Conflict monitoring and cognitive control

Nick Yeung
Department of Experimental Psychology
University of Oxford

(h1) Abstract

Neural mechanisms of cognitive control are hypothesized to support flexible, goal-directed behavior by representing task-relevant information in order to guide thought and action. The conflict monitoring theory proposes that anterior cingulate cortex (ACC) contributes to cognitive control by detecting conflicts in information processing and signaling when increased top-down control is required. This theory provides a computationally specified framework for understanding how cognitive control is recruited, and explains a large literature of human neuroimaging studies reporting ACC activity in conditions of increased cognitive demand. Predictions from the theory have been tested and consistently confirmed in behavioral and neuroimaging experiments with human subjects. However, challenging findings from patients with ACC lesions, and from studies of ACC function in non-human primates, suggest that conflict monitoring may be just one facet of the broader role of ACC in performance monitoring and the optimization of behavior.

(h1) Keywords

cognitive control; conflict monitoring; anterior cingulate cortex; neuroimaging; computational models; neuropsychology; reinforcement learning; performance monitoring.
(h1) Introduction

A central challenge in cognitive neuroscience research is to account for intelligent, purposive human behavior without relying on an unspecified intelligent agent—a controlling homunculus—for explanatory power (Monsell & Driver, 2000). Framed differently, the challenge is to explain how the richness and flexibility of human cognition emerges from interactions among component processes that are themselves fixed and algorithmically specifiable. As a first step toward meeting this challenge, it is often proposed that dedicated mechanisms of cognitive control support flexible cognition by representing task-relevant information and using this information to guide thought and action in accordance with current goals and intentions (Miller & Cohen, 2001). Regions in lateral prefrontal cortex (PFC) are thought to play a particularly important role in exerting this control (Sakai, 2008).

However, the notion that lateral PFC exerts control raises the question of how this region itself knows which information is task-relevant: there is a clear danger here of theoretical regress. Recent years have therefore seen considerable effort devoted to understanding how the inputs to this region might support its functioning. Various functional connections of lateral PFC are relevant in this regard: Interactions with the medial temporal lobe might support the recall of goals set during earlier prospective planning (Cohen & O'Reilly, 1996; Schacter, Addis, & Buckner, 2007), input from orbitofrontal cortex could provide information about the likely payoffs of those plans (Koechlin & Hyafil, 2007; Wallis, 2007), and anterior prefrontal regions might facilitate the coordination of multiple task goals (Badre & D'Esposito, 2009; Koechlin, Ody, & Kouneiher, 2003; Shallice & Burgess, 1991). Complementing these ideas, a major focus of recent research has been the hypothesis that recruitment of control by lateral PFC might depend on input from medial PFC regions that monitor ongoing performance for indications of
success or failure in the chosen task.

The conflict monitoring theory is the most influential articulation of the hypothesis that medial and lateral PFC regions form a regulatory feedback loop (Figure 1). [Figure 1 here]. According to this theory, control is recruited following the detection in medial PFC of competition—or conflict—in information processing (Botvinick, Braver, Carter, Barch, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004; Carter et al., 1998). This chapter surveys recent empirical and theoretical developments in research on this hypothesis. The first section reviews the conceptual and empirical foundations of the theory. The second section reviews empirical debates stimulated by the theory, for example regarding the neural basis of conflict monitoring and the varying forms of conflict that might be monitored in the human brain. The final section places the theory in a broader theoretical context, discussing challenging findings from studies of medial PFC function in other species that suggest a wider conception of the role of this region in the control of behavior.

(h1) Theoretical foundations

(h2) Conflict and competitive interactions

The idea that processing conflicts impose a fundamental constraint on human cognition has been articulated in several influential theories of visual attention (Desimone & Duncan, 1995; Treisman & Gelade, 1980), action selection (Allport, 1987; Norman & Shallice, 1986), working memory (Baddeley, 1996), and executive function (Logan, 2003; Miller & Cohen, 2001). In each domain, top-down cognitive control is thought to be needed to impose coherence on thought and action. In particular, the central role of processing conflict, and the need for top-down modulation, is inferred to follow from three basic and interrelated features of human cognition: that behavior is flexible, that it is goal-directed, and that it is an emergent property of
coordinated activity across distributed neural circuits.

The flexibility of human cognition is apparent even in the artificial simplicity of the laboratory. Presented with words on a screen, for example, subjects can read them aloud, memorize them, classify each as a living or manmade thing, search for a certain word, etc., and can switch fluently between these tasks according to whim or instruction (Logan, 1985; Monsell, 1996). Given that behavior is only loosely constrained by the environment, some form of control must be exerted to select among the available tasks and actions (Duncan, 1986; Miller & Cohen, 2001; Monsell, 1996). However, purposeful behavior need not always rely on effortful control: We often follow established routines that we execute with little thought or effort. The essence of flexible, goal-directed behavior lies in our ability to act effectively when we lack these routines and, perhaps even more critically, when achieving our behavioral goals requires us to override our habits (Norman & Shallice, 1986; Shiffrin & Schneider, 1977). Control is required in the Stroop task, for example, to suppress an established habit of reading words to allow naming the ink color of those words (Cohen, Dunbar, & McClelland, 1990; Stroop, 1935). Cognitive control is therefore needed because our behavioral goals are often in conflict with established response tendencies.

The importance of conflict in constraining our ability to behave coherently arises critically from the distributed nature of neural processing: Even the simplest cognitive task requires the coordination of activity across multiple brain regions. The distributed architecture of cognition is exemplified by the structure of the cortical visual system in which a proliferation of functionally specialized regions feeds into memory and action systems in a massively parallel manner. Within this parallel processing system, coordination is vital to ensure the coherent activation—or binding—of object features represented in distant regions (Treisman & Gelade,
and to ensure that features of selected objects are transmitted to appropriate action systems (Allport, 1980). This coordination is thought to depend on mechanisms of top-down cognitive control implemented in lateral PFC (Desimone & Duncan, 1995).

These considerations converge to suggest that regulating conflict is a core function of cognitive control. It follows that identifying these conflicts as they arise could provide a crucial signal of when increased control is required. This idea is at the heart of the conflict monitoring theory (Botvinick et al., 2001; Carter et al., 1998), which proposes specifically that anterior cingulate cortex (ACC), a region in medial PFC, monitors for the occurrence of conflict between incompatible actions. A canonical example of response conflict comes from the Stroop task, in which the presence of competing responses associated with incongruent color and word attributes (e.g., BLUE in red ink) results in impaired performance relative to the case in which color and word are congruent (e.g., RED in red ink). Response conflict similarly occurs in the flanker task when the central target is flanked by distracting peripheral stimuli associated with a different response (e.g., > > < > >) relative to the case when target and flankers are congruent (e.g., > > > > >). The conflict theory proposes that ACC detects these conflicts as they emerge during the action selection process, and uses this information to signal the need for increased control. Increased control by lateral PFC regions, which serve to enhance task-relevant stimulus and response representations, in turn leads to reduced conflict during subsequent task performance, thus closing the conflict-control loop (Figure 1).

(h2) Anterior cingulate cortex

ACC lies bilaterally around the rostrum of the corpus callosum (Figure 2) [Figure 2 here]. The conflict monitoring theory focuses in particular on dorsal-caudal ACC, Brodmann’s areas 24c’ and 32’ (Devinsky, Morrell, & Vogt, 1995; Vogt, Nimchinsky, Vogt, & Hof, 1995). As well
as being densely interconnected with neighboring premotor regions (Dum & Strick, 1993), this part of ACC shares strong anatomical connections with lateral PFC and parietal cortex (Beckmann, Johansen-Berg, & Rushworth, 2009; van Hoesen, Morecraft, & Vogt, 1993), and co-activates with these regions in human neuroimaging studies during task performance (Duncan & Owen, 2000; Koski & Paus, 2000; Nee, Wager, & Jonides, 2007; Paus, Koski, Caramanos, & Westbury, 1998) and in a slowly evolving manner at rest (Margulies et al., 2007; Raichle, 2010; but see Dosenbach et al., 2007). The region thus forms part of a well-characterized control (or “executive”) network with precisely the connectivity required by a conflict monitoring system: Interconnections with motor structures would provide access to information about co-activation of competing actions—at the level of both abstract motor plans and specific planned movements—while connections with lateral PFC would enable the region to signal the need for increased top-down control when such conflicts are detected.

Evidence that ACC plays this role in response conflict monitoring has come primarily from human neuroimaging studies. Using a variety of methods including PET, fMRI, and EEG, increased ACC activity has been found in conditions characterized by conflict between different actions: when subjects override habitual responses, when they select one of a number of potentially correct responses, and when they make errors (Figure 2; Botvinick et al., 2001; Botvinick et al., 2004). Perhaps the most replicated neuroimaging finding regarding ACC is its increased activity when the required response conflicts with a prepotent or habitual alternative. Thus, ACC activity is observed in the Stroop task when the written word is incongruent with the required color naming response (Bench et al., 1993; Liotti, Woldorff, Perez, & Mayberg, 2000; Pardo, Pardo, Janer, & Raichle, 1990), in the flanker task when peripheral distractors cue a different response than the attended target (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999;
Kopp, Rist, & Mattler, 1996), and in corresponding conditions of many other tasks (e.g., Braver, Barch, Gray, Molfese, & Snyder, 2001; Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003; Paus, Petrides, Evans, & Meyer, 1993; Peterson et al., 2002; Rubia et al., 2001; Taylor, Kornblum, Minoshima, Oliver, & Koepp, 1994).

ACC activation is similarly observed in underdetermined response tasks that induce conflict between multiple potentially correct responses, such as generating verbs associated with particular nouns (Crosson et al., 1999; Petersen, Fox, Posner, Mintun, & Raichle, 1989), or producing random sequences of finger movements (Frith, Friston, Liddle, & Frackowiak, 1991). Importantly, ACC activity in these tasks increases when the response is less constrained by the stimulus, for example when verbs are generated for nouns with many possible actions (e.g., a “ball” can be thrown, kicked, bounced, etc.) compared with nouns with one main associate (e.g., a “bell” tends to be rung) (Barch, Braver, Sabb, & Noll, 2000; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997). Thus, ACC activity varies in relation to the level of conflict between responses even when there is no defined correct response.

However, ACC also strongly activates when subjects make errors. Error-related ACC activity was first found using electrophysiological recordings in non-human primates (Gemba, Sasaki, & Brooks, 1986) and scalp EEG studies in humans (Falkenstein, Hohnsbein, Hoorman, & Blanke, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993). In human EEG, a component labeled the error-related negativity (ERN) is robustly observed within 100 ms of incorrect responses in choice decision tasks such as the Stroop and flanker tasks. Dipole modeling of the ERN suggests ACC as its likely neural source (Dehaene, Posner, & Tucker, 1994), a supposition confirmed using methods with more precise spatial localization (Carter et al., 1998; Debener et al., 2005; Wang, Ulbert, Schomer, Marinkovic, & Halgren, 2005). These converging findings
suggest that ACC is active following errors, when conflict would be expected to occur between correct and incorrect responses.

Evidence that conflict-related ACC activity reflects its role in the recruitment of control has come from studies of sequential adaptation effects. In the Stroop task, for example, interference from an incongruent word is reduced when the immediately preceding stimulus was also incongruent, relative to trials following a congruent stimulus (Kerns et al., 2004). The conflict theory explains this finding in terms of conflict detection on one trial leading to increased top-down control, and hence reduced conflict, on the subsequent trial. Consistent with this hypothesis, neuroimaging evidence indicates that lateral PFC activity increases after high conflict trials (indicating increased top-down control, Egner, Delano, & Hirsch, 2007; Egner & Hirsch, 2005) whereas ACC activity is reduced (reflecting the reduction in conflict, Botvinick et al., 1999). Indeed, conflict-related ACC activity on one trial is predictive of lateral PFC activity and concomitant performance improvements on the next (Kerns et al., 2004). Collectively, these findings support the hypothesis that ACC and lateral PFC have dissociable but complementary functional roles, with the former involved in recruiting control and the latter involved in the execution of control.

(h2) Computational models of conflict monitoring

The two central claims of the conflict monitoring theory—that ACC monitors for the occurrence of response conflict, and that it uses this information to signal the need for increased top-down control—have been formalized in computational models (Botvinick et al., 2001). One such model (Figure 3a) simulates behavior in the flanker task, in which subjects indicate the direction of a central arrow while ignoring task-irrelevant flankers. The model comprises a task network (black) and conflict-control feedback loop (blue). In the task network, units representing
possible stimuli activate corresponding left- and right-hand responses under the influence of top-down attention. Incongruent flankers (e.g., in the stimulus $< > <$) activate the incorrect response, producing conflict with the correct response and, hence, slower and less accurate responding than congruent flankers (e.g., in the stimulus $> > >$) (Cohen, Servan-Schreiber, & McClelland, 1992; Spencer & Coles, 1999). [Figure 3 here]

Importantly, the level of conflict in the response layer of the network accurately predicts empirically observed patterns of ACC activity in neuroimaging studies of the flanker task (Botvinick et al., 1999), and provides a computational basis for recruiting appropriate levels of top-down attentional control (Botvinick et al., 2001). Hopfield energy (Hopfield, 1982) provides a simple formulation of the conflict signal:

$$\text{Conflict} = -\sum_{i=1}^{N} \sum_{j=1}^{N} a_i * a_j * w_{ij}$$

where $a$ denotes unit activity, $w$ the connection weight between pairs of units, and $i$ and $j$ are indexed over competing network units. Thus defined, conflict increases when there is co-activation of units that share a mutually inhibitory connection. This product measure of conflict is not the only possible formulation (Botvinick et al., 2001; Yu, Dayan, & Cohen, 2009) but it formalizes conflict in an intuitive and straightforward way: On incongruent trials, the target and flankers activate different responses such that the product of response activation levels is large and conflict is high. On congruent trials, incorrect response activity is low or zero, the product is correspondingly low, and there is little or no conflict.

Closing the conflict-control loop, detected conflict can be used to adjust the level of attentional control according to the following equation (Botvinick et al., 2001):

$$C(t+1) = \lambda C(t) + (1 - \lambda)(\alpha E(t) + \beta)$$

where $C$ indicates the strength of top-down control, adjusted on a trial-by-trial basis ($t$) as a
function of previously experienced conflict (energy, $E$), scaled by a learning rate parameter $\lambda$ within limits imposed by parameters $\alpha$ and $\beta$. According to this equation, control increases when high levels of conflict are experienced, leading to enhancement of task-relevant stimulus and response representations on subsequent trials, thus enabling the model to account for sequential adaptation effects observed in experimental studies (Botvinick et al., 2001): Strengthening of control after a high-conflict incongruent trial leads to reduced conflict on the next trial (cf. Gratton, Coles, & Donchin, 1992). Equations 1 and 2 together therefore provide a computationally straightforward framework for dynamic and flexible recruitment of control.

As well as the flanker task model described above, this computational approach has been applied to Stroop and underdetermined word generation tasks (Botvinick et al., 2001), and used to simulate both conflict- and error-related adjustments in control (Rabbitt, 1966). In each case, the proposed model implements flexible control without a controlling homunculus: Conflict monitoring uses readily available information about activation levels of competing responses, and the conflict signal is converted algorithmically to set the level of top-down control. Crucially, these models have generated novel predictions that have spurred subsequent empirical work. For example, a core prediction is that increased conflict-related ACC activity on one trial should lead to increased cognitive control and improved performance on the subsequent trial. This prediction has been borne out in fMRI studies (Kerns et al., 2004; Liston, Matalon, Hare, Davidson, & Casey, 2006). Meanwhile, EEG methods have confirmed key model predictions regarding the timing of ACC activity: Whereas ACC activity is predicted to occur after incorrect responses are produced, as reflected in the ERN component of the EEG, conflict on trials with correct responses is predicted occur primarily before the response (Figure 3b), a prediction confirmed in EEG studies of a second component, the N2 (Figure 3c; Nieuwenhuis et al., 2003;
van Veen & Carter, 2002; Yeung, Botvinick, & Cohen, 2004). Predictions derived from computational models of conflict monitoring have therefore been borne out in studies of both the level (using fMRI) and timing (using EEG) of activity in ACC during task performance.

**Section 1: Current directions in conflict monitoring research**

The preceding section summarizes the core claims of the conflict monitoring theory:

- Conflict and competitive interactions place a fundamental constraint on human cognition;
- Activity in ACC varies with conflict in a manner predictive of changes in the level of top-down cognitive control; and conflict monitoring provides a computationally tractable mechanism for the flexible recruitment and allocation of control. Along the way, we have seen a number of claims—for example about the neural basis of conflict monitoring, its relation to error detection, and its role in sequential adaptation—that have each become the focus of considerable research interest in recent years. This section reviews emerging evidence on these topics and others, emphasizing the fertility of the conflict monitoring theory in stimulating new research while also considering important challenges to the theory.

**Section 2: Neural basis of conflict monitoring**

The foregoing discussion identifies dorsal-caudal parts of ACC (areas 24c’ and 32’) as the likely neural basis of conflict monitoring. This region is variously referred to as dorsal ACC (Botvinick et al., 2004), posterior medial frontal cortex (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004), anterior midcingulate cortex (Vogt, Berger, & Derbyshire, 2003), and the rostral cingulate zone or rostral cingulate motor area (Debener et al., 2005; Fink, Frackowiak, Pietrzyk, & Passingham, 1997; Picard & Strick, 1996, 2001). The variability in nomenclature at least in part reflects disagreement about the precision with which conflict-related activity has been localized, and about the functional anatomy of the region of interest. There is nevertheless
broad consensus that if conflict monitoring occurs within ACC, then the relevant region lies in
the caudal-dorsal “cognitive” division rather than the rostral-ventral “affective” division (Bush,
Luu, & Posner, 2000; Devinsky et al., 1995).

However, it has been argued that conflict-related activity may localize outside ACC
within the neighboring pre-supplementary motor area (pre-SMA, Rushworth, Walton,
Kennerley, & Bannerman, 2004; Ullsperger & von Cramon, 2004). Evidence for this view has
come from studies of the flanker, go/no-go and Simon tasks in which contrasts between high-
and low-conflict trials identify regions extending dorsally into the pre-SMA, whereas contrasts
between correct and error trials identify regions only in ACC (Garavan, Ross, Kaufman, & Stein,
Pre-SMA activity has also been observed in some tasks that would be expected to produce
conflict, for example when subjects are required to override a preplanned movement (Nachev,
primates has been found to vary with response conflict in pre-SMA and the supplementary eye
field (Isoda & Hikosaka, 2007; Stuphorn, Taylor, & Schall, 2000) but not in ACC (Emeric et al.,
2008; Nakamura, Roesch, & Olson, 2005).

Attempts to precisely localize conflict-related activity in the medial wall are hindered by
inherent limitations of neuroimaging methods that involve spatial smoothing and group
averaging of data, such that the location of peak activity varies considerably even across studies
with comparable methodologies (Barch et al., 2000; Ridderinkhof et al., 2004), and by the
considerable individual variability in sulcal anatomy in this region, where the dorsal cingulate
areas of interest may lie buried in the cingulate sulcus or may appear in an adjacent paracingulate
region (Paus, Otaky et al., 1996; Paus, Tomaiuolo et al., 1996; Vogt et al., 1995; Yucel et al.,
Adding yet further complication, many functions ascribed to pre-SMA—such as willed action selection (Lau, Rogers, & Passingham, 2006; Passingham, 1993; Passingham, Bengtsson, & Lau, 2010) and behavioral switching (Hikosaka & Isoda, 2010; Rushworth, Hadland, Paus, & Sipila, 2002)—would lead to co-activation with putative conflict monitoring signals in ACC.

Converging evidence nevertheless points to ACC as the consistent source of conflict-related activity in medial PFC. First, despite variability across individual studies, meta-analyses consistently localize the center of mass of conflict-related activity around the cingulate sulcus (Figure 2; Barch et al., 2000; Beckmann et al., 2009; Bush et al., 2000; Paus et al., 1998; Peyron, Laurent, & Garcia-Larrea, 2000; Ridderinkhof et al., 2004; Roberts & Hall, 2008). Second, studies that explicitly consider anatomical variability across subjects—using high-resolution single-subject analyses (Lutcke & Frahm, 2008) or contrasting activation patterns according to sulcal anatomy (Crosson et al., 1999)—localize conflict-related activity to the paracingulate sulcus and gyrus when present, and to the cingulate sulcus otherwise, tracking the known location of ACC areas 24c’ and 32’. Finally, error-related activity is rarely observed in pre-SMA, yet errors should produce high levels of conflict (Botvinick et al., 2001; Yeung et al., 2004), suggesting that observed pre-SMA activity reflects its role in action selection rather than conflict monitoring. In contrast, dorsal ACC is very consistently activated by both conflict and errors (Braver et al., 2001; Carter et al., 1998; Garavan et al., 2003; Holroyd et al., 2004; Kiehl, Liddle, & Hopfinger, 2000; Lutcke & Frahm, 2008; Mathalon, Whitfield, & Ford, 2003; Taylor et al., 2006). Collectively, these considerations converge to identify regions around the cingulate sulcus as the neural basis of conflict signals in the medial wall.

(h2) Conflict and error processing

According to the conflict monitoring theory, error-related ACC activity can be explained
in terms of response conflict that develops following errors, when continued stimulus processing leads to activation of the correct response that conflicts with the initial error (Botvinick et al., 2001). In computational simulations, this post-error conflict signal replicates key properties of the ERN component of the scalp-recorded EEG, such as its timing and sensitivity to factors including stimulus frequency and speed-accuracy emphasis (Yeung et al., 2004). Consistent with the hypothesis that the ERN reflects continuous conflict monitoring rather than explicit error detection as originally proposed (Falkenstein et al., 1990; Gehring et al., 1993; Scheffers & Coles, 2000), ERN amplitude is insensitive to whether or not errors are consciously detected (Klein et al., 2007; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). Subsequent empirical studies have shown, as predicted, that ERN amplitude varies with levels of response conflict (Frank, Woroch, & Curran, 2005), for example being greater when correct and incorrect responses are similar and therefore in more direct competition (Gehring & Fencsik, 2001).

As described above, a key prediction arising from conflict model simulations—that on correct trials ACC should primarily be active before the response—has been borne out in studies of a second EEG component, the N2 (Folstein & Van Petten, 2008; Kopp et al., 1996; Nieuwenhuis et al., 2003; van Veen & Carter, 2002; Yeung et al., 2004; Yeung & Nieuwenhuis, 2009). However, the hypothesis that the ERN and N2 both index conflict monitoring has been challenged by findings that the two components may be affected in opposite ways by cingulate lesions (Swick & Turken, 2002) and some psycho-active drugs (Riba, Rodriguez-Fornells, Morte, Munte, & Barbanoj, 2005; Ridderinkhof et al., 2002). In response, subsequent simulation studies have shown how such dissociations might arise in a conflict monitoring system, for example if cingulate lesions caused attentional deficits by disconnecting ACC from lateral PFC (Yeung & Cohen, 2006). Predictions arising from these simulations—for example, mimicking
the attentional effects of ACC lesions by increasing the salience of irrelevant stimuli in the flanker task—have since been confirmed in empirical work (Danielmeier, Wessel, Steinhauser, & Ullsperger, 2009; Yeung, Ralph, & Nieuwenhuis, 2007).

Also important to address are findings of anatomical dissociations between conflict and errors: Whereas conflict primarily activates dorsal ACC, error-related activity typically extends into the affective rostral-ventral subdivision of ACC (Kiehl et al., 2000; Mathalon et al., 2003; Menon, Adleman, White, Glover, & Reiss, 2001; Taylor et al., 2006; Wittfoth et al., 2008). This anatomical dissociation is not directly predicted by the conflict monitoring theory, but nor is it unexpected given that conflicts and errors may require different compensatory reactions. Thus, whereas conflict detection should primarily drive adjustments in attentional control (Fan, Hof, Guise, Fossella, & Posner, 2008; Kerns et al., 2004), errors may elicit affective responses (Pourtois et al., 2010; Taylor et al., 2006) and strategic speed-accuracy trade-offs (Botvinick et al., 2001; Rabbitt, 1966) that could impose additional cognitive demands (Jentzsch & Dudschig, 2009; Notebaert et al., 2009).

Nevertheless, substantial challenges remain to the conflict account of error-related ACC activity. First, the theory cannot explain evidence of ACC activity following the presentation of feedback indicating that an error has occurred (Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997), as discussed below in relation to reinforcement learning theories of ACC function. Second, the theory predicts that ACC activity on correct trials should primarily occur before the response, yet a correct-response negativity (CRN) is sometimes observed with timing comparable to the ERN (Gehring & Knight, 2000; Pailing & Segalowitz, 2004; Vidal, Hasbroucq, Grapperon, & Bonnet, 2000) although arguably reflecting different neural mechanisms (Coles, Scheffers, & Holroyd, 2001; Roger, Benar, Vidal, Hasbroucq, &
Burle, 2010; Vidal, Burle, Bonnet, Grapperon, & Hasbroucq, 2003; Yordanova, Falkenstein, Hohnsbein, & Kolev, 2004). Finally, some predictions from conflict model simulations have been disconfirmed (Burle, Roger, Allain, Vidal, & Hasbroucq, 2008; Carbonnell & Falkenstein, 2006; Steinhauer, Maier, & Hubner, 2008). For example, the theory predicts greater conflict when errors are quickly corrected, a prediction confirmed in some studies (Rodriguez-Fornells, Kurzbuch, & Muente, 2002) but not others (Burle et al., 2008). Thus, although the conflict theory successfully explains and predicts a range of findings regarding conflict and errors, debate continues as to whether it provides a comprehensive account of error-related ACC activity.

(h2) Conflict adaptation

The conflict monitoring theory claims that detection of conflict should lead to increased recruitment of cognitive control (Botvinick et al., 2001), reflected in increased activation of high-level task representations in lateral PFC (Kerns et al., 2004) and, subsequently, enhancement of task-relevant information in sensory and motor cortices (Egner & Hirsch, 2005). Experimental tests of this claim have focused on trial-to-trial sequential adaptation effects observed in the Stroop (Kerns et al., 2004), flanker (Gratton et al., 1992), and Simon (Sturmer, Leuthold, Soetens, Schroter, & Sommer, 2002) tasks, among others, in which the effects of task-irrelevant information—interference on incongruent trials and facilitation on congruent trials—are reduced following high-conflict incongruent trials (Figure 4) [Figure 4 here]. The theory explains these findings in terms of conflict detected on trial N-1 leading to increased control on trial N that attenuates the effects of task-irrelevant information on behavior.

It has been argued, however, that sequential adaptation effects are problematically confounded with low-level repetition and priming effects (e.g., Hommel, 2004; Mayr, Awh, & Laurey, 2003). For example, performance may improve on successive incongruent trials simply
because this trial sequence includes many exact repetitions of the stimulus and required response (Mayr et al., 2003). Excluding exact repetitions has been found to abolish sequential adaptation effects in some studies (Fernandez-Duque & Knight, 2008; Mayr et al., 2003; Nieuwenhuis et al., 2006). Moreover, sequential adaptation effects have sometimes been found only during the early stages of testing (Mayr & Awh, 2009) and to be somewhat independent of conflict experienced on trial N-1 as measured through EMG activity corresponding to the competing responses (Burle, Allain, Vidal, & Hasbroucq, 2005).

Nevertheless, it seems unlikely that repetition and priming fully account for observed sequential adaptation effects, which may be evident even when repetition effects are factored out via multiple regression (Notebaert & Verguts, 2007) or excluded from analysis (Kunde & Wuhr, 2006; Notebaert, Gevers, Verbruggen, & Liefvooghe, 2006; Ullsperger, Bylsma, & Botvinick, 2005). Critically, repetition trials were excluded in neuroimaging studies reporting that increased ACC activity on high-conflict trials is predictive of increased PFC activity and reduced conflict on subsequent trials (Egner & Hirsch, 2005; Kerns, 2006; Kerns et al., 2004; Larson, Kaufman, & Perlstein, 2009). Feature repetition provides no ready account of these findings, nor of why sequential adaptation should be lost following ACC lesions (di Pellegrino, Ciaramelli, & Ladavas, 2007), transcranial magnetic stimulation of lateral PFC (Sturmer, Redlich, Irlbacher, & Brandt, 2007), the presentation of positive feedback to counteract aversive effects of conflict (van Steenbergen, Band, & Hommel, 2009), or reducing the time between trials (Notebaert et al., 2006). Moreover, experiments favoring the feature repetition account may have overlooked other sequential factors such as negative priming (Tipper, 1985) that could mask conflict adaptation (Bugg, 2008; Davelaar & Stevens, 2009). Taken together, these combined findings provide strong support for the existence of conflict adaptation independent of feature repetition.
Recent studies have extended this conclusion to evaluate the specificity of conflict adaptation, asking whether conflict detection in one task can lead to subsequent control adjustments in another. Although cross-task adaptation is sometimes observed, for example when flanker task conflict on trial N-1 leads to reduced Stroop interference on trial N (Freitas, Bahar, Yang, & Banai, 2007; Kunde & Wuhr, 2006), many studies report that conflict adaptation is task-specific (Brown, Reynolds, & Braver, 2007; Egner et al., 2007; Fernandez-Duque & Knight, 2008; Funes, Lupianez, & Humphreys, 2010a; Kiesel, Kunde, & Hoffmann, 2006). Similarity between two tasks appears to be the critical factor: Cross-task adaptation is observed when tasks share similar stimuli, responses, or stimulus-response mappings (Akcay & Hazeltine, 2008; Cho, Orr, Cohen, & Carter, 2009; Notebaert & Verguts, 2008), but not otherwise. These findings suggest that conflict adaptation is process-specific: Adaptation generalizes when tasks share processes such as attending to particular stimulus features or selecting specific responses. Neuroimaging dissociations between conflict-control mechanisms in the Stroop and Simon tasks are consistent with this hypothesis (Egner et al., 2007). A question for future research is whether such dissociations reflect the existence of dissociable process-specific conflict monitoring systems in ACC, or if instead a common conflict monitoring system interacts with dissociable control systems depending on task structure (Akcay & Hazeltine, 2008).

(h2) Uses of conflict signals

Most research to date has focused on trial-to-trial conflict adaptation, but conflict-control feedback may operate over both shorter and longer time-scales, recruiting control online within a trial and setting levels of control over extended periods. Relevant to the latter, conflict is reduced when incongruent trials occur frequently across blocks of the Stroop and flanker tasks (Gratton et al., 1992; Tzelgov, Henik, & Berger, 1992), and ACC activity covaries with this effect (Carter et
Computational simulations show how accumulated trial-to-trial sequential adjustments in control might account for these block-level adjustments (Blais, Robidoux, Risko, & Besner, 2007; Botvinick et al., 2001; Verguts & Notebaert, 2008). However, block-level and sequential adjustments seem to differ in their generalization across tasks (Funes, Lupianez, & Humphreys, 2010b), persistence over training (Mayr & Awh, 2009), and influence on specific processing stages (Purmann, Badde, & Wendt, 2009). One intriguing hypothesis is that these dissociations reflect the operation of a unitary conflict monitoring system as it interacts with cognitive control systems operating over different time-scales (cf. Braver, Reynolds, & Donaldson, 2003). This issue is ripe for investigation, for example, using neuroimaging designs sensitive to both transient and sustained neural activity (Donaldson, 2004).

Still less is known about within-trial conflict adaptation. fMRI and EEG studies indicate that functional connectivity between ACC and frontoparietal regions dynamically increases on high-conflict trials (Cavanagh, Cohen, & Allen, 2009; Fan et al., 2008; Hanslmayr et al., 2008; Wang et al., 2010), but these interactions might serve primarily to influence future conflict adaptation. Perhaps more conclusive is recent evidence that trials with increased ACC activity are associated with reduced activation of incorrect responses (Pastotter, Hanslmayr, & Bauml, 2010), suggestive of ACC-mediated control adjustments within trials. Future research might usefully study such within-trial effects in other paradigms in which online recruitment of control has been inferred, such as suppression of direct response activation in the Simon and Stroop tasks (Davelaar, 2008; Ridderinkhof, 2002; Yu et al., 2009), reduction of conflict when the presented stimulus has previously been associated with high levels of conflict (Crump, Gong, & Wmiken, 2006; Jacoby, Lindsay, & Hessels, 2003; Wendt, Kluwe, & Vietze, 2008), and reduced sensitivity to task-irrelevant information on trials in which a high level of conflict is experienced.
(Goschke & Dreisbach, 2008). It currently remains open whether ACC conflict monitoring might contribute to these effects.

Recent computational modeling studies have meanwhile suggested uses of conflict signals beyond simple recruitment of control. First, conflict may act as an online “braking signal” that prompts selective or cautious responding. This idea has been incorporated in models of cognitive control in task switching (Brown et al., 2007), as well as models of optimal decision making by the basal ganglia (Bogacz & Gurney, 2007) in which conflict signals from ACC prompt response inhibition by the subthalamic nucleus (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Frank, 2006). Second, conflict has been proposed as a useful signal driving learning (Blais et al., 2007; Botvinick, 2007; Verguts & Notebaert, 2009). For example, conflict adaptation effects have been simulated using models in which conflict modulates Hebbian learning between top-down control and bottom-up stimulus representations (Blais et al., 2007; Verguts & Notebaert, 2008). Conflict might also serve as an aversive signal to drive reinforcement learning of behavioral strategies to minimize cognitive effort (Botvinick, 2007). Specifically, whereas transient conflict indicates the need for increased control, conflict over an extended period might indicate that alternative strategies should be sought (Aston-Jones & Cohen, 2005). Thus, conflict monitoring may provide a valuable source of information that can be used in various ways to optimize behavioral performance, although a coherent and comprehensive account of these varied uses remains to be developed.

(h2) Dissociable forms of conflict

Comparisons between conflict tasks have typically identified overlapping activation patterns in ACC (Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; Rubia et al., 2001; Wager et al., 2005), regardless of the stimuli inducing the conflict (Roberts & Hall, 2008) and
the response modality in which the conflict occurs (Barch et al., 2001). Collectively, these findings suggest that a common ACC system monitors for conflicts among representations of very different kinds. However, one dissociation now well established is between stimulus and response conflict (Davelaar, 2008; Liston et al., 2006; Milham et al., 2001; Schulte et al., 2009; van Veen & Carter, 2002; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001; Wendelken, Ditterich, Bunge, & Carter, 2009). Stimulus and response conflict can be dissociated using flanker tasks in which two stimuli require left-hand responses (e.g., H and P) and two require right-hand responses (e.g., S and M), so that response conflict trials (e.g., HHSHH) can be compared with trials in which the stimuli conflict but their associated responses do not (e.g., HHPPHH). Stimulus conflict has been found to activate lateral frontal and parietal cortex (Liston et al., 2006; Wendelken et al., 2009) but produces little activity in ACC (van Veen et al., 2001), as one might expect given that ACC receives little direct sensory input (Beckmann et al., 2009; Margulies et al., 2007; van Hoesen et al., 1993).

A proposed dissociation between cognitive and emotional conflict has a similarly strong basis in the differential connectivity of the dorsal-caudal and rostral-ventral subdivisions of ACC (Bush et al., 2000; Devinsky et al., 1995). Rostral ACC activity is observed in conflict tasks in which task-irrelevant information is affectively valenced (Bishop, Duncan, & Lawrence, 2004; Vuilleumier, Armony, Driver, & Dolan, 2001; Whalen et al., 1998), for example when subjects judge whether faces are happy or fearful while ignoring an irrelevant superimposed word (HAPPY or FEAR) (Etkin, Egner, Peraza, Kandel, & Hirsch, 2006). This activity is functionally and anatomically dissociable from response conflict-related activity in dorsal ACC (Mohanty et al., 2007). However, dorsal and rostral ACC do not appear to independently monitor response and emotional conflicts, but rather interact during emotional conflict in a way that mirrors the
interaction between dorsal ACC and lateral PFC in resolving response conflict (Egner, Etkin, Gale, & Hirsch, 2008). Thus, dorsal ACC activity is also observed in emotional conflict tasks (Davis et al., 2005; Egner et al., 2008; Mohanty et al., 2007; Wittfoth et al., 2010) and, in emotional conflict adaptation, rostral ACC activity increases rather than reduces following high-conflict trials (Etkin et al., 2006).

Within dorsal ACC, dissociable activity patterns have been reported as a function of the nature of conflicting representations. Most research and theorizing to date has focused on conflict between simple actions, typically key-presses made with different fingers, but efforts have been made to study conflict between more abstract response representations. Thus, decision conflict, which occurs when subjects choose between options that are not (or not yet) associated with particular actions, has been found to activate more anterior portions of ACC than conflict between defined actions (Pochon, Riis, Sanfey, Nystrom, & Cohen, 2008; Venkatraman, Rosati, Taren, & Huettel, 2009). The opposite organization is suggested by studies of conflict between high-level task and semantic representations (Aarts, Roelofs, & van Turennout, 2009; Melcher & Gruber, 2009; Milham & Banich, 2005; Ruff, Woodward, Laurens, & Liddle, 2001; van Veen & Carter, 2005). Task-level Stroop conflict can be identified, for example, by contrasting congruent (e.g., RED in red ink) and neutral stimuli (e.g., XXX in red ink). Response conflict should be low in both cases, but the irrelevant word in congruent stimuli should lead to conflict between the word reading and color naming tasks. Task conflict has been found to activate posterior regions of ACC and pre-SMA, contrasting with anterior activation seen to decision conflict (Melcher & Gruber, 2009; Milham & Banich, 2005; but see Roelofs, van Turennout, & Coles, 2006). Thus, although findings from decision, semantic and task-level conflict suggest subdivisions within dorsal ACC, the precise functional architecture of the region remains unknown.
Finally, recent studies have investigated whether ACC activates solely during the experience of conflict—our focus thus far—or also during anticipation of conflict, using pre-cues that indicate the expected level of conflict prior to stimulus presentation. Despite some evidence to the contrary (Luks, Simpson, Dale, & Hough, 2007; MacDonald, Cohen, Stenger, & Carter, 2000; Nieuwenhuis, Schweizer, Mars, Botvinick, & Hajcak, 2007), it now seems clear that anticipatory activity occurs in ACC (Aarts, Roelofs, & Van Turennout, 2008; Dosenbach et al., 2006; Fan et al., 2007; Luks, Simpson, Feiwell, & Miller, 2002), and more so for cues indicating high conflict to come (Barber & Carter, 2005; Sohn, Albert, Jung, Carter, & Anderson, 2007; Weissman, Gopalakrishnan, Hazlett, & Woldorff, 2005). One recent study found that anticipatory activity dissociates from conflict-related activity in ACC (Orr & Weissman, 2009), with the former lying more posterior and extending into pre-SMA. There is notable overlap between this anticipatory activity and task-level conflict effect discussed above. In both cases, it remains open whether the focus of activity lies within ACC or pre-SMA, and whether the activity is conflict-related or rather reflects processes related to attention and task preparation (Brass & von Cramon, 2002; Orr & Weissman, 2009).

**Neuropsychology of conflict monitoring**

The vast majority of conflict monitoring studies have used correlational neuroimaging methods, but there is obvious value in establishing more causal evidence. However, although ACC lesions are consistently associated with deficits in executive function (Devinsky et al., 1995), evidence specific to conflict monitoring has proven controversial. Several studies have found that Stroop interference is not markedly increased in patients with cingulate lesions (Baird et al., 2006; Fellows & Farah, 2005; Stuss, Floden, Alexander, Levine, & Katz, 2001), and these patients—including some with extensive bilateral damage—may show intact conflict adaptation...
(Fellows & Farah, 2005) even if their subjective experience of conflict is markedly reduced (Naccache et al., 2005). However, there are also several reports of impaired performance in a variety of conflict tasks in patients with ACC lesions (Modirrousta & Fellows, 2008; Picton et al., 2007), including some with very focal damage (Cohen, Kaplan, Moser, Jenkins, & Wilkinson, 1999; Cohen, Kaplan, Zuffante et al., 1999; Ochsner et al., 2001; Swick & Turken, 2002; Turken & Swick, 1999; Yen et al., 2009), and with crucial lesion sites mapping on well to the localization of conflict-related activity in neuroimaging studies (Swick & Jovanovic, 2002). Impaired conflict adaptation and error monitoring have also been observed following damage to dorsal ACC (Modirrousta & Fellows, 2008; Swick & Jovanovic, 2002) as well as in patients with damage affecting more rostral ACC regions (di Pellegrino et al., 2007).

These discrepancies may at least partly reflect methodological issues: Many patient studies use blocked designs in which conflict is predictable, yet ACC is primarily important when adapting to varying levels of conflict (Floden, Vallesi, & Stuss, 2010). Nevertheless, whereas cingulate patients only inconsistently show increased sensitivity to response conflict, the same patients very consistently exhibit general slowing of responding across tasks irrespective of conflict (Alexander, Stuss, Shallice, Picton, & Gillingham, 2005; Picton et al., 2007; Stuss et al., 2005). Such findings have suggested a role for ACC in “energizing” responding (Stuss et al., 2005), but this hypothesis is challenged by evidence that ACC is sensitive to tasks requiring response inhibition (Nieuwenhuis et al., 2003; Rubia et al., 2001; Tsuchida & Fellows, 2009). Instead, the combined findings are intriguing when taken with the recent proposal that conflict signals from ACC may be used to set an appropriate response threshold to optimize decision making (Bogacz et al., 2010; Frank, 2006). If correct, this interpretation would support the view that the role of ACC extends beyond conflict-control regulation to include more general
functions relevant to setting appropriate response strategies.

Conflict monitoring has also been investigated in psychiatric patient populations. Abnormalities in conflict-related activity are inconsistent in mood disorder and attention-deficit hyperactivity disorder, but are well-established in patients with schizophrenia and obsessive-compulsive disorder (Melcher, Falkai, & Gruber, 2008; Ullsperger, 2006). As part of a broader pattern of hypofrontality (Benes, 2000; Cohen & Servan-Schreiber, 1992) and performance monitoring deficits (Malenka, Angel, Hampton, & Berger, 1982), patients with schizophrenia exhibit reduced conflict-related ACC activity (Carter, Mintun, Nichols, & Cohen, 1997; McNeely, West, Christensen, & Alain, 2003) and reduced conflict adaptation (Kerns et al., 2005). These findings are suggestive of conflict monitoring deficits that might contribute to both the positive (psychotic) symptoms and the cognitive deficits evident in these patients. In contrast, ACC tends to be hyperactive in patients with obsessive-compulsive disorder, both in association with errors (Gehring, Himle, & Nisenson, 2000; but see Nieuwenhuis, Nielen, Mol, Hajcak, & Veltman, 2005) and response conflict (Maltby, Tolin, Worhunsky, O'Keefe, & Kiehl, 2005; Ursu, Stenger, Shear, Jones, & Carter, 2003; Yucel et al., 2007), with the degree of ACC hyperactivity predictive of symptom severity. There is obvious intuitive appeal to the notion that patients with obsessive-compulsive disorder have an overactive action monitoring system.

Such findings suggest that performance monitoring deficits may contribute to the symptoms apparent in patient groups and may provide useful diagnostic measures in clinical settings (Melcher et al., 2008; Ullsperger, 2006). Such inferences must be drawn with caution, however, because disruptions in ACC performance monitoring might be secondary symptoms rather than a primary cause of observed deficits (Yeung & Cohen, 2006): ACC function is known to be sensitive to damage to a widespread network of regions, including lateral PFC and
the basal ganglia (Gehring & Knight, 2000; Ullsperger & von Cramon, 2006). Existing findings nevertheless provide grounds for optimism that conflict monitoring might provide a useful framework for investigating cognitive deficits associated with ACC dysfunction. A crucial open question is whether therapeutic interventions targeted at conflict monitoring can ameliorate the cognitive deficits in these patient populations (cf. Edwards, Barch, & Braver, 2010).

(h1) Conflict monitoring in context

The research reviewed above has been conducted within the framework of the conflict monitoring theory, providing extensions and challenges to parts of that framework. This final section considers the theory in the context of wider research on ACC function. As will become apparent, consideration of evidence gathered via other methodologies, in other species, and involving wider conceptions of the functional role of ACC, brings into question the degree to which conflict monitoring provides a comprehensive account of the role of this region in control and in cognition more widely.

(h2) Conflict monitoring in non-human primates

Non-human primate studies have to date provided little support for the conflict monitoring theory. Whereas recordings of spiking activity and local field potentials in human ACC reveal evidence of conflict-related activity (Davis, Hutchison, Lozano, Tasker, & Dostrovsky, 2000; Davis et al., 2005; Wang et al., 2005), no such effects have been observed in monkey ACC (Emeric et al., 2008; Ito, Stuphorn, Brown, & Schall, 2003; Nakamura et al., 2005). Moreover, ACC lesions in primates have been found to have no effect on conflict adaptation, which is instead primarily sensitive to lateral PFC damage (Mansouri, Buckley, & Tanaka, 2007). More promising evidence of conflict-related activity in primates has been observed in the pre-SMA and supplementary eye field (Isoda & Hikosaka, 2007; Nakamura et
al., 2005; Stuphorn et al., 2000). However, conflict effects in these regions typically reflect modulations of movement-selective task-related activity, rather than pure signals of conflict independent of the responses selected (Isoda & Hikosaka, 2007; Nakamura et al., 2005). Indeed, these findings have led to the suggestion that conflict-related activity in human neuroimaging studies may be an artifact of coarse-grained neuroimaging methods averaging activity across populations of neurons that selectively code for conflicting responses (Nakamura et al., 2005).

Given the serious implications this conclusion would have for all of the research considered in this chapter, it is unsurprising that alternative accounts of the between-species discrepancy have been proposed (Botvinick et al., 2004; Cole, Yeung, Freiwald, & Botvinick, 2009). However, several potential explanations can be immediately ruled out. First, the discrepancy is unlikely to reflect the fact that human studies typically use manual responses whereas primate work mostly uses saccadic tasks, because conflict-related activity is observed in human ACC when saccadic tasks are used (Curtis, Cole, Rao, & D'Esposito, 2005; Paus et al., 1993). Nor is it likely that primate studies have simply failed to find conflict-signaling neurons that are nonetheless present, given that these studies sampled widely in ACC and had little trouble identifying error-related activity, suggesting that the regions under scrutiny are homologous (Nakamura et al., 2005). Moreover, although human ACC contains cell types not found in monkeys (Nimchinsky et al., 1999), these large spindle neurons are restricted to the cingulate gyrus rather than the sulcal regions that are the focus of conflict-related activity.

Several possible accounts of the between-species discrepancy remain after excluding these alternatives. First, and perhaps most strikingly, human conflict-related activity localizes to area 32′—an area of cingulofrontal transition cortex interposed between the cingulate motor areas and the pre-SMA and SMA—which has no direct homolog in dorsal ACC in monkeys.
As such, conflict monitoring functions may be substantially more developed in humans compared with non-human primates (Cole et al., 2009). Second, conflict-related activity may be less evident in single-unit spiking activity than in neuroimaging recordings that are sensitive to synchronization and synaptic activity within local neural circuits (Botvinick et al., 2004). Consistent with this interpretation, a recent fMRI study of awake behaving monkeys found strong ACC activity in a conflict-inducing antisaccade task (Ford, Gati, Menon, & Everling, 2009). Third, it may be misguided to expect conflict-related activity in single-unit recordings to occur independent of the particular responses selected: Recent computational modeling suggests that conflict may be detected in a response-specific manner (Brown et al., 2007), and may lead to modulation of control that is similarly stimulus- and response-specific (Blais et al., 2007; Verguts & Notebaert, 2008).

Each of these possibilities identifies important avenues for future research. Thus, if conflict monitoring is uniquely developed in humans, it becomes important to explain why primate behavior is nonetheless sensitive to conflict (Mansouri et al., 2007; Nakamura et al., 2005) and to identify cognitive abilities that might depend critically on this supposedly unique function. Conversely, if conflict signals are apparent in both humans and monkeys but only in regional activity, it becomes vital to identify the functional role of these signals given that spiking activity is required if conflict signals are to be conveyed to other cortical regions (Nakamura et al., 2005). One hypothesis could be that local activity in ACC varies with conflict, as envisioned by the theory, but its spiking output may convey much more than a simple scalar signal of conflict, for example indicating specific changes in control strategy. According to this hypothesis, activation patterns in ACC should correlate with specific behavioral strategies, a prediction ripe for evaluation using multivariate fMRI analysis approaches. Finally, if conflict is
detected in a response-specific manner, as envisioned in recent computational models, the behavior of units in these models should be meaningfully related to single-unit activity recorded in monkey ACC. Thus, although studies in non-human primates have so far provided little support for the conflict monitoring theory, they suggest a number of intriguing avenues for its development in future research.

**(h2) Reinforcement and reward-guided decision making**

Conflict monitoring research typically focuses on ACC activity in tasks in which the required response is clearly defined: The challenge for subjects is to produce this response in the face of conflict from a prepotent or habitual alternative (Miller & Cohen, 2001). However, ACC activity has also been widely studied using tasks in which the correct response is not instructed, but must be learnt by trial-and-error using feedback. In such tasks, robust ACC activity is not only observed as subjects choose their response (Blair et al., 2006; Holroyd et al., 2004; Marsh, Blair, Vythilingam, Busis, & Blair, 2007; Walton, Devlin, & Rushworth, 2004)—a finding that might reflect conflict between the different options—but also as subjects anticipate or receive feedback about their choices (Bush et al., 2002; Gehring & Willoughby, 2002; Holroyd et al., 2004; Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; Miltner et al., 1997; Yeung & Sanfey, 2004). This evidence supports the view that ACC plays a critical role in reward-guided decision making and learning, serving to guide behavior by associating actions with their outcomes (Alexander & Brown, 2011; Holroyd & Coles, 2002; Rushworth, Behrens, Rudebeck, & Walton, 2007). Consistent with this hypothesis, feedback-related ACC activity increases with the unpredictability (Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003) and volatility (Behrens, Woolrich, Walton, & Rushworth, 2007) of experienced outcomes, and is predictive of whether subjects will avoid repeating incorrect responses in the future (Hester, Barre, Murphy, Silk,
This view of ACC function is supported by single-unit recordings in non-human primates where, in contrast to the apparent absence of conflict-sensitive neurons, coding of reward and reinforcement is well-established (Amiez, Joseph, & Procyk, 2006; Shidara & Richmond, 2002). Indeed, when comparable tasks are used in humans and non-human primates, close correspondences are observed between ACC activity across species (Bush et al., 2002; Shima & Tanji, 1998; Williams, Bush, Rauch, Cosgrove, & Eskandar, 2004). ACC neuron activity specifically appears to code the value of experienced outcomes in relation to selected actions (Hayden & Platt, 2010; Luk & Wallis, 2009), for example increasing when unexpected reductions in reward lead to changes in behavior (Shima & Tanji, 1998; Williams et al., 2004). This activity appears to play a causal role in learning: Focal lesions in ACC impair animals’ ability to use reinforcement to learn to select appropriate actions (Amiez et al., 2006; Hadland, Rushworth, Gaffan, & Passingham, 2003; Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006; Shima & Tanji, 1998).

Given these successes, an important question concerns the relation between reward-guided decision making and conflict monitoring. It could be that they are dissociable functions of separate ACC subregions (Marsh et al., 2007; Rogers et al., 2004). However, meta-analyses indicate substantial overlap in conflict- and feedback-related activity within ACC (Beckmann et al., 2009; Ridderinkhof et al., 2004). Moreover, reinforcement learning and conflict monitoring in certain cases provide contrasting accounts of the same phenomena (Alexander & Brown, 2011; Holroyd & Coles, 2002; Yeung et al., 2004), suggesting the importance of reconciling the theories. One suggestion has been that conflict-related activity in ACC is artifactual: a by-product of coding the likelihood of errors or anticipation of negative outcomes that correlate with
conflict (Brown & Braver, 2005). However, this hypothesis struggles to explain ACC activity seen in tasks in which errors are rare (e.g., in the Stroop task; Carter et al., 2000) or absent (e.g., in word generation tasks in which the conflict occurs between multiple correct responses; Barch et al., 2000; Thompson-Schill et al., 1997). Moreover, when conflict and error likelihood are dissociated based on response speed—because slow responses produce high conflict but low error likelihood—ACC activity varies with the level of conflict (Yeung & Nieuwenhuis, 2009).

Recent computational modeling work has attempted to address these criticisms of reinforcement learning approaches, using more complex algorithms that are capable of learning vector representations of expected response outcomes rather than simple scalar predictions of expected reward (Alexander & Brown, 2011). However, this model fails to account for benchmark features of ACC activity seen during reinforcement learning, such as its specific sensitivity to negative outcomes even when those outcomes are as likely as positive outcomes (Holroyd et al., 2003; Yeung & Sanfey, 2004). As such, a reductive model of conflict- and learning-related activity in ACC remains elusive.

Instead, conflict monitoring and reinforcement learning seem to provide complementary rather than competing accounts of ACC function (Botvinick, 2007). In particular, ACC might contribute broadly to the optimization of behavior through its sensitivity to signals indicating the efficiency and effectiveness of actions, both internal (in the case of conflict) and environmental (in the case of reinforcement learning). One specific proposal is that conflict registers as a cost to drive avoidance learning against effortful or error-prone options (Botvinick, 2007). This interpretation is consistent with evidence from animal studies that ACC plays a crucial role in evaluating the effort associated with actions (Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006; Walton, Bannerman, Alterescu, & Rushworth, 2003), and extends this notion
to encompass mental as well as physical effort. More generally, these ideas begin to situate the conflict monitoring theory within the broader context of ACC function as it is understood in a range of task contexts and in a variety of species.

**(h2) Contrasting theories of ACC function**

A broader conception of ACC function might also help relate the conflict monitoring theory to other perspectives on the role of ACC. One influential hypothesis is that ACC supports action selection (Dum & Strick, 1991; Picard & Strick, 1996), particularly when responding is voluntary rather than cued by stimuli in the environment (Paus, 2001). Thus, patients with cingulate lesions produce fewer spontaneous verbalizations and movements than controls, and demonstrate corresponding deficits in unstructured tasks (e.g., when asked simply to draw as many pictures or designs as possible over a fixed period; Cohen, Kaplan, Zuffante et al., 1999). Deficits in voluntary action are seen in extreme form in akinetic mutism, in which patients become immobile and unresponsive to external stimuli following bilateral damage to ACC and nearby cortex (Nemeth, Hegedus, & Molnar, 1988). Neuroimaging studies report greater ACC activity during self-initiated movement—for example, when movements are produced in an irregular rhythm—than when the same movements are made in response to external stimuli (Deiber, Honda, Ibanez, Sadato, & Hallett, 1999; Krieghoff, Brass, Prinz, & Waszak, 2009). Corresponding single-unit recordings in non-human primates demonstrate that this activity may precede movement initiation by more than a second (Shima et al., 1991).

Conflict monitoring provides no ready account of these findings: Although the neuroimaging findings might reflect increased conflict when actions are self-initiated (Barch et al., 2000; Botvinick et al., 2001), it is unclear why a lesion-induced conflict monitoring deficit should reduce spontaneous speech and action. Instead, a more productive reconciliation might be
in relation to the view that conflict monitoring represents one facet of the role of ACC in evaluating the costs and benefits of actions (Botvinick, 2007). Such evaluation would be vital in voluntary action selection, which must normally depend on a cost-benefit analysis of action choices (Holroyd & Yeung, 2012). Within this framework, an important open question is whether ACC itself directly selects or initiates actions (Paus, 2001), or rather supports action selection indirectly via its connections with premotor and prefrontal cortex (Kouneiher, Charron, & Koechlin, 2009). Recent neuroimaging evidence has been taken to favor the former view (Banich, 2009; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008).

The notion that ACC plays a regulative role, directly exerting control rather than simply evaluating performance, has been a longstanding alternative view of ACC function (Posner & Petersen, 1990). Although this hypothesis is challenged by conflict adaptation effects—after high-conflict trials, ACC activity reduces with the level of conflict rather than increasing with the level of control (Botvinick et al., 1999)—recent findings have reignited interest in regulative accounts of ACC function (Banich, 2009; Dosenbach et al., 2008). Evidence that ACC not only activates transiently during conflict, but also during preparation and in a sustained manner throughout task performance (Dosenbach et al., 2006), has led to the view that ACC is part of a core system for maintaining a stable cognitive state over time (Dosenbach et al., 2008). Meanwhile, evidence that conflict-related changes in ACC activity may disappear when between-condition differences in reaction time are controlled for (Grinband et al., 2011) has led to the suggestion that ACC plays a more general role in decision making. Finally, EEG evidence that ACC activity follows lateral PFC activity in the Stroop task, and modulates its influence on behavioral performance (Silton et al., 2010), has been taken to show that ACC operates at a relatively late stage of a control “cascade”, making transient contributions to the selection of
specific actions (Banich, 2009).

These theories present a different view of ACC function to that proposed by the conflict monitoring theory. However, given disagreement about the supposed nature of control exerted by ACC—sustained task-level control versus transient response selection—it remains unclear whether a coherent integration is possible. Moreover, current evidence is far from definitive in challenging the conflict monitoring theory: Preparatory ACC activity could reflect task-level conflict (Aarts et al., 2008; Melcher & Gruber, 2009; Milham & Banich, 2005; Ruff et al., 2001; van Veen & Carter, 2005), sustained activity might reflect the integration of conflict over an extended period (Aston-Jones & Cohen, 2005; Botvinick et al., 2001), activity related to variation in reaction time is well explained by existing models of conflict monitoring (Yeung, Cohen & Botvinick, 2011), and ACC-related modulation of control by lateral PFC might reflect within-trial conflict adaptation (Cavanagh et al., 2009; Fan et al., 2008; Hanslmayr et al., 2008; Wang et al., 2010). Thus, ACC involvement in regulative rather than evaluative aspects of control remains to be determined. It is nevertheless increasingly clear, particularly given evidence of ACC involvement in voluntary action selection and reinforcement learning, that conflict monitoring is just one aspect of the role of ACC in the evaluation and control of thought and action.

**(h1)** Conclusion

The conflict monitoring theory has been a significant success story of cognitive neuroscience research. Countering claims that neuroimaging is of limited relevance to theories of cognition (Coltheart, 2006), the initial impetus for the theory came primarily from imaging work—specifically, emerging evidence of ACC activity associated with a diverse range of cognitive demands—but the theory itself is truly cognitive: It proposes a computationally
specified process that explains not only the original neuroimaging evidence but also established findings in the behavioral literature. In three critical respects, therefore, the theory has proven highly successful. First, it has given substantial impetus to a research field—on performance monitoring and the recruitment of control—that had been largely neglected in prior work. Second, it has provided a powerful set of behavioral, neuroimaging, and computational tools with which to probe these functions. And third, it has proven immensely productive in generating testable predictions about behavior and brain activity that have, by and large, been borne out experimentally: The original empirical (Carter et al., 1998) and conceptual (Botvinick et al., 2001) expositions of the theory have now been cited over 1000 times each, attesting to the influence of the theory in cognitive neuroscience. Ongoing efforts to enrich understanding of the ways conflict might be detected and used (Aston-Jones & Cohen, 2005; Blais et al., 2007; Botvinick, 2007; Brown et al., 2007; Frank, 2006; Verguts & Notebaert, 2009) bode well for the continuing vitality of the theory as a source of new research.

These successes notwithstanding, scientific theories are ultimately judged according to whether or not they are right, and in this respect the legacy of the conflict monitoring theory is far less assured. Several substantive challenges have been discussed, but two issues in particular seem fundamental. The first arises from human neuropsychological studies: Conflict monitoring is proposed as a fundamental mechanism underpinning the recruitment of control, yet patients with ACC lesions only inconsistently exhibit deficits in conflict monitoring, adaptation, and control, and are notably spared in overall intellect and executive function. The second challenge arises from the failure to find evidence of conflict monitoring in non-human primates (and other animals). Again, if conflict monitoring is a fundamental cognitive function, one would expect it to be present in our near relatives, or for there to be obvious between-species differences in the
experience or expression of conflict, yet evidence for either possibility is currently lacking.

The continuing success and influence of the conflict monitoring theory suggest that it would be premature to abandon it in the face of these challenges. Instead, these issues are perhaps most productively viewed as identifying key questions for future research. As discussed above, one such question is the relation between conflict monitoring, action selection, and reinforcement learning. Recent work has begun to reconcile these views (Botvinick, 2007; Holroyd & Yeung, 2012), but it remains unclear whether these integrative accounts can explain the wealth of evidence supporting the conflict monitoring theory as originally presented. Meanwhile, the relative insensitivity of current lab paradigms to lesions in ACC suggests a need to identify tasks that more effectively index this functioning. On the assumption that cognitive control crucially supports the structuring of behavior through time (Duncan, 2010) and the rapid, flexible adaptation to new requirements (Monsell, 2003), the immediate theoretical challenge is to articulate clearly the role that conflict monitoring might play in these abilities. The corresponding empirical challenge is to design tasks that precisely probe this contribution, and to evaluate their sensitivity to hypothesized deficits following ACC lesions and between-species differences in ACC function. Efforts towards these goals would represent an important step in the continuing evolution of the conflict monitoring theory: from an account primarily inspired by findings from constrained laboratory paradigms, towards a mature and detailed theory that makes direct contact with established principles of cognitive control.

**Future directions**

1. Why do recordings of regional activity (e.g., using EEG and fMRI) show evidence of conflict-related activity in ACC while single-unit recordings generally fail to do so? Is the
regional conflict signal merely an artifact of averaging across populations of neurons whose spiking activity is affected by, but does not code for, conflict? Or:

a. Is conflict detected in a stimulus/response-specific manner, such that it would be misguided to search for conflict signals that are independent of the response selected?

b. Does the spiking output of ACC contain more information than a simple scalar signal of experienced conflict, such as instructions about specific changes in behavioral strategy?

2. Can the conflict monitoring theory explain why performance on conflict tasks is only inconsistently affected by ACC lesions in humans? Is conflict-related activity seen in neuroimaging studies epiphenomenal? Or:

a. Do the deficits suffered by these patients provide important clues regarding additional uses of the ACC conflict signal, for example in setting an appropriate response threshold?

b. Are typical laboratory tasks too simple to capture the contribution of conflict monitoring to flexible, structured human behavior? If so, how does conflict monitoring by ACC support rapid behavioral shifting or the structured sequencing of thought and action?

3. How can the conflict monitoring theory be reconciled with the presumed role of ACC in reinforcement learning and voluntary action selection? Can a reconciled view account for the large corpus of existing data collected in support of the theory?

(h1) References


Benes, F. M. (2000). Emerging principles of altered neural circuitry in schizophrenia. *Brain


prefrontal cortex and hippocampus that contribute to planning and prospective memory.


Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex, 42*, 323-331.


course of the Stroop color-word interference effect. *Neuropsychologia, 38*, 701-711.


Melcher, T., & Gruber, O. (2009). Decomposing interference during Stroop performance into


Ochsner, K. N., Kosslyn, S. M., Cosgrove, G. R., Cassem, E. H., Price, B. H., Nierenberg, A. A.,


and superior-rostral sulci: hemispheric asymmetries, gender differences and probability maps. *Journal of Comparative Neurology, 376*, 664-673.


Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to


dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *NeuroImage, 14*, 1387-1401.


imaging probe of the anterior cingulate affective division. *Biological Psychiatry, 44,* 1219-1228.


a. Model architecture:

 ![Diagram of model architecture with nodes labeled for Stimulus, Attention, Conflict monitoring, Response, and connections between them.]

b. Conflict simulation:

 ![Graph showing ERP responses with N2 and ERN peaks, labeled for 100 ms after response.]

c. EEG data:

 ![Graph showing EEG data with N2 and ERN peaks, labeled for 100 ms after response.]

- Green line represents Incongruent minus congruent
- Red line represents Error minus correct

100 ms Resp.
Low response time vs. previous trial conflict:

- High conflict: Low response time
- Low conflict: High response time

Current conflict: