled with the error likelihood model simply by manipulating the learning rate from errors. When the learning rate was higher, the model learned more quickly from errors and showed stronger error likelihood effects in a way that matched the more risk-averse subjects (Brown & Braver, 2008). All of this is consistent with a role for the ACC in driving risk avoidance.

Conclusion

Table 15.1 summarises the essence of a number of computational models that have been proposed over the years (see also Chapter 8 by Verguts in this volume). The ACC shows a number of effects that have attracted significant interest from modellers. The effects variously capture aspects of performance monitoring and cognitive control, including errors, conflict, and more general predictions and prediction errors among many other effects. There is currently substantial controversy around the question of whether the ACC reflects more cognitive signals (such as conflict or choice difficulty) or affective signals (such as action value and reward; Shenhav et al., 2014). Other open questions concern whether predictive timing signals are extrinsic or intrinsic to the ACC, and whether ACC activity correlations with the RT are sufficient to account for other effects such as response conflict. For many years now, the response conflict model in particular has dominated the literature on models of the ACC, to the point that ACC activity has been treated as nearly synonymous with response conflict.

As the above discussion suggests, the modelling literature is much richer now than when the conflict model was introduced, and there is currently a healthy interaction between models and animal and human studies. Future progress will be most efficient if experimental work can be aimed at discriminating among the multiple existing models of the ACC (Alexander & Brown, 2015).

References

Aarts, E., Roelofs, A., & van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience*, *28*, 4671–4678.

Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science*, *2*, 658–677.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*, 1338–1344.

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience*, *8*.

Alexander, W. H., & Brown, J. W. (2015). Reciprocal interactions of computational modeling and empirical investigation. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *Model-based cognitive neuroscience* (pp. 321–338). New York: Springer.

Amador, N., Schlag-Rey, M., & Schlag, J. (2000). Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *Journal of Neurophysiology*, *84*, 2166–2170.

Bao, S., Chan, V. T., & Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, *412*, 79–83.

Barris, W., & Schumann, H. (1953). Bilateral anterior cingulate gyrus lesion: Syndrome of the anterior cingulate gyri. *Neurology*, *3*, 44–80.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221.

Bekolay, T., Laubach, M., & Eliasmith, C. (2014). A spiking neural integrator model of the adaptive control of action by the medial prefrontal cortex. *Journal of Neuroscience*, *34*, 1892–902.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. (1995). 'Where did I go wrong?' A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1312–1322.

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*, 2758–2767.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. C. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Botvinick, M. M., Nystrom, L., Fissel, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179–181.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, M. Kane, A. Miyake, & J. Towse, C. J. (Eds.), *Variation of working memory*. Oxford: Oxford University Press.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307(5712)*, 1118–1121.

Brown, J. W., & Braver, T. S. (2007). Risk prediction and aversion by anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 266–277.

Brown, J. W., & Braver, T. S. (2008). A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. *Brain Research*, *1202*, 99–108.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, *17*, 471–510.

Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task switching. *Cognitive Psychology*, *55*, 37–85.

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: A reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience*, *20*, 1637–1655.

Carp, J., Kim, K., Taylor, S. F., Fitzgerald, K. D., & Weissman, D. H. (2010). Conditional differences in mean reaction time explain effects of response congruency, but not accuracy, on posterior medial frontal cortex activity. *Frontiers in Human Neuroscience*, *4*, 231.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. C., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421.

Cohen, M. X. (2014). A neural microcircuit for cognitive conflict detection and signaling. *Trends in Neurosciences. 37*(9), 480–490.

Cole, M. W., Yeung, N., Freiwald, W. A., & Botvinick, M. (2009). Cingulate cortex: Diverging data from humans and monkeys. *Trends in Neurosciences*, *32*, 566–574.

Crottaz-Herbette, S., & Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: Combined fMRI and ERP evidence. *Journal of Cognitive Neuroscience*, *18*, 766–780.

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science*, *344*, 1481–1486.

Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*, *85*, 628–640.

Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*, 447–455.

Fellows, L. K., & Farah, M. J. (2005). Is anterior cingulate cortex necessary for cognitive control? *Brain*, *128*, 788–796.

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *Journal of Neuroscience*, *32*, 12087–12092.

Fiala, J. C., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *Journal of Neuroscience*, *16*, 3760–3774.

Forster, S. E., & Brown, J. W. (2011). Medial prefrontal cortex predicts and evaluates the timing of action outcomes. *Neuroimage*, *55*, 253–265.

Fukui, H., Murai, T., Fukuyama, H., Hayashi, T., & Hanakawa, T. (2005). Functional activity related to risk anticipation during performance of the Iowa Gambling Task. *Neuroimage*, *24*, 253–259.

Fukunaga, R., Bogg, T., Finn, P. R., & Brown, J. W. (2013). Decisions during negatively-framed messages yield smaller risk-aversion-related brain activation in substance-dependent individuals. *Psychology of Addictive Behaviors*, *27*, 1141–1152.

Fukunaga, R., Brown, J. W., & Bogg, T. (2012). Decision making in the Balloon Analogue Risk Task (BART): Anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. *Cognitive, Affective, & Behavioral Neuroscience. 12*(3), 479–490.

Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports*, *4*, 4816.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: An event-related potential accompanying errors. *Psychophysiology*, *27*, S34.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage*, *57*, 303–311.

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: Neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience*, *31*, 4178–4187.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*, 933–939.

Hohnsbein, J., Falkenstein, M., & Hoorman, J. (1989). Error processing in visual and auditory choice reaction tasks. *Journal of Psychophysiology*, *3*, 32.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.

Holroyd, C. B., & Mcclure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex. *Psychological Review*, *122*, 54–83.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*, 122–128.

Holroyd, C. B., Yeung, N., Coles, M. G., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*, 163–191.

Ide, J. S., Shenoy, P., Yu, A. J., & Li, C. R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *Journal of Neuroscience*, *33*, 2039–2047.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by anterior cingulate cortex during saccade countermanding. *Science*, *302*(5642), 120–122.

Jessup, R. K., Busemeyer, J. R., & Brown, J. W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *Journal of Neuroscience*, *30*, 3467–3472.

Kennerley, S. W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, *9*, 940–947.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, *336*, 95–98.

Kool, W., McGuire, J. T., Wang, G. J., & Botvinick, M. M. (2013). Neural and behavioral evidence for an intrinsic cost of self-control. *PLoS ONE*, *8*.

Krawitz, A., Fukunaga, R., & Brown, J. W. (2010). Anterior insula activity predicts the influence of positively framed messages on decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 392–405.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal cortex and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835–1838.

Middleton, F. A., & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, *266*(5184), 458–461.

Modirrousta, M., & Fellows, L. K. (2008). Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. *Journal of Neuroscience*, *28*, 14000–14005.

Nee, D. E., Kastner, S., & Brown, J. W. (2011). Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. *Neuroimage*, *54*, 528–540.

Nemeth, G., Hegedus, K., & Molnar, L. (1988). Akinetic mutism associated with bicingular lesions: Clinicopathological and functional anatomical correlates. *European Archives of Psychiatry and Neurological Sciences*, *237*, 218–222.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research*, *171*, 213–224.

Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience*, *19*(12), 1994–2004..

Parvizi, J., Rangarajan, V., Shirer, W. R., Desai, N., & Greicius, M. D. (2013). The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron*, *80*, 1359–1367.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*, 692–697.

Roelofs, A., van Turennout, M., & Coles, M. G. (2006). Anterior cingulate cortex activity can be independent of response conflict in Stroop-like tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 13884–13889.

Sallet, J., Quilodran, R., Rothe, M., Vezoli, J., Joseph, J. P., & Procyk, E. (2007). Expectations, gains, and losses in the anterior cingulate cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 327–336.

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology*, *26*, 141–151.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–240.

Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, *17*(9), 1249–1254.

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience & Biobehavioral Reviews*, *46*(Pt. 1), 44–57.

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience*, *5*, 75.

Steinhauser, M., Maier, M., & Hübner, R. (2008). Modeling behavioral measures of error detection in choice tasks: Response monitoring versus conflict monitoring. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 158–176.

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature*, *408*, 857–860.

Walton, M. E., Bannerman, D. M., & Rushworth, M. F. (2002). The role of rat medial frontal cortex in effort-based decision making. *Journal of Neuroscience*, *22*, 10996–11003.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.

Yeung, N., & Nieuwenhuis, S. (2009). Dissociating response conflict and error likelihood in anterior cingulate cortex. *Journal of Neuroscience*, *29*, 14506–14510.