Comments and Controversies

The blind executive

Parashkev Nachev

Institute of Neurology, UCL, Queen Square, London, UK

For reasons we must leave an anthropologist to answer, the mind is irresistibly conceived in broadly tripartite form. Plato's original distinction between the cognitive, the affective, and the conative has found close parallels in such disparate thinkers as Kant ("knowing", "feeling" and "willing"), Freud (Ego, Super-ego, and Id), and many others since. Modern neuroscientific models follow the same fundamental organisation, mapping the mind onto an analogous tripartite brain, with more or less discrete executive, affective, and evaluative components. Valid or not, this conceptual framework creates an immediate problem: how do these subsystems interact to generate unified cognition and behaviour?

The problem is especially acute in relation to the "executive", a system putatively invoked in situations where controlled behaviour would otherwise give way to simpler, more automatic responses (e.g. (Norman and Shallice, 2000)). The executive must either be continually active, in which case it is absorbed into the systems it is supposed to govern and ceases to be worthy of the name, or else it must somehow "know" when to intervene. But where could such information come from? It cannot come from a "super-executive", for that obviously sets up an infinite regress, and it cannot be left unexplained, for knowing when to intervene is clearly a critical part of what the executive is meant to do. In short, the executive is either blind or is not an executive at all.

One simple criterion for intervention might be thought to be the presence of conflict between contingent responses (Botvinick et al., 2001). At any point, if there is no conflict between two or more response tendencies there is clearly nothing for an executive to do; if there is conflict, then without a resolution each response will suffer from interference from the others. It is easy to conceive of a function that relates the degree of necessary control to the degree of conflict between neural ensembles encoding incompatible responses. Even if the executive is blind, a conflict monitor ought to be a good enough guide dog.

Conflict monitoring, then, elegantly solves a problem that otherwise robs the executive of any plausibility. Furthermore, a discrete brain area—the dMFC—has been shown across several conflict paradigms to respond exactly as one would expect a conflict monitor to respond (reviewed by (Botvinick et al., 2004)). Despite the near universal failure to corroborate the theory in lesion studies (Critchley et al., 2003; di Pellegrino et al., 2007; Fellows and Farah, 2005; Kennerley et al., 2006; Mansouri et al., 2007; Swick and Jovanovic, 2002; Turken and Swick, 1999; Vendrell et al., 1995) and neurophysiology(Ito et al., 2003; Nakamura et al., 2005), the imaging evidence has been so abundant, and the conceptual structure so beguiling, that for many researchers the fundamentals of the theory are simply no longer in doubt.

The more general a putative function is, however, the easier it is for essentially correlative methods such as functional imaging to lead us astray. Conflict paradigms are notoriously riven by unacknowledged confounds, many of which are ineliminable because they are *constitutive* of situations we are disposed to label conflicting. For example, we would only call one action more conflicted than another if it is less natural or habitual. Reading a word is more natural than reporting the colour in which it is written (Stroop task), arrows are ecologically more often congruent than incongruent (Eriksen flanker task), we look at novel events more often than look away (antisaccade task), signals tell us to do things more often than to withhold doing them (stop signal task), and so on. Consider formally *the same* task implemented in two different modalities: the prosaccade/antisaccade task and a haptic analogue where the subject either presses down or withdraws in response to a fingertip stimulus. One is considered a conflict task yet the other would not be, simply because withdrawing from a stimulus is natural in the cutaneous sensory domain but not in the visual domain.

Similarly, for conflict to occur between two possible responses the conditions with which they are associated *must* overlap to some degree, for without co-activation of competing response programmes there cannot be any conflict. If so, then the specification of the actual response performed on a given trial must be disambiguated by the *conjunction* of two features—for neither alone uniquely specifies the response—thereby increasing the complexity of the underlying condition-action association compared with the non-conflicted response (Nachev et al., 2008).

Such theoretical considerations aside there is an obvious empirical problem: a task in which the conflict trials are *not* associated with longer reaction times compared with the control trials would not be regarded as a conflict task at all. Thus reaction time is a simple, universal, conceptually-agnostic confound a proponent of conflict monitoring would want to see carefully tested. If reaction time is a

E-mail address: p.nachev@ion.ucl.ac.uk.

better predictor of dMFC activity than conflict then dMFC activation in such tasks cannot be explained by conflict monitoring. Taking the Stroop task as an example, this is precisely what Grinband and his colleagues convincingly show, in a series of analyses that leaves less room for doubt than any of the studies preceding it (Grinband et al., 2011). Critically, slow responses on low conflict trials were shown to be associated with higher dMFC activity than fast responses on high conflict trials, making it impossible plausibly to attribute dMFC activity to conflict. Whatever the dMFC is monitoring in conflict paradigms it is clearly not overt conflict.

A diehard defender of the theory may object that in the context of a conflict task with randomised presentation such as the version of the Stroop employed here, performance on any one trial—whether conflict or baseline—will be modulated by the subject's guess of what the required action is likely to be as much as by the stimuli themselves. Thus, congruent trials the subject incorrectly predicts to be incongruent may be associated with higher conflict than in-congruent trials where he correctly predicts an incongruent response: in short, here expectation-driven *covert* conflict may trump overt conflict. But what, if not the experimental parameters, can we then use to measure conflict? It cannot be assumed to be reaction time, for that is shown to be a poorer predictor of errors than congruence. Treating the dissociation between errors and reaction time as an anomaly here sounds too much like special pleading.

To an outsider it might seem strange that such a basic confound should have taken so long to be examined. Perhaps it is because in our reductionist zeal we too readily take mean reaction times as proxy for distributions: it is the comparison between long and short reaction times—other things being equal—that is the critical manoeuvre in Grinband's study. And yet, the analysis of reaction times distributions lies at the origin of one of the most fertile conceptualizations of action selection and control: the LATER model (Carpenter and Williams, 1995). At the core of LATER lies the realization that variability in reaction times—behavioural "noise"—is not a nuisance such as measurement noise but a window onto the fundamental physiology of action. Action selection is here conceptualised as competition between neuronal ensembles whose activity races stochastically to some fixed threshold at a rate biased by the internal and external environmental context. What aspect of the context we decide to model here—e.g. information, prior probability, urgency, etc—is immaterial: the key point is that response delay is directly related to the process of response selection.

That dMFC activity should be linked to reaction time is therefore profoundly unsurprising: the critical question is what aspect of the context of an action we can find neuroanatomically segregated within the region. The very earliest metanalyses have suggested a smooth rostrocaudal gradient of "difficulty" across the medial pre-motor areas (Picard and Strick, 1996), and more recent studies give no cause for dissent (reviewed by Nachev et al.(2008)). To speak of "difficulty" is of course to say very little, and fractionating the term is no mean task. But if it is to be done we need a radically different approach: not many studies of few paradigms, but few studies of many paradigms, diversified sufficiently to capture whatever contextual aspect is ana-tomically segregated in the underlying physiology.

An important lesson here is the significance of loss-of-function studies, which have near-universally failed to confirm the predictions of the conflict monitoring hypothesis. Damage to medial frontal cortex is very rare by comparison with other cortical areas: lesion data will therefore always be sparse here. Such data is nonetheless more, not less, important than the surfeit of imaging studies which it contradicts, for if a region of the brain is active during a task it *may* do what the experimenter supposes it to do, but it if a region of the brain can be inactivated without impact on the task it *cannot* do what the experimenter supposes it to do. Data ought to be weighted not by numerosity of replication but by inferential power, and here the asymmetry has been striking.

Where does this leave the "executive"? Perhaps closer to the realization that it is a crypto-Cartesian device we must discard if we are to make further progress in understanding the neural basis of complex behaviour and cognition. In any event, it may no longer have a guide dog.

Acknowledgment

References

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychological Review 108, 624.
Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences 8, 539–546.
Carpenter, R.H., Williams, M.L., 1995. Neural computation of log likelihood in control of saccadic eye movements. Nature 377, 59.
Critchley, H.D., Mathias, C.J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.K., Cipolotti, L., Shallice, T., Dolan, R.J., 2003. Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. Brain 126, 2139.
di Pellegrino, G., Ciaramelli, E., Làdavas, E., 2007. The regulation of cognitive control following rostral anterior cingulate cortex lesion in humans. Journal of Cognitive Neuroscience 19, 275–286.
Fellows, L.K., Farah, M.J., 2005. Is anterior cingulate cortex necessary for cognitive control? Brain 128, 788.
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.
Ito, S., Stuphorn, V., Brown, J.W., Schall, J.D., 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 302, 120.
Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rushworth, M.F.S., 2006. Optimal decision making and the anterior cingulate cortex. Nature Neuroscience 9, 940–947.
Mansouri, F.A., Buckley, M.J., Tanaka, K., 2007. Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. Science 318, 987.
Nachev, P., Kennard, C., Husain, M., 2008. Functional role of the supplementary and pre-supplementary motor areas. Nature Reviews Neuroscience 9, 856–869.
Nakamura, K., Roesch, M.R., Olson, C.R., 2005. Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. Journal of Neurophysiology 93, 884.
Norman, D.A., Shallice, T., 2000. Attention to action: willed and automatic control of behavior. Cognitive Neuroscience: A Reader 376.
Picard, N., Strick, P.L., 1996. Motor areas of the medial wall: a review of their location and functional activation. Cerebral Cortex 6, 342.
Swick, D., Jovanovic, J., 2002. Anterior cingulate cortex and the Stroop task: neuropsycho-logical evidence for topographic specificity. Neuropsychologia 40, 1240–1253.
Turken, A.U., Swick, D., 1999. Response selection in the human anterior cingulate cortex. Nature Neuroscience 2, 920–924.
Vendrell, P., Junqué, C., Pujol, J., Jurado, M., Molet, J., Grafman, J., 1995. The role of prefrontal regions in the Stroop task. Neuropsychologia 33, 341–352.

Contents lists available at ScienceDirect

NeuroImage
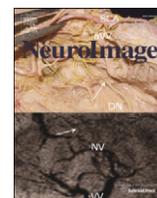
Comments and Controversies

Medial prefrontal cortex activity correlates with time-on-task: What does this tell us about theories of cognitive control?

Joshua W. Brown *

Dept. of Psychological and Brain Sciences, Indiana University, Bloomington IN, USA

ARTICLE INFO

ABSTRACT

A paper by Grinband et al. (this issue) argues that dorsal medial prefrontal cortex (mPFC) activity reflects time-on-task rather than conflict or error likelihood. In this commentary, Brown suggests that the findings are consistent with a new model in which mPFC learns to predict the nature and timing of action outcomes. The new model predicts that time-on-task effects coexist with distinct error likelihood effects, a result which is consistent with other recent findings from fMRI and monkey neurophysiology.

© 2011 Elsevier Inc. All rights reserved.

The medial prefrontal cortex (mPFC) and especially anterior cingulate cortex (ACC) has been the subject of intense study especially in the last 15 years. Empirical findings as well as theoretical and computational models have abounded, but the controversy remains. In this issue, Grinband et al. (2011) provide an important contribution to the range of known effects in mPFC, focusing on the positive correlation between reaction time (RT) and mPFC activity. Their results are striking and yet in many ways consistent with prior findings showing greater activity with longer time on task (Burle et al., 2008; Yeung and Nieuwenhuis, 2009).

The question that arises from these findings is what the results mean for our theories of mPFC. While arguing strongly against conflict and error likelihood accounts, the paper offers little in the way of specific theoretical proposals, aside from passing references to possible general functions of planning, switching, and inhibitory control. This is symptomatic of a larger issue with mPFC research in that theoretical development has been strong overall but has not kept pace with the even greater number of new effects reported. Empirical results from mPFC have multiplied and fractured across species to such a degree that many have questioned whether a unified theoretical account is even possible (Cole et al., 2009).

When we introduced the error likelihood model of ACC six years ago (Brown and Braver, 2005), I hoped that it would provide a way forward to reconcile a larger body of findings. The error likelihood proposal continues to generate much healthy debate. The present Grinband et al. paper argues that in the countermanding task originally used to test the error likelihood hypothesis, RTs are longer in the high vs. low error likelihood conditions, so that general correlations between RT and mPFC activity could in principle

account for the apparent error likelihood effect. This was in principle a possibility, but it was not in fact the case. There was a slightly longer average RT in high error likelihood trials, but we were aware of this as a potential confound and performed additional analysis to covary out the RT. The error likelihood effect remained even when controlling for RT (Brown and Braver, 2005). In a follow-up study, there was actually *no difference* between high and low error likelihood RTs, for the critical correct trials with no countermand signal. Again, the error likelihood effect remained and furthermore correlated with risk avoidance traits (Brown and Braver, 2007). Thus error likelihood effects cannot be accounted for solely on the basis of time-on-task.

Nonetheless, the error likelihood computational model turned out to be limited in that it could not account for various data, such as the larger error effects when error likelihood is low (Brown and Braver, 2005), and it is not clear how it could account for the present finding of positive correlations between time-on-task and mPFC activity. We have therefore proposed a new model that casts mPFC as learning to predict the outcomes of actions (Alexander and Brown, 2010, submitted for publication). We refer to this as the Prediction of Response Outcomes (PRO) model, of which the error likelihood model is a special case. In the PRO model, the probabilities of all possible outcomes (good as well as bad) are represented, and these predictions are compared against the actual outcomes.

Of particular relevance here is that in the PRO model, the outcome probabilities are *temporally structured*, so that cell activities representing the outcome probabilities grow larger as the timing of the expected outcome approaches. If the expected outcome is unexpectedly delayed, then the activity predicting the outcome will continue to increase until the outcome actually occurs, at which point the prediction activity will cease. This leads necessarily to a positive correlation between mPFC activity and the time interval until the outcome occurs, whether it is a movement or some external feedback from the environment (Alexander and Brown, submitted). We have

* Dept. of Psychological & Brain Sciences, 1101 E Tenth St., Bloomington, IN 47405, USA.
E-mail address: jwmbrown@indiana.edu.

tested this model prediction with fMRI and found that not only does greater RT lead to greater mPFC activity, but delayed external feedback about a correct action also leads to greater mPFC activity (Forster and Brown, 2011). Also, this proposal is consistent with previous monkey neurophysiology findings of supplementary eye field cells that show increasing activity toward predictable events such as movement initiation (Amador et al., 2000; Shidara and Richmond, 2002) yet paradoxically do not control the movement (Stuphorn et al., 2010). More strikingly, our simulations show that this positive correlation between time-on-task and mPFC activity *coexists* with both error likelihood effects and the effects of incongruent stimuli that have previously been interpreted as reflecting response conflict (Alexander and Brown, submitted), as well as a large body of other reported effects in mPFC.

Given the above, we welcome the findings of the Grinband et al. paper, as they provide useful and strong constraints on candidate theories of mPFC function.

Acknowledgment

References

Alexander, W., Brown, J., 2010. Computational models of performance monitoring and cognitive control. Top. Cogn. Sci. 2, 658–677.
Alexander, W., Brown, J., Submitted. Medial prefrontal cortex as an action–outcome predictor.
Amador, N., Schlag-Rey, M., Schlag, J., 2000. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. J. Neurophysiol. 84, 2166–2170.
Brown, J.W., Braver, T.S., 2005. Learned predictions of error likelihood in the anterior cingulate cortex. Science 307, 1118–1121.
Brown, J., Braver, T.S., 2007. Risk prediction and aversion by anterior cingulate cortex. Cog. Aff. Behav. Neurosci. 7, 266–277.
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. J. Cogn. Neurosci. 20, 1637–1655.
Cole, M.W., Yeung, N., Freiwald, W.A., Botvinick, M., 2009. Cingulate cortex: diverging data from humans and monkeys. Trends Neurosci. 32, 566–574.
Forster, S., Brown, J., 2011. Medial prefrontal cortex predicts and evaluates the timing of action outcomes. Neuroimage 55, 253–265.
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.
Shidara, M., Richmond, B.J., 2002. Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science 296, 1709–1711.
Stuphorn, V., Brown, J.W., Schall, J.D., 2010. Role of supplementary eye field in saccade initiation: executive, not direct, control. J. Neurophysiol. 103, 801–816.
Yeung, N., Nieuwenhuis, S., 2009. Dissociating response conflict and error likelihood in anterior cingulate cortex. J. Neurosci. 29, 14506–14510.
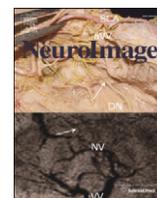
Comments and Controversies

Errors of interpretation and modeling: A reply to Grinband et al.

Nick Yeung [a,*], Jonathan D. Cohen [b], Matthew M. Botvinick [b]

[a] Department of Experimental Psychology, University of Oxford, UK
[b] Department of Psychology and Princeton Neuroscience Institute, Princeton University, USA

ARTICLE INFO

ABSTRACT

Grinband et al., 2011 compare evidence that they have collected from a neuroimaging study of the Stroop task with a simulation model of performance and conflict in that task, and interpret the results as providing evidence against the theory that activity in dorsal medial frontal cortex (dMFC) reflects monitoring for conflict. Here, we discuss several errors in their methods and conclusions and show, contrary to their claims, that their findings are entirely consistent with previously published predictions of the conflict monitoring theory. Specifically, we point out that their argument rests on the assumption that conflict must be greater on all incongruent trials than on all congruent trials—an assumption that is theoretically and demonstrably incorrect. We also point out that their simulations are flawed and diverge substantially from previously published implementations of the conflict monitoring theory. When simulated appropriately, the conflict monitoring theory predicts precisely the patterns of results that Grinband et al. take to present serious challenges to the theory. Finally, we note that their proposal that dMFC activity reflects time on task is theoretically weak, pointing to a direct relationship between behavior (RT) and neural activity but failing to identify any intervening psychological construct to relate the two. The conflict monitoring theory provides such a construct, and a mechanistic implementation that continues to receive strong support from the neuroimaging literature, including the results reported by Grinband et al.

© 2011 Elsevier Inc. All rights reserved.

The conflict monitoring theory proposes that regions within dorsal medial frontal cortex (dMFC), including the anterior cingulate cortex, monitor for the occurrence of competition, or conflict, in action selection to detect and signal the need for increased cognitive control (Botvinick et al., 2001; Carter et al., 1998). This theory provides an account of perhaps the most replicated neuroimaging finding regarding dMFC: its increase in activation in conditions of high cognitive demand. A canonical example of this activation is seen in the Stroop task, in which greater dMFC activity is observed when color and word information are incongruent (e.g., RED in blue ink) than when they are congruent (e.g., BLUE in blue ink) (Bench et al., 1993). According to the theory, this activity reflects the increase in conflict generated by the co-activation of mutually incompatible actions in the incongruent condition (e.g., trying to say "red" and "blue" simulta-neously) relative to the congruent condition. In the target article of this commentary, Grinband et al., 2011 present a novel critique of this proposed relationship between dMFC activity, conflict, and behavioral indices of task demand. Here, we discuss several significant weaknesses in their methods and conclusions, and show that their findings are entirely consistent with previously published predictions of the conflict monitoring theory.

The essence of Grinband et al.'s argument is that conflict is superfluous to explaining observed variations in dMFC activity across conditions: They suggest that conflict explains no residual variance in the data once the effects of RT (time on task) are partialed out. In support of this conclusion, Grinband et al. present three key findings from an analysis of fMRI data collected from the Stroop task. First, they show that dMFC activity does not differ for incongruent and congruent trials that are matched in terms of RT. Second, extending this analysis to dissociate conflict and RT, they show that dMFC activity is low on incongruent trials with fast RTs and high on congruent trials with slow RTs, reversing the typical effects of conflict. Finally, they show that whereas dMFC activity increases with RT for both congruent and incongruent trials, the likelihood of errors tends to decrease with RT—the ubiquitous speed–accuracy trade-off. On the assumption that error likelihood is a reasonable proxy for response conflict—an assumption they claim to validate with simulations of the conflict monitoring theory—Grinband et al. conclude that dMFC activity may even vary inversely with conflict. Thus, Grinband et al.'s data initially appear to seriously challenge the theory that dMFC activity reflects conflict.

We have three principal criticisms of Grinband et al.'s conclusions. First, their argument relies on an intuition that is superficially plausible—that conflict must be greater on all incongruent trials than on all congruent trials—but which is theoretically and demon-strably incorrect. We begin with an exploration of this point. Second, their simulations are flawed and diverge substantially from previously published implementations of the conflict monitoring theory. When

simulated appropriately, the conflict monitoring theory predicts precisely the patterns of results that Grinband et al. take to present serious challenges to the theory. Finally, their proposal that dMFC activity reflects time on task is conceptually limited, since it implies a direct relationship between behavior (RT) and neural activity with no intervening psychological construct to relate the two.

Misleading intuitions

Two intuitive and uncontroversial statements about the conflict monitoring theory are (1) that conflict should be greater on incongruent trials than on congruent trials, and (2) that RT should increase as a function of experienced conflict. In most cases, these two statements run together and the predictions of the theory are clear; in particular, conflict and RT both typically increase on incongruent trials relative to congruent trials in the Stroop task. However, it is less obvious how conflict should vary when the effects of congruence and RT are set in opposition, as is the case when contrasting fast incongruent trials with slow congruent trials. Grinband et al. base their analysis and conclusions on the intuition that conflict should vary with congruence and dissociate from RT in this contrast, being greater for fast incongruent than slow congruent trials. But this intuition is incorrect. As demonstrated in previously published conflict model simulations (Yeung et al., 2004), conflict more closely tracks RT than congruence condition when the two are dissociated.

This important prediction of the conflict monitoring theory follows from two key principles. The first is that activity of the conflict monitoring system should scale with conflict regardless of its source: When the conflict monitor responds to co-activation of mutually incompatible actions, it does not distinguish between conflict caused by incongruent stimulus features and conflict caused by other sources of processing variability and noise. Second, and crucially, these additional sources of processing variability—such as trial-to-trial fluctuations in attentional focus, noise in stimulus processing, and idiosyncratically varying response biases—result in performance variability that dwarfs the variance caused by stimulus congruence (as illustrated in Fig. S1 of the target article). The consequence is that very slow RTs will be observed on some congruent trials—for example when participants incorrectly perceive the stimulus, or happen to be prepared to make the incorrect response rather than the correct one—and, conversely, very fast RTs will be observed on some incongruent trials—for example when participants' attentional focus is high, or when they fortuitously anticipate which stimulus will be presented and thus pre-activate the corresponding response. Crucially, however, slow congruent trials are not slow *despite* having low conflict, and fast incongruent trials are not fast *despite* having high conflict. To the contrary, slow congruent trials are slow precisely *because* conflict is high—a consequence of failing to focus attention, misperceiving the stimulus, preparing the wrong response, etc.—whereas fast incongruent trials are fast precisely *because* conflict is low. Thus, slow RT congruent trials should be associated with greater conflict (and, hence, greater dMFC activity) than fast RT incongruent trials.

To illustrate these points, Fig. 1A replots data from a previously published simulation in which we compared conflict on congruent and incongruent trials as a function of simulated RT (Yeung et al., 2004). Two features of the data are immediately evident: First, within each RT-matched bin, there is little or no difference in conflict as a function of stimulus congruence. Second, simulated conflict is markedly higher on slow congruent trials than fast incongruent trials; for example, conflict is nearly 10 times as high for the slowest congruent trials (RT bin 10) as it is for the fastest incongruent trials (RT bin 1). Thus, for the reasons outlined above, the model actually predicts the pattern of results that Grinband et al. put forward as a challenge to the conflict monitoring theory. As described in our earlier article, "In the model, slow responses to congruent stimuli are marked by high conflict (because of noise in processing) in just the
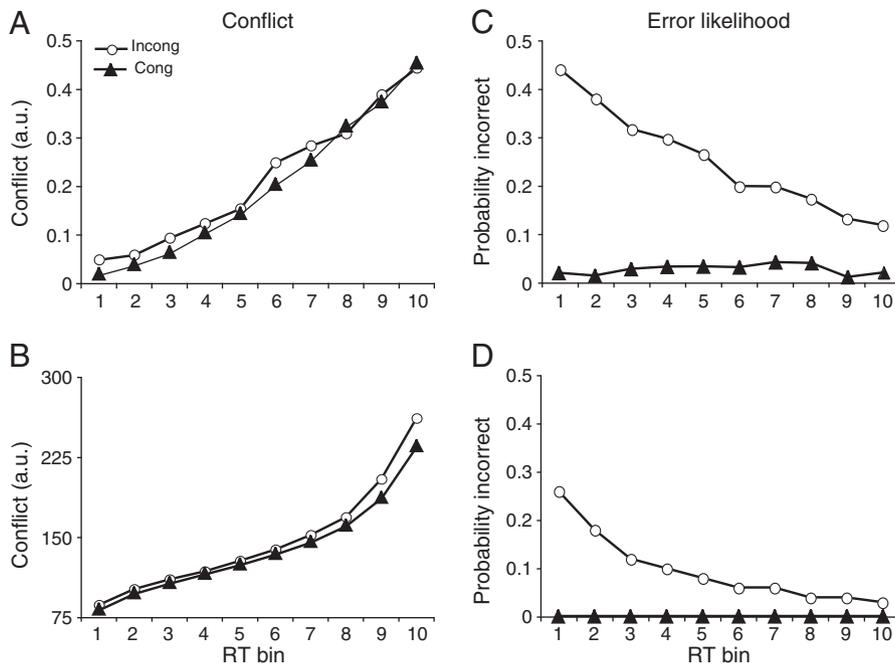

Fig. 1. Conflict model simulation results. Simulated conflict and error likelihood are plotted separately for corresponding models of the flanker task (A,C) and Stroop task (B,D). Data for the flanker task model are reanalyzed from previously published simulations (Yeung et al., 2004), in which trials were divided into bins corresponding to stepwise increases in simulated RT (processing cycles). The Stroop model was based on an earlier implementation of the verbal response version of the task in which errors are very rare (Botvinick et al., 2001), with parameters modified to simulate the increased error rates observed empirically in Grinband et al.'s manual response task. For this simulation, trials were divided into decile bins according to RT. For both models, simulated conflict shows a monotonic increase with RT for both incongruent and congruent trials (A,B), whereas error likelihood shows a monotonic decrease for incongruent trials in which errors were relatively frequent, and no change for congruent trials in which errors were rare or absent (C,D). Replicating Grinband et al. and our previous analyses, simulated conflict is plotted for correct response trials only.

same way as are slow responses to incongruent stimuli. The overall difference in conflict between congruent and incongruent trials reflects the fact that a greater proportion of congruent trials fall in the faster RT bins (as a result of low conflict), whereas incongruent trials tend to have greater RTs (as a result of high conflict)." (Yeung et al., 2004; p. 948).

These simulation results are derived from a model of conflict monitoring in the Eriksen flanker task, a spatial-attention analog of the Stroop task that has been the focus of the majority of our work in the past decade. Nevertheless, corresponding patterns are evident in simulations of conflict monitoring in the Stroop task (Fig. 1B). These simulations were based on the same model employed by Grinband et al., with parameters modified to approximate the error rates of their subjects. (The original model used by Botvinick et al. (2001) was parameterized to simulate performance in the verbal Stroop task, in which errors are vanishingly rare. Indeed, the original parameterization yields no errors at all, a point which leaves us uncertain as to how Grinband et al. obtained their results.) The crucial pattern of results shown in Fig. 1B—with equivalent levels of conflict for RT-matched congruent and incongruent trials, and markedly greater conflict for slow congruent trials than fast incongruent trials—is observed in both the original and updated models. Thus, contrary to Grinband et al.'s motivating intuition, the conflict monitoring theory predicts rather precisely the first two key findings in their data.

Flawed modeling

Grinband et al.'s third key finding is that dMFC activity varies inversely with error likelihood when trials are sorted into RT quantiles within each congruence condition: Trials with the fastest RTs have the highest error likelihood (reflecting the ubiquitous speed–accuracy tradeoff) but the lowest dMFC activity; for trials with slower RTs, error likelihood decreases while dMFC activity increases. Grinband et al. present simulation results indicating that error likelihood, measured in this way, provides an effective marker of response conflict, and thus conclude that their findings provide a further challenge to the conflict monitoring theory.

Before we raise concerns with the validity of Grinband et al.'s simulations, we first note that their empirical results are in fact perfectly consistent with published conflict model simulations. In particular, we have previously shown that error likelihood and conflict dissociate as a function of RT, in an analysis that anticipated precisely the quantile RT plot used by Grinband et al. (Yeung and Nieuwenhuis, 2009). As shown in Figs. 1A and C, the conflict theory predicts that conflict and error likelihood should vary in opposite ways as a function of RT—with conflict increasing and error likelihood decreasing—such that error likelihood is a very poor proxy for conflict. This dissociation is marked for incongruent trials on which errors occur relatively frequently. For congruent trials, error likelihood is very low across all RT bins irrespective of the level of conflict, once again capturing the detail of Grinband et al.'s experimental data. Thus, like dMFC activity in Grinband et al.'s fMRI study, conflict is highest on trials associated with the lowest error likelihood; specifically, trials with the slowest RTs. As we noted in our earlier paper, "It may seem somewhat counterintuitive that trials with the highest conflict should produce the fewest errors. In our simulations, this feature follows from the fact that activity in the target stimulus unit and correct response unit tend to increase over time, as stimulus processing progresses under the influence of attention. As a consequence, although trials with long RTs tend to have high levels of conflict, the responses ultimately made tend to be correct … Conflict and error likelihood therefore vary in opposing ways as a function of response speed." (Yeung and Nieuwenhuis, 2009; p. 14509).

These simulation results contrast sharply with those presented by Grinband et al., but the discrepancy does not seem to reflect the specific task being simulated (flanker task vs. Stroop task): Once again,

simulation data from our Stroop model replicate closely the patterns seen in our flanker task model (Fig. 1D). We have been unable to replicate Grinband et al.'s simulation results using this model, and we are unsure why their model should produce results that are inconsistent with basic features of all evidence accumulation models of human decision making. Specifically, the simulation data plotted in Fig. S2C of the target article indicate that their model exhibits error free performance at the fastest RTs, with error rates rising to chance levels at the slowest RTs. This perplexing pattern is inconsistent with a fundamental feature of human decision making—the ubiquitously observed speed–accuracy trade-off—that is evident in their empirical data (their Fig. 4A) and is replicated in all of our previous models of conflict monitoring (including those presented here). In the absence of an adequate model of this basic feature of human cognition, the value of Grinband et al.'s simulation results remains unclear.

Conceptual limitation

An important strength of the conflict monitoring theory is that it provides a mechanistic account of the observed relationship between dMFC activity and task demand (as reflected in behavioral measures of RT and accuracy) in terms of an intervening psychological construct (detection of response conflict). In contrast, Grinband et al. offer no corresponding psychological or mechanistic account of this relationship, and instead simply conclude that dMFC activity "is correlated with time on task" and "is predicted by trial-to-trial differences in response time". These limited statements risk, in turn, conceptual confusion: Is the proposal that neural activity in some way directly relates to RT? Or that dMFC monitors RT and thus scales monotonically as RT increases? Or that dMFC serves another cognitive function that, in some unspecified manner, shares a consistent relationship with RT? Thus, while we agree with Grinband et al. that it is analytically useful to consider RT as an independent variable in analyses of behavioral and neuroimaging data— and are gratified that the results of their analyses concur with our own in validating key predictions of the conflict monitoring theory (Yeung et al., 2004; Yeung and Nieuwenhuis, 2009)—we would argue that it is vital to provide some account of the relationship between RT and observed neural activity.

This theoretical story is important in linking the reported findings to the broader literature on dMFC function. For example, there is now substantial empirical support for a central claim of the conflict monitoring theory, that detection of conflict in dMFC should lead to increased recruitment of control in lateral frontal cortex (Kerns et al., 2004; Liston et al., 2006). The time-on-task view offers no ready explanation of this critical observation. Even more directly challenging to the time-on-task view are observed dissociations between dMFC activity and measured RT. For example, dMFC activity in the Stroop task is sometimes found to be greater on congruent trials (e.g., BLUE in blue ink) than on neutral trials (e.g., XXX in blue ink), even when RT is lower in the former condition (Kadosh et al., 2008). This finding is inconsistent with the time-on-task view, but can be explained in terms of conflict between color naming and word reading at the level of whole tasks, which is increased in the presence of irrelevant stimulus attributes (Botvinick et al., 2001; Herd et al., 2006). Activity in dMFC is also consistently increased on error trials, despite the fact that errors typically occur on trials with the fastest RTs. Again, the time-on-task hypothesis provides no obvious explanation of this finding, whereas the conflict monitoring theory has been shown to account for detailed aspects of this error-related activity such as its temporal dynamics (e.g., post-response stimulus processing) and sensitivity to key experimental manipulations (Botvinick et al., 2001; Yeung et al., 2004). Thus, whereas the conflict monitoring theory provides a theoretically articulated link from the observed association between task demand and dMFC activity to the wider literature on the functional role of this region, the time-on-task

view offers only a conceptually empty restatement of the bare empirical facts.

Conclusion

Grinband et al.'s critique of the conflict monitoring theory of dMFC function has several important strengths: It presents a serious and thoughtful examination of the conflict monitoring view that is accompanied by detailed analyses of the relationship between behavioral and neuroimaging measures and by a formalization of the key predictions tested. However, as detailed above, the research also has critical weaknesses: The analyses are motivated by an intuition that is superficially plausible but that is theoretically and demonstrably incorrect, and all of the presented results are entirely consistent with the conflict monitoring theory (and, indeed, were anticipated by our previously published simulation results). Given that the conflict monitoring theory predicts all of the presented results, and provides a mechanistic account of these findings that encompasses other key features of observed dMFC activity, we conclude that this account continues to provide a productive and accurate account of the contributions of dMFC to decision making and cognitive control.

References

Bench, C.J., Frith, C.D., Grasby, P.M., Friston, K.J., Paulesu, E., Frackowiak, R.S.J., et al., 1993. Investigations of the functional anatomy of attention using the Stroop test. Neuropsychologia 9, 907–922.
Botvinick, M.M., Braver, T.S., Carter, C.S., Barch, D.M., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychol. Rev. 108, 624–652.
Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280, 747–749.
Grinband, J., Savitsky, 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.
Herd, S.A., Banich, M.T., O'Reilly, R.C., 2006. Neural mechanisms of cognitive control: an integrative model of Stroop task performance and FMRI data. J. Cogn. Neurosci. 18, 22–32.
Kadosh, R.C., Kadosh, K.C., Henik, A., Linden, D.E.J., 2008. Processing conflicting information: facilitation, interference, and functional connectivity. Neuropsychologia 46, 2872–2879.
Kerns, J.G., Cohen, J.D., MacDonald 3rd, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S., 2004. Anterior cingulate conflict monitoring and adjustments in control. Science 303, 1023–1026.
Liston, C., Matalon, S., Hare, T.A., Davidson, M.C., Casey, B.J., 2006. Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. Neuron 50, 643–653.
Yeung, N., Nieuwenhuis, S., 2009. Dissociating response conflict and error likelihood in anterior cingulate cortex. J. Neurosci. 29, 14506–14510.
Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol. Rev. 111, 931–959.
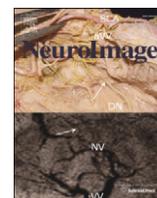
Comments and Controversies

Conflict, error likelihood, and RT: Response to Brown & Yeung et al.

Jack Grinband [a,*], Judith Savitskaya [b], Tor D. Wager [c], Tobias Teichert [d], Vincent P. Ferrera [d], Joy Hirsch [b]

[a] Departments of Radiology & Neuroscience, Columbia University, New York, NY 10032, USA
[b] Program for Imaging & Cognitive Sciences (PICS), Columbia University, New York, NY 10032, USA
[c] Department of Psychology, Muenzinger D261D, University of Colorado, Boulder, CO 80309, USA
[d] Department of Neuroscience, Columbia University, New York, NY 10032, USA

We would like to thank our reviewers and the authors of the commentaries on our paper. Though we may have scientific differences, the criticisms raised in the commentaries will undoubtedly improve our understanding of the neural basis of decision making and provide a springboard for future experiments. In our response, we clarify why our data are incompatible with both the error likelihood and the conflict monitoring models presented by Brown and by Yeung et al., respectively.

The error likelihood model

Brown (2011) argues that time on task is not a likely confound in the experimental design of previous error likelihood studies that activated the dorsal medial frontal cortex (dMFC, often referred to as "anterior cingulate"; Brown and Braver, 2005, 2007). Specifically, in the countermanding task used by Brown and Braver, differences in RT between conditions were small or non-existent and, when present, were controlled by incorporating the RTs into the general linear model used to analyze the BOLD data. However, even in the absence of RT differences between conditions, the critical issue in countermanding tasks is the "time on task" as measured by the time the subject is actively monitoring for the countermanding cue. When the cue is delayed in these tasks, working memory and attentional demands are extended for longer durations, which should result in larger BOLD responses. When the cue is not delayed, shorter cognitive demands should result in smaller BOLD responses. RTs may or may not be related to this duration depending on factors such as the subject's speed/accuracy tradeoff, motivation, and strategy. This example illustrates how error likelihood could be confounded with time on task despite the lack of RT differences

between conditions. Moreover, we welcome Brown et al.'s Prediction of Response Outcomes (PRO) model and are interested to learn whether it is consistent with our finding that, in the Stroop task, no relationship between error likelihood and dMFC activity exists.

The conflict monitoring model

The goal of our study was to critically evaluate the predictions of the conflict monitoring model. Thus, we started with the definitions provided by the original authors of the model. For example, Carter et al. (1998) stated that "response competition arises when a task elicits a prepotent but inappropriate response tendency (manifested as activity in the incorrect response channel) that must be overcome to perform correctly. These conditions are also more likely to elicit incorrect responses." Similarly, Botvinick et al. (2001) defined conflict as the "simultaneous activation of mutually inhibiting [response] units" and predicted that "ACC activation would be greater on incompatible trials, as this is where conflict is greatest." Such statements suggest that conflict (1) involves the co-activation of competing response options; (2) is greatest when stimulus features are associated with incompatible responses; and, critically, (3) is linked to increased error likelihood under most conditions, particularly on incompatible trials. These descriptions of the conflict monitoring model assert that conflict and error likelihood are correlated.

To test that the computational model of conflict we used was compatible with these predictions, we simulated the effects of conflict on RT and error likelihood. We used Matlab code written by Matthew Botvinick (Botvinick et al., 2001) and applied by Siegle et al. (2004) to simulate conflict differences in depression. The model (Fig. S2 of Grinband et al., 2011) makes predictions that appear consistent with the conclusions made by Botvinick et al. (2001), Carter et al. (1998), and Botvinick et al. (1999). Specifically, it predicts that both RT and error likelihood are greater on incongruent trials and that these effects are due to the presence of conflict.

In their commentary on our article, Yeung (2011) present a conflict monitoring model (a variant of the basic model discussed previously, but presumably with different parameter settings) that makes different predictions from the Botvinick model. They simulate a case in which conflict closely tracks RT on both incompatible and compatible trials, and shows no effect of congruency (Fig. 1A). Thus, for any given RT, it predicts that congruent and incongruent trials will have equal levels of conflict, a result that appears consistent with our data.
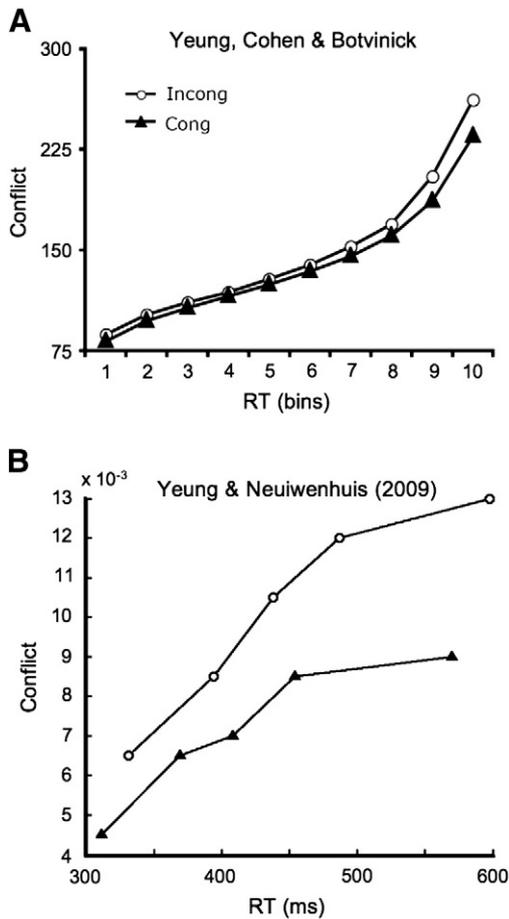
Fig. 1. Predictions of Yeung and Nieuwenhuis (2009) model. (A) Model replotted from Yeung (2011) reply to our article. There are no differences in conflict between congruent and incongruent trials. (B) Data replotted from Fig. 2A of Yeung and Nieuwenhuis (2009). Large differences in conflict exist between congruent and incongruent trials even after controlling for RT differences. Each point represents an RT quantile. Because bin position was different between conditions, data was plotted against ms rather than bin number.

Although there may exist conditions under which compatibility effects are completely mediated by RT, as in Yeung et al.'s simulation, it is not clear that such conditions hold in a wide variety of experimental situations. For example, Yeung et al.'s simulation appears to contradict results from another previously published conflict monitoring model (Yeung and Nieuwenhuis, 2009; Fig. 1B). Unlike the simulations by Yeung et al. (Fig. 1A), the Yeung and Nieuwenhuis model estimated conflict to be substantially greater for incongruent than congruent trials (Fig. 1B). The relationship between stimulus congruency and conflict appears to depend entirely on which implementation of the model is used. Given that the Botvinick et al., Yeung and Neuwenhuis, and Yeung et al. models all make different predictions, it is unclear whether a single unified model of conflict monitoring exists.

A second feature of the Yeung et al. model (Fig. 1A) is its prediction that greater activation of the incorrect response results in fewer errors. Though this appears counterintuitive, Yeung et al. claim it can be easily explained. Conflict can stem not only from the stimulus but also from a variety of independent sources, such as sensory noise, attentional fluctuations, and momentary response biases. In fact, they claim that variance from these types of sources dwarfs congruency-related conflict effects, which explains why "conflict" can be as strong or stronger on compatible trials as on incompatible ones. Thus, their model does not differentiate conflict effects that arise from situations with clear competing responses – which the incongruent- vs. congruent-trial comparison was designed to identify – from speed/

accuracy trade-offs due to fluctuations in sensorimotor noise, changes in task goals or rules, motor readiness, and other processes.

Insofar as conflict is defined as "any sensorimotor or cognitive process that lengthens RT", as Yeung et al. appear to do, their model is compatible with our data. However, we believe this definition trivializes the idea of conflict. Conflict is no longer defined as competition between response options, but rather arises from a less well-specified set of processes. This definition leads to strange conclusions, for example: RT variability in a simple signal detection task, with only a single response option, is due to the "conflict" created by sensory noise. Most studies of conflict monitoring, including the earlier studies by Cohen, Carter, Botvinick, and colleagues (Botvinick et al., 1999; Botvinick et al., 2001; Carter et al., 1998), explain RT variability due to incompatible stimulus–response relationships, not the speed/accuracy trade-offs present in all reaction-time tasks. If neither the stimulus conditions (i.e. congruent vs. incongruent), nor the internal causes, that generate 'conflict' can be specified precisely, the ability to test and falsify the conflict account of dMFC is diminished.

Alternative models of dMFC activity

Electrophysiological studies (Deco and Rolls, 2003; Histed and Miller, 2006; Olson, 2003) have demonstrated that dMFC neurons encode spatial and sensorimotor task demands consistent with a role in working memory and spatial attention. All tasks requiring a response from the subject place a load on the working memory/attention systems from the onset of a stimulus until a response is made. Thus, it is possible that dMFC stores task-related stimulus/response parameters or maintains attentional focus on the stimulus.

To test this hypothesis we asked 17 subjects to view a flashing checkerboard of variable duration and press a button when the stimulus disappeared. In this paradigm, the stimulus was unambiguous, only one response was possible, no choice decision was required, and no errors were made. Thus, error likelihood was uniformly zero and no competition between response alternatives existed. The only common feature between the checkerboard and Stroop tasks was an identical temporal structure. Nevertheless, the checkerboard task produced the same pattern of neural activity in dMFC as the Stroop task (Fig. S3 of Grinband et al., 2011). Specifically, the dMFC response was proportional to the duration of each checkerboard (i.e. time on task). To explain this result in the context of the conflict monitoring model, one would need to postulate that long duration checkerboards generate greater conflict than short checkerboards. We believe this is not a parsimonious account.

Summary

The fundamental problem in the original studies of conflict is the lack of an adequate null hypothesis, specifically, that the dMFC is driven by task general processes (e.g. working memory, spatial attention) rather than conflict. Tested against this null hypothesis, we found no evidence in favor of dMFC's role in conflict monitoring – dMFC activity was uncorrelated to stimulus congruency or error-likelihood. The Yeung et al. model (2011 — current issue) proposes that conflict is independent of stimulus congruency and highly correlated to RT; that is, when RT is held constant, congruent and incongruent trials have equal conflict. However, if conflict is independent of congruency, then tasks that manipulate congruency, such as those used in the original conflict studies (Botvinick et al., 1999; Botvinick et al., 2001; Carter et al., 1998), are neither necessary nor relevant for arguing in favor of this model. The model further implies that the large differences in error likelihood between congruent and incongruent trials are unrelated to conflict, an implausible assertion. Finally, this model makes identical predictions for dMFC activity whether it is specialized for detecting conflict (alternative hypothesis)

or involved in task general processes (null hypothesis), making it impossible to test using standard fMRI methods.

Conflict was first defined as the competition between mutually exclusive response options. This definition helped explain greater RT and error likelihood on incongruent vs. congruent trials in a wide variety of experimental paradigms. Yeung et al.'s definition of response conflict does not clearly differentiate conflict from other sensorimotor, memory, and attentional processes, weakening its usefulness as a psychological construct. The notion that dMFC is a conflict detector is difficult to maintain unless one accepts this diffuse definition. We welcome further discussion on these issues.

References

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

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

Brown, J.W., 2011. Medial prefrontal cortex activity correlates with time-on-task: what does this tell us about theories of cognitive control? NeuroImage 57, 314–315.

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

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

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

Deco, G., Rolls, E.T., 2003. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. The European Journal of Neuroscience 18, 2374–2390.

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.

Histed, M.H., Miller, E.K., 2006. Microstimulation of frontal cortex can reorder a remembered spatial sequence. PLoS Biology 4, e134.

Olson, C.R., 2003. Brain representation of object-centered space in monkeys and humans. Annual Review of Neuroscience 26, 331–354.

Siegle, G.J., Steinhauer, S.R., Thase, M.E., 2004. Pupillary assessment and computational modeling of the Stroop task in depression. International Journal of Psychophysiology 52, 63–76.

Yeung, N., 2011. Errors of interpretation and modeling: a reply to Grinband et al. NeuroImage 57, 316–319.

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