Modelling the Correlation Between Two Putative Inhibition Tasks: A Simulation Approach

Richard P. Cooper (R.Cooper@bbk.ac.uk) and Eddy J. Davelaar (E.Davelaar@bbk.ac.uk)
Department of Psychological Science, Birkbeck, University of London
Malet Street, London WC1E 7HX, UK

Abstract

Behavioural studies of individual differences have shown mild but significant correlations in performance on tasks that require the withholding of a response to a prepotent stimulus, i.e., on so-called *response inhibition* tasks. Several computational models of response inhibition tasks have been developed, but the dominant models of such tasks have been produced in isolation of each other. Consequently they fail to present a coherent unitary picture of response inhibition. In this paper we consider two established interactive activation models of distinct response inhibition tasks – the stop signal task and the Stroop task – and explore potential mechanisms within those models that might underlie the observed behavioural correlation. Only one plausible account of the correlation emerges: that it results from shared mechanisms of attentional bias. This account does not map onto the classical concept of response inhibition. It is concluded that either the accepted models are flawed or that the concept of response inhibition as applied to these tasks is misleading (and hence counterproductive). More generally the work may be taken to support an architectural approach to modelling, albeit at the level of interactive activation models, rather than the more traditional production system models.

Keywords: Executive processes; cognitive control; response inhibition; individual differences; Stroop task; Stop signal task.

Introduction

The construct of "response inhibition" is frequently invoked when attempting to explain behaviours in tasks or situations that demand the withholding of a strongly prepotent response. Response inhibition is held to be a separable task-general executive or cognitive control function, the efficacy of which varies across individuals.

In the laboratory response inhibition is standardly explored in variants of the stop signal task (Logan & Cowan, 1984). This is a form of simple reaction time task in which subjects are normally required to respond as quickly and accurately as possible. However, on a small number of trials a compound stimulus is presented and on these trials and these trials only the subject is required to withhold their response. Such trials are referred to as "stop trials". Typically the compound stimulus consists of a standard stimulus that might occur on any normal trial followed almost immediately by an auditory beep. Stop trials are rare in comparison to normal "go trials". This and the need to respond on go trials as rapidly as possible ensures that the go response is prepotent. Performance is measured in terms of the number or proportion of stop trials on which a response is (incorrectly) produced. This measure varies reliably between subjects. Good response inhibitors produce few stop responses, while poor response inhibitors produce many.

There is substantial behavioural and neuroscience evidence, as well as good theoretical reasons, for supposing that response inhibition is a task-general control function. From the theoretical perspective, response inhibition fits clearly within the supervisory system/contention scheduling framework of the control of thought and action of Norman and Shallice (1986). On this influential account, a system for the control of routine or well-learned behaviours, *contention scheduling*, is modulated by a deliberative system, *the supervisory system*, when routine behaviour is inappropriate and must be overridden. Contention scheduling is appropriate for generating the prepotent response, whatever the situation. If this is not appropriate, as in stop trials of the stop signal task, the supervisory system must override contention scheduling. A plausible way for this to be operationalised is in terms of response inhibition acting as a sub-function of the supervisory system.

From a neuropsychological perspective, patients have been reported who are well-characterised in terms of a deficit in response inhibition. Thus, utilisation behaviour patients tend to exhibit behaviours that are driven largely by environmental contingencies rather than their stated intentions (Lhermitte, 1983). Alien hand patients show similar problems, but they are restricted to one hand (Goldberg et al., 1981). Both deficits may be seen as arising from a failure in response inhibition.

One source of behavioural evidence for the task-general nature of response inhibition comes from a large individual differences study of Miyake et al. (2000). In this study, 137 subjects were each tested on a total of 14 tasks. Performance on 3 of these tasks was argued, on a priori grounds, to specifically require response inhibition. Subsequent factor analysis of subject performance across the tasks supported this view, with performance on the response inhibition tasks being related to a single factor that differentiated those tasks from others in the study, which were held to primarily tap other executive functions (namely the functions of *set-shifting* and *memory monitoring and updating*).

The three response inhibition tasks of Miyake et al. (2000) were a) a forced-choice decision variant of the stop signal task, b) the Stroop colour naming task, and c) an anti-saccade task. Our focus in this paper is on the first two, and so we described these in detail. In the stop signal task, subjects were required to indicate with a button press whether a (visually presented) word was an animal or a non-animal. The first block of 48 trials were all "go" trials.

These were used to establish a mean response time for each subject. One quarter of the trials in the second block (of 192 trials) were stop trials. In these trials, a beep was sounded shortly after presentation of the word (225ms prior to the subject's mean response time, as determined in block 1), and subjects were required to withhold their response. The dependent measure was the proportion of stop trials on which a response was given. In the well-known Stroop colour naming task, subjects were presented with a "word" written in one of six colours. They were required to name the colour of the stimulus word. On neutral trials the word was a string of asterisk symbols, while on incongruent trials it was the name of another colour. The dependent variable was the difference in mean response times for incongruent and neutral trials.

For our purposes, the critical result of this individual differences study was mild but significant positive correlations ($r \approx 0.20$) between performance on the stop signal task and the Stroop task (and in fact between all pairs of response inhibition tasks). In general, these correlations were stronger than those between any single response inhibition task and any of the non-response inhibition tasks explored in the study. However, while the study is impressive in its scale, interpretation of the results is limited because Miyake et al. fail to provide process accounts of the various tasks. While it is perhaps unreasonable to expect such models of all 14 tasks, the absence of process models leaves unexplained the mechanism that is, on the account proposed by Miyake and colleagues, shared by the response inhibition tasks. Similarly, it leaves open the issue of why that function is *not* significantly involved in successful performance of the other tasks considered in the study.

The purpose of the work presented here is to begin to address this limitation by exploring potential common mechanisms within established process models of two of Miyake et al.'s response inhibition tasks. We focus on models of the stop signal task and the Stroop task because there are established models of each task (due to Boucher et al., 2007, and Cohen & Huston, 1994, respectively) that bear some correspondence. This correspondence offers the possibility of relating the models to each other and thereby identifying a shared response inhibition mechanism. For such a mechanism to be explanatorily adequate, it must be parameterisable, with the observed behavioural correlations between tasks arising in part from variation in a shared parameter. To foreshadow, simulation findings derived from reimplementations of the existing published models suggest that directly shared parameters fail to yield the required correlation in performance. However, an appropriate correlation is forthcoming if attentional biasing mechanisms are yoked. Unfortunately, attentional biasing is not normally related conceptually to response inhibition. We conclude that either a) response inhibition is not the mechanism underlying the behavioural correlation in these tasks, or b) one or both of the accepted models requires updating. These simulation results extend those of a complementary analytic study (Davelaar & Cooper, 2010).

The Stop Signal Task

The Model

Early work with the stop signal task demonstrated that behaviour on the task could be well accounted for by a race model consisting of two stochastic processes, a "go" process which is slow to activate but has a head start, and a "stop" process which is faster to activate but starts late (Logan & Cowan, 1984). Successful performance on a stop trial requires that the stop process reach threshold before the go process. Boucher et al. (2007) note that despite this model's strengths, it is inconsistent with neural evidence of interaction between stop and go processes. They present the *interactive race model*, an update of the original model in which the stop and go processes compete through mutual lateral inhibition. The model, as applied to Miyake et al.'s semantic categorisation variant of the stop signal task, is shown in Figure 1.

The model is extremely simple, consisting of just three nodes: one for each response and one for the stop process. Processing in the model is cyclic with each node operating as a leaky competing accumulator (Usher & McClelland, 2001). On each cycle, the activation of a node is increased by an amount proportional to its external input, less an amount proportional to the activation of its competitors (corresponding to lateral inhibition), less an amount proportional to its current activation (its leakage), plus normally distributed random noise. Parameters control the contributions of the various sources to this accumulation. For default behaviour we assume ballpark parameters scaled from those of Boucher et al. to give a response threshold of 1.0. Thus, we assume lateral inhibition, β, of 0.025 between all pairs of nodes, leakage of 0.0 (i.e., the accumulators do not leak), and the standard deviation of noise, σ, of 0.025 units per cycle.

In addition, it is assumed that on any trial external input to one of the response nodes (*animal* or *non-animal*) is provided by a semantic categorisation process (which is not modelled). The level of input is controlled by the parameter μ_{go}, set to 0.005 units per cycle by default. It is assumed that the other response node receives zero external input. On stop trials it is assumed that at some point during the trial

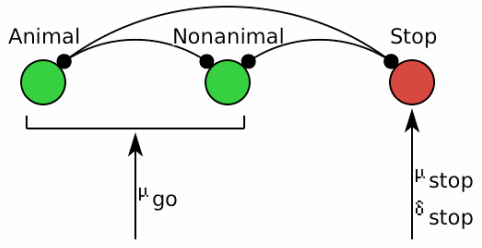

Figure 1: The interactive race model of the stop signal task. On any one trial, either the animal or the nonanimal node receives activation from a semantic categorisation process. On "stop" trials, the stop node also receives activation, though this activation is delayed relative to the activation from the semantic categorisation process.

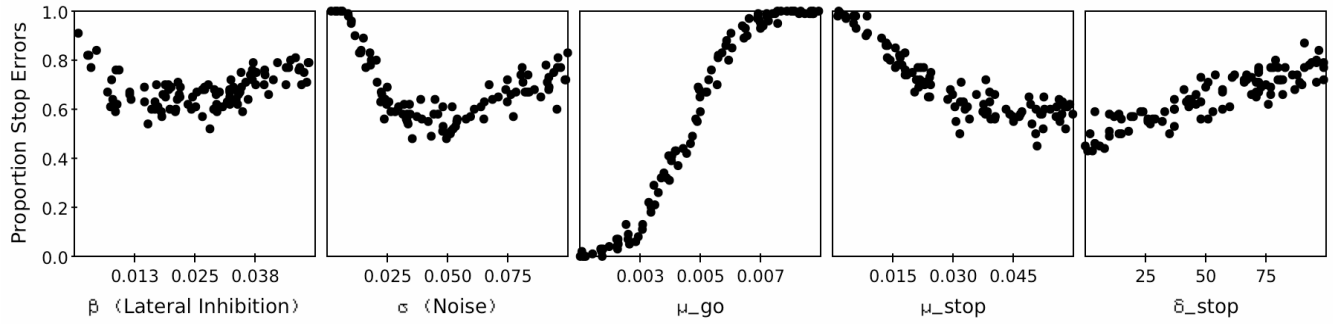

Figure 2: Effects of varying key parameters on the proportion of stop errors produced by the interactive race model of the stop signal task.

external input is provided to the stop node. The level of this input is μ_{stop}, set to 0.030 units per cycle by default. Finally, we assume that the delay between input to the response nodes and input to the stop node is 250 cycles. This delay is the sum of the actual delay between presentation of word and stop stimuli, *SSD*, and the time to initiate the stop process, δ_{stop}. With these parameters, the model performs as desired – on go trials its response accuracy is approximately 99% (with noise and lateral inhibition occasionally leading to error) while on stop trials it fails to stop on approximately 65% of occasions. This compares well with mean subject performance of 67% as reported by Miyake et al. (2000).

Simulation Results

An initial set of simulation studies was performed to determine the relation between the model's performance and the key parameters that could reasonably be argued to vary across individuals, that is: μ_{go}, μ_{stop}, β (lateral inhibition), σ (standard deviation of noise) and δ_{stop}.[1] Each parameter was varied about the default value (with the other four parameters fixed at default values) to determine the effect of that parameter on the proportion of stop errors. Figure 2 summarises the results, based on 100 blocks per parameter, each of 100 trials.

As can be seen from the figure, there is a slight non-monotonic relation between β (lateral inhibition) and the model's performance, with fewer stop errors at intermediate values. Similarly there is a non-monotonic relation between σ (noise) and stop errors. Perhaps surprisingly, when noise is very low there are more stop errors than when noise is at moderate values. This is because noise may delay the model's decision, causing it to respond more slowly on some trials (but more quickly on others). On stop trials when noise acts against the go process this gives the stop process more time to affect behaviour. There is an optimal value for noise, however, and if it is too high successful stopping again becomes rare. Increasing μ_{stop} also reduces stop errors, though here the relation is monotonic and the explanation is more obvious: with stronger excitation of the stop node it is more likely to reach threshold on stop trials before one of the go nodes. Stop errors correlate positively with μ_{go} and δ_{stop}. In both cases the effect of the parameter is transparent. With faster excitation of the go process or with greater delay, the stop process has less chance of reaching threshold before the relevant go process. Consequently stop errors are more likely.

Relating the results to the concept of response inhibition, it appears that good inhibitors are those who either have a) near optimal levels of lateral inhibition or noise, b) slow go processes or short stop process delays, or c) fast stop processes. Miyake et al. (2000) do not report the behavioural data that would help to discriminate between these options.

The Stroop Task

The Model

Many models have been developed of the Stroop task. We focus on the well-known model of Cohen and Huston (1994), as its principal functional mechanism, interactive activation, is shared with Boucher et al.'s interactive race model. The model, shown in Figure 3, consists of four sets of nodes, with nodes within each set competing for activation through lateral inhibition. There are two task demand nodes, three word input nodes, three colour input nodes, and two response nodes. One task demand node corresponds to the colour naming task while the other corresponds to the word reading task. The colour naming task demand node is connected to all colour input nodes, while the word reading node is connected to all word input

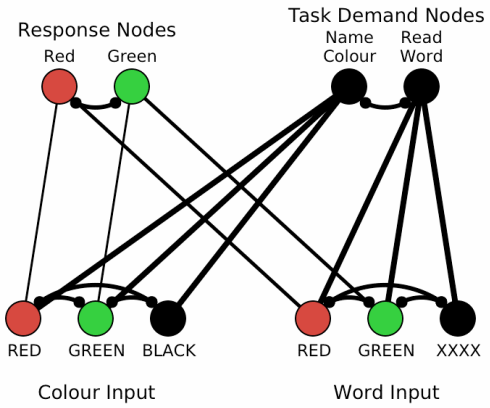

Figure 3: The Stroop model of Cohen and Huston (1994).

[1] Indeed, Boucher et al. (2007) consider how their model may be fit to data from different monkeys by varying these parameters.

nodes. Colour input nodes and word input nodes are each connected to one response node. Crucially, the connections from word inputs to response nodes are stronger than those from colour inputs to response nodes. This is justified on the grounds that word reading is the more practiced of the two tasks. As in the stop signal model, the operation of the network is cyclic with activation accumulating over time. However, the accumulation functions differ. For the Stroop model activation accumulates according to the logistic function of the time-averaged input to a node. (See Cohen & Huston, 1994, for details.)

Processing on any given trial occurs in two stages. First, input is provided to one of the task demand nodes (based on the task instructions). This causes that node to become active and the other task demand unit (through lateral inhibition) to become depressed. As a task demand unit becomes active, it excites the input nodes to which it is connected, raising the resting activation of either the colour input nodes or the word input nodes. The network settles into this temporary state, which, it is assumed, corresponds to a subject who is prepared for either a colour naming or word reading Stroop trial. Input is then provided to one colour input node and one word input node. If, for example, the trial was to name the colour of the word "RED" printed in green ink, then input would be provided to the GREEN colour node and the RED word node. In this case the colour nodes would already be moderately activated, and so the additional input to one colour node would tend to excite the appropriate response node (i.e. GREEN). At the same time, the less active word node for RED would also be receiving input and this would be tending to excite the RED response node. Hence both response nodes will receive excitation, and the balance of this excitation, plus the degree of lateral inhibition between the response nodes, will determine how quickly either response node reaches threshold.

As is clear from the architecture, there is no dedicated parameter of response inhibition. Thus, verbal descriptions of performance on the Stroop task are at odds with the computational details of the models. Nevertheless, what may be interpreted as response inhibition may well have a different label at the computational level.

Simulation Results

As in the case of the stop signal model, an initial set of simulations was performed to determine the relation between the model's performance and key parameters that could plausible be related to individual differences. Paralleling Miyake et al.'s study, the dependent variable was the difference in processing time between incongruent and neutral colour naming trials. Once again, five parameters were varied: lateral inhibition (β), the standard deviation of normally distributed noise (σ), the strength of the task demand units (μ), the gain of the activation function (γ) and the response threshold (τ). γ controls the rate at which a node's activation accumulates. It is included because Cohen and Servan-Schreiber (1992) suggest that it corresponds to an attentional modulation parameter. τ controls the sensitivity of the network to produce a response. It is fixed at 0.60 in the Cohen and Huston (1994) simulations, but we consider varying it here as it has a demonstrable affect on Stroop interference and might reasonable vary across individuals. We do not consider varying the weights from input nodes to response nodes, as these are intended to capture learned contingencies which, while possibly varying across individuals, should not vary systematically with any specific executive function.

The results of these five sets of simulations are summarised in Figure 4. The model is more complex than the stop signal model, and consequently the relations between the parameters and the relevant dependent measure – Stroop interference – are less intuitive. Nevertheless, four of the five relations are monotonic, with Stroop interference correlating negatively with β (lateral inhibition) and γ (gain), and positively with σ (noise) and τ (threshold). That is, good inhibitors correspond in the Stroop model to high lateral inhibition, low noise, optimal task demand weight, high gain or low threshold.

Yoked Simulation Studies

Recall the purpose of considering the effects of the various parameters on the performance of the two models: we are concerned with understanding the source of common variance in the tasks to which the models relate. It is hypothesised that this might be achieved by identifying a parameter that could plausibly vary across individuals and, in so doing, could underlie the observed behavioural correlation between Stroop colour naming interference and stop signal errors.

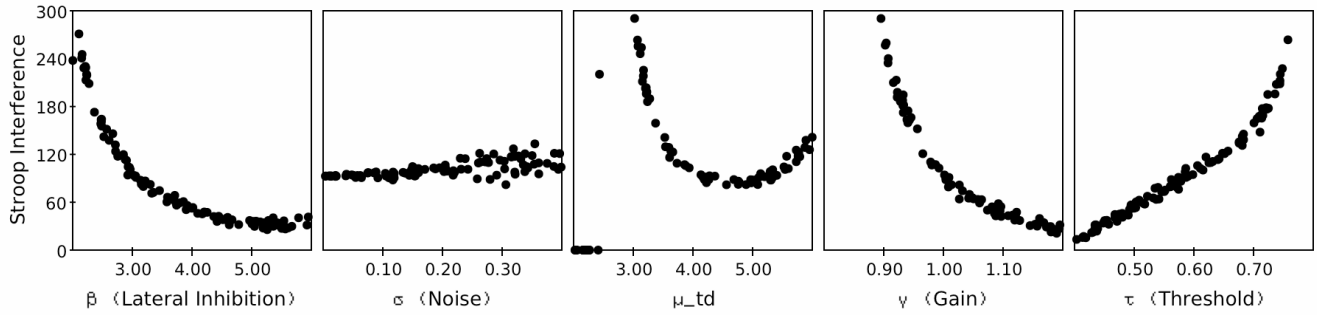

Figure 4: Effects of varying key parameters on the difference in processing time for correct incongruent and neutral colour naming trials produced by the interactive activation model of the Stroop task.

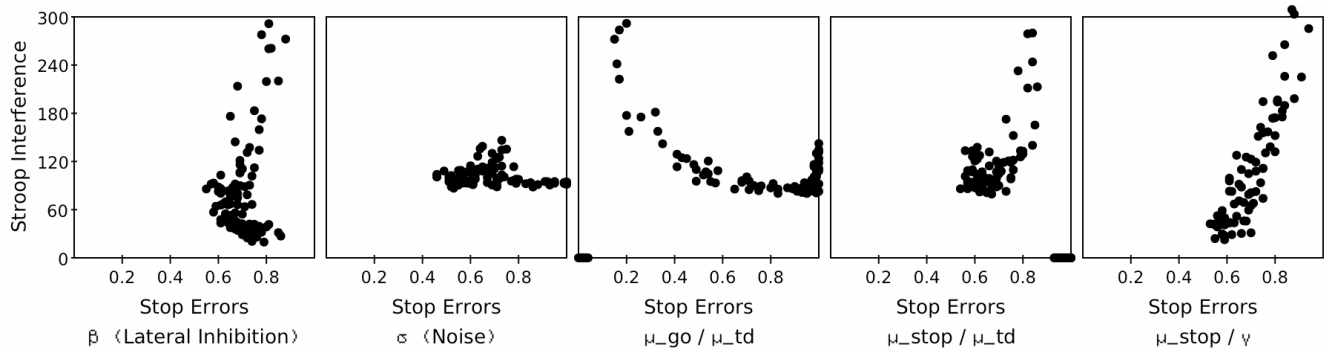

Figure 5: Effects of varying key parameters in a yoked fashion on the correlation between Stroop interference and the proportion of stop errors produced by the two models.

We are now in a position to consider candidate parameters. For example, both models share a mechanism of lateral inhibition, and pre-theoretically one could suggest that it is this mechanism, and individual differences in the shared parameter β, that underlies the behavioural correlation. The left-most panels of Figures 2 and 4 suggest that this is implausible. The issue is not the absolute size of the parameter's default value (0.025 for the stop signal model and 3.0 for the Stroop model). One can envisage re-engineering the models so that lateral inhibition in both is of a similar magnitude. The issue is that relatively high values of β lead to a reduction in Stroop interference accompanied by, if anything, a slight increase in stop errors, i.e., a negative correlation between the tasks. This is in direct contrast to the observed positive correlation.

In fact, because the relation between β and stop errors is non-monotonic, low values of β can yield a positive correlation between the tasks. This is shown in Figure 5 (left-most panel). The figure shows simulation results from 5 studies in which the value of a parameter in one model is yoked to the value of a corresponding parameter in the other model. In all 5 cases the relevant parameter values vary across the full ranges explored in Figures 2 and 4. Thus, the data in the left-most panel was generated by random sampling a dummy variable uniformly distributed between 0.0 to 1.0, and mapping the value of this onto a) the interval 0.00 to 0.05 to give a value of β for the stop signal model, and b) the interval 2.0 to 6.0 to give a yoked value of β for the Stroop model. This procedure was repeated 100 times for each of the five scatter-plots in Figure 5.[2]

From the figure we may immediately rule out several potential factors underlying the observed correlation between performance on the tasks and hence several candidates for the response inhibition function. Neither of the parameters shared by the models – lateral inhibition (β) or noise (σ) – produce correlations of the appropriate form. Hence, it would seem that individual differences in these parameters cannot underlie the observed correlations. Equally, as shown by the third plot in Figure 5, yoking the strength of the go process and the strength of task demand weights – an account not immediately related to any conceptual mechanism of response inhibition but one which, nevertheless, relates two parameters with similar functionality – fails to yield a positive correlation between the relevant dependent measures.

The desired positive correlation is shown, however, in the two right-most plots of Figure 5. Thus, the models predict that performance on the two tasks will correlate positively if a) the strength of the stop process and the strength of task demand weights are (positively) correlated, or b) the strength of the stop process and the gain in the Stroop model are (positively) correlated. There is no apriori reason to suppose the latter, but the former is plausible as both parameters concern the strength of deliberative or attentional bias. Thus, these simulation results fail to provide support for the idea that the positive behavioural correlation between Stroop interference and stop signal errors is due to a shared mechanism of response inhibition. Rather, they suggest that the correlation arises because subjects who are able to provide stronger activation to the stop process in the stop signal task are also able to provide stronger attentional bias to the colour naming task in Stroop. This suggestion is backed up by the right-most plot which shows a positive correlation resulting from yoking μ_{stop} and γ (gain). Recall that γ was also associated (positively) with attentional bias by Cohen and Servan-Schreiber (1992).

Discussion and Conclusion

In a companion paper (Davelaar & Cooper, 2010), we consider closed-form approximations to the same two models discussed here. It is demonstrated that the explanation of the behavioural correlation in terms of a shared process of response inhibition is suspect, and an attentional biasing account is proposed as a plausible alternative. The simulation results reported here corroborate both of these conclusions.

Our suggestion of attentional biasing, rather than response inhibition, as the locus of shared variability on the tasks resonates with the approach to response conflict

[2] One can envisage other approaches to yoking the parameters, e.g., by restricting attention to sub-ranges of a parameter in which its effect on the relevant dependent variable is monotonic. A further alternative focuses on the ranges of parameter values chosen. As yet there is no principled way of selecting the ranges other than through a cognitive architecture approach. Due to space limitations we do not consider these approaches here.

management of Botvinick et al. (2001). They demonstrate, within the context of three models including the Cohen and Huston Stroop model, how trial-by-trial regularities in behaviour might be accounted for in terms of a mechanism of conflict monitoring which measures the degree of conflict in the network's output nodes and modulates attentional bias, increasing it under conditions of high conflict and decreasing it under conditions of low conflict. Thus, rather than addressing response competition through response inhibition, Botvinick et al. (2001) do so through attentional biasing.

We are reluctant to fully endorse this account, however. Critically, the account is not fully consistent with the results of Miyake et al. (2000). They hold that while stop signal errors and Stroop interference are dependent upon response inhibition, they are also not dependent on two other putative executive functions – task shifting and memory monitoring and updating. Thus, if we are to account for the behavioural correlation between these tasks in terms of attentional bias, it is also necessary to show that attentional bias does *not* systematically affect behaviour on the other tasks of Miyake et al. which were held to tap these other two functions and not to tap response inhibition. Here there is reason to be cautious. Gilbert and Shallice (2002) consider performance on a task switching variant of the Stroop task in which subjects switch between colour naming and word reading. They model the critical behavioural affects by using essentially the same mechanism proposed here (i.e., by biasing task demand units) in exactly the same model (the Cohen and Huston model). Yet these are effects that, on the decomposition of Miyake and colleagues, should be explained in terms of a distinct task shifting function. Moreover in the study of Miyake et al. (2000) all correlations between putative task shifting tasks and putative response inhibition tasks were non-significant.

The concept of response inhibition held by Miyake et al. (2000) to underlie good performance in the stop signal and Stroop tasks was also held to underlie good performance in the anti-saccade task. Thus, a fuller analysis of response inhibition requires also consideration of process models of the anti-saccade task. This remains to be attempted. We would hypothesise, however, that performance in this task will also correlate with an attentional bias parameter.

Returning to the two models considered, it should also be noted that while they share principles of interactive activation, there are also major differences between them. For example, different equations govern the accumulation of activation in each model. Whether these differences are substantive or cosmetic remains to be demonstrated. However, these differences really only serve to reinforce our primary conclusion, namely, that until we have unified process models of the various putative separable executive functions, any theoretical account of their supposed unity and diversity is incomplete. By extrapolation, to understand the executive functions which underly the battery of tasks used by Miyake et al. (2000), we need to develop, within a single unified framework, models of all of those tasks. Such models must, of course, demonstrate the hypothesised shared mechanisms. Only then can we be confident that we have a plausible account of the various executive functions that contribute to the control of complex behaviour. This is, of course, one of Newell's arguments for the utility of Unified Theories of Cognition (Newell, 1990).

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*, 624–652.

Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*, *114*, 376–397.

Cohen, J. D. & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In *Attention and performance 15: Conscious and nonconscious information processing*. C. Umiltà & M. Moscovitch (Eds) (pp. 453–476). Cambridge, MA, US: The MIT Press.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *99*, 45–77.

Davelaar, E. J., & Cooper, R. P. (2010). Modelling the correlation between two putative inhibition tasks: An analytic approach. In Catrambone, R., & Ohlsson, S. (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Cognitive Science Society Incorporated, Portland, OR, USA.

Gilbert, S. & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*, 297–337.

Goldberg, G., Mayer, N.H., & Toglia, J.U. (1981). Medial frontal cortex infarction and the alien hand sign. *Archives of Neurology*, *38*, 683–686.

Lhermitte, F. (1983). Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain, 106*, 237–255.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press, Cambridge, MA.

Norman, D.A. & Shallice, T. (1986). Attention to action: willed and automatic control of behaviour. In R. Davidson, G. Schwartz, and D. Shapiro (eds.) *Consciousness and Self Regulation, Volume 4*, pp. 1–18. Plenum: NY.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.