Integrating psychological and neuroscientific constraints in models of Stroop processing and action selection

Thesis submitted in March 2003

for the degree of Doctor of Philosophy

Tom Stafford

Department of Psychology,

University of Sheffield

DEDICATION

For my grandfather

ACKNOWLEDGEMENTS

Above all I would like to thank Kevin Gurney for his generous and astute supervision.

Many thanks are also due to Jackie Andrade, for supervision and advice, especially in the early stages; to the Basal Ganglia research group, for support of all kinds (especially Mark Humphries and Ric Wood); Neil Porter for help with Matlab code; others who have provided their time, expertise and suggestions: Jim Stone, Jon Porill, Stuart Booth, Jon May, Richard Lane, John Frisby, Tom Simpson, The Cognitive Section reading group, Tom Spencer, Rod Nicolson, Lisa Lynch, Andrew Brown, Tim Gamble, Pete Furness, Karen Briggs; all the other postgrads for being available for coffee breaks, especially Mike Bywaters, most stalwart of coffee drinking-companions; Nicol Harper, for providing helpful suggestions on the research and for being the best of friends; Dick Eiser, for paying me while I wasn't doing my thesis; all my housemates and friends, especially Kevin Gillan, for company, chats, food, drinks and essential sustenance of all kinds; Jon Stafford, who provided some mathematical expertise; the rest of my family, for nurturance;

"Here's to the few that forgive what you do,

and the fewer that don't even care"

- Leonard Cohen

ABSTRACT

This thesis concerns the investigation of the principles of action selection when applied to a cognitive task. Action selection is the task of mediating between competing potential behaviours. A connectionist model of the Stroop task (Cohen, Dunbar, & McClelland, 1990) is improved by combination with a biologically plausible model of action selection which is based on the functional anatomy of the basal ganglia (Gurney, Prescott, & Redgrave, 2001a). A schematic model of word-reading (Ellis & Young, 1988) is also incorporate into the combined model. These models show that the basal ganglia model provides an improvement upon response mechanisms based on choice models of reaction time, as well as allowing a wider range of data to be accounted for. The combined model allows the prediction of the pattern of results when using non-standard stimulus-response mappings in the Stroop task. The predicted result is experimentally verified. These models and findings support a decomposition of the concept of automaticity within a connectionist framework. Action selection is an important component of cognition and it is essential to consider the properties of response mechanisms when constructing psychological models. Proposals concerning the use of connectionist models in psychology are made.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1. Overview

The main theme of this thesis is an investigation of the principles of action selection when applied to a cognitive task. I have adapted an existing connectionist model of the Stroop task (Cohen et al., 1990) to make use of a biologically plausible model of action selection. Action selection is the task of mediating between competing potential behaviours. This model of action selection is a system level network model of the functional anatomy of the basal ganglia (Gurney et al., 2001a; Gurney, Prescott, & Redgrave, 2001b). A schematic model of word-reading (Ellis & Young, 1988) was also incorporate into the combined model.

The idea of action selection provides a unifying framework for thinking about behaviour, and is related to Allport's idea of 'selection for action' (Allport, 1987). I show that psychological models and experiments can be investigated within this framework, which allows the integration of theories from disparate fields and the cultivation of new perspectives on old problems. The investigation of action selection models, in a limited sense, is already a topic of research within psychology, in the form of models of decision mechanisms for predicting reaction times (Luce, 1986). I was motivated by the question of whether the basal ganglia model could perform as well, or perhaps better, than existing decision mechanisms.

The use of computational modelling requires consideration of the role of modelling in the scientific process. Can models explain, or just describe? What structure or system are we attempting to model? At what level of description? Using this modelling work as a test-bed for these and other ideas allows some general lessons concerning the use and value of connectionist models within psychology to be drawn out.

1.2. The Action Selection Problem

> *A selection problem arises whenever two or more competing systems seek simultaneous access to a restricted system* (Redgrave, Prescott, & Gurney, 1999).

The action selection problem is the fundamental problem for any cognitive system; that of deciding 'what to do next'. A cognitive system, by definition, must have available a variety of possible responses and must adaptively select those responses depending on internal and external stimuli. The appropriate resolution of conflicts between competing responses forms the *sine qua non* of adaptive behaviour. Considered from an intra-organism perspective the selection of a single response is the resolution of the competition between multiple systems or impulses for access to the effectors of the organism. So, the action selection problem is also the problem of mediating between disparate functional units. Different nervous system areas have access to different information and have responsibility for processing with respect to different functions. Selection is necessary because there is a finite number of simultaneous actions possible.

Behavioural action selection involves both a) deciding on the priority of actions and b) enacting that prioritisation. For complex situations the majority of cognitive effort occurs in deciding on the relative priority of possible actions. However the second part of action selection - enacting the decision - is in itself non-trivial, and cannot be subsumed within the mechanisms for performing the first.

In the human case, action selection will consist of multiple, interacting, components; perceptual mechanisms, attention, decision making, learning & memory and behavioural switching – in short the whole gamut of cognitive functions. A key part of action selection is the final behavioural switch, the final gating mechanism which allows only the most important response to be enacted. It is fundamental to this thesis that the functionality of this switch is non-trivial. The design of a suitable

behavioural switch is significantly more problematic when considered in the context of an organism that must make complex, successive and real-time choices.

The importance of the action selection problem has been recognised for some time in robotics and ethology, (this is discussed in Prescott, Redgrave, & Gurney, 1999). Action selection is a priority for ethology and robotics because both have come to deal with the problem of reconciling multiple competing systems within the behaving unit. An initial cause of the explicit focus on action selection in these fields is their occupation with the co-ordination of behavioural sequences. Rather than consider single decisions, models of the complete organism must address the need to perform multiple actions in the appropriate order.

Action selection is a crucial issue for these disciplines which are closely related to psychology. It is also related to other contemporary issues within psychology. The modular organisation of the brain and mind (Coltheart, 1999; Coltheart & Langdon, 1998; Fodor, 1983; Shallice, 1988) is related to action selection by virtue of selection being a co-ordinating function between competing modules. This, along with the growing recognition that cognition is fundamentally an embodied function which, in all likelihood, is based upon the same organisational principles as behaviour (Clark, 1999), should encourage the investigation of the importance of action selection in human cognition.

1.2.1. Switching mechanisms

In order to illustrate this difficulty of enacting behavioural decisions, consider the simplified case where, in any one particular instance, there are a number of mutually exclusive behaviours, each associated with a signal which defines the urgency or *salience* of that behaviour. Two general schemes can be envisaged for mediating the selection of a single winner. All the neural circuits instantiating the competing options could interact with each other, with the most salient action emerging from

the universal interaction. Alternatively there could be a centralised switching device that takes inputs from all the competitors, and which itself decides on a winner.

These two general schemes each have advantages and disadvantages. Firstly, a universally connected switching scheme has wiring costs that grow disproportionally with each functional area added to the scheme, which is an evolutionary handicap (Redgrave et al., 1999). Secondly, a central switch makes an organism more vulnerable to a single debilitating injury than a distributed switch. Parkinson's Disease, in which the loss of dopaminergic innervation to the basal ganglia (our putative central switch) causes global movement difficulties, is an example of such an injury. However, such a scheme also makes it easier to add additional modules without the need for integration with all existing aspects of the architecture. So, the choice of switching *architecture* involves both costs and benefits. On balance the benefits of a central switch were averred by Redgrave et al. (1999).

The *functionality* of the switching algorithm is also a non-trivial design choice. *It is not the case that a successful switch could operate merely by selecting the competitor that provides the strongest salience signal*. Firstly, signals are likely to be noisy, and hence, momentary values may not be truly indicative of salience superiority. Some relative threshold of distinction needs to be achieved before one signal can be selected over others. Secondly, and related to this, there needs to be an absolute salience threshold that should be reached before a behaviour is selected. Without a threshold any momentary 'behavioural whim' might be enacted. Three other desirable characteristics of a switching device have also been identified (Redgrave et al., 1999). *Clean switching*; As soon as it is clear which is the winning behaviour, that behaviour should be selected rapidly. Evolutionary selection has presumably put a priority on fast, 'good-enough' action selection rather than the calculation of the optimum behaviour at a prohibitive time cost; "better a good plan today than a perfect plan tomorrow". *Absence of distortion*; those actions which lose the competition for expression should be suppressed so that they do not interfere with the expression of the selected action. Effective behaviour needs to be decisive.

Persistence; once selected, a behaviour should be carried through to completion. The execution of most behaviours results in the reduction of the corresponding salience, either due to satiety mechanisms (in the case of biological drives like hunger) or via affecting a change in the environment. Where two behaviours are competing for expression, both producing negative feedback on their respective saliences, chronic dithering will result; action A reduces the salience of A fractionally below the salience of B, causing the selection of action B, which then reduces the salience of B minutely below the salience of A, causing the selection of action A, and so on. The medieval logician St. Thomas Aquinas hypothesised an ass that starved to death because it was unable to select between the two identical bales of hay it was positioned exactly equally between. A modern adaptive behaviour version of this fable might have the ass dying of starvation (and exhaustion) as it runs between a water trough and a food trough. Each stalk of straw makes the animal's hunger less than its thirst, and in turn each sip of water makes its thirst less than its hunger. Persistence of selection is needed to ensure useful completion of action. This persistence, or hysteresis, needs to be balanced against the openness of the switching device to interruption by genuinely important salience differences. Obviously if, while the ass is eating, a lion wanders into the barn it is adaptive for the ass to forget about eating and drinking, and to run away instead. In conclusion, there are a number of non-obvious factors to account for and conflicting needs to reconcile in the design of the response mechanism.

1.2.2. The Basal Ganglia as a central switch

The basal ganglia (BG) are a collection of subcortical nuclei, which are common to all vertebrates. We, the Adaptive Behaviour Research Group (ABRG) at the University of Sheffield, have proposed that, in vertebrates, the 'central switch' of action selection is located in the basal ganglia (Prescott et al., 1999; Redgrave et al., 1999).

Inspired by this, a computational model of the basal ganglia, based on their known connectivity, has been developed (Gurney et al., 2001a; Gurney et al., 2001b; Humphries & Gurney, 2002). Henceforth this will be known as the BG model. The model successively implements the characteristics of a successful switching mechanism (see section 1.2.1). Additionally the BG model provides a suitable switching mechanism when placed within an embodied robot model (Montes-Gonzalez, Prescott, Gurney, & Redgrave, 2000). A fuller review of the BG model is given in chapter 3.

If there is a general switching mechanism, as we believe there is, then it's characteristics will be evident across a wide range of tasks. It is this consideration which prompts my attempt to simulate the results from a classical reaction time paradigm with a model which includes a biologically plausible response mechanism designed to produce effective switching. Conversely, I also believe in the validity of comparing models of reaction time with respect to the general criteria for effective switching by response selection mechanisms (section 1.2.1).

1.3. Response selection; the psychological perspective on action selection

A concept from cognitive psychology that appears similar to action selection is that of response selection. This refers to the central processes involved in choosing between responses based on stimulus information (Pashler, 1998). Much work has gone into the investigation of how different task requirements affect response selection; identified factors include stimulus-stimulus conflict, stimulus-response conflict, and simple response conflict. A mainstay of the investigation of these effects is the Stroop task (Stroop, 1935).

Much of this thesis concerns efforts to include a biologically based model of action selection (the BG model) into models of response selection in the Stroop task. The nature of the Stroop task is outlined below in section 1.6.

1.4. Attention

Attentional function in humans plays a role in action selection by filtering or attenuating irrelevant stimuli and enhancing the processing of stimuli relevant to adaptive behaviour. One approach to attentional function that attempts to recast it in the context of the need to make appropriate behavioural choices is that of 'selection for action' (Allport, 1987; Allport, 1993). Sharing something in common with Gibsonian theories of perception (Gibson, 1979), this framework views attentional selection as dependent on the need to have one stimulus driving behaviour at any one time, rather than as the result of some intrinsically limited resource or of a structural bottleneck in processing detached from the output requirements of the system. This behaviour-based perspective on attention has found useful application in explaining changes in the nature of attentional processing in similar tasks which have different response requirements (Brown, 1996) and in providing a general framework for understanding attentional phenomena (Allport, 1993; Styles, 1997).

Much debate in the psychology of attention focussed on the location of a putative bottleneck in processing – whether it occurs early or late in stimulus processing. This concept is complicated by the idea that the location of the bottleneck might depend on task, and/or modality, and that the filtering it provides may not be complete, but instead an merely an attenuation of to-be-ignored stimuli (see Pashler, 1998, for a very clear presentation of these ideas). Within this context, the basal ganglia-as-switching-mechanism can be seen as a final, absolute, late bottleneck. A tentative start is made in this thesis to the investigation of the 'hard' and late BG gate and it's interactions with the 'soft', early, gate of attentional control. The possibility has been raised that separate pathways or systems may exist for conscious perception and for action-based behaviour (Goodale & Humphrey, 1998; Goodale & Milner, 1992). This, at the very least, makes it more problematic to insist on any fixed bottleneck

Consideration of the issues surrounding attention must include reference to automaticity (discussed in sections 1.6.2. and 7.3). One convenient definition of

automaticity is the development of capacities that can operate without attentional supervision (see Pashler, 1998, for a critique). This learning dynamic, and its interaction with attention, provides an additional complication to the application of the action selection concept to the human case.

An important distinction is between preparatory and dynamic attention. Preparatory attention affects the baseline activation of a region involved in processing signals before, or as, they arrive. Dynamic attention is invoked by the existence of activity that it modulates, and so occurs after some initial signal processing. The Cohen model uses preparatory attention and this has found interesting parallels in subsequent functional imaging work (see section 7.2.2), whereas more recent models (Botvinick, Braver, Barch, Carter, & Cohen, 2001) utilise dynamic attentional control.

1.5. Choice models of response conflict

All PDP models of Stroop processing explicitly or implicitly express a theory of response selection. A mechanism is needed to compare and select potential actions, even if only mediating between two simple choices. In a sense all stimulus processing could be considered response selection, so we define 'the response mechanism' as the process(es) which enact the switching between actions once their relative importance has been established. This mechanism, which translates the output of the network into actions, will play a role in determining reaction times. Two basic approaches to models of choice situations are the mathematically abstract (henceforth 'mathematical') and the neurobiologically and ethologically inspired (as illustrated by our BG model, henceforth 'neuro-ethological'). Mathematical approaches, as exhaustively reviewed by Luce (1986), involve constructing an algorithm which provides reaction times given the input of parameters which represent such things as the number of choices, the endogenous noise, a fixed encoding time, etc. Neuro-ethological approaches involve constructing or analysing choice behaviours in real-world animals or animats (Prescott et al., 1999). Existing models of the Stroop task universally use abstract mathematical and thereby highly

artificial models of response selection. Response mechanisms of this type benefit from being contrasted with insights from the study of the 'action selection problem' in neuroscience, robotics and ethology. Research in these areas has recognised the complexity of the problem of switching between behaviours, and so response mechanisms developed in this context possess additional features which have hitherto been neglected in the in the development of mathematical models of response selection.

1.5.1. Response selection models

Response mechanisms may consist of a simple threshold (Cohen & Huston, 1994; Zhang, Zhang, & Kornblum, 1999) or a more complex signal processing algorithm based investigations of reaction times in decision paradigms (Cohen et al., 1990; see Luce, 1986, for a review of the extensive literature; Phaf, Vanderheijden, & Hudson, 1990).

Arguably the most successful mathematical model of response times for two-choice decisions is the diffusion model (Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). This model belongs to the general class of random walk models, which are closely related to accumulator models such as that used by Cohen et al (1990). They differ mainly in that they contain only a single counter or accumulator, which is incremented or decremented towards positive and negative thresholds representing the two competing responses. Both classes of models have a long history of investigation in the context of choice reaction time studies (Luce, 1986). In the diffusion models, within each trial, the drift is stochastic. However, it is possible to define the mean 'drift rate' as the mean rate of approach to the threshold, and which may be considered to reflect the relative strength of evidence for a response.

A somewhat different model due to Reddi & Carpenter (2000) - Linear Approach to Threshold with Ergodic rate, or LATER - uses a constant drift rate within each trial but varies this rate randomly from trial to trial. This model is based on studies of

saccade generation latency in humans and other primates (for reviews see Gold & Shadlen, 2001; Schall, 2001). The LATER model is formally equivalent to the diffusion model without the addition of noise.

Ratcliff, Van Zandt & McKoon (1999) compared connectionist models which accounted for reaction times against diffusion models of reaction time. They found that connectionist models were not as good as diffusion models in accounting for the known patterns of reaction time distribution. It is critical however, that the connectionist models assessed all used a simple response threshold as their response mechanism; when the activity of a unit crosses a fixed threshold it was taken to indicate a response. So simple threshold response mechanisms are unable to account for the wealth of data available using reaction time methodologies.

To provide a more sophisticated response mechanism than a simple threshold for their connectionist model Cohen et al (1990) used an 'evidence accumulation' algorithm to select a response based on the difference between signal strengths integrated over time. Using this response mechanism the model accounts reasonably well for the pattern of reaction times in the Stroop task. However, the algorithm is adapted from models developed in the context of choice reaction time experiments, and thus designed specifically and solely for the purpose of accounting for response times. The algorithm calculates the time to produce a single response starting from an arbitrary 'reset' condition that initialises the start of the period. In addition the algorithm necessarily selects a response if allowed to run on long enough from time zero; crucially, there is no provision for doing nothing. The consequences of this for modelling stimulus onset asynchrony (SOA) variants of the Stroop task are discussed later (section 2.4.2). These are general flaws of the mathematically abstract response time models. They are disembodied from the rest of processing involved in a task, attempting instead to describe the pattern of reaction times found without reference to the internal structure of the mechanism doing the processing. Because of this disembodied focus they are not designed to simulate the continuous nature of response selection. That is, they do not account for the timing of actions which interrupt others already ongoing, but instead are designed to provide a

response times to stimuli that appear is isolated 'trials'. They are conceived and investigated entirely separately from the rest of the behaving organism – they are dislocated physically ('in space') as well with respect to actions occurring in succession ('in time'). Because of this they are impractical for use as response mechanisms for real-world, real-time operating agents.

An ideal model of the human response mechanism needs to account for patterns of reaction time data (as classic decision mechanisms aim to), while also providing a ethologically functional action selection device (as discussed in section 1.2.1). This thesis is an attempt to show that the basal ganglia model (section 3) can fulfil both these requirements. Not only this but it moves beyond phenomenological description of reaction time data to suggesting a neural locus for the response mechanism and this gives corresponding mechanisms underlying the production of reaction times. By connecting the basal ganglia model of response selection with the Stroop model we begin the task of re-connecting response mechanisms with the rest of cognitive processing in the chain between perception and action.

1.6. The Stroop Task

1.6.1. Task description

The well-known Stroop task (Stroop, 1935) involves responding to either the word or the colour of a coloured-word string. This word string can, itself, be the name of a colour. Examples of the main combinations are given in Table 1. There are two basic tasks, word-reading and colour-naming. Within each there are three possible general classes of stimuli. For *congruent* stimuli the word and the colour match (e.g. the word 'red' in red ink), for *conflicting* stimuli the word and the colour are at odds (e.g. the word 'red' in green ink), and for *control* stimuli the irrelevant dimension is, at least nominally, neutral with regard to the target dimension (so, for example, the word 'red' in standard black ink, or the string 'XXXX' in green ink).

CONDITION	TASK			
	Word Reading		Colour Naming	
	Word	Colour	Word	Colour
Control	"red"	black	"xxxxx"	red
Conflict	"red"	green	"green"	red
Congruent	"red"	red	"red"	red

Table 1: A taxonomy of the possible combinations of basic Stroop stimuli, with examples. The correct response to all six stimuli is 'red'.

The main result found with such stimuli is that word information interferes with the naming of colours, while colour information does not interfere with the naming of words. This is the basic 'Stroop effect'. Subjects find it almost impossible to ignore the word while they are supposed to be naming the colour – as if word reading is automatic or involuntary. The slowing of reaction time, or increase in error rate, due to the information on the irrelevant dimension of the stimuli is termed 'interference'. When reaction times are decreased via the influence of the irrelevant dimension, as happens with congruent stimuli, it is termed 'facilitation'. There are, however, other replicable patterns in the data. Firstly, word-reading is faster than colour-naming in all conditions. A second pattern of results is that the interference effect for colour-naming is notably larger than the facilitation effect[1]. These features can be seen in Figure 1, which shows a typical set of reaction times from a Stroop task (Dunbar & MacLeod, 1984). These are the basic features that any model of reaction times in the Stroop task needs to mimic.

[1] At least when XXXX-controls are used. Brown (2002) discuss contrary results when neutral word controls are used as the baseline.

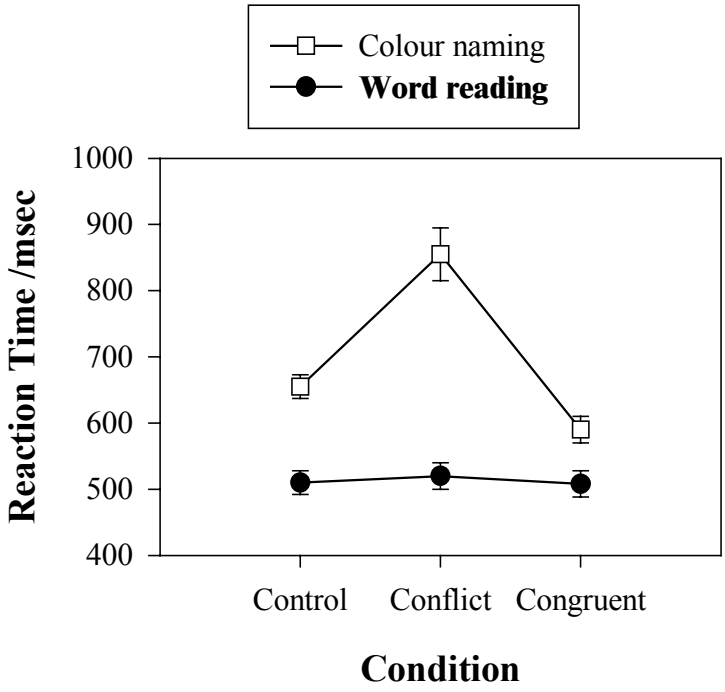


Figure 1: Reaction Times for all fundamental conditions in the Stroop task, after Dunbar & MacLeod (1984, p. 630). Standard error bars are shown.

Research using a stimulus-onset-asynchrony (SOA) paradigm in which word and colour information appear at separate times (Glaser & Glaser, 1982; Schooler, Neumann, Caplan, & Roberts, 1997; Sugg & McDonald, 1994) has provided an additional phenomenon. Essentially, when the colour dimension of the stimulus is shown before the word dimension in the word reading condition *interference of colours on words does not occur*. This is not what would be expected if interference was merely due to word information being processed faster than colour information and arriving first at a decision-making bottleneck – a 'horse-race' model. If this were so, the 'head start' given to colour information in an SOA experiment would allow it to interfere with the (non-simultaneous) word information. Conversely, in the SOA experiments, interference on colour naming by words persists, even when the colour information significantly precedes the word information. Although interference in SOA experiments is less than when the two stimulus dimensions are presented simultaneously, word information is not suppressed completely even when subjects are given hundreds of milliseconds in which to adjust to ignoring the irrelevant word information. These experiments show that speed of processing is not the defining difference between word and colour processing. So it must be that word processing possesses some other characteristic with leads to its preferential processing. This preferential processing is consistent with a mechanism which weights word information as somehow 'stronger' than colour information, as postulated by Cohen et al (1990). In addition, parallel processing of information – a key element of connectionist models – appears to be a necessary properties of any successful account of the Stroop task (see Cohen, Servan-Schreiber, & McClelland, 1992; Hommel, 1997; MacLeod & MacDonald, 2000). For reviews of the extensive research on the Stroop effect see MacLeod (1991) or, more recently, MacLeod & MacDonald (2000).

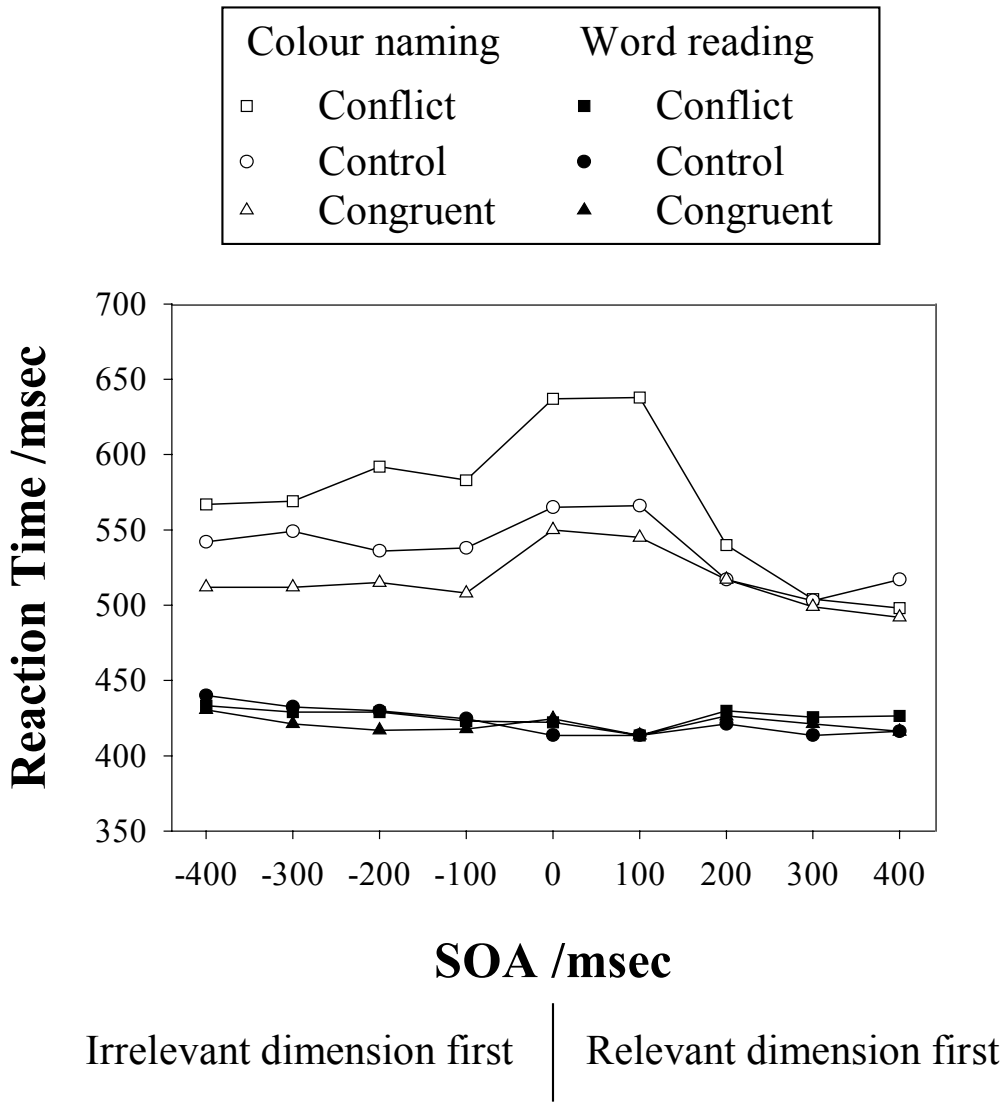


Figure 2: Effects of varying stimulus-onset-asynchrony (SOA) between word and colour stimuli in the colour-naming and word-reading tasks. Data from Glaser & Glaser (1982).

Note that interference is not a monotonic function of SOA, it is maximal between 0 and 100ms SOA and decreases to 0 above 300ms SOA, and decreases to around 25-50 ms below –200 ms SOA.

1.6.2. The psychology of the Stroop task

The Stroop task has traditionally been interpreted in terms of the automatic-controlled processing distinction (e.g. Posner & Snyder, 1975; reviewed in MacLeod, 1991). Word-reading is seen as an automatic process, occurring involuntarily and without effort. Colour-naming is a controlled process, which requires both effort and supervisory attention. As discussed above, a simple 'horse-race' model, which accounts for the differences as due to different intrinsic speed of information processing for word and colour information, has been falsified by the SOA experiments. Word information is obviously privileged in a way beyond speed of processing alone. Words may be processed faster only as a consequence of that privilege. Automatic processes do not require attention for their performance, and the Stroop task seems to suggest that word reading is performed even when attention is being used to ignore or actively suppress word information. The exploration of this phenomenon requires a theory, or model, which can deal with the quantitative interaction of attention, processing and learning. Connectionist modelling, in the form of unit activations and weighted connections, provides a common – quantitative – currency for exploring the interaction of these factors.

The invariant nature of automatic word processing has been questioned by a number of recent studies. Durgin (2000) shows that a reverse Stroop effect (i.e. when colours interfere with word reading, not vice-versa) can be obtained by altering the nature of the response required to Stroop stimuli. Asking subjects to indicate their response by moving a computer mouse to a colour patch (i.e. matching a visual stimulus to a visually guided motor response) produced the opposite pattern of interference effects compared to the standard Stroop (i.e. which involves subjects matching a visual stimulus to a verbal response). Thus the automaticity of word

reading is not invariant, a property of stimulus processing in isolation, but contingent on the nature of the stimulus-response relationship as well. Further MacLeod & Dunbar (1988) have provided evidence that automaticity is a relative property that exists on a continuum of automatic-controlled possibilities. Finally, the automatic word reading effects in the Stroop task have been shown to be affected by task context (Besner & Stolz, 1999; Dishon-Berkovits & Algom, 2000), attentional set (Besner, Stolz, & Boutilier, 1997), the compatibility of the stimuli with each other and with the responses required (Durgin, 2000; Zhang & Kornblum, 1998) and even social context (Huguet, Galvaing, Monteil, & Dumas, 1999). This evidence encourages a view of automaticity as relative rather than absolute, stemming from the encoding of particular stimulus-response mappings that can be primed by context and attention, rather than as an invariant and involuntary response to a stimulus. Connectionist models are an ideal vehicle for exploring this formulation of automaticity. In such models the activities of the components have continuous values, so they can easily capture the putative continuous nature of automaticity and the continuous nature of variables associated with it such as interference, reaction time, etc. Additionally the internal structure that connectionist models have, and focus they place on input-to-output mappings, allows their natural extension to accounts of the differential automaticity for different input-output requirements (as chapter 5 shows).

The unity of the automaticity concept has received a number of critiques (Bargh, 1989; Duncan, 1986; Logan, 1987; Pashler, 1998; Ryan, 1983). Inconsistencies in the manifestation of supposedly automatic properties point to the existence of a number of component parts covered by the term. Dimensions along which automatic-like processes seem to vary include whether the action is involuntary, outside of awareness, effortless, independent of other demands, and/or ballistic (i.e. runs to completion without the need for, or opportunity of, intervention).

Detailed computational models of automatic processes can provide useful insight into these tangled issues. The influence of context and learning can be varied continuously in these models, and so properly explored. The use of computational

models makes it possible to investigate whether the components of automaticity necessarily appear together, or are somewhat independent. Insight into the model mechanisms which underlie an operational definition of automaticity will illuminate the use and meaning of the term in psychological theory. It is to computational models of Stroop processing that we turn in the next section.

1.6.3. Models of the Stroop task

There exist a number of explicit computational models of Stroop task (Cohen et al., 1990; Kornblum, Stevens, Whipple, & Requin, 1999; Phaf et al., 1990; Zhang et al., 1999) and related tasks (Zorzi & Umilta, 1995). The model of Cohen et al (1990), henceforth 'the Cohen model', is the most celebrated and is the starting point for much of the work in this thesis.

The Cohen model (dealt with extensively in section 2.1) provides an explicit, computational account of the processing mechanisms involved in the Stroop task, within a connectionist framework. The nature of processing in connectionist models – i.e. continuously graded signals passed between units with continuously graded connection strengths ('weights') – means that the Cohen model can simulate the continuously graded properties associated with automaticity of processing, and can account parsimoniously (i.e. naturally within the existing framework of the model) for learning phenomena in the Stroop task. An important feature of the model is the way attention operates. Attentional control is controlled by input units which are connected to the network like any other unit and pass weighted activity to them in the same way. So the operation of attention is couched in the same terms as all other processing in the model, and is not separate or exogenous to the model's function. This grounding of attention is a fundamental starting point for the investigation of attentional function and the establishment of cross-talk between cognitive neuroscience and computational modelling.

The Cohen model is similar to other models of Stroop processing. Both the models of Phaf et al (1990) and Zhang et al (1999) are essentially similar, if more complex

and/or more general. The success of the Cohen model lies in the way it captures the essential features of these other models and is presented relatively clearly.

Although there are only a small number of models of processing in the Stroop task, it is clear that the Stroop task contains many elements which are themselves the subject of computational investigation. While Cohen et al (1990) titled their paper 'On the Control of Automatic Processes', Phaf et al (1990) considered their model to concern 'attention in visual selection tasks'. In addition to attention and automaticity, the Stroop task also involves stimulus-stimulus and stimulus-response compatibility issues (Zhang et al., 1999). The choice of the mechanism which mediates response-response conflict – i.e. the response mechanism - also turns out to be of significant importance.

1.7. Modelling cognition

In a famous paper, the Nobel Laureate Francis Crick lampooned the psychological modelling community for the 'physics envy' that it is often accused of:

> *"I also suspect that within most modellers a frustrated mathematician is trying to unfold his wings. It is not enough to make something that works. How much better if it can be shown to embody some powerful general principle for handling information, expressible in a deep mathematical form, if only to give an air of intellectual respectability to an otherwise rather low-brow enterprise."* Crick (1989).

Contrary to Crick's sarcasm, I hold that making 'something that works' is indeed a wholly inadequate function for the role of modelling in psychology. Connectionist models in particular have come to play a poorly understood but vital theoretical role. These models provide a functional description of cognitive mechanisms and can bridge the cognitive and biological levels, and it is therefore essential to understand their correct use vis-à-vis psychological and neuroscientific theory.

1.7.1. Connectionist models in psychology

The beginning of the recent excitement about neural networks is often marked from the publication of the two volumes of 'Parallel Distributed Processing: Explorations in the microstructure of cognition' by Rumelhart, McClelland and the PDP research group (1986b). Combined with the spread of computing technology, this successful exposition of fundamental methods and theory led to modelling becoming a lynchpin of psychological research. Although some were, and are, resistant to the application of these ideas in psychology (Crick, 1989; Marcus, 1998; Pinker & Prince, 1988), Parallel Distributed Processing (PDP) or connectionism can claim a *de facto* victory for acceptance within psychology. I will now discuss some of the issues surrounding Parallel Distributed Processing models.

contentions

Despite the widespread kudos of modelling work not all are agreed on *why* modelling work is done or *how* it should be done. At best this represents healthy variety, at worst it is indicative of the unwarranted celebration of an esoteric art, an art which to be questioned critically requires more time and training than the majority of psychologists have.

PDP models, in utilising many, multiply connected, simple units, are often feted as carrying out processing in a similar way to the brain. However, because it fails to fully match the biological details of the brain, PDP has been criticised for its lack of biological plausibility (Crick, 1989; Crick & Asanuma, 1986). For those who wish to model cognitive processes without explicit reference to the underlying biology, this is a distraction, while others – computational neuroscientists – have attempted to correct this lack of biological plausibility by including increased levels of biological detail in their models.

The issue of what level of detail to pitch models at (Bechtel, 1994; Broadbent, 1985; Cleeremans & French, 1996) is closely related to the debate over the nature of

representation in cognition (Page, 2000; Smolensky, 1988). It is apparent that those who believe that symbolic representations are appropriate for their area of study of cognition, even if they do accept connectionist modelling (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Young & Burton, 1999), prefer localist models which instantiate theories derived from psychological theory. Those who favour subsymbolic representations as appropriate (O'Reilly & Farah, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996) use models with distributed representations and believe that psychological-level explanations will emerge from biologically-styled mechanisms (Seidenberg, 1993).

A broader set of criticisms has more recently been levelled at the 'cognitive' bias of psychological modelling (Beer, 2000; Clark, 1999). Most models in psychology, it is asserted, assume an environment which merely provides static inputs and is unchanged by the model outputs. Although models of cognitive processes are becoming increasingly sophisticated they ignore interactions between the mind and world and between the mind and body which can drastically alter, and often simplify, the nature of a cognitive task. As well as altering the nature of the problem, including the world (situating) and including the body (embodiment) in a model can reveal the true nature of the task facing an organism. This holistic approach focuses on model which behave in an evolutionary adaptive and real-world, real-time manner (Brooks, 1991).

Our contention is that methodological debates over the use of modelling work can only be answered by considering the theoretic purpose for which a particular model is constructed and level of description at which it is aimed. There is no right approach to modelling independent of the purpose of a particular model. In this thesis, where possible, I have tried to make explicit the choices I have made, and explain and justify them.

scientific function

Roberts & Pashler (2000) have criticized the practice of using a good fit between the model outputs and the empirical data as a criterion for judging the adequacy of a model. They are quite right in pointing out that the 'good-fit' method of assessment ignored the flexibility of the model and the variability of the data. It can be that a model which provides a perfect fit in one instance would fail to provide an adequate fit if the data were to be re-acquired. Conversely, it can be that a model has sufficient degrees of freedom to fit any plausible set of data, and therefore says nothing special. Roberts & Pashler (2000) single out for particular mention the Cohen et al (1990) model as being assessed, wrongly, merely on the basis of a good fit to data. I feel that other criteria are often used implicitly by the authors of a model, but they need to be made explicit. An adequate account needs to establish the range of possible outputs that the model *could* produce, rather than just report the specific outputs that the particular final version of the model does produce. Let us call this range the 'explanatory range' of the model. Establishing what results are outside of the explanatory range of the model is just as important as establishing what is inside the range of the model. Knowing what the model *cannot* predict is vital for establishing criteria for falsification of the model.

To establish the explanatory range it is necessary to know which features and which parameters can be varied. This relates back to the overall theoretical context within which the model is constructed. Specifically it is important to distinguish *theory-specified* from *theory-unspecified* features of a model. McCloskey (1991) claimed that a model isn't a theory any more than a black-box which provided outputs similar to human performance would be a theory. This is true so long as the principles inspiring the presented model are not articulated. The difficulty arises because constructing a working model requires the specification of many details that may not be explicit in the theory motivating the model. This is both a blessing and a curse. Although constructing the model forces the theorist to consider in detail how a function is performed, it is also necessary to pick, somewhat arbitrarily, specific forms and values for elements of the model. These theory-unspecified elements take

on precisely specified form once the model is constructed and can obscure which aspects of the model are specified by the theory and form the core assumptions that the model is based on. These hazards are discussed by Lewandowsky (1993).

So, uncertainty exists about why and how to use computational models. In the next section I discuss the main reasons to use computational models, while the work presented in the rest of the thesis provides material which will be used in Chapter 7 to draw out some of the strengths and weaknesses of the modelling approach, and to provide some guidelines on efficacious use and presentation of models.

1.7.2. Why use explicit computational models?

It is apparent that there are debates still current within psychology concerning the use of connectionist models. Despite the lack of resolution of some of these debates, there is undoubted value in the enterprise of computational modelling in general and in connectionist modelling in particular. A brief sketch of some of the benefits of computational models will illustrate my reasons for feeling this, and help make distinct the reasons why it was thought that computational modelling would be productive in the present context.

i - problem elaboration

A model must be complete mechanistic description of how some function is carried out. This 'functional closure' of a process can reveal the exact nature of the problem being solved by the process, since modelling a solution entails the full specification of the problem. This full specification is not always entirely clear for many psychological fields of investigation. Marr (1982) made stating the problem and the constraints available for its solution the first level of his hierarchy for investigating a psychological mechanism. Once a problem has been elaborated, analogies to other problems and/or processes may become apparent. In this way modelling a process/problem can reveal connections between diverse phenomenon (McClelland, 1988).

ii - theory-testing – 'opaque thought experiments'

In the context of simulations of the dynamics of biological evolution, models have been described as 'opaque thought experiments' which in turn require their own exploration before they can be related back to the theory level (Di Paolo, Noble, & Bullock, 2000). To this we can add that models are *rigorous* thought experiments, which take advantage of their explicit mathematical form to provide quantitative results and compel us to provide a full articulation of the theory, which may perhaps reveal previously hidden inconsistencies or inadequacies.

As well as being a clarifying abstraction from messy reality, models require that their guiding theory be rigorously articulated (McClelland, 1988). Assumptions need to be clearly stated if they are to be incorporated into the model. It should, however, be said that models also require a great deal more detail than is usually provided by psychological theory. For this reasons the core assumptions of the theory that is being incorporated in model form need very explicit articulation, lest they become lost in a morass of necessary, but theory-unspecified implementational choices.

Complex theories require quantitative, computational models for their proper exploration because the consequences of the interactive behaviour of many entities over time can be highly opaque; even the inclusion of a single feedback loop in an information-processing model confounds most people's attempts at intuitive understanding. Models can help reveal the implications of proposed information processing principles in psychology (McClelland, 1988). In many cases unexpected system-level properties emerge from the interaction of component parts, a phenomenon which has received much attention (e.g. Chiel & Beer, 1997; Clark, 1997; Hofstader, 1979; Holland, 1998; Newman, 1996). The widespread awareness of the possibility of 'emergence' was made possible by advances in both the theory, and technology, of computational simulation. Work presented later in this thesis

provides an example of emergence in the context of interactions between training environment and innate architectural constraints (section 2.3).

By elucidating the relationships between formally stated entities, modelling can allow the discovery of general computational truths, that is, the generic correlates of particular information processing systems. Thus, we may find that all systems of type X (i.e. that possess a certain set of formal properties) also possess property Z. It is of course a matter of empirical inquiry to discover whether the cognitive system in question is also of type X. Two common and related uses of this type of proof are demonstrations of *sufficiency* and of *existence*. Models can be used to show that a minimal set of features is sufficient to generate a certain property. An example is the work of Seidenberg & McClelland (1989) which claims that a PDP model of word reading can account for non-word reading without needing to utilise a 'dual route' architecture On the other hand, an existence proof establishes that some property is indeed possible given some restricted set of features. For example Hinton and Nowlan (1987) demonstrate that lifetime learning can guide evolution - the so-called 'Baldwin effect' (discussed in Maynard Smith, 1987; see McNellis & Blumstein, 2001; and Riolo, Cohen, & Axelrod, 2001, for more recent examples of an existence proof being offered). These two types of proof are two sides of the same coin. Existence proofs show that a property exists given a set of features that are fixed, sufficiency proofs show that a minimal set of features can generate a property of a system that is fixed (i.e. because it is known to exist).

iii - explanation

A model which matches empirical results can help provide insights into the processes involved, as well as the behaviours exhibited by experimental participants. So, for example, the Rumelhart & McClelland (1986) model of past-tense acquisition (Bullinaria, 1997) provides a potential explanation of *why* children have a stage of over-generalising the past tense of English verbs[2], in addition to

[2] I.e. it is because of the nature of the interleaved learning of regular and irregular verbs in a single set of connection weights.

successfully mimicking the empirical description of the phenomenon (i.e. the *what*). It is the claim of more radical connectionists that many psychological explanations will stem from this biologically-constrained computational level (Seidenberg, 1993). Cleeremans (1996) provides a useful handle on the difference between description and explanation. He asserts that computational models can provide explanations of phenomena if the primitives of the model are of a lower descriptive level than the phenomena to be explained.

The use of models provides new computational hypotheses about the function of cognitive and neural systems. They are part of the scientific cycle of theory construction and testing. Computational hypotheses are therefore essential to make sense of the overwhelming mass of neurobiological and cognitive data.

iv - metaphor

A computational framework provides a powerful metaphor for the consideration of problems in cognition. Although computational modelling seeks to provide more than metaphors alone (as discussed above) this should not belittle the very real, if nebulous, value of metaphors derived from computational investigations. Although most work in modelling will consider finished models, even the act of beginning to think about how to model a cognitive process can illuminate hitherto latent facets of a theory. Computational metaphors can be part of the 'context of discovery' (Vallacher & Nowak, 1997) in science, as well as the 'context of proof'. Scientific models allow the cultivation of intuition, so that all phenomenon of a particular class can benefit from the insights gained in the modelling of a single phenomena (see Marton, Fensham, & Chaiklin, 1994, for a discussion of the importance of scientific intuition).

v - conclusion

The work discussed in the rest of this thesis provides an opportunity to illustrate all of these themes, and to elucidate the connections between computational modelling, theory advancement and empirical research in psychology and neuroscience.

1.8. Thesis outline

This introduction has established the theoretical (sections 1.1 to 1.6) and meta-theoretical (section 1.7) motivations for my investigations. Chapters 2 and 3 establish the technical background to this work, giving the details of the Cohen et al (1990) model of the Stroop task and of our model of the basal ganglia. These chapters also provide illustrations of the methodological and conceptual issues that are dealt with in the final discussion (e.g. sections 2.2 to 2.4 and section 3.3). Chapters 4 to 6 present the bulk of technical work; the simulations and experiments. Chapter 4 concerns the use of the basal ganglia model as a response mechanism for the Cohen et al (1990) model. Chapter 5 concerns integration of theories of word reading into the model. The improved model makes a prediction that is experimentally confirmed. Chapter 6 discusses the addition of dynamic attentional control to the models. Chapter 7, the discussion, provides the theoretical context for these findings (sections 7.2 to 7.5) and draws out conceptual and methodological issues surrounding the use of connectionist models in psychology (section 7.6). This section concludes the arc of meta-theoretical speculation begun in section 1.7. The thesis concludes, sections 7.7 and 7.8, with an overview of the limitations and potential of the work adumbrated here.

2. A CRITIQUE OF THE COHEN MODEL AND ITS RESPONSE MECHANISM

The purpose of this chapter is, firstly, to validate and fully understand the behaviour of the model of Cohen et al (1990). The model is key to the models subsequently developed in this thesis. Secondly, the exploration and validation of the model suggests further experiments. These investigations of the nature of representation in the model, and, most importantly, of the crucial nature of the response mechanism are pivotal to the subsequent model development presented in this thesis. Thirdly, the Cohen model has become a highly cited piece of work, appearing in textbooks and becoming an obligatory reference in any review of the Stroop task. The other work in this chapter makes it natural to include a critical evaluation of the model in light of advances made since it's first publication.

2.1. The Cohen et al (1990) model of Stroop processing

2.1.1. Overview

The original Cohen model (Cohen et al., 1990) is a feed-forward network in which word and ink-colour information are processed in two, initially separate, pathways which converge at the response stage. The architecture of the model is shown in Figure 3. Features of the stimulus are presented to the input layer of the network, which generates a cascade of activation change throughout the units of the network. 'Task-demand' units simulate the effect of attention by modulating activations in the hidden nodes of the network's pathways. The differential automaticity of the two tasks is mimicked by the network having stronger connections in the word reading pathway. This difference is evoked through giving the network a higher exposure to word-reading patterns than colour-naming patterns during the training stage of the network. For each set of inputs the reaction time is computed via a response algorithm which takes outputs from the two response units of the network.

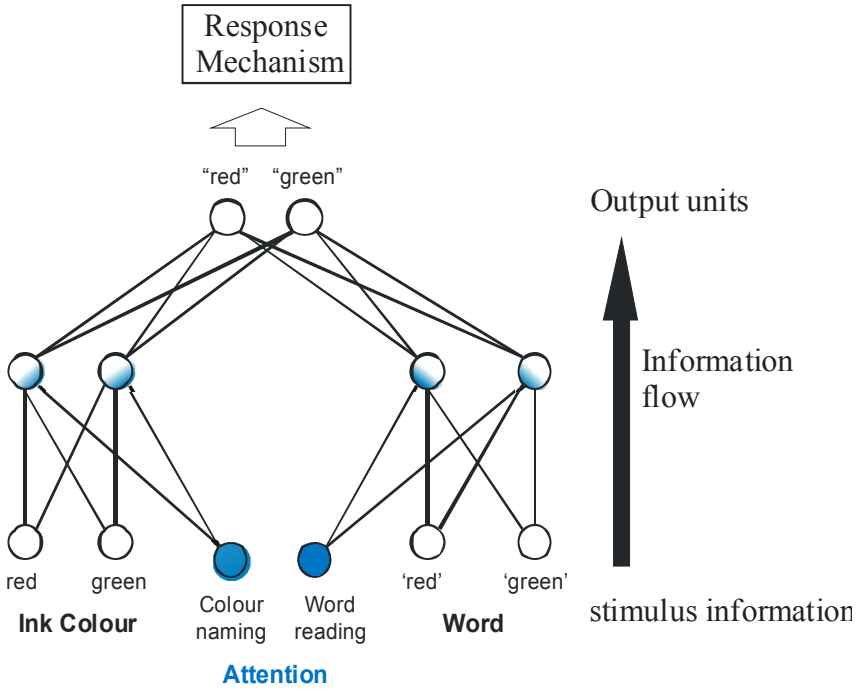


Figure 3: Architecture of the Cohen model, after Cohen et al (1990, figure 3, p. 339). The sites and sources of attentional modulation are shown shaded.

2.1.2. Stimulus input

The stimuli are represented as patterns of activation over the input units of the network. In the model attention, in the form of the task-demand inputs, operates in the same qualitative manner as any other stimulus input. Each dimension of the input (colour, task/attention and word) has two nodes, which between them allow a localist representation of the two possible features in that dimension. For example, if the physical stimulus is the word ‘red’ written in black ink (the neutral colour) combined with the task-demand being to read the word then the forth and fifth units (numbering from left to right) will have an output clamped to 1, while the others will have an output clamped to 0. This is illustrated in Table 2A. Another possible input is illustrated in Table 2B.

A: Stimulus = **RED** Task = word reading

(word is 'red', ink is black. Control condition).

Colour Input		Task Demand Input		Word Input	
Red	Green	Word	Colour	Red	Green
0	0	0	1	1	0

B: Stimulus = **RED** Task = colour naming

(word is 'red', ink is green. Conflict condition).

Colour Input		Task Demand Input		Word Input	
Red	Green	Word	Colour	Red	Green
0	1	1	0	1	0

Table 2: Two illustrations of how stimuli are coded into inputs for the network.

2.1.3. Dynamics

Activation throughout the network is determined by a cascade mechanism (McClelland, 1979). This approximates to a discrete-time version of leaky-integration (see Gurney, 1997). Hence, with time, the stimulus information has an increasing influence on the activations of the output inputs, rather than having its full affect immediately. However, with respect to the differences in simulated reaction times between the different conditions, the effect of this is negligible. This is because the parameter determining the rate of leaky integration, τ, is very much smaller than the reaction times produced. Of more importance is the function of the model response mechanism (see section 2.2).

2.1.4. Response mechanism

The network response is determined by the outputs of the response layer units. Each output unit has a corresponding 'evidence accumulator', whose value increases based on the relative output of its corresponding response unit compared to the output of the other response unit(s). The model is considered to have given a response when the value of any evidence accumulator exceeds a fixed threshold (in this case 1). The network functions in discrete time, with the two evidence accumulators updated each cycle. To do this, the difference between each unit and it's nearest competitor is calculated (see Cohen et al., 1990, p. 338). For this network, which has only two possible responses, the two differences in output will each be the opposite of the other. For the simulations presented here noise (as used by Cohen et al., 1990) was omitted, since this did not affect the underlying mean reaction times produces. To convert from simulation cycles to simulated reaction time in milliseconds a linear regression is carried out on the empirical RTs provided by Dunbar & MacLeod (1984), shown in Figure 1. This maintains the proportions between the model response times, but provides human like absolute values.

In calculating response time based on relative evidence the Cohen et al (1990) response mechanism is similar to those based on mathematical models of decision processes (e.g. Luce, 1986; Reddi & Carpenter, 2000).

2.1.5. Training

The weights which determined the connections between units in the network were trained using the standard back-propagation algorithm with batch training (Rumelhart, Hinton, & Williams, 1986a). The network was trained only on the control patterns (i.e. no conflicting or congruent stimuli) and was also trained more on word patterns than colour patterns. This, of course, is supposed to mimic the human situation, where we read colour words far more often than we name the colour of colour words. The key feature of our learning experience is not merely greater experience with words than with colours, but greater experience with the pairing of word inputs with spoken outputs.

Rather than stop training when the error was had ceased to decrease significantly, the training regime was stopped when the network responded correctly to all experimental input patterns (i.e. the control, conflict and congruent stimuli for both word and colour naming) within a time limit (in this case 50 cycles). The weights from the attention units and the biases on the hidden units were fixed.

2.1.6. Operation

Before the stimulus information is presented to the network, the appropriate pattern is presented on the task-demand units and the network is allowed to settle into a stable output. Thus the operation of attention in the model is preparatory. Without input from the appropriate task-demand unit the signals through the ignored pathway are severely attenuated. However this attenuation affects the word and colour information differently.

The weights between units in a pathway determine the strength of the signal carried along that pathway. This strength of processing determines the evidence that is accumulated by the response mechanism and hence the response time. The weights in the two pathways are dependent on the training of the network and this is asymmetric for word and colour patterns. The training causes the weights in the word pathway to be of a greater magnitude than the weights in the colour pathway. When the task is to respond to the word, the signal attenuation, due to the lack of input from the attention units, prevents colour information signals influencing the output. However, even in the absence of attention, the superior strength of the word pathway word information has an influence on colour naming. This causes the asymmetry of interference between the two dimensions – i.e. that word information interferes with colour-naming but not vice-versa.

2.1.7. Key results

Figure 4 shows that the Cohen model, as replicated by myself, clearly matches the empirical findings with the basic Stroop task (as shown in Figure 1). Note that reaction times have been transformed from the simulation values, which are in arbitrary units, into millisecond values. The transformation maintains the same ratios between times, but makes them comparable to the human data. Details are given in Cohen et al (1990, p.340). All simulations reported in this thesis use the same method.

Figure 5 shows that the Cohen model results are similar to the results found with SOA methodology, as shown in Figure 2. Not, however, that for the colour-naming task the reaction times in the conflict and congruent conditions increasingly diverge as SOA becomes more negative. In contrast, in the empirical data, the reaction times for the conflict and congruent conditions seem to converge as SOA becomes more negative.

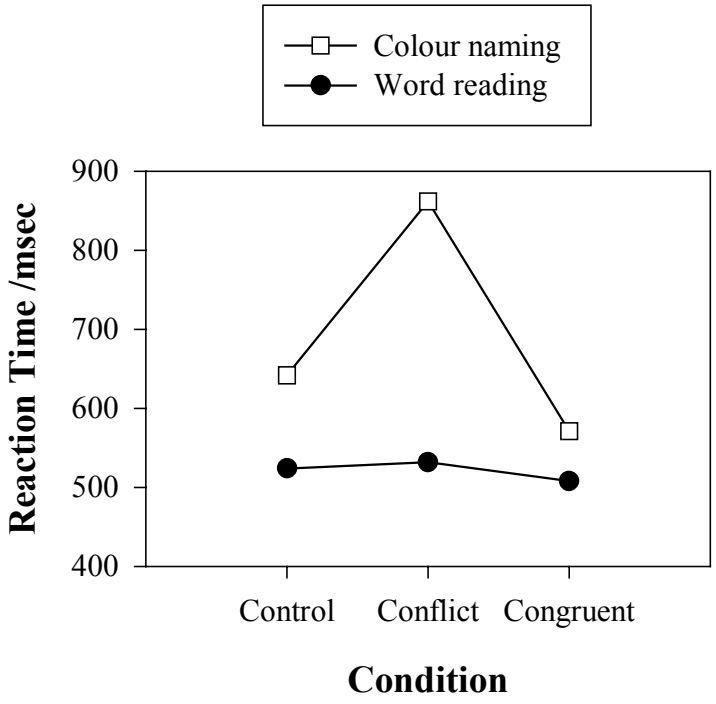


Figure 4: Simulation of fundamental Stroop conditions using my replication of the Cohen model.

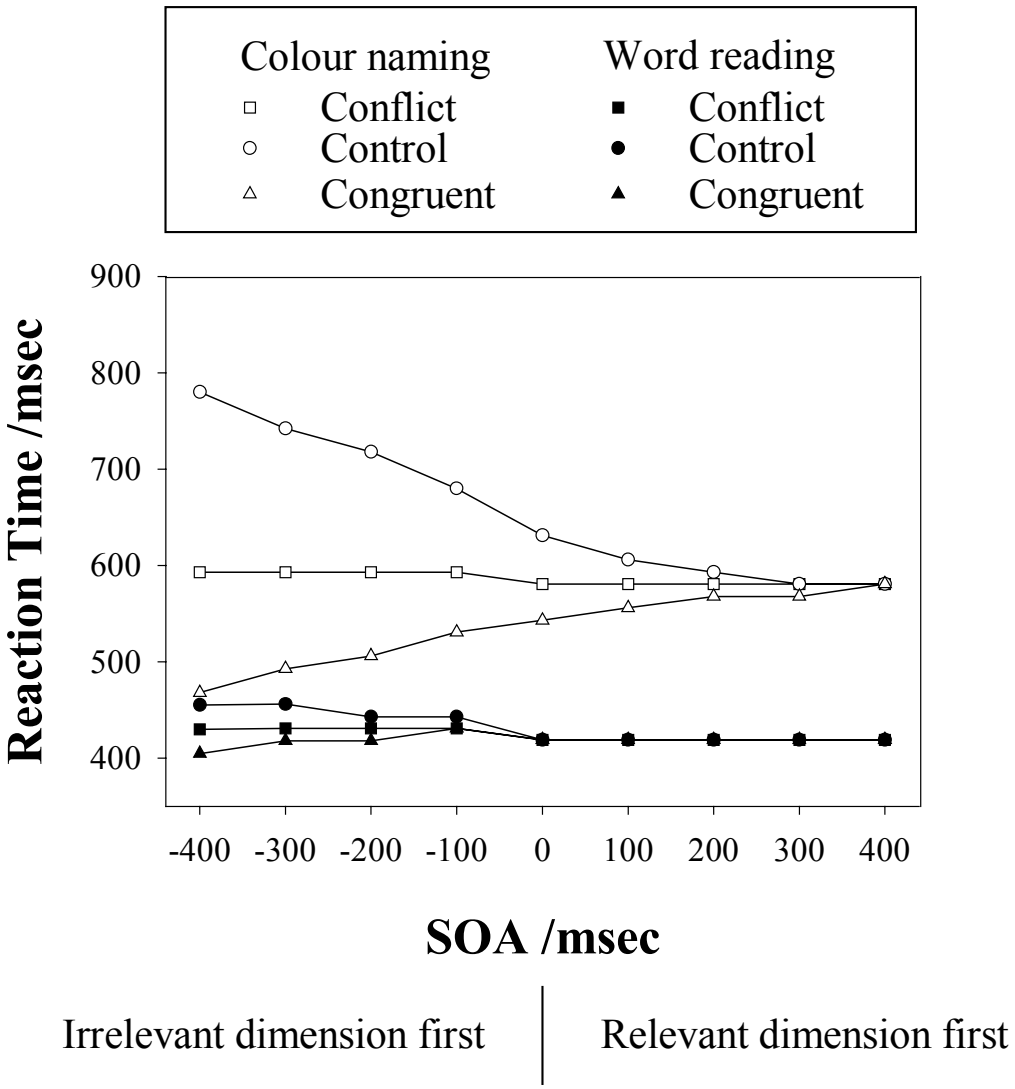


Figure 5: Simulation of SOA experiment using my replication of the Cohen model.

2.2. importance of the response mechanism

An important feature of the basic Stroop effect is that interference is greater than facilitation (MacLeod, 1991). Cohen et al (1990) simulate this and base their explanation of the model performance on the properties of the logistic function, which relates unit-input to unit-output. As Figure 6 shows, relative to a control-condition baseline which is greater than zero, increasing the input (as occurs for facilitation) does not produce as large an increase in output as the decrease in output produced by an equal but opposite decrease in input (as occurs for interference).

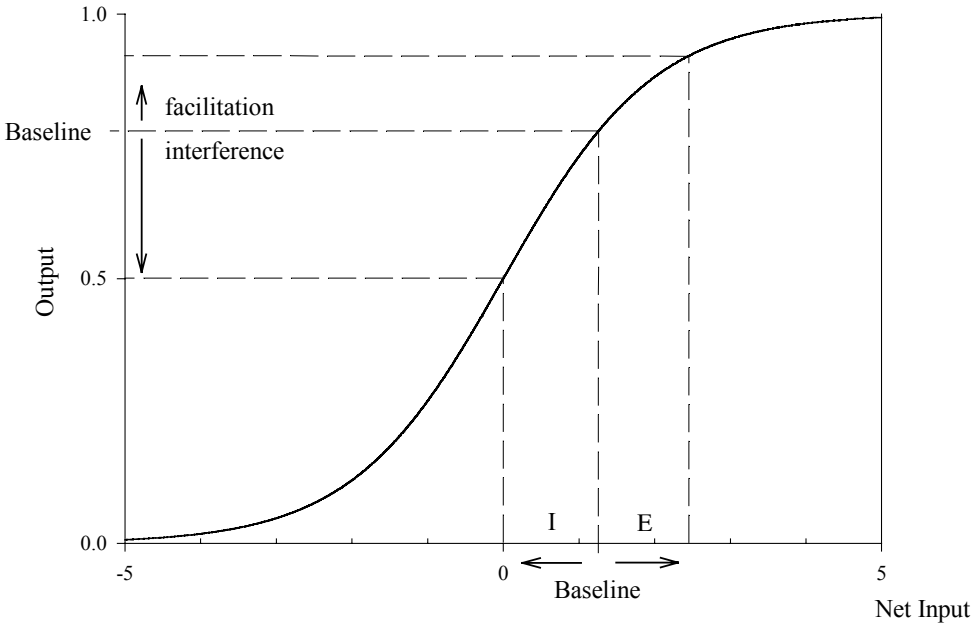


Figure 6: The logistic activation function, with annotation showing how excitation, 'E', and inhibition, 'I', of the baseline input affect output. The effect of equal excitation and inhibition is asymmetrical for the logistic function, the putative source of the difference between interference and facilitation.

This explanation is used to support the view that interference and facilitation are products of the same mechanism (Cohen et al., 1990; Cohen et al., 1992). This explanation also accompanies the exposition of the model in textbooks (e.g. Ellis & Humphreys, 1999, p252; Sharkey & Sharkey, 1995) and critical reviews (notably MacLeod, 1991). This 'single mechanism' explanation has, however, been criticised because of evidence which shows that interference and facilitation can be differentially affected by experimental manipulations (MacLeod, 1998; MacLeod & MacDonald, 2000). I show here that, while the explanation offered by Cohen et al (1990) requires revision, a single mechanism account *can* explain how interference is normally larger than facilitation, in a way which does not preclude the two measures varying in a seemingly independent way under other conditions.

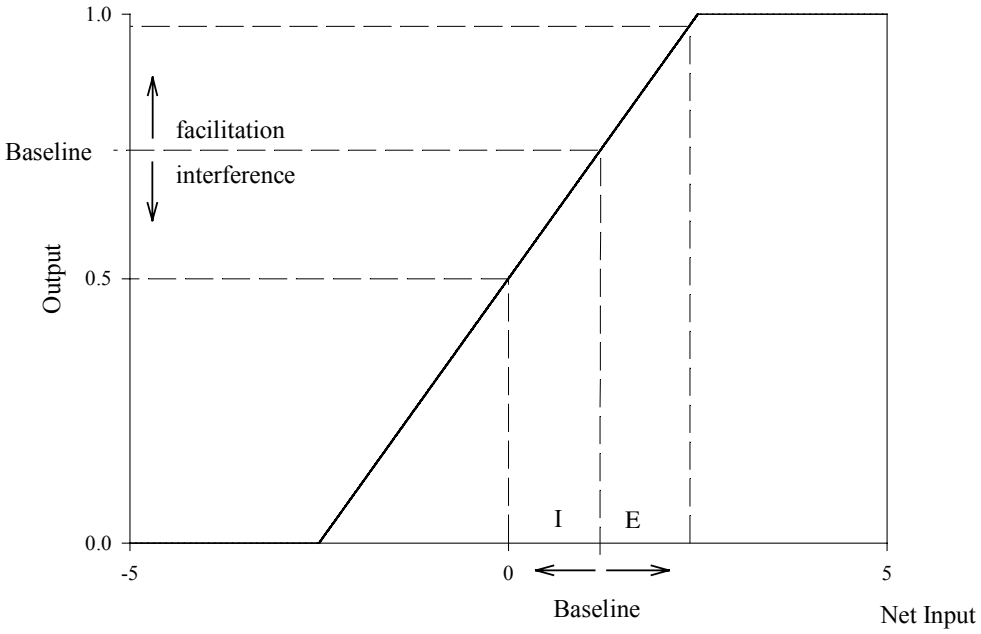


Figure 7: The piecewise linear activation function with annotation showing how excitation, 'E', and inhibition, 'I', of the baseline input affect output. The effect of equal excitation and inhibition is symmetrical.

I have found that the non-linearity of the activation function on its own cannot explain the difference between interference and facilitation in the model. This is best demonstrated by a replication of the original basic Stroop effect, but using a piecewise linear activation function instead of a logistic activation function (Figure 7). The rationale for this is that, since the slope of the piecewise linear function is constant, it ceases to be the case that increases in a unit's net input (associated with facilitation) result in smaller output changes than decreases in the net input (associated with interference). Despite this the model with the piecewise linear activation function still correctly simulates the basic Stroop effect, including the greater ratio of interference to facilitation (Figure 8). The similarity of the results from the model using the different activation functions shows that the decreasing slope of the logistic function cannot be the primary source of the difference between interference and facilitation. The equations defining the two activation functions are:

$$y = \frac{1}{1 + e^{-(x-\theta)/p}}$$

The logistic function

$$where\ \theta = 0\ and\ \rho = 1.$$

$$y = \begin{cases} 0 & : x < 0 \\ m\,(x - \varepsilon) & : otherwise \\ 1 & : x > 1 \end{cases}$$

The piecewise linear function

$$where\ m = 0.2\ and\ \varepsilon = -2.5$$

For both simulations the weights for the network after training, as given by Cohen et al (1990), were used.

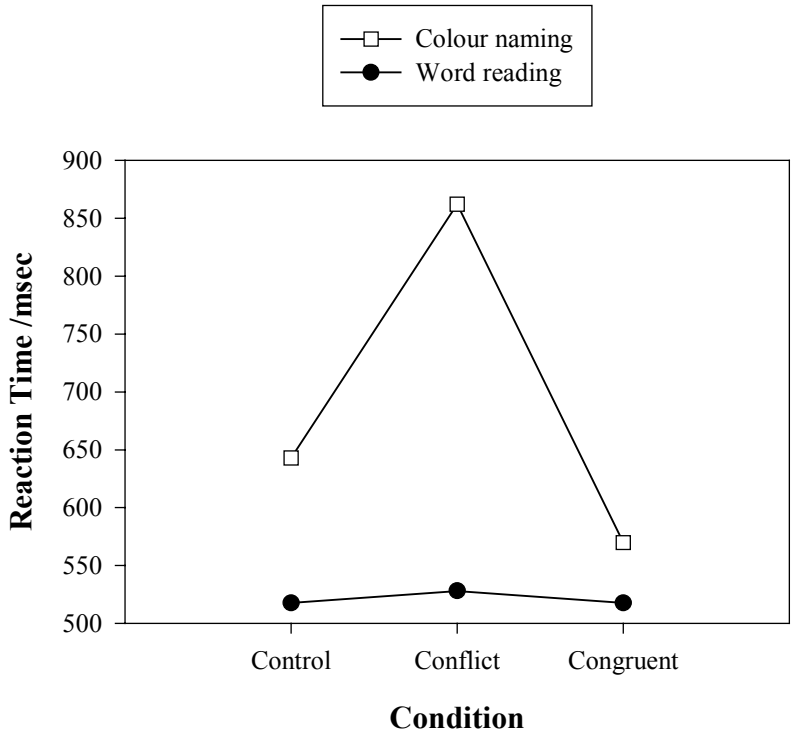


Figure 8: Reaction times from my replication of the original Cohen et al (1990) but using a piecewise linear activation function.

Notwithstanding these results, the argument given by Cohen et al (1990) is clearly *potentially* valid, and even with a piecewise linear activation function, it would be possible for interference to be greater than facilitation, if the upper-bound of the function is reached and the increase in input raises the output to the maximum. This saturation of the output function does not occur in the current model (see the Appendix, which shows the numerical values of the output nodes in my replications of the model), but it could occur in a similar model and/or in the corresponding neurobiological mechanism (although the possibility of the neurons involved reaching a firing-rate ceiling in the control condition of the task seems unlikely). However, having established that the difference between facilitation and interference is not accounted for in *this* model using the mechanism invoked by Cohen et al (1990) it remains to be seen what feature of the model *does* create this phenomenon.

The model correctly simulates the fact that interference is greater than facilitation, but clearly some additional factors must be involved. Cohen et al (1990) note that the time constant in the leaky integrator neuron plays a role here. However, removing the neuron temporal dynamics altogether has very little effect; the explanatory gap remains. To fill this explanatory gap, I have demonstrated that the response mechanism of the model is the main cause of interference being greater than facilitation. This mechanism is based on evidence accumulation. In a basic version of this scheme, each of the two possible decisions is associated with an evidence 'bin' and, at each time-step, each bin has its value altered by an amount proportional to the difference between the network output for its corresponding decision and that of the alternative decision. Thus, introducing decision indices i,j = 1,2, if μ_i is the change in evidence for decision i and y_i the associated network output then

$$\mu_i = \alpha\,(y_i - y_j)$$

where $i \neq j$ and α is a scaling parameter less than 1 which determines the rate of evidence accumulation. The counters are initialised to zero at the start of each trial

and a decision is signalled when the counter for either decision crosses some threshold. Cohen et al (1990, p.338) finesse this basic scheme by adding zero-mean gaussian noise to the evidence μ_i before accumulating it in each counter.

The role of the response mechanism may be elucidated by examining the functional relationship between the RT and the strength of evidence $E = y_i - y_j$, under the approximation that *E* is fixed for the duration of the response. The resulting function is shown in Figure 9, which shows that response time is a negatively accelerating function of the *E*. I now show how the form of this function allows a full explanation of the difference between interference and facilitation in Cohen et al's (1990) model.

As shown in Figure 9, increasing the relative strength of evidence above baseline for a decision does not speed the response time as much as an equally sized decrease slows response time. This is exactly what is required to explain the fact that interference is greater than facilitation in the model.

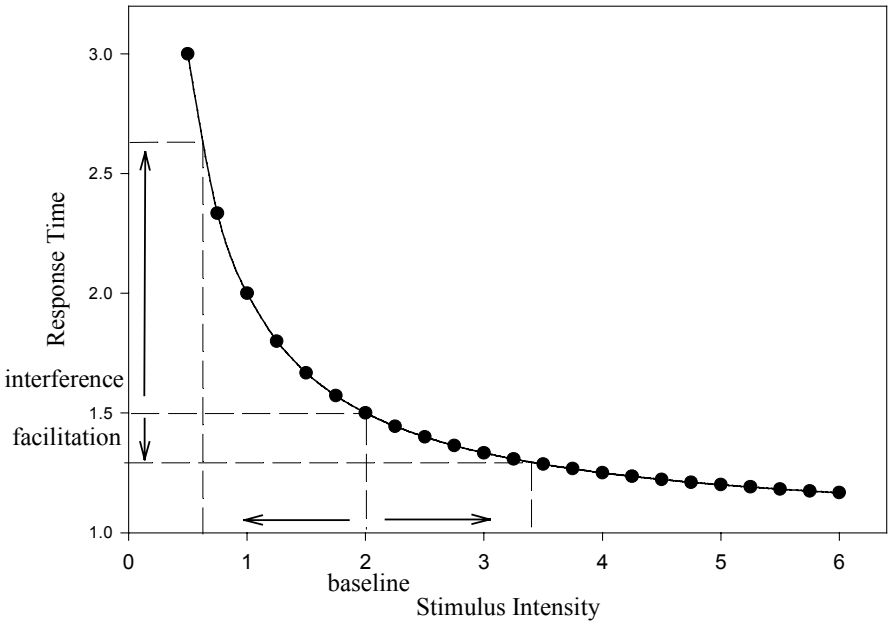


Figure 9: Response time as a function of strength of relative evidence in Cohen et al's response mechanism. The change due to increased intensity is greater that the change due to a decrease in intensity.

Further insight about this function may be obtained by quantifying its analytic form. Let $b_i(n)$ be the value of evidence bin i at timestep n. Without loss of generality, assume bin 1 forces a decision by reaching the threshold θ. At each time step, bin 1 is incremented by αE so that $b_1(n) = n\alpha E$. Let n_I be the smallest integer n such that $b_1(n) \geq \theta$ then, if αE is much less than θ (or equivalently, n is much greater than 1), $n_I \alpha E \approx \theta$. Rearranging and taking the log of both sides

$$\log n_I \approx \log(\theta / \alpha) - \log E \qquad (1)$$

Now, n_I is the analogue of a reaction time, RT, so that (1) may be written

$$\log RT \approx \log k - \log E \qquad (2)$$

where, $k = \theta / \alpha$. This is a special case of the more general form

$$\log(RT - R_0) \approx \log k - \beta \log E \quad (3)$$

where $\beta = 1$, and $R_0 = 0$. This, in turn, may be written as

$$RT \approx R_0 + k.E^{-\beta} \quad (4)$$

which expresses the reaction time as a decreasing function of the strength of evidence with an asymptotic response time R_0.

If strength of evidence is replaced by stimulus intensity then equation (4) corresponds to Pieron's Law (Pieron, 1914; Pieron, 1920; Pieron, 1952) which describes an early finding from psychophysics that intensity of a stimulus is related to the latency of response by an exponentially decaying function. Pieron's Law has been found to hold for both visual and auditory stimuli (reviewed in Luce, 1986), for gustatory reaction times (Bonnet, Zamora, Buratti, & Guirao, 1999) and for simple and choice reaction time tasks (Pins & Bonnet, 1996). From equation (3) the law may be expressed in an affine form (linear with non-zero offset) with slope -β and intercept log k. Straight line plots of this kind provide a convenient method of assessing the extent that other functions follow a form analogous Pieron's Law.

Such a plot for the Cohen et al (1990) response mechanism is shown in Figure 10. The asymptote for the data is found using a standard non-linear function fitting routine. Full details of this procedure are given in the appendix.

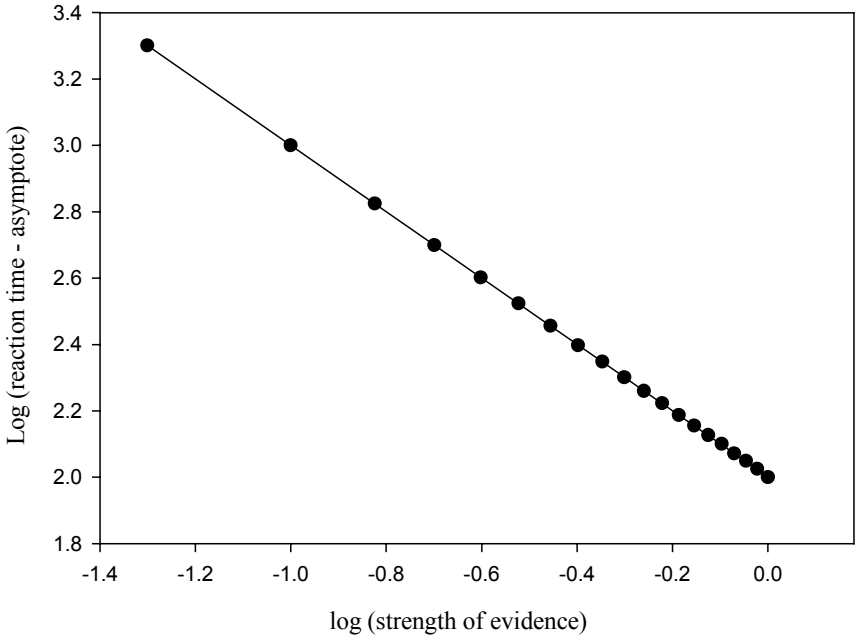


Figure 10: Log-log showing that the Cohen response mechanism follows Pieron's law.

The response mechanism does not have direct access to stimulus intensity, but instead deals with outputs from the connectionist 'front-end' of the model. Effectively, in this context, strength of evidence in the Stroop task can be said to be equivalent to the intensity for simple stimuli.

Two issues arise from the reconsideration of Cohen et al's (1990) original explanation of the relative magnitudes of interference and facilitation. Firstly, the activation function explanation has been used to suggest that interference and facilitation might be caused by the same mechanism. This has been strongly criticised (MacLeod & MacDonald, 2000), on the grounds that task manipulation can affect interference, without seeming to affect facilitation. This does not, however, invalidate 'single mechanism' explanations. The non-linearity of the response mechanism, in conjunction with equal increases and decreases in the relative strengths of evidence provides an alternative parsimonious explanation for the relative size of the interference and facilitation effects in the standard model simulation. There is no principled reason, however, why the increase in evidence in the conflict condition and the corresponding decrease in the congruent condition cannot vary independently. It is easy to imagine that, in a system more complex than the original Cohen model, the conflict and congruent conditions need not produce equal but opposite changes in the 'evidence' supporting a response.

Additionally, because facilitation relies on change over the least rapidly changing part of the activation function, tasks which evoke changes in interference might not be associated with significant changes in facilitation. Facilitation is small relative to absolute response times and relative to the amount of noise in empirical measurements. Therefore changes in the amount of facilitation may often be too small to detect reliably. It has not been shown experimentally that the two phenomena vary independently, only that interference can decrease without significantly decreasing measured facilitation (e.g. MacLeod, 1998). To truly demonstrate the independence of the two phenomena it would be necessary to show an increase in facilitation without a corresponding increase in interference.

Secondly and more importantly, this shift in locus of explanation emphasises the importance of response mechanisms for cognitive tasks. All PDP models of Stroop processing explicitly or implicitly contain a model of action selection. A mechanism is needed to compare actions and switch between selected actions, even if it is only mediating between two simple choices. Existing models of the Stroop task use artificial models of response selection determined solely by their mathematical form. The later chapters of this thesis consider the importance of using biologically plausible response mechanisms in the Stroop task.

2.3. Learning & representation

2.3.1. Mutual interference

Cohen et al (1990) report empirical data of MacLeod & Dunbar (1988) from training on a new task; that of naming shapes with colour names. They show that shape-naming is initially interfered with by colour-naming, but this effect reverses with training, passing though a point where both processes interfere with the performance of the other. The Cohen model simulates the two single-process interference conditions, but appears to fail to simulate the mutual interference phenomenon.

However this failure to simulate mutual interference is not a necessary aspect of the model. Rather it is incidental to the particular form of the simulation they ran, a result of the time during training at which they assessed the model outputs. All that is required for a mutual interference effect is for the weights for both shape-naming and colour-naming to be strong enough in relation to the attentional inhibition that comes from being ignored. Do the weights of the network pass through this part of weight space during training, as described by Cohen et al (1990, simulation 4)?

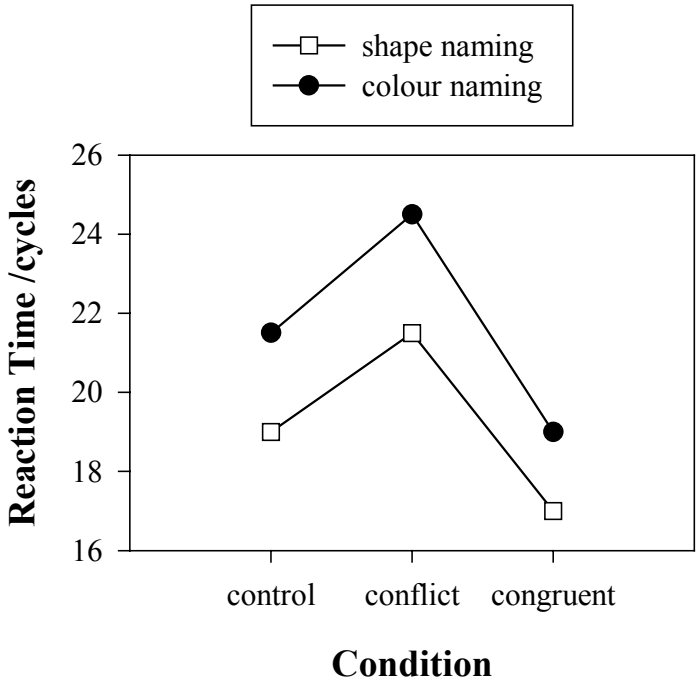


Figure 11: Mutual interference simulated by the Cohen model.

My replication of Cohen et al (1990, simulation 4) produced a set of weights, at 250 epochs which showed mutual interference of the two processes, as shown in Figure 11. This shows, as supposed, that the model is capable of producing the mutual interference phenomenon.

Cohen et al (1990) look for mutual interference after 504 training epochs. They arrive at this number after assuming a one-to-one correspondence between a human training trial and a model learning epoch. Although this ratio (1:1) is possible, it isn't the only conceivable ratio, given that the back-propagation algorithm is used primarily because of its efficacy of arriving at a functional set of weights (what it does), rather than for its functional correspondence to human learning (how it does it). The speed with which the algorithm functions, and in other words the point in training at which mutual interference will occur, depends on the parameters specified for the algorithm. These parameters, the learning rate and momentum, do not have direct human equivalents and, for the Cohen model, no approximate values have been established by comparison with human learning data. Indeed, it could be argued that the mutual interference simulations are a valid forum within which to experiment with different parameters if it was desirable to establish which back-propagation parameters brought the best correspondence with human learning in this context. It can be assumed that, via parameter manipulation, we could get the mutual interference phenomenon to manifest at 504 epochs. However, theoretically, this would be uninteresting. It is enough to show that the phenomenon can exist within the model framework. Further simulations, although perhaps producing a 'better-fit' would only be producing a better fit with (in this context) essentially arbitrary parameters.

2.3.2. Modularity, attention, training set interactions

The training of the model involves only the control condition stimuli, but the model's responses are tested for all possible conditions; control, conflict and congruent. As discussed, the model provides the correct responses for all possible

inputs, despite the restricted training set. This is possible because of the interaction of the training set with the pre-specified architecture of the network.

To explore the way the network architecture and initial weights allow training with a restricted training set I trained several variations of the original Cohen model. The following items were varied, all in combination with each other:

- the starting weights from the input to the hidden units.
- the division of the network into two pathways, with no weights crossing between the two until the output stage.
- the fixity of the biases and the weights from the task-demand units.

These three items are the model formulations of the following items, respectively:

- the initial internal representation of the inputs.
- the modularisation of word and colour information in separate pathways.
- the pre-specified attentional gating mechanism.

Setting the starting weights from the input units, and thus removing the clear initial representation of the possible inputs on the hidden units, does not greatly affect learning in the network. Without the starting weights used by Cohen et al (1990) the network takes longer to learn and, in a small minority of cases, becomes stuck in a local minimum and does not complete training correctly.

Out of the modularisation and the attentional gating mechanism, only the attentional gating mechanism was found to be crucial to the correct learning of the full output set while only using a restricted training set. If the weights from the task-demand units and the bias unit are not pre-specified the network does not learn the correct response for all inputs. The preponderance of word-examples in training causes word information to override colour information in the colour-naming conflict condition, thus eliciting an erroneous response.

It seems that the influence of the attentional gating mechanism is so strong that even without enforced modularisation the network is still effectively divided into two pathways. The weights that develop from the word information inputs to the colour pathway, and vice-versa, are negligible; the inhibition that results without input from the task demand units effectively stops any significant activity in the 'ignored' pathway caused by the cross-pathway weights. Conversely, without the attentional gating mechanism, the modularisation of the network is insufficient to overcome the undue influence of word information in the conflict condition.

Unsurprisingly if both the gating mechanism and modularisation are removed the network fails to learn the correct responses for all conditions. Like the modularisation without attention variant, it fails to learn the correct responses in the colour-naming conflict condition. It is interesting to note that training on the full training set is enough to overcome the lack of attentional gate, with or without the presence of modularity. This shows that the variants of the network which fail to complete training correctly with the restricted training set can, with the right conditions – i.e. the full training set – learn the correct responses. Essentially there is a trade off between network pre-specification and training set. For a restricted training set, more pre-specification, in the form of an attentional gate in this case, is required. For a full training set less pre-specification is required.

This specific example is a nice illustration of the principle expounded by Clark & Thornton (1997). They discuss the trade-off between representational naïveté and computational power required to solve a problem. Pre-training biases in representation can help solve problems in which vital regularities in the training data are attenuated. In the Stroop task model, the vital information is the input from the task-demand units, which, although irrelevant in the control conditions which comprise the training set, is vital in the conflict conditions which are present in the test set. Without pre-specified attentional weights the network fails to learn the importance of these inputs, and hence cannot learn the correct response to all the test patterns.

Pre-specified modularity, as discussed by Jacobs, Jordan & Barto (1991), is a way of effectively reducing task complexity. In this case modularity allows the extraction of vital, but attenuated regularities, from training data. Without modularity the preponderance of word training examples causes word information to inappropriately dominate some responses. It could be that a bias towards modular architectures is enough to allow the learning of problematic tasks, such as the Stroop task, without it being necessary to pre-specify what the modules are (Elman, 1994; Jacobs & Jordan, 1992).

For this model and problem, the solution to the task problem resides neither in the data, nor entirely in the architecture. It is not explicit in the training patterns what the correct responses to all the test patterns should be. Nor does the modular architecture of the network fully constrain what the responses to all the patterns should be either. Only in combination does the solution emerge, from an interaction between the network architecture ('nature') and the training set ('nurture'). This sort of 'rethought innateness' is discussed at length by Elman and colleagues (1996).

2.3.3. Hidden unit representation

The internal representation of the inputs is highly constrained in the original Cohen model by the starting weights of the network. As mentioned above, without these starting weights the network takes longer to learn the correct outputs and a small minority – about one in twenty – of the random starting weights lead to the network becoming stuck in a local minimum and failing to successfully complete training.

However those networks with small random starting weights which do complete training arrive at the same representation of the inputs on the hidden units as the networks with the more constrained starting weights. Specifically, each hidden unit comes to represent one of the four possible colour-task combinations, i.e. a single unit responds maximally to one of red-word, red-colour, green-word, green-colour. Thus the nature of representation in the network is localist, rather than distributed.

To find out if this is necessarily the case an additional four hidden units were added to the network so that it would be conceivable for a distributed representation to develop. With small random starting weights from the input units, the representation is free to develop within the possible bounds determined by the back-propagation algorithm. The network with an expanded hidden layer trained successfully, and indeed produced the basic Stroop effect comparably to the original Cohen model.

By inspecting the weights which developed between the input units and the hidden units it is possible to uncover the internal representation of the network. On inspection of these it is obvious that the network still develops an essentially localist representation of the inputs, despite the fact that a localist representations are a minority possibility out of the set of representations that are conceivable. Plots of the hidden unit response to possible inputs reveal that each input causes a large response in two, and only two, hidden units (Figure 12). Each hidden unit responds to only one input, meaning that the interpretation of any one hidden unit activation does not require reference to the activation of the other units, so we can say that the representation is not distributed, in the technical sense, although it is redundant (i.e. covering more that one unit). The back-propagation algorithm guides the network into exploiting the representational redundancy provided by the expanded hidden layer, although it settles on a simple localist representation. This is probably because the nature of the input coding and output coding, i.e. localist input and output representation, which restricts which possible internal representations will be efficacious and can be easily found by the back-propagation algorithm. The pattern of reaction times is the same for networks with eight hidden nodes as for the standard network with four hidden nodes. Adding extra nodes turns out to be an implementational detail rather than one that affects on the results or theory.

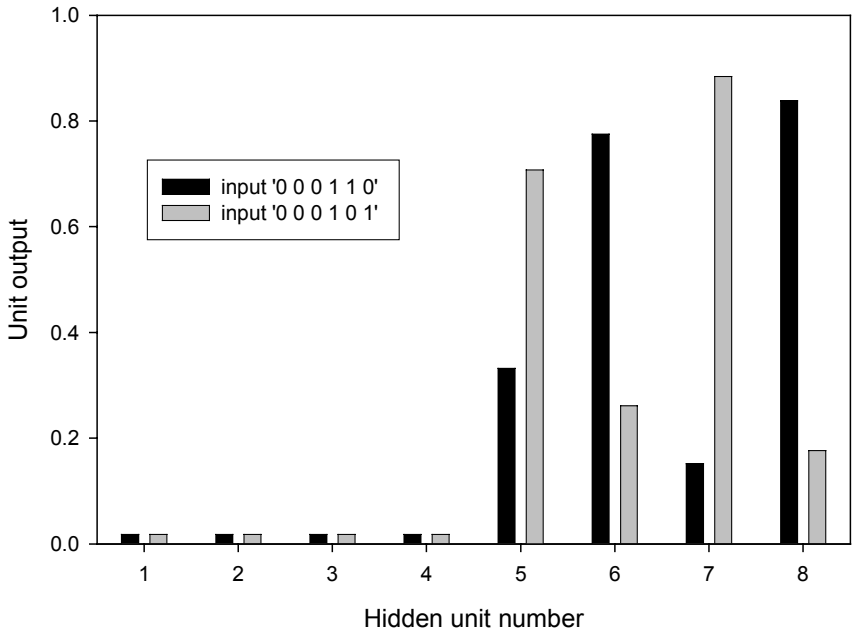


Figure 12: Plot of outputs of hidden units in the Cohen network with an expanded hidden layer. Output across all eight hidden units is shown for two input patterns.

2.4. A critical evaluation

2.4.1. An accepted standard

The Cohen model has become an accepted standard, both of Stroop models and as the prime exemplar of a general class of models of Stroop-like processing (Ellis & Humphreys, 1999; Lu & Proctor, 2001; Monsell, Taylor, & Murphy, 2001). It has received much attention, in the form of extensions (Cohen, Botvinick, & Carter, 2000; Cohen, Braver, & Oreilly, 1996; Cohen & Huston, 1994; Cohen & Servan-Schreiber, 1992; Cohen et al., 1992; Cohen & Usher, 1996; Cohen, Usher, & McClelland, 1998), and criticisms (Kanne, Balota, Spieler, & Faust, 1998; Mewhort, Braun, & Heathcote, 1992; Schooler et al., 1997; Tzelgov, Henik, & Berger, 1992). The Cohen model as also influenced other researchers, being used or imitated by them (Dejong, Liang, & Lauber, 1994; Dixon, Brunet, & Laurence, 1990; Matthews & Harley, 1996; Sugg & McDonald, 1994; Wiles, Chenery, Hallinan, Blair, & Naumann, 2000; Williams, Mathews, & MacLeod, 1996).

2.4.2. Inadequacies

Despite it's popularity, the Cohen model suffers from a number of inadequacies. Notably, not all the reported simulations in the original paper match well with the corresponding empirical results (Cohen et al., 1990, simulations 2 & 4). These, and other inadequacies are discussed below.

Simulation 4 of Cohen et al (1990) concerns the development of automaticity. MacLeod & Dunbar (1988) investigated the development of automaticity with a shape-naming task, in which various shapes were given colour names, so that a shape-colour Stroop could be performed (cf. the conventional word-colour Stroop). They showed that at a point in the development of automaticity of shape-naming, shape-information and colour-information both interfere with the performance of the converse task. As Cohen et al (1990) acknowledge, their model fails to replicate

this. As demonstrated in section 2.3.1, there is no principled reason why such results could not be accommodated with the framework of the model. The fact that the number of human learning trials does not translate one-to-one into model training epochs is no cause for concern, given that the back-propagation algorithm being an analogue to human learning mechanisms is not an assumption of the model. So, although Cohen et al (1990) do not report the successful simulation of this phenomenon, it should be regarded as falling within the explanatory range of the model, or of the family of related possible models.

Although the Cohen model accurately simulates mean reaction times in the different conditions of the Stroop task, it does not reflect the change in the shape of the underlying distributions across the congruent, control and conflict conditions of the task (Mewhort et al., 1992). Mewhort et al (1992) observed that the distribution of response latencies was more skewed in the congruent and conflict conditions than in the control conditions. The Cohen model predicts more skew in the control conditions than the congruent condition, and more skew in the conflict condition than in the control condition. They conclude that this data suggests that the Cohen model is an inadequate model of Stroop processing. I discuss the strength of this criticism, and how it might be addressed in the context of the model of Stroop processing in conjunction with the basal ganglia model, in section 3.6.

Simulation 2 of Cohen et al (1990) concerns stimulus-onset-asynchrony (SOA) effects. The Cohen et al (1990) simulation (Figure 5) roughly matches the original data (Figure 1), although in the colour-naming conflict and congruent conditions the reaction times at negative SOAs fail to stabilise as they appear to in the empirical data. When, however the simulation was replicated with an extended SOA range it became clear that the colour-naming conflict and congruent condition reaction times *never* stabilise (Figure 13).

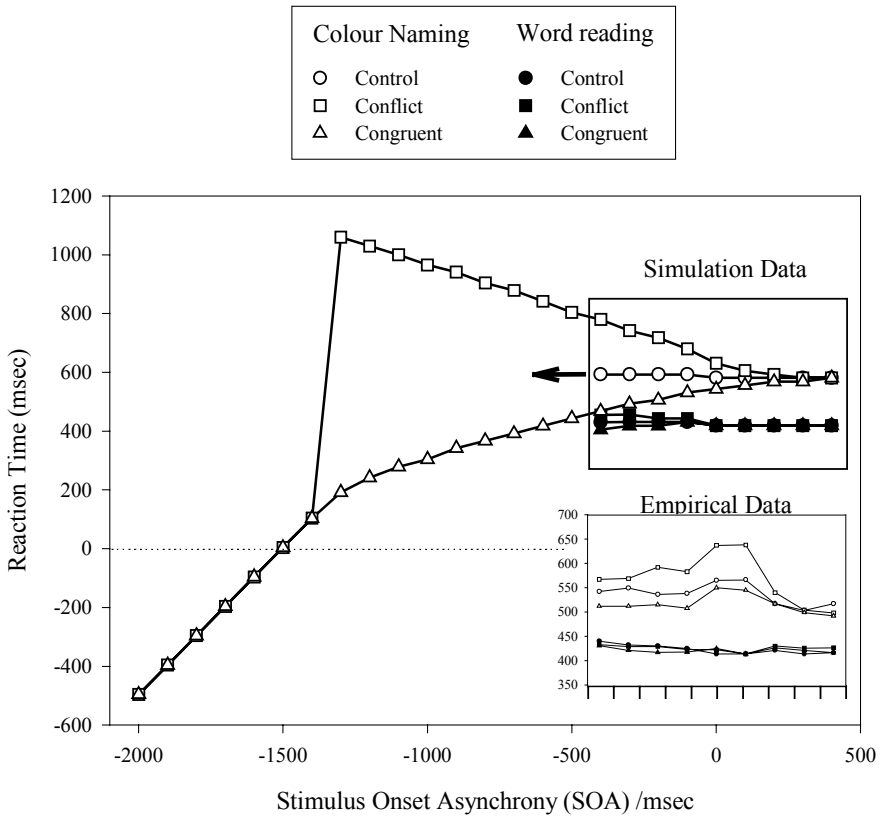


Figure 13: SOA effects with the original Cohen model, and with extended range of SOA values. The range of results reported in the original paper is shown by the box. The empirical data are also shown inset below the simulation results. The simulation results are not shown on the full range for the colour naming control condition and the word reading conditions; reaction times for these conditions remain approximately constant for the range shown.

Beyond –1320 ms SOA the original model begins to make the wrong colour naming responses in the conflict condition. This is because, while the word dimension is presented and the colour information is not, the response mechanism accumulates evidence for the response compatible with the word dimension; attentional control is not strong enough to prevent *some* activation associated with the word information. Indeed, if it were there would be no Stroop interference in the 0 ms SOA condition. After sufficient time the response mechanism accumulates enough evidence to signal a response, even though the weight of evidence at each time step is very small. The converse problem is found for the congruent condition of the colour-naming task. In this case the word dimension does not cause a wrong response, because the word dimension prompts the same response as the colour dimension. However, the response mechanism is driven by signals based on the input from the word dimension; at an extreme enough negative SOA the reaction time falls below zero – i.e. the response comes before the colour dimension has even appeared. Although not shown here, the same pattern occurs for the word reading condition, although at a much greater SOA range, because the difference between the two inputs before the relevant dimension is present is very much smaller than for the colour naming task.

Indeed, the Cohen model makes erroneous selections even on the colour-naming task within the –400 ms to +400 ms SOA range if the same parameters are used as Cohen et al (1990) used in the simulation of the basic Stroop effect (simulation 1). For the SOA simulation they increase the resting bias on the hidden units from –4.0 to –4.9. Without this increase in bias wrong selection occurs at –400 ms SOA for the colour-naming conflict condition.

These extended SOA results reveal two essential flaws of the response mechanism, and of evidence accumulation mechanisms in general. Firstly, they accumulate evidence for a response even if that evidence is insignificantly small; given long enough, the evidence level crosses the response threshold, even when the amount of evidence per time step is not commensurate with any action being taken. Secondly,

the evidence accumulation mechanism has trouble dealing with switching between actions. Evidence accumulates progressively until a single action is selected; this hinders the sequential selection of actions or the interruption of actions. In the colour conflict condition, as evidence is accumulated for the response to the irrelevant dimension, it takes progressively longer for the response to the relevant dimension to interrupt. Suppressed responses affect the time to select the correct response - the mechanism does not possess the property of clean switching (see section 1.2.1). The flat shape of the empirical SOA data shows that the human response mechanism does possess this clean switching property, at least when the two stimulus dimensions are separated by more than 100 ms in time.

2.4.3. Attentional mechanisms and fMRI findings

The attention scheme in the Cohen model has a number of important properties. As well as being essential to the correct function of the model (as discussed above), the attentional mechanism is notable in that attention functions as an input, qualitatively the same as any other. Casting attention within a connectionist framework allows a bridge to be made between computational studies and the investigation of attention in neural tissue.

Functional imaging studies have revealed some important correspondences between the neural correlates of attention and the operation of attention in the model. Directing attention to part of a visual scene changes the baseline activity of striate and extrastriate regions involved in processing target information (Brefczynski & DeYoe, 1999; Chawla, Rees, & Friston, 1999; Gandhi, Heeger, & Boynton, 1999; Heinze et al., 1994; Kastner, DeWeerd, Desimone, & Ungerleider, 1998). This modulation of activity is truly due to mental intention, rather than the direction of gaze or correlated visual stimulation (Kastner, Pinsk, DeWeerd, Desimone, & Ungerleider, 1999). It can also occur in regions associated with auditory sensation (Jancke, Mirzazade, & Shah, 1999). Modulation of activity, independently of input, which then biases competition between representations (Duncan, 1998) is exactly how the Cohen model operates, a fine example of a modelling solution to a problem

being subsequently observed empirically. It is thought that gain control, as well as additive bias modulation, plays a role in attention (Hillyard, Vogel, & Luck, 1998; Rees & Frith, 1998). This attentional control mechanism is not included in the model, although this class of models may prove extremely fruitful for investigating the control of attention via multiple processes in distributed networks.

Attempts to model the SOA data have shown that preparatory attention alone is not sufficient to prevent erroneous selection. Within the existing framework it is necessary that word information is strong enough to break through the attentional bias to some extent. If this were not so then no interference would result in the conflict condition. However, this fact means that at long negative SOAs (where the irrelevant dimension of the stimulus is presented first) the existence of word information provokes activity which is strong enough (in conjunction with an evidence accumulation response mechanism), even in the absence of attention, to cause selection. Some kind of secondary attentional control is required to explain the correct selection of responses at long negative SOAs, even though conflicting stimuli still cause interference.

2.4.4. Recent advances

Recently Cohen and his colleagues (Botvinick et al., 2001; Cohen et al., 2000) have suggested a reformulation of their basic network, neuro-anatomically locating the various components and augmenting it with a conflict monitoring module, putatively representing the anterior cingulate cortex (ACC). This helps relate the modelling work to imaging and other investigations in cognitive neuroscience. However, one of the advantages of the original Cohen model was its simplicity and, hence, its capacity to represent a generic network for the processing of conflicting stimuli. The task of the neuro-anatomical grounding of the model, combined with the exploration of the generic traits of stimulus and response conflict, is a considerable one. Cohen and colleagues have made a beginning with these recent publications, and the work presented in this thesis also addresses this.

3. THE BASAL GANGLIA MODEL AS A RESPONSE MECHANISM

The basal ganglia model is based the hypothesis that the BG perform action selection (section 1.2.2). Two key points regarding the model are, firstly, that it uses model neurons essentially the same as those used in other connectionist models (including the Cohen model) and, secondly, that it is faithful to a systems level analysis of the functional anatomy of the regions it purports to represent.

3.1. Connectivity

We (the ABRG) have proposed that the basal ganglia complex is the 'central switch' that mediates action selection in all vertebrate species (Prescott et al., 1999; Redgrave et al., 1999). We have also formulated a computational model of the basal ganglia based on this hypothesis (Gurney et al., 2001a; Gurney et al., 2001b). Further work has incorporated the thalamus and cortex into a complete functional sensorimotor loop (Humphries & Gurney, 2002; Montes-Gonzalez et al., 2000). These studies show how the basal ganglia architecture is suited to its operation as a central switching device, satisfying the desirable properties of such a switching device outlined in the previous section.

Anatomically, the basal ganglia are well placed to fill this role, receiving inputs from virtually the entire cerebral cortex, limbic system structures such as the hippocampus and the amygdala, and, notably, the anterior cingulate cortex (Masterman & Cummings, 1997; Redgrave et al., 1999). It has been well established that within the basal ganglia there exist parallel functional loops which deal with different competencies; motor, oculomotor, prefrontal, associative and limbic loops (Alexander, Delong, & Strick, 1986). Within these loops signals from the basal ganglia are relayed back to the cortex via the thalamic nuclei, which produces feedback to modulate activity in the cortex, and hence indirectly modulates subsequent basal ganglia inputs. Within each loop we posit a further parallel subdivision into separate *channels* (Redgrave et al., 1999), with each channel

representing an action for a particular motor region. The basal ganglia perform competitive selection on the channel inputs within these loops.

The output nuclei of the basal ganglia are the globus pallidus internal segment (GPi) and substantia nigra pars reticulata (SNr). These nuclei tonically inhibit their target (motor) outputs, so that, in effect, all currently active requests for motor action are ready to be expressed but 'held back' (rather than sufficient input to activate them needing to be initiated from cold, as it were). Basal ganglia inputs can, via internal selection mechanisms, selectively inhibit GPi, which has the effect of turning off the tonic inhibition and thus releasing the selected motor output.

3.2. BG model functionality

This section outlines how the basal ganglia model (see Figure 14) performs its selection function. This sketch is included for the sake of completeness and the full details of how the basal ganglia model performs action selection are presented elsewhere (Gurney et al., 2001a; Gurney et al., 2001b; Humphries & Gurney, 2002).

Salience information arrives from cortex; that is, in our model we presume that the cortex provides the results of stimulus processing to the BG in a form suitable for evoking a response. The selection by the basal ganglia is indicated when the output for a particular channel in GPi drops to zero.

The main input nucleus of the basal ganglia is the striatum, which provides the first processing of incoming saliences (see section 1.2.1). The projection neurons of the striatum (medium spiny neurons) are by default quiescent (in a 'down-state'), and do not do not respond to low levels of input. Only after a substantial and coordinated excitatory input do they move to an 'up-state', in which they produce significant output which may subsequently be affected by smaller changes in input (Wilson, 1995). In the model this property is interpreted as a minimal input threshold, below which inputs are simply filtered out. This has important

consequences for the operation of the basal ganglia as a response mechanism, since it means that small magnitude saliences, such as those caused by noise or by ignored stimuli, cannot cause selection.

Two types of striatal projection cells are distinguished in the model, those with D1-type dopamine receptors and those with D2-type dopamine receptors. Dopamine has an excitatory effect upon D1 receptors and an inhibitory effect upon D2 receptors[3]. The different types of cells also have different projection targets; D1 neurons project to GPi and D2 neurons project to globus pallidus external segment (GPe.).

The subthalamic nucleus (STN) also receives excitatory cortical input from diverse cortical sources. The output from the STN provides diffuse ('on-surround') excitation to GPi, the output nucleus of the model, and GPe.

Because connections from striatum to GPi are inhibitory, higher saliences will tend to inhibit (i.e. select) their corresponding channels in GPi. In contrast, the diffuse projections from STN increase the output level on non-selected channels. The other pathways in the basal ganglia modulate the selection function of this core cortical-striatal-GPi pathway. For example, the GPe-STN negative feedback loop limits the overall level of activity.

[3] The level of dopamine within the model can be adjusted, but no simulations reported here manipulate this parameter.

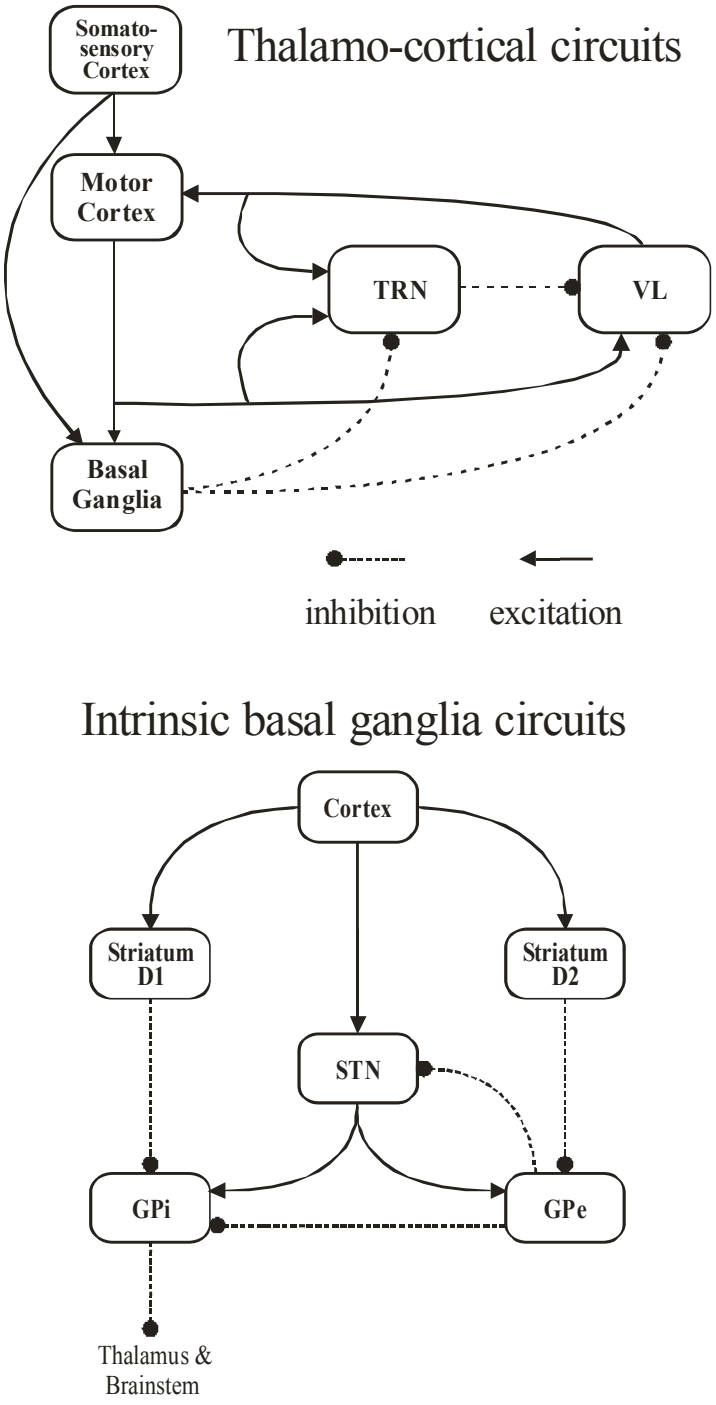


Figure 14: The architecture of the basal ganglia model (after Humphries & Gurney, 2002, figures 2 and 3). See text for explanation of abbreviations.

The action selection function of the basal ganglia alone is augmented by the thalamocortical loop comprising the ventrolateral thalamus (VL), the motor cortex and the thalamic reticular neucleus (TRN). This circuit acts as a gated positive feedback loop. The loop between motor cortex and the thalamic complex (via basal ganglia) amplifies cortical signals which reach above a threshold. Outputs from the basal ganglia are used to modulate the positive feedback loop so that only the signals of a single selected channel are amplified. Amongst other selection benefits, this enables selection and effective switching between channels over a wider range of salience values. In particular we make the hypothesis that the TRN functions as part of a 'clean-up' circuit, sharpening selection signals (Humphries & Gurney, 2002).

3.3. Functional role

The basal ganglia has often been seen as a motor area, although there is increasing evidence of the basal ganglia's involvement in a range of cognitive functions (Doya, 2000); including attention (Hayes, Davidson, Keele, & Rafal, 1998; Jackson & Houghton, 1994; Koski, Paus, Hofle, & Petrides, 1999), learning (Graybiel, 1998; Jog, Kubota, Connolly, Hillegaart, & Graybiel, 1999), memory (Beiser & Houk, 1998; Levy, Friedman, Davachi, & GoldmanRakic, 1997) and executive functions (Brown, Schneider, & Lidsky, 1997). The basal ganglia's connections to the frontal cortex (Wise, Murray, & Gerfen, 1996) make it difficult to ascribe to them an independent functional role separate from the frontal cortex, but this does serve to underline the importance of the basal ganglia to higher cognitive processes that traditionally have been thought to be mediated by the frontal cortex.

Involvement of the basal ganglia with multiple and varied functions is consistent with the basal ganglia's proposed role a central switch which gates access to motor resources (in the case of the skeleto/oculomotor loop) and modulates the expression of activity in other areas of cortex. The hypothesis that the basal ganglia performs a switching function is supported by the motor, sensory and cognitive difficulties of those with Parkinson's Disease (Brown et al., 1997; Holthoff-Detto et al., 1997;

Inzelberg et al., 1996; Mink, 1996), which involves a loss of dopaminergic innervation to the basal ganglia. Disorders of basal ganglia switching functions may be involved with a number of neuropsychological dysfunctions; Tourette's syndrome (Brito, 1997; Mink, 2001), cognitive disorder in Schizophrenia (Calabresi, DeMurtas, & Bernardi, 1997), and Huntington's Chorea (Crossman, 1987).

Furthermore, the basal ganglia's motor functions may be closely related to the attentional difficulties inherent in the Stroop task. Houghton, Tipper, Weaver and Shore (1996) proposed a model of selective attention, which they explicitly relate to action selection. The attentional control in the model is generated by a match-mismatch unit, which compares object inputs with targets. The basal ganglia are a suggested location for this match-mismatch unit (Jackson & Houghton, 1994). Brunia (1999) discusses the functional similarities between anticipatory attention and motor preparation, and their common anatomical substrate - frontal-thalamic loops of the sort involved in our model of basal ganglia processing. These loops can provide feedback from potential motor outputs to cortical signals, perhaps implementing the kind of dynamic control necessary in motor selection.

3.3.1 The basal ganglia and automaticity

The basal ganglia have been implicated in the learning of automatic responses (Knowlton, Mangels, & Squire, 1996; Salmon & Butters, 1995; White, 1997). Graybiel (1998) proposed that the striatum of the basal ganglia is involved the recoding of motor and cognitive sequences in to performance units that can then be implemented as single 'chunks'. This, 'chunking', hypothesis is supported, for example, by experimental work with Parkinson's Disease patients (Doyon et al., 1998), which shows that this patient group has trouble with the automatisation and retention of motor sequences. Positron Emission Tomography (PET) of the brains of a non-clinical sample during a procedural learning task shows significant basal ganglia activation (Ghilardi et al., 2000). Single cell recordings from rats learning to

run mazes show that codes in the striatum of the basal ganglia signal the beginning and end of automatised sequences (Jog et al., 1999).

While learning is clearly a key aspect of basal ganglia function, we view this as being prior to the overall function of action selection. Learning is thought of as being a modification of the action repertoire of the basal ganglia, and will create the association between saliences and responses that is assumed to exist within our model. On the time scale of action selection in the Stroop task, the learning function is beyond the present scope of our investigations, and thus can be provisionally ignored.

3.4. The BG follows Pieron's Law

As discussed above, the Cohen model produces the correct ratio between interference and facilitation because of the response mechanism, not – as was previously assumed – because of the sigmoidal activation function used. The Cohen model response mechanism follows exactly Pieron's Law, which relates the intensity of input to reaction time via a negatively accelerating function. Figure 15 shows that the basal ganglia model also closely follows Pieron's Law, if salience is equated with stimulus intensity[4]. The time to selection was gauged for a range of salience inputs (shown as dots on the graph) using the same model as used by Humphries et al (2002). Although the model makes provision for selection across six channels simultaneously, input was only provided on one channel. The best-fit line was obtained using the procedure described in the appendix. Luce (1986) notes that fitting Pieron's Law to data provides "an estimation problem of some delicacy". An important factor is whether the fit is carried out before or after the transformation to log-log coordinates. The transformation to log-log coordinates exaggerates the discrepancy between the data and the best-fit line at lower RTs. Hence fitting in log-log space can provide the illusionary appearance of a better fit.

[4] Strictly, of course, Pieron's Law can only apply to functions that relate stimulus intensity to reaction time. Response mechanisms, which take as their input 'evidence' or 'salience', can only conform to an analogue of Pieron's Law in which evidence or salience is taken to equate to stimulus intensity.

In Figure 15 this fit is also shown, and although it appear closer to many of the points on the line, if the fit is considered on linear coordinates it is less good than the fit performed on linear space.

Note that the basal ganglia response times do not cleanly match Pieron's Law in the same way that the Cohen response mechanism does. This is because, unlike the Cohen response mechanism, the basal ganglia model is not a simple mathematical formulation for calculating response times. The basal ganglia model is a complex system in which response times are emergent from the time-dynamic properties of the processing units. Unlike the Cohen response mechanism the basal ganglia goes beyond describing the behaviour of response times, it makes a connection with the underlying neuroanatomy and provides a possible biological basis for Pieron's Law.

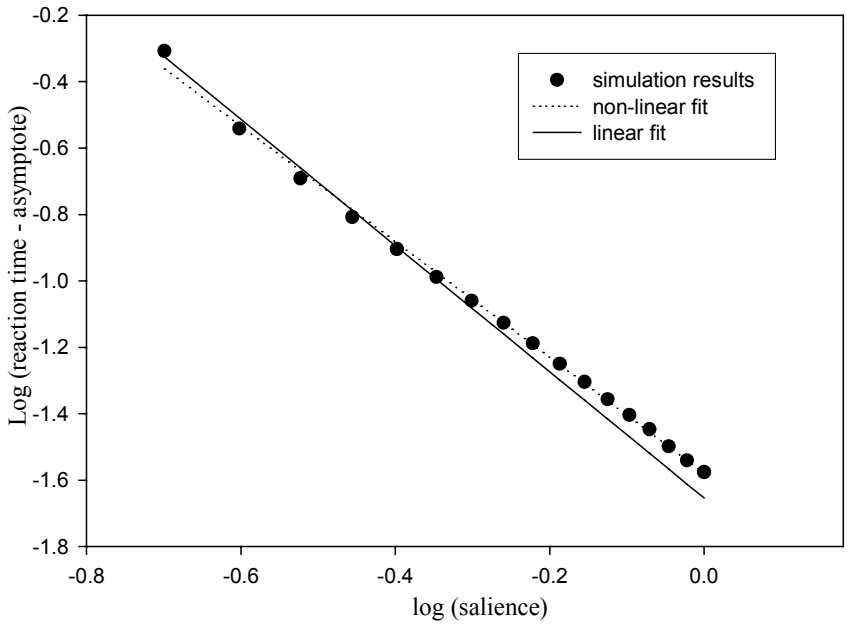


Figure 15: The basal ganglia model closely follows Pieron's Law. The input-RT function for the basal ganglia model shown on a log-log plot. The best-fit line derived by the standard procedure (as defined in the appendix) is shown as a solid line. The best-fit line obtained after the transformation to log-log space is shown as a dashed line.

It is possible to explain the trends observed for the basal ganglia model reaction times by the functionality of the units which make up the model. Like many other neural network models (including that of Cohen et al, 1990) the basal ganglia model consists of *leaky integrator* neurons. Such model neurons represent the simplest possible approximation to a dynamic neural membrane and, in a way similar to real neurons, adjust their output gradually to be commensurate with their input. A representation of the function of such a neuron may be given in the following form

$$\frac{da(t)}{dt} = -ka(t) + cI(t) \quad (5)$$

where *I(t)* and *a(t)* are, respectively, the weighted sum of inputs, and activation of the neuron at time *t*, *k* determines the characteristic time constant $\tau = 1/k$, and *c* is a constant which affects the influence of the input. The dynamics mean that the neuron is continually integrating ('accumulating') information over time and therefore has some of the characteristics of the response mechanisms discussed above. We can now derive a relationship between the response time of the neuron t_θ and a constant input *I*, where t_θ is defined as the time for the activity *a*, to cross a critical threshold, θ. Suppose that the neuron is at rest and receives a step input *I* at *t=0*. It is then straightforward (see, for example Kaplan, 1952) to solve equation (5) to obtain

$$a = \frac{cI}{k}(1 - e^{-kt}) \quad (6)$$

When $t = t_\theta$ then $a = \theta$. Substituting these into (6) and solving for t_θ in terms of *I* gives

$$t_\theta = -\frac{1}{k}\log\left(1 - \frac{k\theta}{cI}\right) \quad (7)$$

This function is shown in Figure 16 together with a regression line based on fitting Pieron's Law.

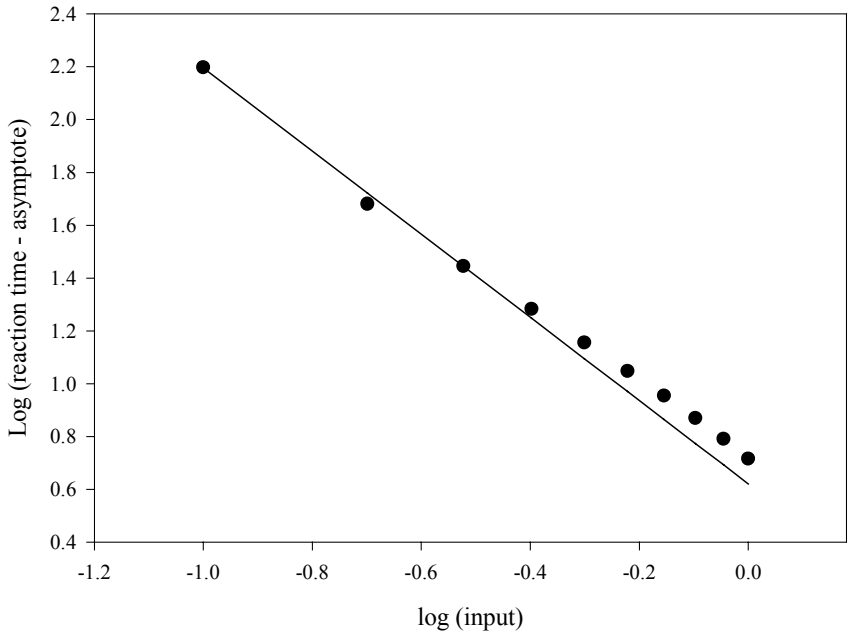


Figure 16: Time for model neuron output to cross threshold for different intensities of input, from equation (7).

The similarity to the basal ganglia input-response time function suggests that the essential characteristics of the latter reflect the basic response properties of its component model neurons. Because these investigations have been exploratory – as opposed to analytic – and based on simulations it is not possible to be make definite pronouncements, but it would appear that, as a default, any network based on processing units with these properties will manifest Pieron's Law-like selection functionality.

3.5. Pieron's law and relative saliences

Further analysis of the basal ganglia response time function shows that it is influenced by both the relative levels of salience inputs (the 'evidence' in the context of the Cohen response mechanism) and by the absolute levels of salience inputs. For the Cohen response mechanism two inputs of 0.3 and 0.5 generate the same response time as two inputs of 0.7 and 0.9. The basal ganglia response is dependent on the absolute level of inputs and would respond quicker to the 0.7 / 0.9 pair than the 0.3 / 0.5 pair. This is because the absolute level of the saliences are higher in the first case, even though the relative difference between the two saliences is the same in both conditions. For saliences of similar values, differences between saliences are the more important factor for determining reaction times, rather than absolute levels of the saliences.

In isolation, the BG model instantiates a reaction time function which, accounts for both relative salience and absolute level of competitor. Figure 15 shows that the BG model follows Pieron's law for a single input (i.e. the salience of the competitor is zero). Figure 17 shows that Pieron's law holds for relative saliences with different absolute levels of the competing salience. Log-log plots are shown, as for the other Pieron's Law analogue-plots. The three conditions are for different levels of the competitor / target saliences, varied to produce different amounts of relative salience. It is this relative salience that is shown on the x-axis. Note that it is possible to fit a Pieron-like law approximately to all three curves.

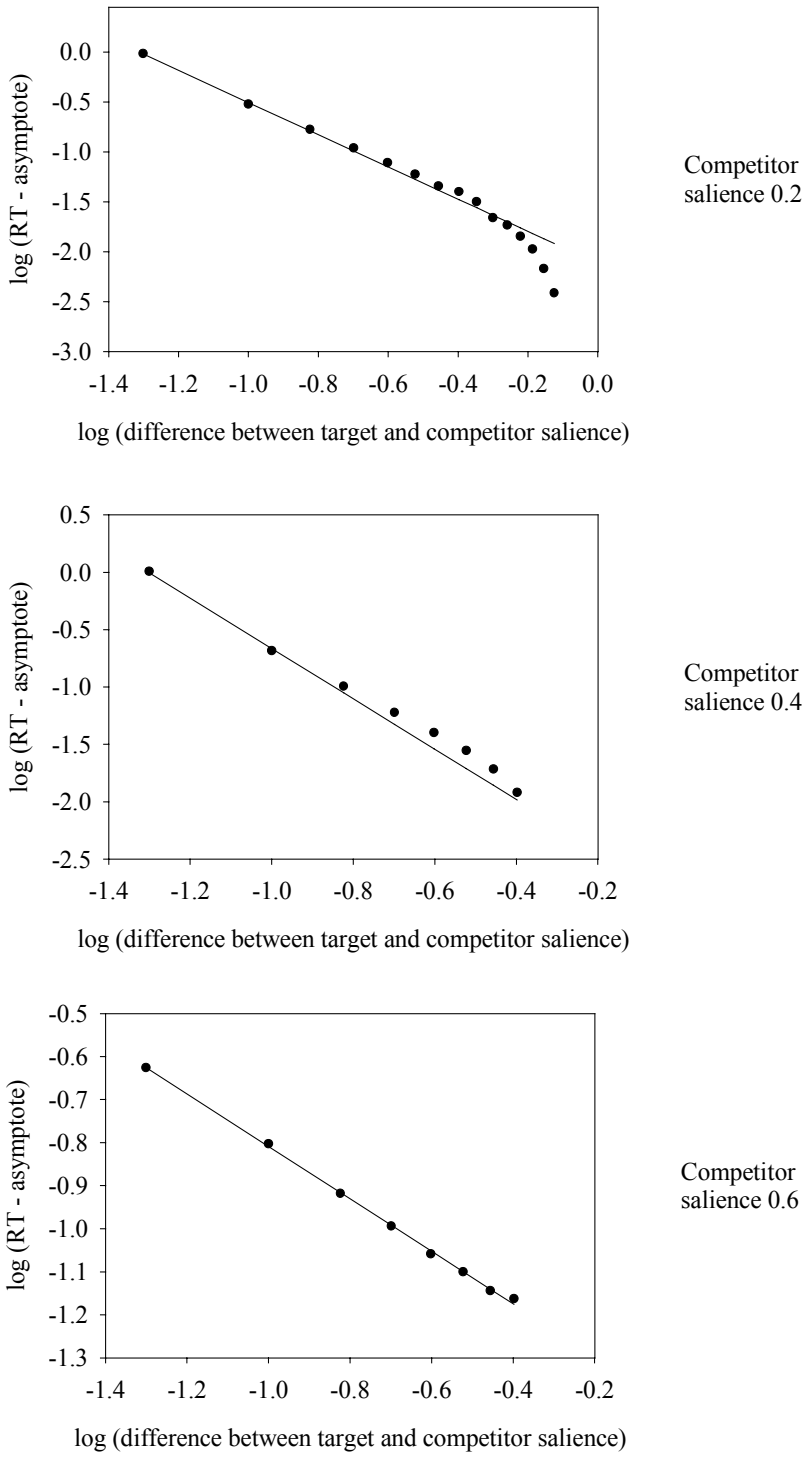


Figure 17: The BG model follows Pieron's Law for different absolute levels of the competing salience.

The Cohen model response mechanism provides reaction times solely as a function of the difference between the two input saliences. It doesn't account for absolute levels of the competitor in the same way as the BG model.

3.6. Adding noise to the BG model

Hitherto results from a deterministic BG model have been presented. However, it is possible to add stochastic function to the BG model. We would expect the BG model to be robust under conditions of noise, damage and ambiguity, as it is supposed to simulate a biological mechanism, By investigating BG function under the addition of two different sources of noise I attempted to reveal this robustness.

Noise can be added to the BG model in two ways:
1. exogenously; inputs to the model (saliences) are noisy
2. endogenously; signals within the model, between or within modules, are noisy

These two possibilities were investigated by adding uncorrelated, gaussian distributed noise with a mean of zero to either the saliences or to each channel emanating from each module in the BG model (BG proper, VLT, TRN, and Motor Cortex) respectively.

The principle finding was the same for both endogenous and exogenous noise. The addition of noise resulted in a positively skewed reaction time distribution (Figure 18). Positively skewed distributions are characteristic of reaction time data, and it promising that for the BG model the addition of gaussian distributed noise leads to a skewed distribution of outputs.

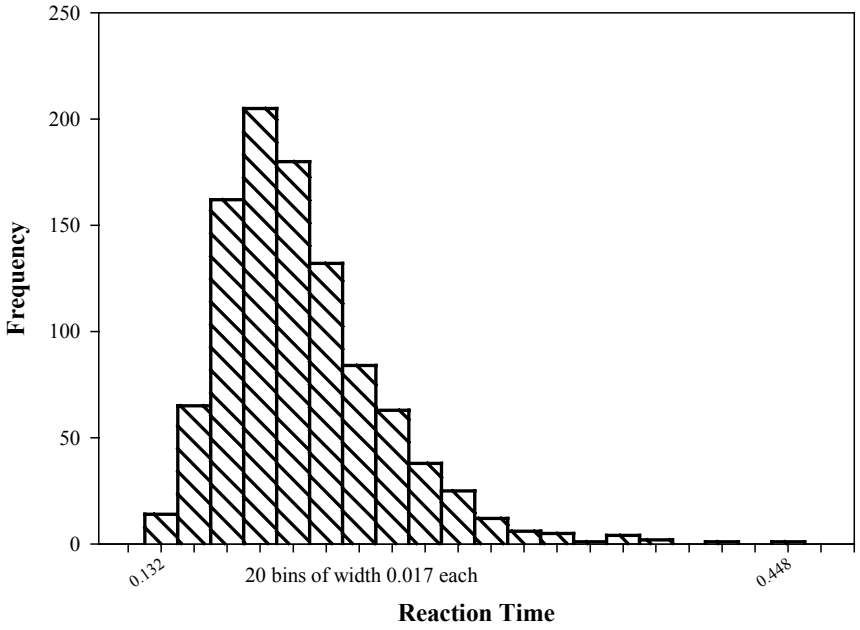


Figure 18: Histogram of reaction times produced by basal ganglia model with endogenous skew (variance of noise added to between module connections was 0.01, target salience was 0.5, salience on other channels was 0.0, number of trials was 1000).

A secondary important finding is that increasing the variance of the noise increases the extremity of the skew for the reaction time distribution. Small amounts of noise were sufficient to induce skew without causing a significant amount of erroneous selection. The addition of noise with a large variance created larger skew, and also made erroneous selections more frequent. The reaction time distribution was skewed even if the time from erroneous selections were removed.

Mewhort et al (1992), as discussed in section 2.4.2., criticised the model of Cohen et al (1990) on the grounds of the shape of reaction time distributions produced. Simulations with model 1 revealed that the BG model fares no better than the Cohen response mechanism; reaction time distributions from the simulations are least skewed in the congruent condition, rather than in the control condition, as is found empirically.

This is a weakness of both the model's presented here and of Cohen et al's models. However the shape of the reaction time distributions, as discussed by Mewhort et al (1992, see Spieler, Balota, & Faust, 2000, for a more recent treatment), may provide an insight into the nature of processing in the different conditions of the Stroop task. If the congruent and conflict conditions are significantly more skewed than the control condition, then it may be that the processes involved in these conditions are more noisy, or that a greater number of noisy processes contribute. Cowan (1998) notes that the product of processes containing gaussian noise is a process with log-normal noise, in other words, the characteristic positive skew of reaction time data. This supports the supposition that the congruent and conflict conditions are more skewed because they involve additional noisy processes. Indeed a recent report by Destoto and colleagues of neuroimaging of a variant of the Stroop task (Desoto, Fabiani, Geary, & Gratton, 2001), suggests that conflicting stimuli activate two, rather than one, regions of motor cortex. Given this it then it might be reasonable to assume that the sources of noise are two rather than one.

There is another possible way of accounting for the different skew of the reaction time distributions in the control compared to the other ('bivalent') conditions. If the noise variance is related to the magnitude of the saliences or the signal levels involved then it would follow that there was more noise variance – and hence more skew – in the congruent and the conflict conditions. Within the framework used by the models of this thesis, in the conflict condition there is significant activation on the competing channel (and hence more noise from this source) and in the congruent condition the target response has a heightened salience. So, if the variance of noise was related to signal magnitude then both of these conditions would result in more skewed results than the control condition.

Regardless of these speculations, these initial investigations show that the BG model is amenable to the addition of noise, and that it can function if provided with uncertain information or if its internal operation is stochastic. Although we have focussed on a deterministic BG model for the bulk of the work presented in this thesis, these findings indicate that it is not necessary to be restricted to the artificial situation of a noiseless, deterministic, model for future investigations.

4. MODEL 1; USING THE BG RESPONSE MECHANISM WITH THE COHEN MODEL OF STROOP PROCESSING

Explorations of the function of the Cohen model suggested that the nature of the response mechanism is an important, and limiting, factor on the performance of the network (section 2.2 and 2.4.2). The background theoretical analysis to the development of the basal ganglia model (e.g. Redgrave, 1998; Redgrave et al., 1999) suggested that the action selection problem is an important one and that designing a competent response mechanism for a real-time, real-world agent is a non-trivial task. Moreover, we assume that the importance of the action selection problem is such that the influence of the human response mechanism will be manifest in all tasks, including those that involve complex cognitive processes. The Stroop task is just such a task. I have therefore integrated an existing model of processing in the Stroop task (Cohen et al., 1990, as discussed above) with our model of action selection in the basal ganglia and thalamo-cortical loops (Gurney et al., 2001a; Gurney et al., 2001b; Humphries & Gurney, 2002, as discussed above).

Replacing the response mechanism of the Cohen model with the basal ganglia model corrects previous inadequacies of the model and shows that the basal ganglia model is an adequate response mechanism, even for cognitive tasks. This model was originally formulated to perform the function of switching between actions, in the context of motor control. Its extension to a cognitive task shows its general applicability as a model of behavioural switching. In addition, it is worth noting that the basal ganglia model was at no point designed to simulate reactions times, in the way the Cohen response mechanism was. The time-related properties of the model simply emerge from the dynamic properties of the functional units. Another advantage of the basal ganglia model over purely mathematical formulations of response mechanisms is that the basal ganglia model has internal structure which is neuro-anatomically localised. This allows the easy exploration of the effects of lesions and neurotransmitter manipulations on switching behaviour.

4.1. Model construction

The architecture of this model, henceforth 'Model 1', is shown in Figure 19. Essentially it consists of the concatenation of a modified Cohen model, less the response mechanism, with the basal ganglia model. Outputs indicating the salience of each possible response feed into the basal ganglia model from the modified Cohen model. The basal ganglia model then acts as a response mechanism, selecting the response with the highest associated salience. The reaction time is the time from the first presentation of stimulus information to the time when a response channel is selected by the basal ganglia (see above for details of basal ganglia selection). The basal ganglia model replaces the response mechanism of the original Cohen model which worked via evidence accumulation.

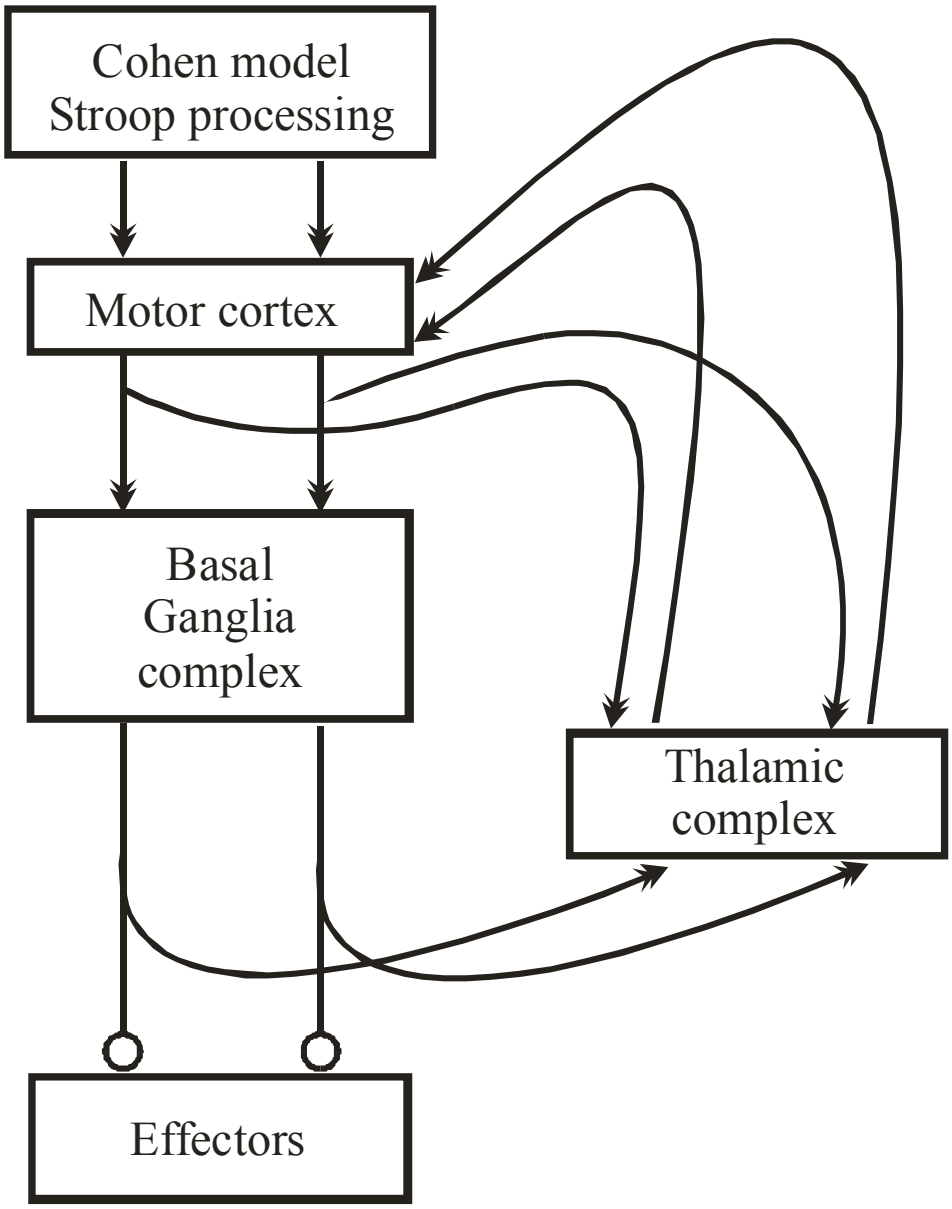


Figure 19: Architecture of Model 1. The modified Cohen model (Figure 3) provides inputs to the basal ganglia and thalamo-cortical loops (Figure 14).

Inputs to the basal ganglia are interpreted as salience signals, the magnitude of which is directly comparable to the urgency of the action they represent. It is assumed that low urgency responses will have low salience signals, or, in other words, that under quiescent conditions inputs are close to zero. In the original Cohen model the output signals have a resting activation of 0.5, and rise or fall about that point after stimulus input (see Figure 20). This default would represent tonic selection for the BG model. The signal representation of the Cohen model outputs was therefore adjusted so that the outputs of the front-end were compatible with the basal ganglia input scheme. Thus each output had a low resting level (around 0.1) and increased to reflect the confidence that that output corresponded to the required decision (see Figure 20). This was done by shifting the unit activation function along the input-axis, which in turn required the adjustment of the pre-training weights.

The parameter changes necessary to make the Cohen model compatible with the basal ganglia model are given in Table 3. All original values, and details of unchanged parameters, are given in Cohen et al (1990). The initial weights were changed to be entirely positive, so that non-selected responses remained at the resting activation level, rather than falling further below it. Removing the competitive forward projections of the original Cohen model had the effect of leaving the mediation of response competition entirely to the basal ganglia back-end.

Additionally these models use a continuous, variable step, time function, rather than a discrete, or fixed step, time function. This has no affect on the operation of the model, but means that it uses a system of differential equations to calculate the changes in activation levels at each point of time. A discrete time model has a fixed time step and thus can produce discontinuities in activation levels if changes are very rapid. The simulations were run without noise, since I was not interested in modelling reaction time variability phenomenon. The addition of noise to both models does not change the mean reaction times.

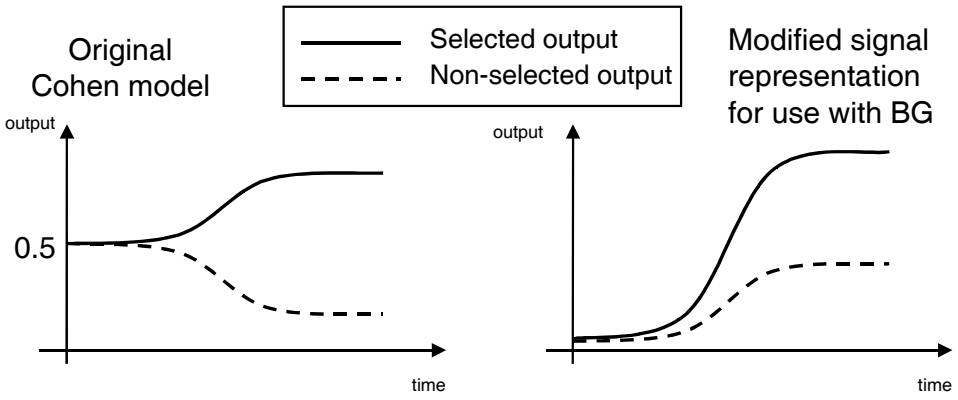


Figure 20: Signal representation in the original and modified front-end.

Parameter	Previous value/s	New value/s	Rationale
The output of the logistic function for zero input, θ.	0	2.2	Change resting activation from 0.5 to approx. 0.1
Target outputs for back-propagation.	[0, 1] or [1, 0]	[0.1, 1] or [1, 0.1]	0.1 = the resting activation, given θ = 2.2
Initial weights from input units to hidden units	+ 2 or - 2.	Original values + θ. i.e. + 4.2 or + 0.2	To preserve symmetry around a point which provokes activation of 0.5
Initial weights from hidden units to output units	Small random values	Small random values + θ	Ditto

Table 3: Summary of parameterisation changes to original Cohen model for its inclusion in Model 1

When integrated with the model of Stroop processing, does the combined BG-Cohen model perform as well or better than the original Cohen model with its mathematically based response mechanism? The combined model was assessed on the first three simulations reported by Cohen et al (1990).

4.2. Simulation 1 – basic Stroop results

Figure 21 shows that Model 1 replicates the basic Stroop condition data, both with respect to the empirical data (Figure 1) and with respect to the Cohen model simulation (Figure 4).

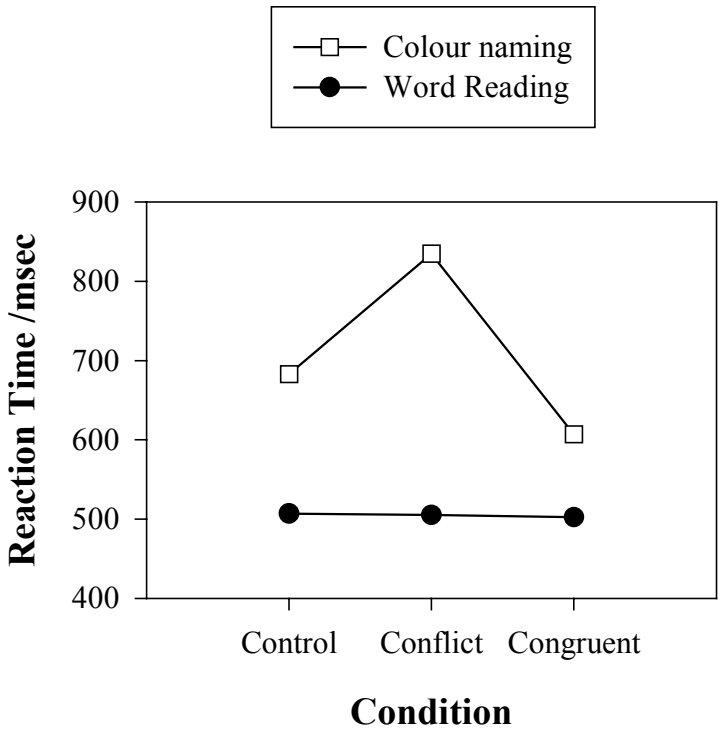


Figure 21: Simulation of the basic Stroop conditions Model 1.

These results show that, at the very least, the basal ganglia model is an adequate substitute for the response mechanism used by Cohen et al (1990). The two mechanisms are both adequate in this context because they both share similar functions to relate relative strength of evidence to response time. The time for selection with the Cohen response mechanism is wholly dependent on the relative strength of evidence. In the basal ganglia model, the time for selection depends both on the absolute level of the evidence for the to-be-selected output and the relative level of evidence between the to-be-selected output and its competitors. Crucially both mechanisms follow Pieron's law[5], exhibiting a negatively accelerating relationship between magnitude of input and response time (see above, Figure 9 & Figure 15).

The adjustment of the signal representation in the front-end of the model does not affect its proper function. The choice of signal representation in the front-end appears not to be intrinsically important, but rather can be constrained to be compatible with other components of the model as required.

4.3. Simulation 2 – learning follows the power law

It has been shown that, for many tasks, the decrease in reaction times with practice follows a power law (e.g. Logan, 1988; but see Heathcote, Brown, & Mewhort, 2000; Palmeri, 1999). Cohen et al (1990, simulation 3) demonstrated that their model of Stroop processing also follows a power law with training. To assess whether the combined model retains this feature I adopted the same method used by Cohen et al (1990) and trained the network front-end on the colour naming task, recording the reaction time at regular intervals. Figure 22 shows that the new model does indeed follow a power law with training.

[5] In the case of the BG model this means that the relationship between the magnitude of the selected salience and the reaction time follows a Pieron's Law-like function while the magnitude of competing saliences are held constant.

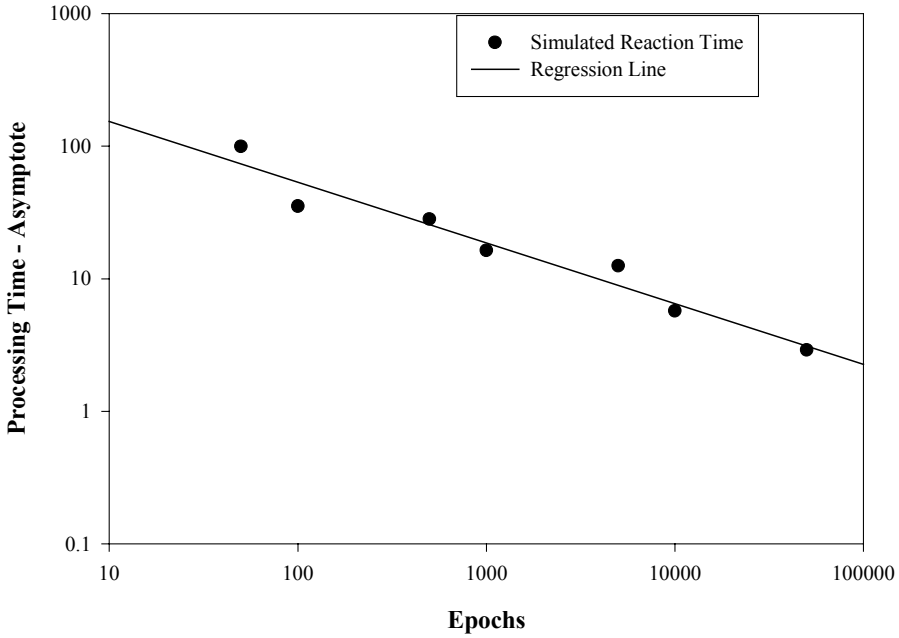


Figure 22: Model 1 conforms to the power law of practice. Both axis use a log scale. Simulation results are shown as dots. The simple regression for the data is shown as a straight line and follows the form $\log_{10}$ (Processing Time) = 2.645 – 0.459. $\log_{10}$(Epochs). R_2 = 0.948.

The fit between the reaction times and the regression line as training progresses shows that Model 1 follows the power law of practice. A key advantage of connectionist models is that they can parsimoniously account for learning phenomenon. Cohen et al (1990) have already demonstrated that their network architecture can exhibit a human-like learning dynamic. This result shows that using the basal ganglia model as a response mechanism does not interfere with the expression of that learning dynamic. In a similar way to the Cohen response mechanism, the basal ganglia model translates changes due to learning in the front-end output into changes in reaction time which follow the power law.

4.4. Simulation 3 – SOA results

Figure 23 shows the simulation of SOA results using Model 1. The basal ganglia model prevents selection based on the distorting stimulus and, in addition, the influence of that stimulus reaches a limited maximum. Because of this the amount of interference or facilitation remains constant beyond –200 ms SOA (compare with Figure 5 and Figure 13)

The basal ganglia response mechanism is able to overcome the action selection problems posed by the SOA experiment because it has a minimal absolute salience threshold, below which saliences do not prompt response (see section 4.5). When the irrelevant dimension is present but the relevant dimension is not, it does provoke a non-zero salience, but that saliences does not prompt a response because it is not above the minimal threshold.

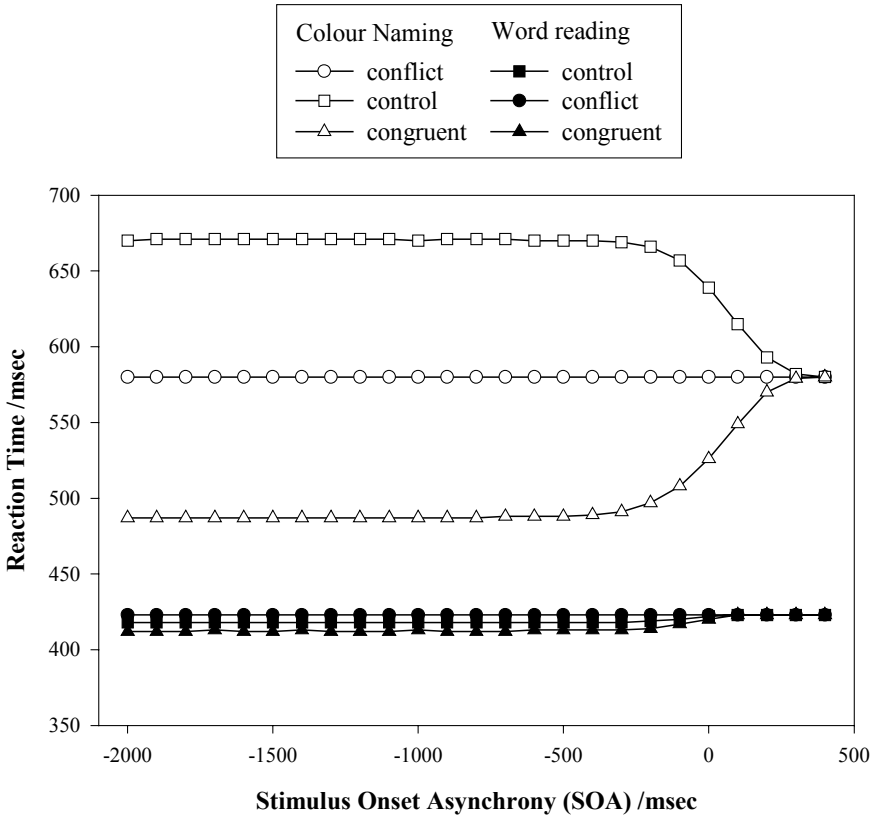


Figure 23: Model 1 simulates SOA results well within original empirical range, and never makes wrong selections at long SOAs.

4.5. Discussion

The basal ganglia model is a suitable response mechanism for acting on the outputs of the Cohen et al (1990) architecture for Stroop processing. Simulation 1 shows that the basal ganglia model performs as well as the original Cohen response mechanism when replicating basic Stroop interference and facilitation effects (or their absence) for colour naming and word reading. In addition, changing the signal representation in the front-end of the model does not affect the functionality of the model. The initial choice of signal representation by Cohen et al (1990) was not forced by a core assumption of the model-theory and its exact nature turns out not to be crucial to behaviour of the model. This implementational choice can therefore be constrained by making the signal representation consonant with the response mechanism used. In this case the basal ganglia response mechanism also requires the signal representation to be neurobiologically plausible.

Simulation 2 shows that the basal ganglia model does not interfere with the plausible learning dynamics of the front-end. Those learning dynamics are due to the continuous nature of the connection weights in the model. Graded changes in the front-end are translated by the basal ganglia model into appropriately graded changes in response times.

Simulation 3 shows the superiority of the basal ganglia as a response mechanism over the original response mechanism. Over long negative SOAs the Cohen response mechanism makes wrong selections, due to the small but significant influence of the distracting stimulus dimension. The input units of the basal ganglia model filter out small salience inputs (as discussed section 3.2). This creates a minimal salience threshold, below which inputs are ignored. Thus, using the basal ganglia response mechanism, Model 1 makes the correct selection at all SOA values.

This minimal threshold was included in the basal ganglia model because of the neurobiology of medium spiny neurons in the striatum – the input nucleus of the basal ganglia. These neurons possess upstate / downstate functionality, which means that they only start to release action potentials if their input is above a certain threshold. This feature also has the effect of filtering out noise in the inputs which is below threshold. The sub-threshold input still alters the resting level of the GPi nuclei, via connections from SNr, and thus can prime, positively or negatively, the response to the relevant dimension when it occurs. However the amount of this priming is limited because the model uses units with an output range restricted between 0 and 1 (cf. the Cohen response mechanism which contains the evidence accumulation counters with the capacity to retain infinitely large values).

The effect of this threshold of required minimal salience for selection in the SOA simulation is to stop the reaction times for the conflict and congruent conditions changing beyond –200 ms SOA. The inputs due to the irrelevant stimulus do not prompt a response at these negative SOAs and, because of the minimal threshold, they do not increasingly affect the selection time of the correct response. Human participants may be primed by the presence of the word dimension in the colour naming condition, but the priming (reflected in the amount of interference or facilitation) does not continue to increase as the SOA becomes more negative. In fact, after a brief time window, centred on 0 ms SOA, in which interference is at a maximum, it decreases and then levels out to a constant value (Glaser & Glaser, 1982; Schooler et al., 1997). This pattern is better reflected by a model which uses a response mechanism with a minimal threshold, such as the basal ganglia, and which is based on signal processing in units with realistic output ranges. There remains some discrepancy between simulation and empirical data in the SOA experiments. Although the combined model simulation shows interference and facilitation stabilising at negative SOAs, they do not decrease from their maximum values as in the empirical data. Chapter 6 shows one way of addressing this.

The competence of the basal ganglia model as a response mechanism lends credence to the view that the basal ganglia architecture may be performing a selection

function. The combination of the basal ganglia model with a model of sensory processing allows the exploration of the influence of the human response mechanism on cognitive tasks like the Stroop task. It may be that the properties of the human response mechanism have a ubiquitous influence where ever fundamental response qualities are used to assess cognitive function.

5. MODEL 2; INCORPORATING MODELS OF WORD READING

The simulations discussed above demonstrate that the basal ganglia model is potentially an improvement on previous models of response selection. Incorporating constraints from theories of action selection and from neuroscience into the 'back-end' of the Stroop model improves the range of findings which can be accounted for. Does incorporating additional constraints from the psychology of word-reading to the front-end further improve the function of the model? Is the processing performed by this architecture homologous to the processing performed by the original Cohen model? It would be heartening if the independently formulated cognitive architecture of word reading provides a working account of word processing in a connectionist framework. The following simulations show that it does.

5.1. Cognitive theories of word reading

Conventional accounts of word-reading (e.g. Ellis & Young, 1988; Eysenck & Keane, 1995) incorporate the existence of multiple routes between the encoding of the visual form of the word (the graphemes) and the production of the associated sounds (the phonemes). These box-and-arrow accounts are based on studies of patients with acquired dyslexias (although see Fiebach, Friederici, Muller, & von Cramon, 2002, for a recent report of fMRI evidence for multiple route models of word reading). A 'sublexical route' allows the conversion of graphemes to phonemes without regard (or knowledge) of the meaning of the word. Within the 'lexical route' semantic and non-semantic routes exist. The separation between the two is supported by incidents, in patients and controls, of word recognition but without access to the corresponding word meaning. Figure 24 shows the standard 'box-and-arrow' information-processing model of word reading.

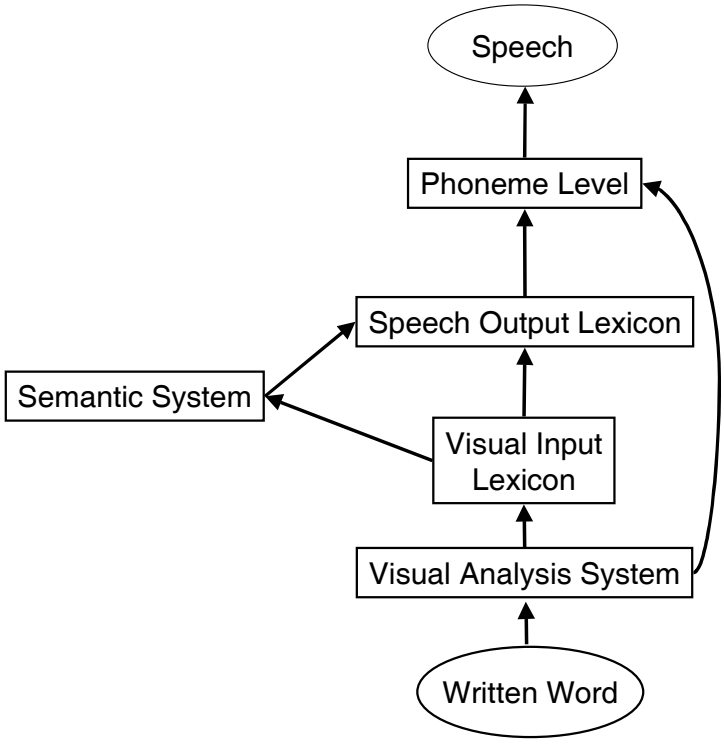


Figure 24: The standard information processing model of word reading (after Ellis & Young, 1988, figure 8.1, p. 192).

Although it is agreed that there are multiple processes involved in word-reading, there have been disagreements over their implementation in computational models. Seidenberg & McClelland (1989) proposed a model which involved a single route from print to speech. This PDP model used the same weights to account for words with both regular and irregular pronunciations. This account was heavily criticised by Coltheart, Curtis, Atkins, & Haller (1993) who proposed a model which is based more closely upon the schematic accounts and includes all the proposed schematic routes implemented in separate pathways. Subsequent work by Seidenberg and colleagues (Plaut & Gonnerman, 2000; Plaut et al., 1996) has now implemented a semantic route and so the difference between the two approaches is really over the nature of the processing between the stages, rather than the existence of multiple routes per se. The 'radical connectionist' account emphasises the importance of distributed representations and their acquisition via the back-propagation algorithm, while Coltheart's model uses localist representations and includes elements of a symbolic rule-based system. The debate between these two camps exemplifies two very different approaches to the use of connectionist models; two approaches which can be seen to be at odds over other topics as well (this will form one of the topics of the general discussion, see section 7.6.5).

The architecture of the front-end of Model 2, based upon the cognitive neuropsychology schematic of Figure 24, is shown in Figure 25. To adopt a simpler (albeit more powerful) scheme vis-à-vis the stages involved, as exemplified by Plaut et al (2000), would be incongruous with the simplistic nature of processing and representation required in the present context. We are only modelling the reading of two words and are not using realistic input or output forms. We are not setting out to compare irregular and regular mappings, nor do we wish to investigate the development of the internal representations of the mapped words. Our model utilises a localist coding of the two word possibilities, because our purpose is the investigation of the strength and speed of the possible responses, and of their competition, only.

The colour naming components of the front-end are based on the simplest conceivable generic model of colour-naming: colour-encoding → semantics → output path. Note that, after the semantics stage, the colour-naming path is coincident with the word-reading path. These stages are specifically restricted to producing verbal outputs. Non-verbal outputs (such as pointing to the correct response) do not involve these stages. Our model accounts for manual responses in a way comparable to the schematic depicted for verbal outputs, but fed by only a single route. In this model manual responses are driven by information from a module which receives outputs from the semantic stage.

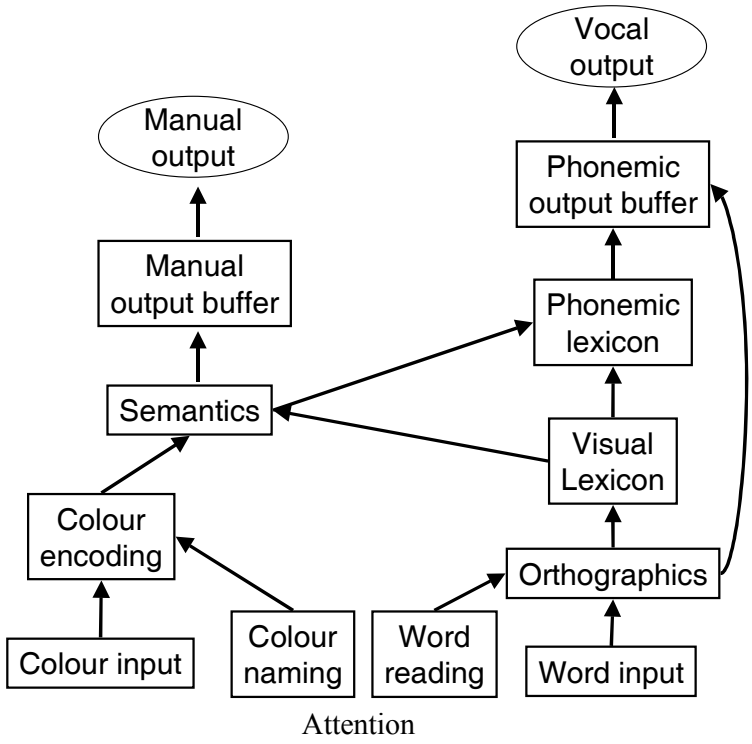


Figure 25: Processing stages based on word-reading theory. In Model 2, each modules contains two units which connect, only, to their corresponding units in modules forward and backward in the model.

This model mirrors the, independently developed, architecture proposed by Coltheart and colleagues for their Dual Route Cascaded (DRC) model of reading aloud (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001; Coltheart, Woollams, Kinoshita, & Perry, 1999). For the purposes of transparency of function the feedback connections that are included in Coltheart et al's model (Coltheart, 1993; Coltheart et al., 2001) are omitted. The full front-end model involves the synthesis of the pathways for the two processes – colour naming and word reading. They join at the semantics stage and over-lap for the production of outputs.

In contrast to the original Cohen model, in which the differential strength of processing ensued from the asymmetrical *training* of the network, for Model 2 we are primarily interested in the influence of the architecture on strength of processing. In contrast therefore, all the weights are set to the same value, so that, although strength of processing remains a key element of this model, it is instantiated as the existence of multiple pathways in the word-reading route, rather than as greater weight magnitude in the word-reading pathway. In addition, it was required that both word and colour inputs produce some activity even when not supported by the attentional input. This was true for Cohen et al's (1990) model and if this was not the case then no interference effects would be possible. To ensure this, given the choice of output function (see below) all network weights were set to 1.1 and the bias fixed on the first layer of units at –1.0. Since attentional input was either 0 or 1, the net input to a first layer unit in the absence of attention was 0.1, which provokes an increase in unit activation which is marginally above zero.

5.2. Unit output functions

In developing this model I discovered that, while it could reproduce the main trends in the data, the strength of these trends was often weaker than when observed in Model 1. The origin of this phenomenon was traced to the way in which signals propagated through the network front-end. The sigmoid activation function tends to attenuate low strength signals. The signals are of low value because of the need for

the output saliences in the without-attention condition to be below the selection threshold. One approach to remedy this problem might be to increase the gain on the sigmoid function. An alternative, and preferable, solution is to use a Weibul function rather than a sigmoid. The two functions are shown and compared in Figure 26. The equation defining the Weibul function is:

$$y = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{(-x/b)^a} + \theta & \text{otherwise} \end{cases}$$

Where y is the output, x the input and a,b,θ are parameters, which for all work presented here take the following values $a = 1$, $b = 1$ and $\theta = 0$.

The sigmoid is often quoted as resembling a veridical neuron activation function. In fact it is a crude approximation when compared to a Weibul function, and has gained widespread adoption due in part to the fact that it was used in the original very influential PDP books (Rumelhart et al., 1986b) and to the fact that it is continuously differentiable, which is important for the implementation of the back-propagation algorithm. The Weibul function is not continuously differentiable around $x = 0$. While the use of the Weibul function may appear to merely be expedient in this context, it is supported by neurobiology, since the Weibul function more accurately represents the relation between neural input and output (compare Figure 26 with Figure 27).

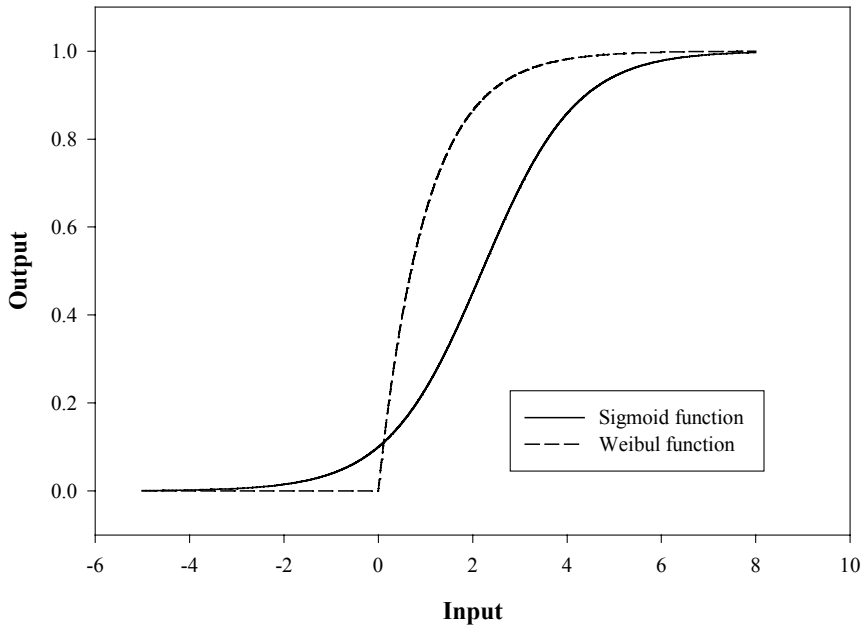


Figure 26: The Weibul function and the sigmoid function.

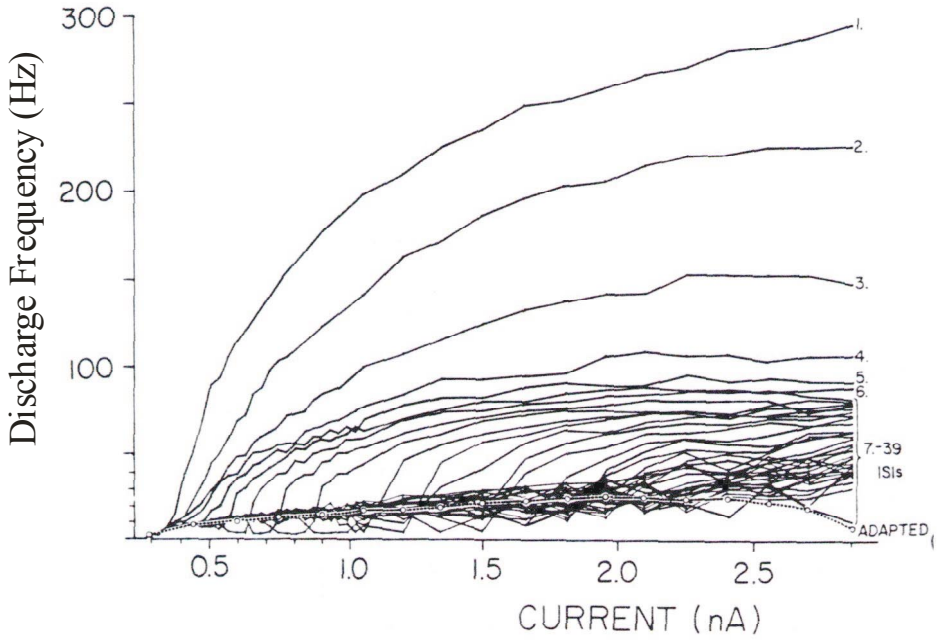


Figure 27: Frequency-current plot from Lanthorn (1984, fig 6C). The firing frequency in response to injection of 1.5s long, rectangular depolarising current pulses has been plotted against current strength for one CA1 pyramidal cell with a continuous firing pattern.

Notice that the Weibul function lacks the 'rounded heel' of the sigmoid function, and begins a rapid rise in output as soon as input crosses a threshold to bring it above zero. The second derivative of the rate of change of the Weibul function is consistently negative along it's entire non-zero range, whereas for the sigmoid function the rate of change switches from positive to negative. At low unit input levels the sigmoidal activation function, as parameterised in the model, further attenuates the inputs. The Weibul function was introduced so that the low inputs to some units in the model were not so reduced as to be irrelevant to processing in the model.

5.3. Model 2 results

5.3.1. Sim1 & SOA

As shown in Figure 28 and in Figure 29, Model 2 successfully replicates the Basic Stroop and SOA data. Note that interference is greater than facilitation in the basic Stroop simulation and that for the SOA data the colour naming reaction times stabilise quickly, and are constant beyond –100 ms SOA, as with the empirical data. These results are comparable to those for Model 1 on Sim 1 (Figure 21) and better for the SOA results (Figure 23 cf. empirical results in Figure 2).

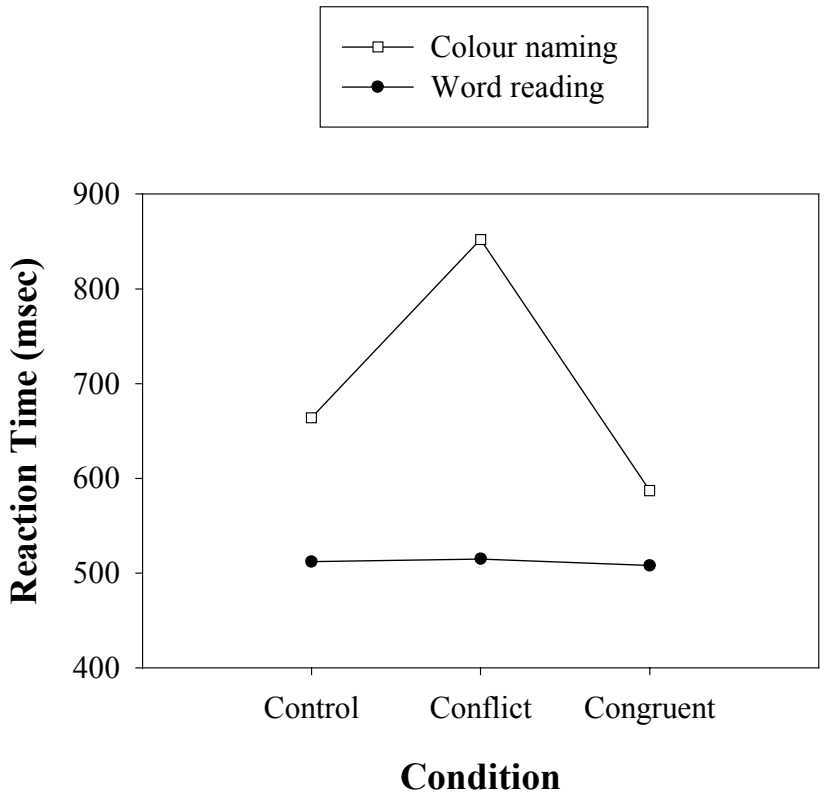
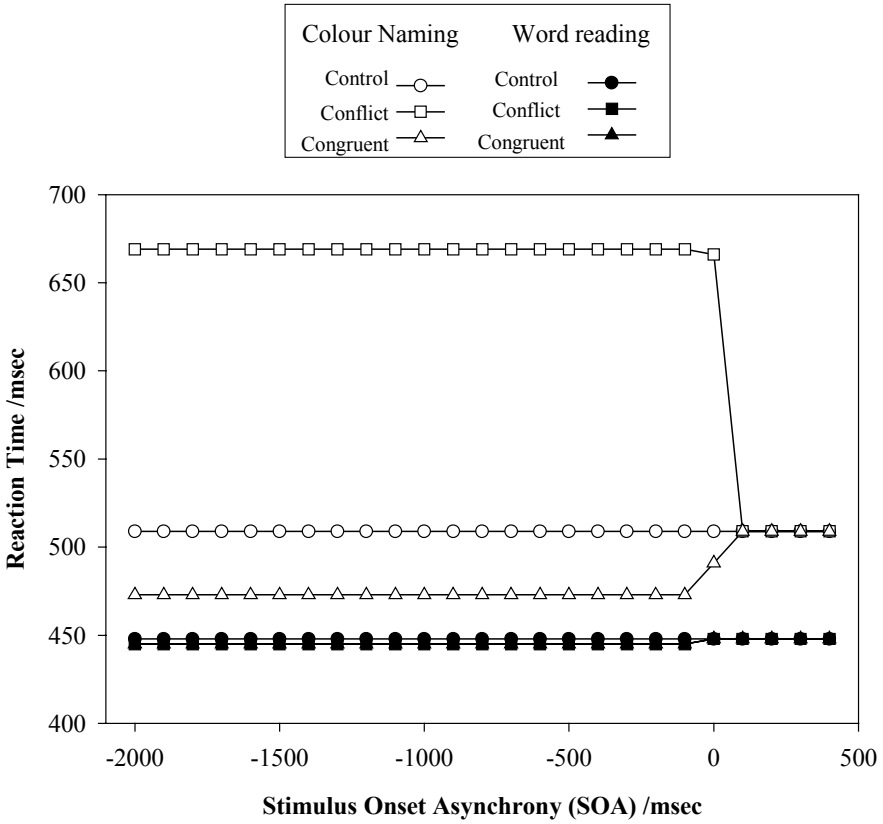


Figure 28: Simulation of the basic Stroop conditions with Model 2.



.

Figure 29: Simulation of SOA results with Model 2.

The original Cohen architecture (see Figure 19) has been shown to capture the essential dynamics required, based on the assumption that activations associated with word reading are stronger than activations associated with colour naming. It is shown here that the same effects can be gained by assuming a greater number of pathways contribute to word reading, rather than involving higher activation in just one pathway. This addresses the criticism that word reading and colour naming could not be performed by identical architectures (Kanne et al., 1998). In addition, the introduction of a cognitively constrained architecture allows the range of data accounted for to be extended, as shown in the next simulation.

5.3.2. Manual results and the predicted reverse Stroop effect

The original Stroop task (Stroop, 1935) recorded the time for the subject's verbal responses. Subsequent studies have compared Stroop reaction times for verbal response and for manual responses, such as pointing at a label indicating the correct response. A comprehensive review (MacLeod, 1991) concluded that *'although still significant, interference (but perhaps not facilitation) is reduced when response modality is switched from oral to manual'*. This finding is confirmed by a number of more recent studies (Henik, Ro, Merrill, Rafal, & Safadi, 1999; Sharma & McKenna, 1998; Weekes & Zaidel, 1996). These studies also show that manual responding is in general quicker than vocal responding, by around 50 to 150 ms in the colour-naming control condition.

It is not clear how the original Cohen architecture could be adapted to simulated manual responding. In contrast, because the connectionist architecture of Model 2 has been derived from a psychologically plausible information flow diagram, it is clearer how manual output may be incorporated into the overall scheme. Thus it is natural to suppose that the manual output is driven from the core semantic representation, rather than being reliant on the post-semantic and verbal-response orientated representation of colour or visual word material (Figure 25). Once saliences for the basal ganglia response mechanism are taken from this 'manual

output buffer', simulated reaction times can be generated in the standard way. The minimal number of data points to verify the observed differences between manual and vocal responding are shown in Figure 30.

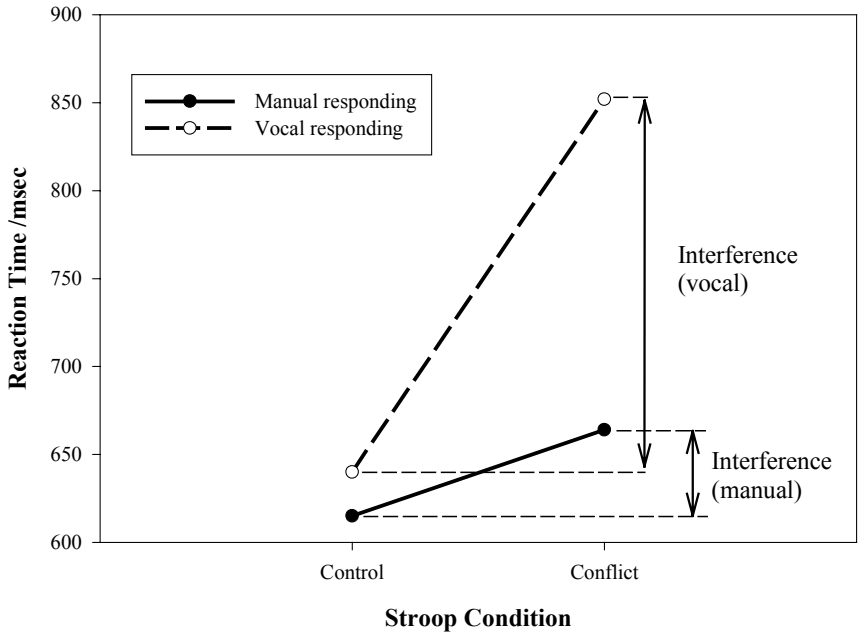


Figure 30: Simulation of the colour-naming task in the control and conflict conditions for both manual and vocal responses by Model 2. Annotations mark the size of the interference effects in the two response conditions.

Both aspects of the experimental data are replicated in the simulation. First, reaction times are faster when the response modality is manual responding. This is because the greater number of stages between inputs and vocal outputs means that responses take longer than for manual output. This is not because the activations are weaker but because the signals simply take a longer time to cascade from the input to the vocal output units. The manual response units take signals directly from the semantics stage and thus there is less 'distance', in terms of processing stages, between inputs and outputs. Second, interference between the control and conflict conditions is smaller with manual responding than with vocal responding. This is because the manual output units come from the semantics module, and hence they are unaffected by the multiple pathways involved in word reading. Ignored word information does not increase in strength because of the cascade of the same information through multiple pathways. This is the cause of the (relatively) reduced interference on colour-naming from word information.

Our model also provides predictions that had not been directly experimentally verified so far. Figure 31 shows the simulated reaction times for all possible combinations of response modality, congruency condition and task, including those data points shown in Figure 30. It is evident that the model predicts a 'reverse Stroop effect' for manual responding. A reverse Stroop effect, as shown in the figure, is where, contrary to the standard Stroop effect, colour information interferes significantly with word reading, not vice-versa. Model 2 predicts that a reverse Stroop effect will exist for manual responding (but not vocal responding) on the word reading task.

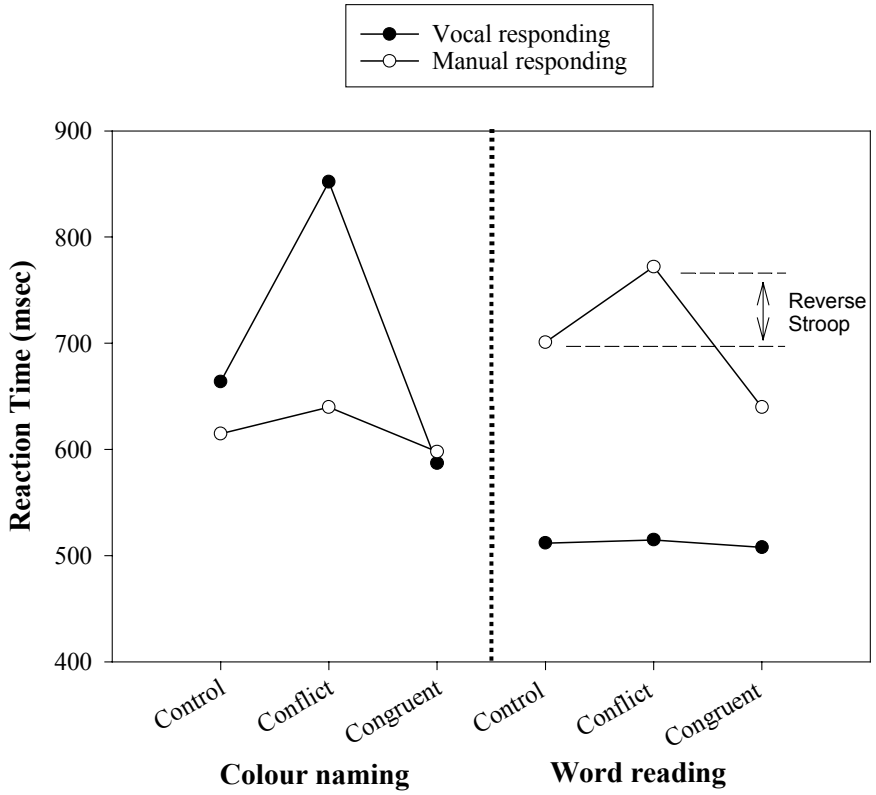


Figure 31: Simulation of all possible results for both manual and vocal responses with Model 2. Annotation shows the predicted reverse Stroop effect

Our model also makes the prediction that word reading time will be slower in the manual condition than in the vocal responding condition. For the same reason that the interference is less (less word activation) the manual reaction times for word reading are slower in the model than both vocal reaction times for word reading and slower than manual reaction times for colour naming.

Both of these predictions have not so far been directly experimentally verified, although there exist some previous findings which suggest that this may indeed be the case. Treisman & Fearnley (1969) used a card-sorting paradigm and demonstrated that cross-attribute matching produced slower response times than within-attribute matching for both word and colour determined responses. Morton & Chambers (1973) confirm this result, and they demonstrate (experiment IV) that within-attribute matching *is* affected by the properties of the irrelevant dimension. Their discussion of interference effects turns on the differential speed of naming. This is another example of the 'horse-race' model of interference (section 1.6.2) which 'strength of processing' models, as exemplified by the simulations presented in this thesis, supersede.

More recent pertinent findings are those of Melera & Mounts (1993, especially figures 5 & 7) and Durgin (2000). The evidence from Melera & Mounts (1993) is equivocal because their experiments involved controlling baseline discriminability and practice effects and rather than directly comparing vocal and manual colour and word reading in the same way as, for example, Dunbar & MacLeod (1984). Durgin (2000) shows that a reverse Stroop effect is obtainable for manual responding, although he does not compare the two possible response modalities.

These results anticipate the predictions made by our model, but do not represent a definite experimental test. The next section details the experimental verification of the predictions.

5.3.3. Experiment 1: Test of the Predictions from Model 2

Method

Subjects

14 undergraduate psychology students from the University of Sheffield participated in the experiment, as part of a course requirement. All had normal or corrected to normal vision. The mean age was 21, and 7 of the participants were female.

Stimuli & Apparatus

The stimuli consisted of three colour words ('Red', 'Green' and 'Blue') and the string XXXX, presented in the colours red, green, blue and black. Black colouring of the words was considered the control for the word-reading task, while the string XXXX was the control for the colour-naming task.

Stimuli were presented on an Apple Power Macintosh (Mac 8500 with 64MB Ram), using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993). Participant's responses and response times were recorded using a Psyscope external button box with red, green and blue buttons (for manual responding conditions) and a microphone headset (for vocal responding conditions).

Procedure

The participant sat directly in front of the computer and read the experimental instructions, as well as having them verbally summarised by the experimenter. The four experimental tasks (colour naming, verbal and manual responding, and word reading, verbal and manual responding) were each assessed twice, producing eight blocks. Instructions for each particular task appeared before each block. A block was a set of 24 stimuli for the relevant task. Within a block, the stimuli appeared in a random order with at least 1.5 seconds between the participant's response and the

appearance of the next stimulus. Each task-block had 12 possible word-colour combinations which were presented 4 times each in a random order. Hence 12 conflict and 12 control condition stimuli, and 24 conflict condition stimuli were in each block. Subjects were told verbally when they were halfway through the experiment, reassured not to worry if they got any answers wrong and to just move onto the next stimulus.

Piloting

The experiment was piloted and a number of alterations were adopted. The response set size was increased from 2 to 3 possible responses. This raised the mean RT and made the conventional Stroop effect clearly manifest. For the colour naming task both word (e.g. the word 'DOG' in a colour ink) and non-word (i.e. 'XXXX' in a coloured ink) controls were piloted. The word-control was taken out because:

a) it made the colour naming task symmetrical with the word-naming task (i.e. just one control condition).

b) over a number of subjects there was not a statistically significant difference between the two control conditions.

c) the XXXX control was well established in the literature.

Results

Reaction times from incorrect responses, and those quicker than 300ms or slower than 1500ms, were discarded from the analysis.

The mean response times across all conditions are shown in Figure 32.

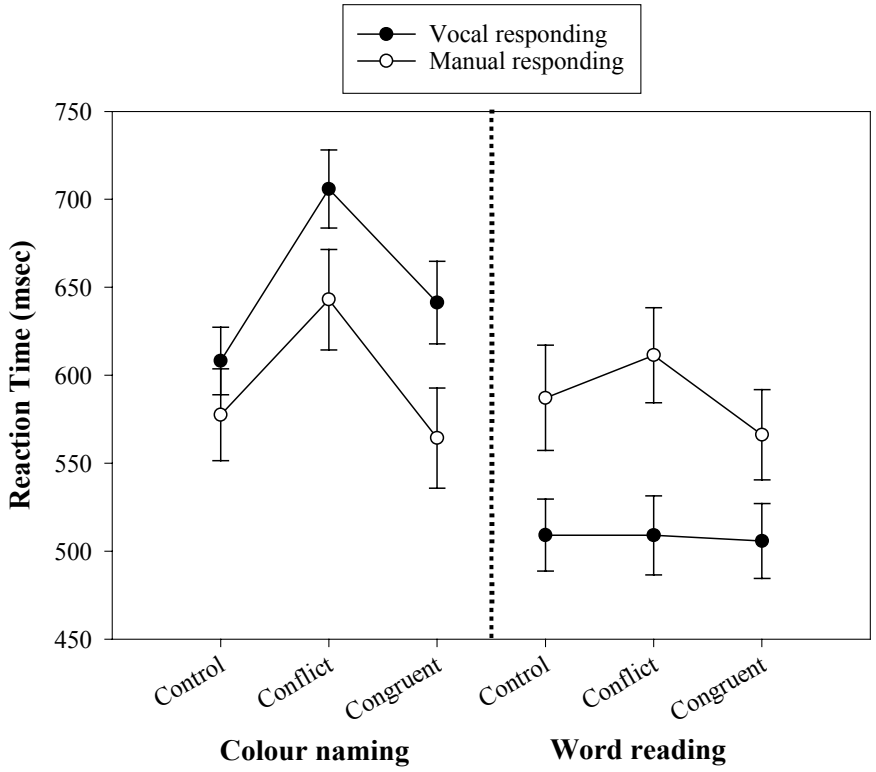


Figure 32: Full results for experimental test of Model 2 predictions. n=14. Standard error bars shown.

Comparing Figure 32 with the predictions of the model shown in Figure 30 and Figure 31 we can see a partial agreement of the empirical results with the model predictions. The relative speeds of the four groups of response times (manual and vocal colour-naming, and manual and vocal word-reading) match, and the predicted reverse Stroop exists. These effects are verified by paired t-tests on the relevant means. Single-trailed significant values are used because the model provides a specific direction in which the means should differ.

Comparing colour naming control conditions for vocal and manual responding, the difference in the means is 30.6 ms which is approaching significant ($t=1.63$, $df=13$, one tailed $p = 0.06$). So, in the basic condition, vocal responding is slower than manual responding.

The difference between the manual word-naming control condition and the manual word naming conflict condition is significant ($t=2.07$, $df=13$, one tailed $p < 0.05$) which supports the idea of a reverse Stroop effect (colour information interferes with word naming for manual responding). However a significant difference was also found for manual colour-naming, between the control and conflict conditions ($t=4.78$, $df=13$, $p < 0.01$).

For vocal responses the Stroop effect is unambiguously present. Word information interferes with colour naming (control versus conflict condition, $t=10.6$, $df =13$, $p < 0.01$) but not vice-versa (control versus conflict condition for vocal responding word naming, $t=0.01$, $df = 13$, ns).

The main disparity from the predicted RTs is that the reduced interference in the manual colour-naming condition compared to the vocal colour-naming condition is not present. Although the interference is less, it is still significant (i.e. the conflict condition is significantly slower than the control condition).

Further analysis shows that the pattern of results is more complex. Using Tukey's HSD post-hoc test on the data from each of the 14 individuals it is possible to see if the interference effect with vocal and manual responding is significant across individuals or merely on aggregate. Table 4 shows the significance values for the 14 subjects. It can be seen that for 9 of the 14 subjects the individual Stroop effect was significant for vocal responding, whereas for manual responding the individual Stroop effect was significant for just 2 of the 14.

Subject No.	Vocal Sig.	Manual Sig.
1	0.001*	0.001*
2	0.018*	0.054
3	0.001*	0.925
4	0.321	0.126
5	0.177	0.791
6	0.002*	0.463
7	0.081	0.894
8	0.049*	0.965
9	0.000*	0.588
10	0.306	0.310
11	0.017*	0.247
12	0.046*	0.385
13	0.331	0.016*
14	0.007*	0.185
Total p<0.05	9 / 14	2 / 14

Table 4: Tukey's HSD significance values for each subject, comparing colour naming control condition versus colour naming conflict condition for vocal and manual responding. Values of $p<0.05$ are indicated by a *.

Discussion of experiment 1

Overall the experimental results (Figure 32) match well with the simulation predictions (Figure 31). The main discrepancy is the existence of an interference effect for manual responding in the colour-naming task. Manual responding leads to interference in both the colour-naming and the word-reading conditions. This mutual interference effect was found by MacLeod & Dunbar (1988, experiment 3)

when they observed the progression of interference due to practice. The effect in this experiment is highly subject dependent.

A second, albeit minor, discrepancy of the experimental data from the simulation predictions is the lack of facilitation for vocal responding in the colour-naming condition. This could be because of the nature of the control condition used. Although piloting showed no clear difference between an 'XXXX' control and a 'word' control, there was a trend towards the 'XXXX' control generating faster reaction times. Lack of a facilitation effect is not overly important given the weakness of the effect in general (as noted by MacLeod, 1991). Increasing the amount of data points collected from each subject could conceivably reverse the (non-significant) difference between the control and congruent condition for vocal-responding/colour-naming, while leaving the general pattern intact.

5.4. Competitor facilitation in the BG model

Whilst exploring the function of the BG model as part of Model 2 I discovered a phenomenon which I have termed 'competitor facilitation'. Briefly, in the conditions where the BG is switching from a previously selected response to mediating the competition between two new responses, a closer competition between the two new responses can cause a quicker selection than a more clear cut competition.

To define this situation, let us term the originally selected response the pre-competition response, and the associated input to the BG as the pre-competition salience. At a certain point the pre-competition salience is removed and a simultaneous competition begins between two new saliences, each representing two incompatible responses. Let the larger of the two new saliences be called the target and the smaller be called the competitor. The selection time is the time from the appearance of the competing saliences to the selection of the target. If we fix the values of the pre-competition salience and of the target salience we can assess the influence that the value of the competitor salience has on selection time. Without a

pre-competition salience, reaction time increases exponentially as the competitor salience become closer in magnitude to the target. However, in cases where the pre-competition salience is non-zero the reaction time function takes a counter-intuitive form (Figure 33); as the value of the competitor increases the selection time *decreases* up to a certain point and thereafter selection time increases rapidly.

This non-monoticity is present across a wide range of pre-competition saliences and a wide range of target saliences. It occurs because, although the presence of a non-zero competing salience slows selection time, the *de-selection* of the pre-competition response is assisted by the *sum* of the competing saliences. This phenomenon is not reliant on an instantaneous switch from the pre-competition response to the competition between the two new responses, but does dissolve as the gap between the pre-competition salience turning off and the competition saliences turning on increases.

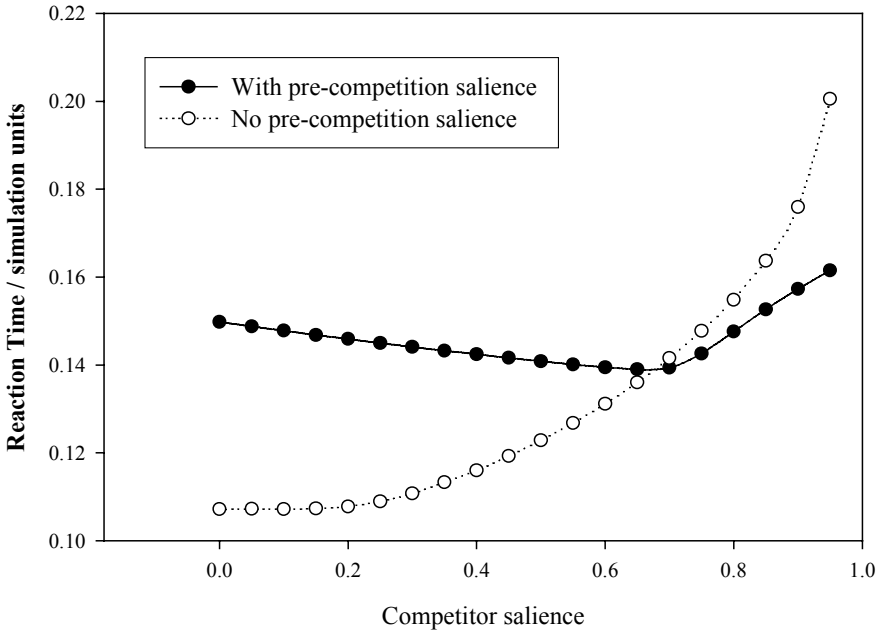


Figure 33: Competitor facilitation in the BG model

This phenomenon forms an unverified prediction from the BG model. Because it is extremely counter-intuitive, and not predicted by any other theory, it is one of the best kinds of prediction to make. Popper said that scientific theory is advanced by the falsification of safe predictions and the verification of unlikely ones (Popper, 1963). Competitor facilitation - if verified! - would be a good example of this second kind of prediction. For this prediction to be verified a task needs to be found which conforms to the assumptions of the model. It must involve task switching which is endogenously motivated; the participant must stop doing the pre-task in order to respond to a new stimulus.

Initial pilot studies were inconclusive. A standard Stroop task was used as the competition task. Simulations showed that the difference between control and conflict stimuli in the colour-naming task, with and without a pre-competition task, would be suitable to reveal whether the competitor facilitation phenomenon was manifest. Choosing the pre-competition task was problematic. The task needed to be sufficiently demanding, but not involving motor or attentional demands that would interfere with the selection of the Stroop response in a way that would contaminate our assessment of the competitor facilitation phenomenon. It is difficult to know what salience any pre-competition task has when compared to the competition (Stroop) task. If the pre-competition task involves the same motor region as the Stroop task, the vocal chords in this case, then we cannot be sure that the physical demands of switching are not interfering with reaction times. If the task involves a different motor region from the Stroop task then it is less certain that the resulting saliences are comparable, or even that they would compete directly in the basal ganglia (See Gurney et al., 2001a; Redgrave et al., 1999). It is also less than certain which exact prediction the model makes for the Stroop conditions used. Without a pre-competition salience the salience-RT function is monotonic. This means that the exact values of the saliences used to provide predicted reaction times are not crucial; the qualitative pattern would be the same even if the saliences provided by the model front-end were of a different magnitude but maintained roughly the same relative proportions. In the condition where there is a pre-competition salience,

however, the salience–RT function is non-monotonic. This means that the exact salience values affect the prediction qualitatively as well quantitatively. We provisionally used the salience values from a standard Model 2, however the absolute salience values, as opposed to the relative salience values, are in this context crucially under-constrained; the model could be constructed in multiple ways that would fit the rest of the data but would provide different predictions on a test of the competitor facilitation hypothesis using the Stroop task.

The Cohen response mechanism does not predict competitor facilitation. Selection time in the Cohen response mechanism is a simple monotonic function of relative evidence (i.e. the difference between the target and the competitor salience). There is no reason why the existence of a pre-competition salience would affect the selection time differentially for different levels of competition. This relates to the fact that (as discussed in section 2.1.4), the response mechanism is designed for the somewhat artificial situation of a competition between two responses which finishes once one response is selected. There is no provision for the successive selection of actions and hence no need to deselect previous responses.

It is conceivable however that the Cohen response mechanism would produce competitor facilitation if operating in conjunction with the gradual, rather than instant, rise and fall of saliences from the front-end. Although the leaky integration function of the front-end had a negligible influence on reaction time for the Cohen et al (1990) simulation of the Stroop task, it is conceivable that this feature might lead to competitor facilitation in the same circumstances that the BG model produces the phenomenon. To test for this I simulated the gradual rise of the competition saliences and the gradual decay of the pre-competition salience in conjunction with the Cohen response mechanism. This meant that at the start of competition the saliences for all three responses co-existed at sub-maximum levels. This did not effect the behaviour of the Cohen response mechanism; no competitor facilitation resulted. This result would apply equally to the diffusion model (Ratcliff, 1978).

5.5. Discussion

The previous chapter demonstrated the validity of using the basal ganglia model as a response mechanism. Having confirmed this I was free to alter the front-end of the architecture to see, first, if it could be improved, and, second, to help illuminate what the essential features of the Cohen architecture are. The emphasis on differential strength of processing between word and colour processing was retained, but instantiated by having multiple routes in the word pathway, making it non-symmetric with the colour pathway. The success of this new architecture highlights the essential nature of differential strength of processing, but the inconsequentiality of how this is instantiated. Kanne et al (1998) have criticised the original Cohen model for utilising identical pathways for word and colour processing. This is, they point out, unlikely to be true in vivo, with word reading and colour naming likely to be using multiple different processes. They cite, for example, the work on word reading that informs the choice of architecture for Model 2. The Model 2 simulations show that the essential assumptions of the Cohen model are sound and provide a possible route for the extension of the model while addressing the criticisms of Kanne et al (1998).

The cognitively constrained front-end architecture also allows the simulation of manual response results, and it is interesting to see the predictions partially confirmed, as the findings of Melera & Mounts (1993) and Durgin (2000) seemed to suggest that they would. The experimental test of the Model 2 predictions validates the model and confirms its usefulness in making predictions that can then be verified by experiment.

As it stands, Model 2 combines constraints from two distinct levels of analysis. The basal ganglia model was developed independently, based on a systems level analysis of the neuroanatomy of that region and on the hypothesis that it functions as a selection mechanism. The front-end incorporates constraints from cognitively based theories of word reading. The ability to join the two is contingent on the assumption

that they share a common signal representation – namely that of saliency represented by a unit activity proportional to urgency of corresponding response. In addition it is pleasing to note that the function of the front-end is improved when the activation function of the individual units is made more neurobiologically plausible and changed from a sigmoid to a Weibul function.

6. DYNAMIC ATTENTIONAL INHIBITION

6.1 Motivation

The SOA simulations produced by Model 1 (Figure 23) and Model 2 (Figure 29), although an improvement on the original Cohen model simulation (Figure 13), still do not match the empirical results (Figure 2) in all aspects. Particularly, it can be seen that, for the colour-naming conditions, reaction times are longest between 0 and +100ms SOA, and decrease to constant, and smaller, values at SOAs below –100ms. Although Models 1 and 2, unlike the Cohen model, replicate the flattening out of reaction times at low SOAs, they do not produce the 'hump' around 0 ms that appears in the empirical results. Given the replication of the original SOA data of Glaser & Glaser (1982) by others (Schooler et al., 1997; Sugg & McDonald, 1994) it can be assumed that this limited time-window of maximal interference is a reliable phenomenon. It is therefore justified to expect any model of the Stroop paradigm to produce this time-window effect.

Not only do the previously discussed models in this thesis fail to replicate this transient rise in reaction times, but they *cannot* replicate it, in their current form. In the models the modulation of signals evoked by the stimuli occurs concurrently with the passage of the signals along the pathways. No possible mechanism exists which could reduce the signal strength at a subsequent time point. Output signals from the front-end could not be less at longer SOAs, which is what the decreases in interference in the empirical data seems to suggest needs to happen. A sufficient feature for this to occur would be delayed inhibitory feedback.

The model of Phaf et al (1990) did replicate the relevant SOA data. Inspection of the model reveals that this is due to the activation function of the units used by Phaf et al (1990). As can be seen from Figure 34 the activation function of this model produces an initial response to an input (stimulus) that is greater than the equilibrium response to that input. This momentary initial response allows the

model to produce greater interference on colour naming by word reading for a limited time window.[6] Including such an activation function is one way of replicating the SOA phenomenon, and perhaps could be reviewed as reflecting adaptation processes in biological neurons. The models in this chapter, on the other hand, do not have intrinsic unit activation functions that follow such a pattern, but rather have patterns of such activation that emerge as a feature of feedback dynamics.

[6] The duration of this time window in the (Phaf et al., 1990) model is determined by the arbitrary time assigned to each network iteration, in this case each iteration is taken to be 25 ms.

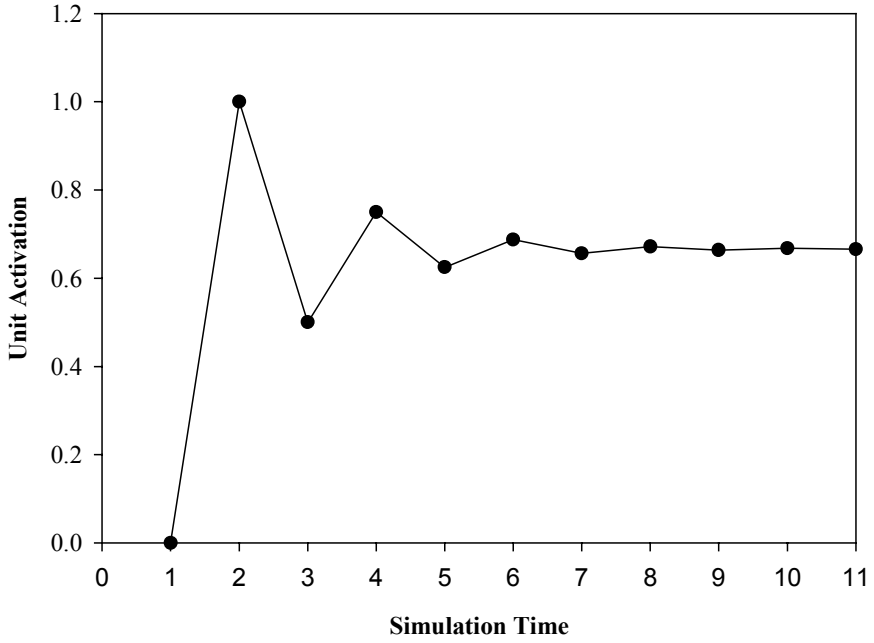


Figure 34: Unit activation function used by Phaf et al (1990).

I supposed that this transient hump reflected the operation of some kind of delayed attentional inhibition. The co-occurrence of the two stimulus dimensions presents the condition in which it is hardest to ignore the irrelevant dimension, whilst at longer negative SOAs it is possible to adjust to the presence of the irrelevant dimension and therefore minimise its effect on reaction times. Obviously at long positive SOAs the irrelevant dimension appears significantly after the response has been initiated, based on the relevant dimension, and so has no affect on the response. Within the context of our connectionist models this kind of attentional inhibition is best characterised as increased inhibition on the irrelevant pathway, and hence on the to-be-ignored stimulus representation.

There is supportive cognitive and neurophysiological evidence for the development of exactly this kind of stimulus-evoked attentional inhibition. Furthermore, this evidence also seems to suggest that 50 to 100ms is the likely time-scale of such attentional inhibition. I discuss this evidence below.

6.1.1. Cognitive level: negative priming

Negative priming is caused when a stimulus that is not the focus of selection on a probe trial subsequently causes an increased latency of response to that stimulus on a target trial. Debate exists in the literature about the mechanism behind negative priming (see Fox, 1995; and Neill, Valdes, & Terry, 1995, for reviews). One proposed mechanism is that of selective inhibition. Thus ignoring a stimulus (known as the 'probe') generates inhibition which persists to produce negative priming when an identical stimulus property, such as shape or location, is later the target. The locus of this inhibition is also in debate. Inhibition of response output cannot account for negative priming on its own (Neill et al., 1995). Thus, that inhibition at a perceptual level – i.e. as part of some sort of attentional process – is involved in negative priming is a reasonable assumption.

There is, however, evidence to contradict this view. For example Wood & Milliken (1998) found evidence of negative priming in a paradigm which did not involve the act of ignoring. Fox (1995) provides a review of competing explanations for the phenomenon of negative priming, including the attentional inhibition hypothesis. Tipper (2001) provides a more recent discussion of the evidence which it purported to call into question the inhibitory mechanisms account.

The phenomenon of negative priming was actually first demonstrated using a Stroop task (Dalrymple-Alford & Budayr (1966) cited in Neill (1995)). Lowe (1985, cited in Neill (1995)), using a SOA-Stoop paradigm similar to that of Glaser & Glaser (1982), found that negative priming was not present in the 50ms SOA condition but was for longer SOAs up to 400ms. This time course is interesting. If inhibitory mechanisms are involved in negative priming then the absence of negative priming at 50ms SOA indicates an absence of inhibition at that interval as well. Other work with negative priming paradigms supports the view that inhibition typically comes on after initial processing of stimulus, after about 100ms (May, Kane, & Hasher, 1995).

So, from these studies, we may conclude that inhibition of an irrelevant perceptual feature, response or location can persist after the evoking stimulus has disappeared, and that this inhibition takes over 50ms from the first appearance of the irrelevant stimulus to come into effect.

6.1.2. Neural level: visual attention

There is evidence from single-cell recording of cells involved in perception that attentional inhibition takes a similar amount of time to develop. Chelazzi et al (1998) report the initial activation of cells responsive to non-target stimuli before the development of inhibitory suppression after approximately 100ms. The authors note that this delay in inhibition is in line with the Theory of Visual Attention (Bundesen, 1990) which proposes initial parallel activation of units before biased competition between units mediated by attention (see Duncan, Humphreys, & Ward,

1997, for recent review). Usher & McClelland (2001) discuss transient peaks in the activation of to-be-ignored stimulus representations and ascribe the delayed inhibition to recurrent inhibition which is only activated by input to the representing module. Such stimulus-evoked inhibition exists because inhibitory connections tend to be local, within-module, unlike excitatory connections which are more likely to be between modules (Crick & Asanuma, 1986).

If attentional inhibition relies on within-module feedback mechanism it would necessarily be delayed with respect to the stimulus appearance. Single cell recordings (Chelazzi et al., 1998) and neurally based attentional theory (Duncan et al., 1997; Usher & McClelland, 2001) indicate that such inhibition occurs approximately 50ms to 150ms after the occurrence of stimulus related excitatory input to a module.

6.2. Implementation

Currently the attentional mechanism in the front-end of the models - the first ('soft') selection gate – consists entirely of pre-stimulus, preparatory, attentional inhibition. If we augment this preparatory attention with stimulus-evoked, dynamic, attentional inhibition, it seems natural to locate the two at the same point in the model. In the simulations described below the dynamic attentional inhibition consists of a stimulus-dependent change in the bias on the first stage units in the irrelevant-stimulus pathway. 100 (simulated) ms after the appearance of the irrelevant stimulus the inhibitory bias was increased from its normal value by a constant amount (0.9 in Model 1 and 0.05 in Model 2). The results shown below could have been finessed if the delayed attentional inhibition gradually developed, instead of coming on as a step function, but this seemed an unnecessary complexity. We are only interested in establishing that the basic trend can be captured in this manner at this stage.

6.3. Results

The addition of dynamic attentional inhibition, as discussed above, improves the SOA simulations in both Model 1 (Figure 35) and Model 2 (Figure 36). Reaction times in the colour-naming conflict condition now peak around the 0ms SOA point, and flatten-off at a lower level, as occurs in the empirical results.

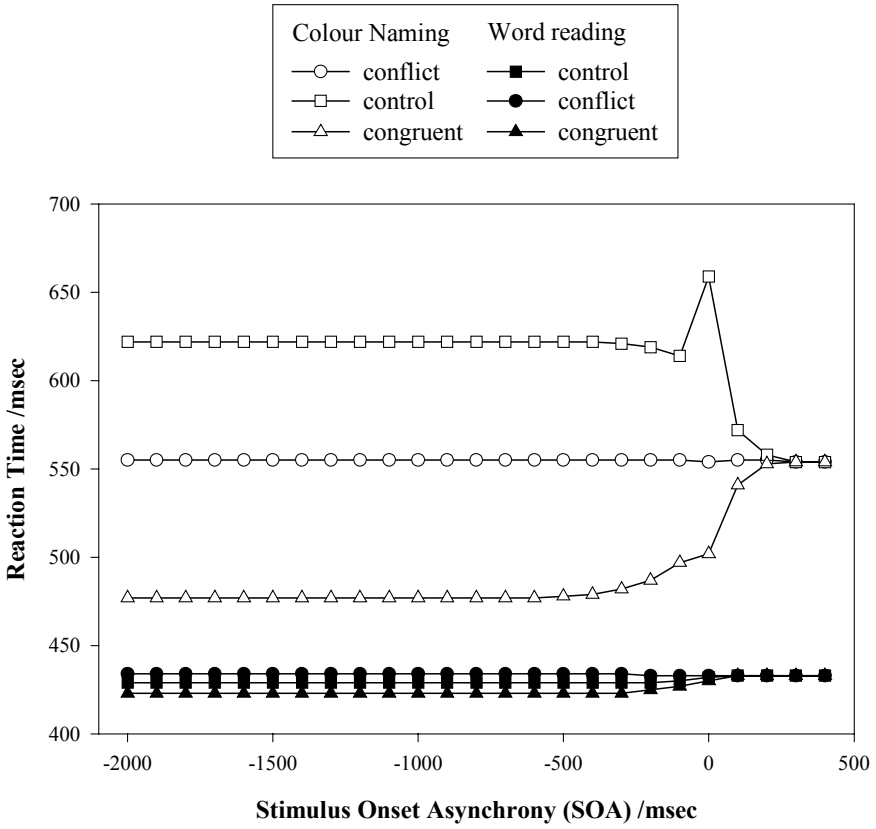


Figure 35: Simulation of SOA results with Model 1 and the addition of dynamic attentional inhibition.

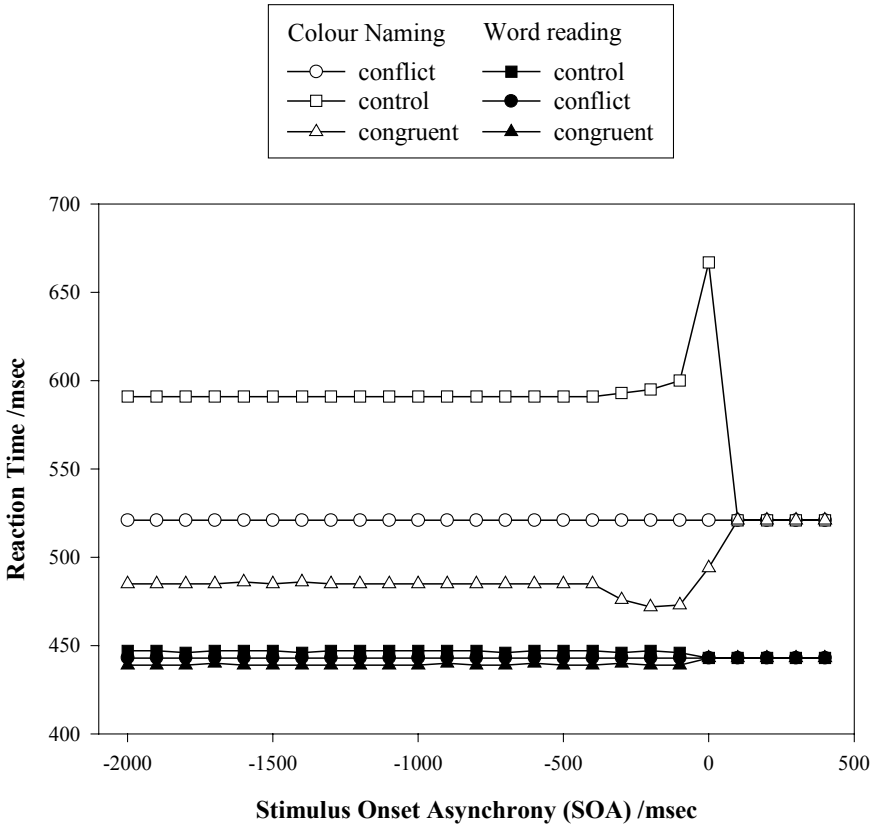


Figure 36: Simulation of SOA results with Model 2 and the addition of dynamic attentional inhibition.

It is worth noting that the addition of dynamic attentional inhibition does not improve the SOA simulation with the original Cohen model (Figure 37). Although the delayed attentional inhibition produces a peak in the colour-naming conflict condition around 0ms SOA, the colour-naming condition reaction times do not level-off beyond 100ms. This is because the response mechanism (discussed above in section 2.4.2) still causes the reaction times continue to be influenced by the irrelevant stimulus, and at long enough SOAs come to provoke erroneous responses, just as in the original Cohen model.

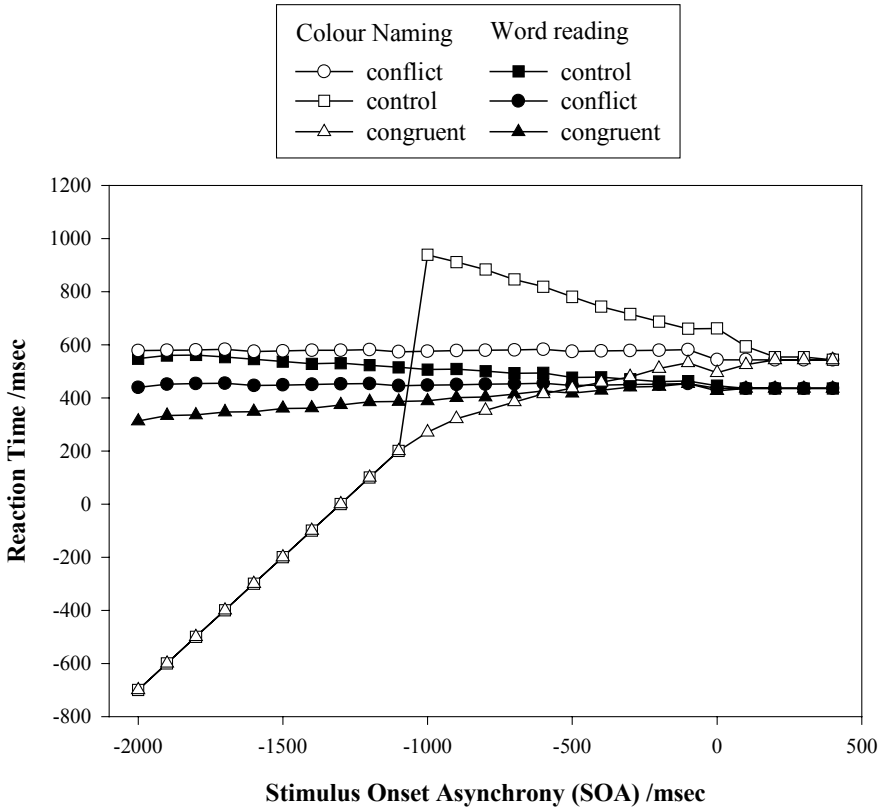


Figure 37: Original Cohen model SOA simulation over the extended range and with the addition of dynamic attentional inhibition.

6.4. Discussion

These simulations show that the addition of stimulus-evoked attentional inhibition is a possible way of introducing the kind of non-monoticity into the reaction time patterns of the SOAs simulations that is present in the empirical results. Although this non-monoticity could be introduced in other ways (such as in the manner of Phaf et al (1990)), stimulus-evoked inhibition seems supported by the cognitive evidence on negative priming and the neurophysiological evidence on visual attention.

In addition, it makes functional sense for inhibition to be tied to external events rather than to be a mandatory property intrinsic to the units themselves. Usher & McClelland (2001) discuss the connectivity typical of brain-architecture. Specifically they note that between-region connections tend to be excitatory, while the majority of inhibitory connections are within-region. This means that any inhibition is reliant on input to the region first evoking activity in a population of inhibitory inter-neurons, and thus it will consequently suffer a delay relative to excitatory input. It has been suggested that in the Stroop task the anterior cingulate cortex performs this function (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Carter et al., 1998; MacDonald, Cohen, Stenger, & Carter, 2000; Paus, 2001) while the prefrontal cortex provides a preparatory attentional set that gates information via changes in baseline activation levels (Banich et al., 2000; Buchel & Friston, 1997). Cohen has specifically suggested that his original model requires augmentation by a conflict monitoring function (performed by ACC) which could increase attentional control upon the occurrence of conflicts (Botvinick et al., 2001).

Introducing a stimulus evoked inhibition not only allows correct selection for all SOA values, but also has the side effect of making the SOA data more closely fit the empirical data. This indicates that the transient peak of interference in the SOA paradigm may be related to the delay in the activation of attentional inhibition. The

appearance of a to-be-ignored stimulus provokes activity which in turn evokes it's own suppression.

Dynamic attention inhibition does not fulfil the same function as the minimal salience threshold introduced by the basal ganglia response mechanism (section 3.2), although they produce similar effects. For small SOAs dynamic attentional inhibition can prevent wrong selection due to irrelevant stimuli, just as a minimal threshold can (if the preparatory attention suppresses irrelevant stimuli sufficiently to cause the saliences to be below the minimal threshold). However, because the units of the model use a continuous activation function, even when preparatory and dynamic attention are combined there is still some residual salience caused by irrelevant stimuli. Without a minimal threshold (or a decay factor, see section 7.2.3, iii) this, given enough time, will provoke a response.

7. GENERAL DISCUSSION

In the course of investigations into the Stroop task I have developed and tested several different models, and multiple variations on each of these models. It is now appropriate to discuss how the modelling work has clarified issues concerning Stroop processing and action selection (sections 7.2 to 7.5). I will also attempt to derive some general lessons for the use of computational modelling in psychology; how models should be developed, their exposition and their comprehension (sections 7.6). Finally, I will summarise the limitations and potential for development of the work presented in this thesis (sections 7.7 to 7.8). I begin with a summary of each of the chapters in this thesis.

7.1. Summary

Often a single main conclusion can be derived from the investigations covered by each chapter. Obviously, each chapter also contains some lesser notes for which the space cannot be afforded to mention here. The chapters of this thesis detail the evidence for the following conclusions:

- Chapter 1: The problems of implementing selected responses are under-estimated by psychologists, and adequate solutions need to be included in any proposed model of the Stroop task.
- Chapter 2: It has not previously been recognised that Cohen et al's (1990) model relied on the response mechanism to correctly simulate reaction times. Their model provides a good starting point for the investigation of response selection in the Stroop effect using the BG model.
- Chapter 3: The basal ganglia can be considered as a response mechanism. Our model of the BG follows Pieron's Law in relating strength of input to reaction time.

- Chapter 4: The basal ganglia model provides an suitable response mechanism for the simulation of reaction times in a model of the Stroop task; this is despite the fact that it was developed independently of this purpose based on the known neuroanatomy of that region.
- Chapter 5: Information processing models of word-reading can be co-opted into a connectionist model of Stroop processing, and demonstrate that the essence of these models is greater strength of processing for word-reading than colour-naming, aside from the specifics of whether this greater strength is instantiated in single or multiple pathways.
- Chapter 6: Preparatory attention, although well supported by a variety of evidence, is insufficient to deal with the attentional problems created by the Stroop task. The human attentional system probably involves stimulus-evoked inhibition of the representation of distracting stimuli.
- Chapter 7: Computational modelling can be embarked upon with multiple purposes, and the failure to be explicit about the purpose of a modelling venture can lead to misplaced criticism of the model. Most important to the clarity of modelling work is making explicit the core assumptions on which the model is based and which constrain its construction. The value of a model lies not in any particular form it takes, but on the relationship of its core assumptions to possible behaviours and on the potential for development of the model.

These conclusions lead, respectively, to the following speculations:

- The properties of the human response mechanism have a pervasive influence on the performance of selection task.
- The basal ganglia play a key role in the selection of behaviours.
- Manual responding in the Stroop task will lead to a 'reverse Stroop effect'. This is confirmed experimentally.
- Under the right conditions the competitor facilitation phenomenon could distinguish between the basal ganglia model and evidence accumulation models.
- Attention is closely related to behaviour, and involves multiple, distributed, mechanisms. Because of the close coupling of perception and behaviour

suggested by the 'selection-for-action' perspective (Allport, 1987) we expect the basal ganglia, traditionally thought of as a motor control region, to be closely involved in attentional and other cognitive tasks.

7.2. Essentials of models presented

It is important to draw out the core assumptions of the theory that a model is based on (as discussed in section 1.7.1). Often these will be explicit, but it can be that investigation of the function of a model brings to the fore critical features it possesses which were hitherto neglected. Below I discuss the core theory-relevant features that my models incorporate.

7.2.1. Pathways and channels

The implementations of the two pathways in models 1 and 2 are relatively neutral on the neuroanatomical basis of word and colour processing. The simplest possible generic architecture was adopted by Cohen et al (1990) and used here in model 1. However, as the simulations with model 2 demonstrate, this architecture can be replaced by a more sophisticated one based on theories of word reading. The only essential features are two pathways with separate attentional modulation and in which word and colour information can be processed at differential strengths - either by multiple routes or greater weights in a single pathway.

For model 1 the division of the front-end into two pathways allows learning with a reduced training set (section 2.3.2). The fixed division of the PDP front-end into two pathways allows the Cohen model to learn the correct response for the full set of Stroop stimuli, despite being trained on a reduced set of stimuli. These training simulations show that the attentional gating mechanism is more important than the weights in dividing the network into two halves.

The crucial features of the model are contingent on there being two separate pathways for colour naming and word-reading. Firstly, the division of the word-reading and colour-naming allows the differential operation of attentional modulation in the two pathways. Secondly, it is possible for the two functions to have separate weight values, so that the strength of processing therein is different. This is the instantiation of automaticity within the model theory. Model 2 shows that there is more than one way to implement this difference in relative strength. The multiple routes model is equivalent to the original Cohen model front-end, not a conflicting account. In the absence of other considerations, parsimony would suggest favouring the minimalist front-end of Cohen et al (1990). However, wider theoretical considerations have led to the incorporation of models of word reading. Extending the model front-end allows greater contact with a body of data on word reading, so that, for example, Stroop performance from people with specific lesions could be predicted. In addition I have successfully simulated manual Stroop results.

7.2.2. Attention

The original front-end relies upon preparatory attention, which gates activation in the two pathways independently of the level of activation in those pathways – i.e. there is a fixed bias and a fixed weight from the task-demand input. For the purpose of the model it was a theory-unspecified choice whether the attentional input was inhibitory or excitatory. It was done so that attentional input was excitatory on units whose resting activity was low. This seems neurologically plausible and supported by the results from functional imaging discussed above (e.g. Brefczynski & DeYoe, 1999; Gandhi et al., 1999; Heinze et al., 1994; Kastner et al., 1999). For the model, the key feature was that the attentional input primed the activity of one pathway relative to the resting activity of the alternate pathway.

While functional imaging studies have provided support for this preparatory view of attention, it is apparent that in our models the closest replication of all the empirical results required an additional kind of attention, which I call dynamic or stimulus-evoked. Others have described similar phenomenon which they call ‘on-line

adjustment of cognitive control' (Botvinick et al., 2001) or 'brief attention' (LaBerge, Auclair, & Sieroff, 2000).

Introducing a stimulus evoked attentional inhibition not only allows correct selection for all SOA values, but also has the side effect of making the SOA data more closely fit the empirical data. This indicates that the transient peak of interference in the SOA paradigm may be related to the delay in the activation of attentional inhibition. The appearance of a to-be-ignored stimulus provokes activity which in turn evokes it's own suppression.

During the preparation of this thesis, two of the authors of the original Cohen et al (1990) paper independently suggested just such a mechanism (Botvinick et al., 2001; Usher & McClelland, 2001). Usher & McClelland (2001) discuss the neurophysiological data from the same lab (Chelazzi et al., 1998) as I do. Drawing upon this, and upon the observation that inhibitory connections tend to be within-module, they propose that lateral inhibition is a general feature of models of perceptual choice. This results in the transient rise, and then suppression, of the activity of the unit(s) associated with the representation of the to-be-suppressed response. This suppression presumably plays a role in negative priming. Although I approach the topic from the other end – models of response choice – the conclusions coincide to a large degree (discussed in more detail below in section 7.2.3,iii).

Botvinick et al (2001) present a model of conflict monitoring and cognitive control. They argue that the anterior cingulate cortex functions to detect conflicts and dynamically implement cognitive control. In their model the co-activation of incompatible representations triggers additional attentional inhibition. As with my models, this additional attentional inhibition has the same locus as the preparatory attentional input, and is effectively just an increase in the 'task-demand' signal. Botvinick et al (2001), following from (Cohen et al., 1996) locate the source of the task-demand signal in the prefrontal cortex, to which they ascribe the function of maintaining a general context signal or 'attentional set'.

7.2.3. Response mechanism features

In the models presented here, the Stroop effect arises because of a response mechanism which uses the relative strength of conflicting decisions, not because word or colour information are processed at different speeds. In my models, as in the empirical data, the speed of processing before the response mechanism bottleneck can have an affect on reaction times. The operation of attention ensures that the nature of the final response is determined by the relevant dimension. The flattening of the reaction times for long negative SOAs shows that the human response mechanism, like the basal ganglia model, possesses the property of clean-switching. Thus, although irrelevant information may arrive first, it has a limited distorting effect on the selection of the response due to the relevant information. Outside of a small range of temporal coincidence between the appearance of the irrelevant and relevant stimuli, this distorting effect is constant.

Consideration of the wider action selection problem provides a number of additional constraints on the performance of response mechanisms.

The crucial feature of the basal ganglia - that which allows it to function in the context of simulating reaction times for the Stroop task - is that the time to selection is affected by the relative magnitude of the competing input saliences. Both the basal ganglia and the Cohen model response mechanisms follow Pieron's Law, but with relative evidence instead of intensity of stimulus determining response time. The difference between the inputs was the sole determinant of reaction time in the original Cohen model response mechanism, and although the basal ganglia model is a good deal more sophisticated, the difference between inputs is still the major determinant of response times. It is promising that, although the basal ganglia model was developed independently of any idea of simulating reaction times, it conforms to Pieron's Law.

Further, I have shown that certain features of the model of Cohen et al (1990) can only be properly understood if attention is paid to the response mechanism.

Specifically, it may be that the response mechanism is the locus of the fact that interference is greater than facilitation. The time for the two conditions (conflict and congruency) to be processed is equal in the front-end, but the different salience values produced create different response times in the response mechanism. The recently proposed LATER model (Reddi & Carpenter, 2000) also conforms to Pieron's Law. This model springs from investigations on the influence of urgency on decisions to make eye movements, and from investigations of the underlying neurophysiology. In fact the Cohen response mechanism can be considered a reduced version of the LATER model with some parameters fixed. The BG model, however, also accounts for absolute salience levels in calculating response times (and importantly has an absolute salience threshold below which no response is made). As discussed in section 3.5, the BG has different response time functions for the same relative salience between different absolute saliences. However these functions all fit Pieron's Law. I have also shown that the Phaf et al (1990) response mechanism - along with the later and diffusion model response mechanisms - conform to Pieron's law (Stafford & Gurney, manuscript submitted to Psychonomic Bulletin & Review).

Our analysis therefore suggests two things: that Pieron's law may be explained in terms of an underlying response mechanism; conversely, any successful response mechanism must accommodate Pieron's law. This is no less true for response mechanisms incorporated in models of the Stroop task. The basal ganglia model is not an exception to this, producing an input-RT function closely fitting an analogue of Pieron's Law. More generally, since Pieron's Law seems to be a common feature of successful models of the human response mechanism, the human response mechanism may have an ubiquitous influence on response times.

i - The Single Mechanism contention

The reinterpretation of the major extant models of the Stroop task (Cohen et al., 1990; Phaf et al., 1990) allows a reconciliation of the seeming contradiction between these models and the criticisms of 'single mechanism' explanations of interference and facilitation by MacLeod (e.g. 2000). If the asymmetry of interference and facilitation is due to the response mechanism (which operates on inputs after they have been affected by attention and the integration of visual information) then it is possible to parsimoniously explain the asymmetry while still allowing the dissociation of the two phenomena by experimental manipulations which (putatively) affect the initial processing of the Stroop stimuli.

The asymmetry of the magnitude of the two phenomena can be seen as a natural consequence of a response mechanism which follows Pieron's Law – increases in reaction time from some baseline are obtained with less change in the input than corresponding decreases in reaction time. Results which show the dissociation of interference and facilitation can be considered to reflect the differential influence of the experimental manipulation on the initial processing of the stimuli. Tzelgov, Henik & Berger (1992) found that interference, but not facilitation, increased as the proportion of colour-word trials decreased. MacLeod (1998) found that two manipulations – practice and integration versus separation of stimuli – strongly affect interference but not facilitation. In the context of the Cohen et al (1990) model these manipulations would be thought of as affecting the processing done before the response mechanism, i.e. that concerned with strength of processing (due to learning) and attentional function. They do not affect the assertion that, in the basic Stroop task, interference and facilitation are of different magnitudes because of the response mechanism. The non-linearity of the response mechanism input-RT function, in conjunction with equal increases and decreases in the relative strengths of evidence provides a parsimonious explanation for the relative size of the interference and facilitation effects in the standard model simulation. There is no principled reason, however, why the increase in evidence in the conflict condition

and the corresponding decrease in the congruent condition cannot vary independently.

Consideration of Pieron's Law makes clear an additional reason to be unconcerned about the seeming dissociation of interference and facilitation. Because facilitation relies on change over the least rapidly changing part of the reaction time function, any further increases in facilitation will be diminishing, in contrast to interference effects which show an accelerating increase. Therefore tasks which evoke changes in interference might not be associated with significant changes in facilitation. Facilitation is small relative to absolute response times and relative to the amount of noise in empirical measurements. Thus, changes in the amount of facilitation may often be too small to detect reliably. To the best of my knowledge it has not been shown experimentally that the two phenomena vary independently, only that the amount of interference can vary without significantly affecting the amount of facilitation (MacLeod, 1998). To truly demonstrate the independence of the two phenomena it would be necessary to show an increase in facilitation without a corresponding increase in interference.

The current resolution between the dissociation of interference and facilitation, on the one hand, and the parsimonious explanation of the asymmetry of their typical magnitude, on the other, relies on distinguishing multiple, but integrated, stages of processing in the Stroop task. My contention is that interference and facilitation may involve different processes in the first stage, but the fundamental source of the asymmetry in their magnitude is common to both phenomenon – namely, a (second stage) response mechanism which follows a Pieron's Law-like function to relate inputs to reaction time.

ii - Pieron's Law and information integration at the neuronal level

In the biologically grounded basal ganglia model, high-level Pieron's Law-like behaviour is based on the properties of the component units. This raises the possibility that Pieron's Law itself may be based on the information integration

properties of individual neurons. Note that the particular shape of the function for the BG model (shown in Figure 15) resembles that of the model neuron function (shown in Figure 16). This gives some support for the hypothesis that the Pieron's Law properties of the basal ganglia model are due to the properties of its fundamental units, rather than being an anomalous result of the particular connectivity of the system. I propose that any system comprised of units with the same dynamics as model neurons such as these will, by default, follow a Pieron-Law like selection function. I include in this category, of course, the neural systems involved in decision making in the human brain.

It is usually acknowledged that the function of neuronal elements is substantially more complex than that expressed by the simple equation (5) used to model the units in this paper. However Koch (1999) points out that the complexity of multiple non-linear intra-neuronal processes can combine to create a simple linearity in the input to mean-firing rate relationship. Thus, the function of some biological neurons may be approximately equivalent to that of the units described here.

The neuronal explanation of Pieron's Law can be contrasted with an emergence based explanation. Rather than systems-level properties emerging from the interaction of fundamental elements that don't themselves follow Pieron's Law, the systems-level property is carried through from the properties of the fundamental elements. We could call this kind of explanation 'transparent'. Such explanations are more robust to minor modifications of systems-level features of the model than emergent explanations.

iii - Additional necessary functions of a response mechanism

The erroneous selection produced at long SOAs shows that a response mechanism must not make selections based on inconsequentially low inputs. Our basal ganglia model avoids this by having a minimum salience threshold, below which no action is selected. Cohen at al's evidence accumulation mechanism has no such threshold, and no decay of accumulated evidence, and because of this it always made a

selection if left for long enough. Without an input threshold, the evidence accumulation counter increases and increases towards the selection level. Usher & McClelland (2001), in their model of perceptual choice, present an alternative strategy to a minimal input threshold, but one which has the same functional role. They argue that models of perceptual choice – they discuss the same kind of choice algorithms that are the basis for the Cohen et al (1990) response mechanism (Luce, 1986) – require the addition of activation decay on the choice representations. A decay mechanism can fulfil the same role as a minimal difference threshold, since for situations where input is less than the decay that input is effectively filtered out.

Putting response mechanisms in the wider context of adaptive control provides additional features a suitable response mechanism must possess. As discussed above (section 1.2.1) a response mechanism needs to work in real-time, continuously and dealing with the successive selection of actions and interruption of old actions by new. It is precisely because the BG model is designed to operate continuously that it has equilibrium final states. The basal ganglia model possesses equilibrium states in which no action is selected. All patterns of input, if unchanging, eventually produce unchanging output states (although such a situation is unlikely to arise). For some patterns of input, the final output state indicates that no action is selected. The evidence accumulation response mechanism, on the other hand, has only one type of final state – that of selecting an action – and it continuously moves towards this state. The existence of equilibrium final states allows the successive switching between actions, without those actions interfering more with the selection of new actions the longer they have been selected. That the basal ganglia response mechanism does not allow previous signal values to carry potentially unlimited weight when selecting new actions allows another solution to the selection problems involved in the SOA task; a 'no-go' signal can be provided by the front-end to the basal ganglia on a third channel. This 'no-go' signal, by being itself selected, can prevent selection until the relevant stimulus dimension has appeared.

In negative SOA conditions the interference on reaction time does not get progressively longer with increasing SOA, but instead levels off – there is a

maximum amount of interference that a distracting stimulus can produce on reaction times. So not only is the wrong response not selected, but also the right response is selected efficiently. This is an example of the clean switching property which the basal ganglia possesses. The minimal input threshold of the basal ganglia model also plays a role in this.

7.3. Automaticity

I fully endorse the notion of a continuum of automaticity (MacLeod & Dunbar, 1988) and the PDP framework for exploring it (Cohen et al., 1990). I have shown how the notion of strength of processing can be varied and extended whilst retaining its essential function within the model of the Stroop task. These models suggest that an increase in the automaticity of processing does not involve a shift in the underlying loci of the function, but instead a quantitative shift in the efficiency of the performance of the function. This is supported by recent functional imaging work (Jansma, Ramsey, Slagter, & Kahn, 2001).

Previous results (Cohen et al., 1990; Cohen et al., 1992) show that a connectionist account of automaticity can incorporate contextual and relative influences on what is putatively automatic processing. The simulations in this thesis of manual responding and the reverse Stoop effect show that the connectionist account of Stroop processing can be extended to account for stimulus-response compatibility.

Paying proper attention to both the stimulus processing and the response mechanisms in a model of Stroop processing makes this more sophisticated account of automaticity possible. The continuous nature of activations and weights in the front-end allows different stimuli, and different stimulus-response combinations, to be processed in differing degrees of strength. In conjunction with the emphasis of the response mechanism on relative saliences, this allows the manifestation of relative automaticity as found experimentally by Dunbar & MacLeod (1988) and discussed theoretically by Cohen et al (1992).

The manual results reported in simulation 5 support the view that automaticity is not just an intrinsic property of the stimulus, but depends upon the stimulus-response mapping required by a task (Durgin, 2000; Zhang & Kornblum, 1998). The strength of connection between arbitrary stimuli and responses depends on number of times they have been paired. Durgin's (2000) results show that the pairing of colour-words and vocal responses does not necessarily confer any automaticity on the pairing of colour-words and manual responses.

Other results (Besner et al., 1997; Dishon-Berkovits & Algom, 2000; Huguet et al., 1999) demonstrate that manipulation of attention affects automaticity in the Stroop task. Besner et al (1997) affected attentional set, for example, by colouring only a single letter of the target stimulus rather than the whole word, and found that this reduced Stroop interference. Dishon-Berkovits & Algom (2000) showed that the frequency of correlations between the value of the two dimensions in the stimulus affected Stroop interference, showing the influence of expectations on attention. Huguet et al (1999), showed that the presence of others can reduce Stroop interference, though what they term 'social facilitation inhibition'. Together these results question the context-invariant nature of automaticity. It is also worth noting that such results can easily be accounted for without this framework, in which attentional gating interacts with inputs of different strengths. The experiments mentioned above could all be considered to produce their effects via the manipulation of attention.

In summary recent results suggest a deconstruction of the notion of automatic *processes* per se. However, with the help of computational models, the notion of automatic *processing* retains value. Automaticity is not a property intrinsic to a mental task like word-reading, and is not unaffected by the vagaries of context. Instead automatic processing is an emergent property of the interaction of training, attention and task within a specific context. Whether or not the particular properties associated with automaticity manifest themselves depends on the interaction of these factors – automaticity is not automatic!

This 'emergent' hypothesis of automaticity fits with the evidence, discussed in the introduction, which suggest that automaticity is not a unitary phenomenon (e.g. Bargh, 1989). The concept of automaticity can be sensibly decomposed if attention is paid to the functional organisation of the nervous system. Some actions are automatically elicited by a stimulus, due to the strength of connection between the stimulus and the response. More specifically the action is involuntary, rather than merely being an automatic action. Other actions may proceed without intervention once they are deliberately initiated, but not before then. Such actions are probably those for which a distinct motor program has been developed. This 'motor chunking' allows the actions to be selected as one unit, rather than merely the sum of the component actions that make it up. The acquisition of a composite response which can be evoked by the signal of one of these channels is assumed to be part of what it means to learn an automatic response. This chunking brings with it enhanced fluidity, speed and efficiency. According to this hypothesis the existence of a motor program to be activated is a precondition for the presence of other distinct, but associated, elements of an automatic action – namely, being involuntary, effortless and ballistic. Because automatic actions are executed as single learnt units by a previously associated cue they are vulnerable to disruption due to changes in either the content of the output required (e.g. changing the response modality from oral to manual) or changes in the nature of the stimulus which evokes them. Less automatic, controlled, processes are less fluid but more flexible precisely because they are not implemented as indivisible units.

One putative locus for motor chunking is the basal ganglia (Graybiel, 1998), although it is not covered in our model, which only deals with the expression of motor outputs, leaving aside the complexities of including the learning dynamics in the basal-ganglia-cortical circuit. Our model assumes that there are channels which are able to transform cortical processing of colour-related and semantic information into the motor actions required for a response in the task (uttering a name or pressing a lever). Since we can, indeed, execute such responses with facility when there is no interference, the existence of such channels would seem to be a

reasonable assumption; the details of their acquisition as a developmental process, while of real interest, does not concern us here.

The framework for automatic processing which was established by Cohen et al (1990) makes little progress towards understanding the nature of controlled processing. Other than being non-automatic processes, controlled processes have no characterisation in the model. Controlled processes can be considered as cases where parallel distributed models of cognition break down, or come to be equivalent to serial, symbolic, conscious processes, which require the continuous imposition of rapidly changing attentional control (see, for a discussion, Cohen et al (1990), simulation 4).

7.4. The biological basis of action selection

We have previously proposed the basal ganglia as the neural substrate of action selection (Redgrave et al., 1999). Obviously, however, any subcortical central switch which controls access to motor resources must work in tandem with earlier stage cortical processing. This is what the simplistic connectionist model front-end is supposed to represent. We have demonstrated that a model of the basal ganglia based on a reinterpretation of the functional architecture (Gurney et al., 2001a; Gurney et al., 2001b; Humphries & Gurney, 2002) possesses the desirable characteristics of a switching device. Further, in the work presented here, it is demonstrated that the basal ganglia model has further characteristics which are sufficient to simulate empirical data from Stroop reaction time paradigms. The analysis of how the basal ganglia model performs the role of response mechanism in our model of the Stroop task is revealing of the general requirements of the human response mechanism, independent of any thesis about the anatomical loci of that response mechanism. This is the benefit of assessing generic computational features. Conclusions about the necessary functions of a response mechanism will hold whatever the eventual consensus on the neural substrate of action selection.

Given this, it is rewarding that Usher & McClelland's (2001) recent work on perceptual choice models has converged on our conclusions for the successful attributes of a response mechanism, whilst remaining within the classic stochastic information accumulation framework. Usher & McClelland (2001) introduce a decay factor into their evidence accumulator, which ultimately has a similar effect to having a minimal threshold – rates of evidence less than the decay are cancelled out. Additionally they use self-excitation that parallels the positive feedback loops in the basal ganglia model, whilst retaining the basic principle that selection is based on relative strength of inputs. The principle of lateral inhibition in their formulation has the same effect as stimulus-evoked inhibition in our models: the representation of the to-be-ignored stimuli rises transiently and is then suppressed.

Recent work has focussed on the role of the anterior cingulate cortex (ACC) in response selection, especially in the Stroop task. There is converging evidence from imaging work (Bush, Luu, & Posner, 2000; MacDonald et al., 2000), neuropsychology (Turken & Swick, 1999) and neurophysiology (Davis, Hutchison, Lozano, Tasker, & Dostrovsky, 2000). It has been proposed that the ACC monitors cognitive conflicts, for example, as with the Stroop task where two stimuli demand different responses (Botvinick et al., 1999; MacDonald et al., 2000). It therefore seems likely that the ACC is a functional part of the brain's decision making apparatus, although it probably plays an early–stage (cognitive) role, more concerned with stimulus-stimulus conflict than response conflict (see, for example, Kornblum et al., 1999). If the ACC provides a control signal evoked by conflict, as suggested by the model of Botvinick et al (2001), it would be ideally placed to produce the kind of delayed attentional inhibition required to accurately simulate the transient peak of reaction times under SOA conditions, as discussed in section 2.1.1.

These studies should be considered against the caveat that imaging of the basal ganglia is problematic. The majority of functional magnetic resonance imaging (fMRI) studies use machines of strength 1.5 Teslar. This level of power makes it difficult to detect changes in basal ganglia activation that might be occurring (Tom Farrow, personal communication), although is sometimes reported (e.g. Gitelman et

al., 1999). The statistical methods used to analyse fMRI data mean that if changes in basal ganglia are not expected they may not be detected. In general sub-cortical regions are more difficult targets for fMRI, not least because of their size relative to the spatial resolution of fMRI. Additionally, it is not uncommon to save time while scanning by limiting the area of the brain scanned to regions above the sub-cortex.

Given this, the current focus – based on neuroimaging studies - on ACC involvement in response selection in the Stroop task should not be seen as contradictory to efforts in modelling basal ganglia involvement in the Stroop task. The modelling studies of stimulus-evoked inhibition (chapter 5) support the need for a conflict monitoring function, and show it working in tandem with the basal ganglia response mechanism. In effect the models of this thesis have two attentional gates. An early gate which operates soft selection, i.e. letting some information though from unattended stimuli (albeit with attenuated strength or for a limited time only), and a late (response) gate which operates hard selection, i.e. only one response is selected at a time.

These models lend further support to the idea that 'attention' is not a unitary function (Allport, 1993) and is due to the action of multiple brain loci (Duncan et al., 1997). Some of the classic dichotomies of attention research, early versus late selection and attention to objects versus attention to locations (Styles, 1997), seem to have ambiguities, contingencies and/or compromises as their answers. This seems to open one to a theoretical situation where 'anything goes'. Models such as the ones presented in this thesis provide a concrete basis for the investigation of these compromise positions. The switching of attention between stimuli which are driving action may be hard to distinguish from action switching. Indeed there are good reasons to suppose that perception and action are more closely intertwined than the stage divisions of traditional psychology would have us believe (Allport, 1987; Gibson, 1979; Goodale & Humphrey, 1998). An advantage of explicit computational models is that they allow theories developed in different fields, using different experimental methods, to be combined and integrated whilst retaining

theoretical rigour. Explicit models allow a firm footing in the middle of the road, so to speak.

The development of sophisticated models has great potential if used hand-in-hand with functional imaging studies. As well as unifying results and providing possible hypothesis, neuroanatomically grounded models could be verified against, and used to make predictions of, functional imaging work. Within PDP models the general activity, and change in activity, in each simulated region could be assessed, and so the model directly interpretable in terms of the changes revealed in imaging scans (see Horwitz, Friston, & Taylor, 2000; Horwitz, Tagamets, & McIntosh, 1999). As well as showing where the most activity in a system is predicted for a particular task, the incorporation of other neurobiological constraints may explain situations where a particular task does not increase region activity (for example in the case of inhibitory inputs, or of spatially distributed representations). One example of a connection between imaging and modelling suggests itself from the current work: the implementation of attention in the model, as an input like any other, brings to the surface a distinction that has been recently drawn by Coull & Nobre (1998). They highlight the distinction, based on functional imaging work, between the sources of attention (i.e. those regions which provide 'task-demand' input) and the sites of attention (i.e. those regions, notionally the hidden units in the models considered here, where attention operates). The recognition of this distinction becomes a necessary corollary of a anatomically grounded connectionist theory of attention.

7.5. Other theories of BG function

The basal ganglia has been ascribed a role in diverse cognitive functions (e.g. for a review see Wise et al., 1996). However, because of the close integration of the BG with frontal areas, it is not always clear how to ascribe distinct functions to the two distinct regions. The basal ganglia have been related to attentional function, both in preparatory attention (Casey et al., 2000) and in shifting attentional set (Ravizza & Ivry, 2001). Amos (2000) and Dominey & Boussaoud (1997) have proposed computational models of BG-frontal cortex interaction in the maintenance of a 'context signal' from the cortex which allows the production of appropriate behaviour based on the situation. Neither of these models include sophisticated BG architectures, instead being limited to modules for the BG input and output nuclei, omitting the 'indirect' or control pathway. Neither model includes feedback loops between the BG and cortex, via the thalamus.

Related to the maintenance of a context signal are models of BG involvement in working memory function (Beiser & Houk, 1998; Frank, Loughry, & O'Reilly, 2001; Monchi, Taylor, & Dagher, 2000). Similarly, none of these models include more complex architectures of the basal ganglia, instead being restricted to a single input and a single output nucleus. Frank et al (2001) make the interesting observation that BG architecture points to the occurrence of information compression between the input and the output; the number of striatal neurons is much larger than the number of neurons in GPi. They contend that this is because information from the cortex is integrated in the striatum and outputs from the BG are only used to signal 'when', rather than 'what'[7]. They propose that frontal cortex is responsible for the active maintenance of representations (in other words, for working memory) and that the BG selectively gates these representations and triggers their updating. This is compatible with our hypothesis of basal ganglia function as that of action selection (Redgrave et al., 1999). Our model focuses on

[7] This assertion should be balanced against the fact that striatum is topographically organised, which suggests that some minimal information ('where') is conveyed from cortex.

the motor aspects of action selection, but it is reasonable to assume that the cognitive aspects of BG function are due to the involvement of the BG in frontal and limbic loops (Alexander et al., 1986). Frank et al (2001) point out that extant selection-based models of BG function suffer because the striatum does not appear to perform a competitive function among inputs as most models contend[8]. However, our model enacts competitive function via two pathways between striatum and GPi. This demonstrates the benefit of a more comprehensive modelling of the BG nuclei.

The model of Djurfeldt, Ekeberg & Graybiel (2001), although computationally unremarkable, draws out the point that if the BG is involved in attentional set (via connections with pre-frontal cortex) *and* with selection then it may play a role in shifting cortical activity into different 'attractor wells' (i.e. attentional sets). This in turn may play a role in sequence learning. Sequence and/or procedural learning is also the focus of Nakahara, Doya & Hikosaka (2001). Their model of the BG is as simple as the others discussed above, but include a more differentiated model of PFC, allowing the interaction of visual and motor loops in that locus.

The importance of the basal ganglia in learning and automatic processing (see section 3.3.1) reinforces the potential benefits of including functional models of these structures in any modelling framework requiring decision processing. While we do not address the specific problem of learning in our model, we have laid the groundwork for future work that will.

[8] See Jaeger, Kita & Wilson (1994) for evidence against lateral inhibition in the striatum. See Tunstall, Oorschot, Kean, & Wickens (2002) and Koos, Tepper, Goldman-Rakic, Wilson (2002) for more recent evidence in favour.

7.6. Connectionism in psychology

PDP Models possess a large number of parameters and features which can be varied, often with little effect on the performance on the model. Some parameters will only affect performance if varied outside of a certain range, or if varied along with some other feature. Usually only a single model is presented in a published work, but that model will be part of a family of possible models. The possible variations of a model will all be linked by a set of common theoretical principles. It is these principles which allow the relation of the model to psychological theory. Like Dawson (1997) I believe more emphasis should be placed on the interpretation of model function, so that it can be related back to the theory level. Furthermore the theoretical motivation, and implications, of each model feature should be elucidated as much as possible.

According to this view we might say that a model is*,* in fact, the theoretical principles that motivate it, rather than its specific instantiation in a simulation. Not only should the theory-proscribed implementation choices be distinguished from the theory-neutral implementational choices, but the findings from the model should, by definition, be robust under theory-neutral variations of the model. Further, if possible, it should be made clear which features of the model cause which properties. The interactions of features and the effects of varying each theory-relevant parameter should also be made clear. The properties which necessarily result from the core assumptions of the theory – and therefore are the conclusions drawn from the model – should always be explicit. Necessary properties are the hard-core of the model, they constitute the axioms which the model is involved in constructing a proof from.

The onus should be on the authors of a model to present it in a form which makes clear what aspects of the architecture and which parameters are proscribed by theory, and which aspects can vary, and within what range that can vary. It is vital to establish what the essentials of a model are. Because connectionist models are

complex systems it is not always apparent what properties depend on which features. Plaut & Shallice (1993) provide an excellent example of good practice. They present a model of word reading, which relies on the existence of attractor dynamics in the mapping between orthographics and semantics. To demonstrate that the attractor dynamics are responsible for the interesting higher-level phenomena, rather than the specific modular arrangement of the model, the authors test alternate architectures for mapping from orthographics to semantics. By doing this they show that the theory-relevant implementational choice is attractor dynamics, rather than the theory-irrelevant choices of the arrangement of functional modules in their model. This goes some way to addressing the criticisms of Roberts & Pashler (2000) concerning the value of good fits in model testing. By experimenting with the possible alternative forms of the model the explanatory range is elucidated and it becomes clearer what other outputs are possible from the model. The ideas of Roberts & Pashler (2000) are extended, and their relation to the current work is discussed, in section 7.6.3.

7.6.1. Models as theories in the Lakatosian framework

Lakatos (1978) proposed a way of viewing progress in science that has application to the nature of computational models in psychology. Research programmes, Lakatos observed, are rarely abandoned completely due to some new piece of evidence. Instead they tend to be 'progressive' or 'degenerative', either gaining support or declining in popularity. What determines the health of a research programme is its capacity to account for new data whilst retaining the core assumptions of the theory. In addition to these core assumptions, a research programme is constituted by a periphery of 'soft' assumptions. These are open to modification, addition and deletion as circumstances require. The theory can be aligned with contradictory evidence by adjusting the peripheral assumptions of the theory, rather than forcing the rejection of the core assumptions. Lakatos recognised that a theory, whilst having hardcore commitments, must also make auxiliary assumptions that aren't always vital to the nature of the theory. Using the current investigations as an example, it can be seen that a number of elements of the original Cohen model have been altered, but the original assumptions of the model concerning automaticity and strength of processing remain valid.

In this Lakatosian framework we can begin to view models as clusters of related possibilities derived from a theory, rather as their single fixed form in any one instantiation. Thus, models can be seen as comparable to research programmes in the sense used by Lakatos (1978). Rather than being a single hypothesis that can be absolutely falsified by a single datum, models consist of a number of linked hypotheses. The model as a whole is subject to 'soft falsification'. On the one hand some elements of the model form a 'hard core', the unmodifiable basic assumptions of the model. On the other hand some elements are peripheral and can be modified in the light of new evidence whilst retaining the essential premises of the model. As discussed by Lakatos the success of a research programme (in this context read 'model') is not judged by the complete absence of contradictory data, but by the suitability of the programme to accommodate changes without modifying the hard

core of assumptions. Progressive research programmes can be extended, albeit in modified form, to new situations and to encompass new evidence. The parallel with computational models is clear. Most models are constructed with a specific task in mind, but the extent that the same model can be applied to different tasks we can say that the model principles are good. For example, the basal ganglia model was developed at a level of systems theory of the neuroanatomy of that region. Its application to the simulation of reaction times in conjunction with a model of Stroop processing gives a promising indication of its general validity. Heuristics for the valid extension of models are provided by broader biological and psychological theory. Additional criteria from the descriptive level of the model, either cognitive or neuroscientific, can be drawn in to guide extension of the model.

The emphasis on extension and modification of the model in line with the core assumptions makes it critical to try to identify in advance which features are central and must be retained in any future modifications of the model. Different elements will have different theory-centralities, but it becomes clear that there are 'grey-areas' between theory-relevant and theory-irrelevant choices. Indeed, part of the purpose of modelling is to identify what features of a model are relevant to its function in a theoretically interesting way. Grey areas may be discovered to have hitherto unsuspected importance – an example is the current work on response mechanisms – or their irrelevance may be more clearly established. Some parameters may be free to take arbitrary values within a certain range, only coming to alter the behaviour of the model at extremes. An example from the current work might be the speed of function of the individual units. Small values leave the behaviour of the model unaffected, but if the individual units took an extremely long time to adjust to new input, in the order of seconds, then the model would fail to accurately simulate the SOA data. The speed of function of the individual units is an example of a parameter that can arbitrarily take any *reasonable* value (although as constrained by known membrane properties).

An example of model extension is the inclusion of cognitive theories of word reading into the model. The essential property of the original front-end was that

word information was processed stronger than colour information. This property can be retained, whilst changing the architectural connections between the front-end units to reflect a theory from cognitive psychology. Whilst the original Cohen model front-end has the advantage of parsimony – being perhaps the simplest generic network architecture for this task – the cognitive model front-end allows contact to be made with neuropsychological work on lesion studies and data from the manual response task.

Only by taking the view that models should be open to improvement, and that their validity can be in part assessed by their amenability to this, can PDP modelling be a cumulative endeavour. This view establishes the continuity between versions of the same family of models, and, by relating models to their motivating theoretical concerns, establishes a deeper connection between models of different types.

7.6.2. On the falsity of Bonini's paradox

McClosky (1991) has stated that network models are not psychological theories, due to their opaqueness. I have attempted to show the models can be explicit extensions of theories – opaque, rigorous, thought-experiments. Computational simulations could be viewed as similar to animal models, not a precise human analogue, but revealing none-the-less.

Critiques such as McClosky's (1991) rely on the assumption that computational models can defy scientific investigation by virtue of their complexity. This is based, in part, on the non-transparent nature of distributed representation. So, it is thought, an investigator designs a model of a complex process which he or she can subsequently then not understand the function of. This is *Bonini's paradox* (Dutton & Starbuck, 1971) – that the model becomes as hard to understand as the system being modelled. Although it may sometimes appear this way to overly-fatigued researchers, this paradox is universally a false one. Not only are models generally

simpler than the systems they model, omitting irrelevant detail[9], but they have the virtue of being intentionally and knowingly constructed. This provides the researcher with knowledge of exactly how each part operates and exactly what parts are involved in the model. Moreover, even in the hypothetical limiting case where the model is as complex as the system modelled – i.e. a one-to-one mapping exists between the simulation and physical entities – models will still not suffer from Bonini's paradox. This is because models are easier to explore than natural systems. Manipulations of models are unconstrained by limitations of resources or ethics and are easier, faster and of greater scope than manipulations of living systems. So, although some systems, such as those using distributed representations, may possess an apparent opacity, this is no reason to avoid their exploration. The correct course of action with models that threaten to become incomprehensible is to redouble one's efforts at analysis.

7.6.3. Criteria of model comparison

> *...the development for good criteria for adjudicating between competing models is urgently needed* (Cohen, 2000, p.446)

The transparency of a model's working is an obligation of the investigators, not a requirement of the model. That said, parsimony should limit the complexity of models, and should be a criterion for the comparison of different models (or model families as I have termed them).

In line with the Lakatos's theory of science, the ease with which a model can be developed is one of the main criteria for its evaluation. Models can be extended along one dimension, to increase their descriptive adequacy, or along another to make contact with higher or lower-level phenomenon. The models presented in this thesis attempt to do both these things, making contact with the biology of action selection, on one hand, and the psychology of word reading on the other.

[9] Abstraction can be said to be a defining feature of models, and indeed science in general.

Incorporating constraints from independently motivated principles is another mark of a good model. Models which sit comfortably with the general principles of their field, as well as just adhering to their particular theory, are more likely to be valid in the longer-run. This consideration motivates the application of principles from action selection to the general problem of response mechanisms. The ethological, evolutionary and neurobiological considerations all inform the nature of the problem of modelling response behaviour in humans.

Normally models are assessed on their descriptive adequacy, although as Roberts & Pashler (2000) have pointed out, it is also important to define what other outputs the model could produce and what outputs the model can't produce. Models which are too powerful, and could produce any output, cannot be falsified and hence are scientifically useless. Multilayer neural networks have been accused of this, in that, with enough units, a multilayer network can learn any function. However, since the models presented in this thesis include additional constraints, and, critically, are assessing the network on aspects which the network was not explicitly trained to produce, this accusation does not hold in this case. Elaboration of the explanatory range is the best preparation for the correct comparison of the model with data. Roberts & Pashler (2000) also note that the possible range of the data should be known. Knowing the possible and plausible alternate forms that an empirical data set could have taken reveals whether the collected data contains the possibility of contradicting the model.

7.6.4. Combining multiple levels of analysis

Criticisms, such as that of Kinsbourne (1994, cited in Cohen, 2000, p.442) that connectionist models are "so unconstrained that a simulation could hardly fail" need to be answered. It is true that connectionist techniques are extremely powerful, and too general to explain psychological phenomena on their own, devoid of the context of theory. The incorporation of additional constraints allows the construction of falsifiable models for specific contexts. 'Explanatory connectionists' (Seidenberg, 1993) have asserted that a number of computational properties which can be easily

enacted in connectionist networks – parallelism, distributed representations, noise – will be found to underlie important psychological phenomena. While I admire this approach, I have taken another tack and incorporated specific cognitive and biological constraints which are already established in those fields into my models.

Like Seidenberg (1993) I hope that important phenomena can be found to emerge from the low-level properties of the model. An example is that the minimal input threshold that the neurobiology of the basal ganglia imposes has functional use when the network is considered on the higher level as response mechanism. The idea of emergence helps links the multiple levels of analysis that are necessary when attempting to scientifically study human cognition.

7.6.5. The role of modelling in theory advancement

Our discussion of the purposes of modelling (chapter 1) shows several possible roles a model can play in theory advancement; problem elaboration, theory elaboration, explanation and as metaphors. The work of this thesis has fulfilled functions in each of these areas. The project to model a response mechanism which fulfils the requirements of an embodied, situated, agent *problematises* the nature of response selection, which has previously been restricted to the simulation of reaction times and error rates in artificial experimental conditions. So, as well as reifying the nature of a problem, the task of modelling can also, in a related manner, identify problems. An apocryphal tale tells of one of the founders of modern artificial intelligence, Marvin Minsky, setting the task of 'vision' to one of his students as a summer project! The point being that until attempts were made to design seeing machines the nature of visual perception was severely under-estimated as a problem. The counter-side to problem-elaboration is theory-elaboration. The work considered in this thesis draws out the essentials of theories of Stroop processing, namely strength of processing and attention. The analysis of the response mechanisms points, for example, to the importance of Pieron's Law and, another example, of minimal thresholds in the basal ganglia.

The reconceptualisation of automaticity, as an emergent rather than intrinsic property (section 7.3), illustrates the role modelling may play as a metaphor. In cases like these the function of the model is often nebulous, and perhaps does not seem to be essential to the substantive part of the suggestions made; instead the model plays a role in the trajectory of thought that leads to the new suggestions being made. It belongs to the 'context of discovery', rather than the 'context of proof' (see section 1.7.2, iv).

However, it is in no way mandatory for a model to fulfil all the possible functions of models in theory advancement. In many cases models with tightly defined aims are better than those with too broad a scope. The purposes, in terms of the likely theoretical benefits, of the model help guide how the model should be constructed and/or extended. There is no correct level of description for a model without reference to the purpose of that model. Simple instantiations of explicitly formulated theories, such as Young & Burton (1999) which implemented a localist theory of face recognition, need to aim for transparency of representation and function. Those models which aim to provide explanation in psychology which stem from a lower level of computational principles, such as O'Reilly (1999), pitch themselves at a lower level and so transparency becomes an aim of the analysis of the model, not of its construction.

In both cases, modelling work is required because the theory has become too complex to interface directly with data. By this I mean that the nature of the theory, whether a set of explicit proposition or a set of hunches about the function of biological computations (O'Reilly, 1998), has become an entity which defies straight-forward, intuitive, comprehension.

There have been attempts to derive general computation principles for the guidance of computational modelling (McClelland, 1993; O'Reilly, 1998). This is admirable, since the recognition of, sometimes implicit, principles allows their conscious rejection or acceptance. The background of common principles unifies a group of models into a model-research programme of the kind alluded to above. However, it

should be recognised that these 'general' principles are the principles of a specific research project, and need not invalidate models of a different kind which are constructed with different purposes. Again, it is of vital importance that the purposes of a model, both theoretically and meta-theoretically, are made explicit at the outset of a modelling endeavour and that researchers put all due effort into the transparent and honest exposition of the function of the model.

Modelling currently enjoys a respect which sometimes obfuscates the minor scientific relevance of some models. However the potential for computational modelling to unify diverse phenomenon and disparate levels of description and analysis means that its use will become more and more popular and yield greater and greater returns.

7.7. Limitations

In this section I address the limitations of the research described in this thesis. These limitations can be conveniently grouped under three headings. Firstly, there are problems – mostly of a meta-theoretic nature – which beset any investigation which uses computational modelling. These are addressed under the heading 'problems faced by any model'. The discussion is grounded by the fact that, although the questions posed are general and can be asked of any modelling work, each model must provide individual answers. Secondly, there are those problems that are specific to my models and which result from the particular way the psychological issues are addressed. These are addressed under the heading 'specific problems of these models'. Finally there are problems which relate not so much to a model, per se, but to the theoretical and empirical background which informs the construction of the models. These, if you will, are difficulties with the very starting point for investigation of these phenomena. Such difficulties are addressed under the heading 'problems with the theoretical and empirical background'.

7.7.1. Problems faced by any model

Roberts & Pashler (2000) ask the question 'how persuasive is a good fit?' and provide a good starting point for assessing the success of a model. More than a simple match to data, we must ascertain how representative that match is. Specifically; how variable is the data? How flexible is the model? Are there data patterns that the model *cannot* fit?

Roberts & Pashler (2000) specifically cite the Cohen et al (1990) model as an example of naïve data fitting of little theoretical value. Although I agree completely with their general point, in defence of Cohen et al (1990) it should be said that wider theoretical considerations informed the construction of their model, so it was not simply an arbitrary way of matching the empirical data. That said, it is true both for Cohen et al, and for this thesis, that more could have been done to explore the range of possible results from the models.

Beyond the fit to the data, as McCloskey (1991) discusses, there is the issue of *how* the model simulates the data. This inspired the discussion (see section 7.6.1) of the nature of models with respect to theory and the conclusion that models have scientific value because they belong to a model-family of possible models and are part of a programme of development. As with scientific theories, choice between models requires reference to meta-theoretical principles. The reliance on these meta-theoretic principles blurs the boundary of falsification and means that, like theories, it is possible (even desirable) to have multiple, competing, contradictory models extant at any one time. These points need to be born in mind when considering the issues in the following paragraphs.

Although simple models can produce clear predictions, they forfeit the scope that more complex models allow. There is a tendency for more complex models, which cover a wider range of phenomenon, to lack the clarity of prediction that simple models possess. This trade-off is present whenever a choice needs to be made about

how to develop a model. The current model, although simplistic by many standards, is significantly more complicated than, say, just as a convenient benchmark, the Cohen et al (1990) model. This weakens the clarity of the model, both in terms of transparency of function and certainty of prediction. As more assumptions, and with them more parameters, are added to a model, so it become less clear what model features cause what results and whether the errors in data-fitting are fundamental to the model or merely due to peripheral assumptions.

There are always additional features that could be built into a model, and additional empirical findings that the model could be extended to cover. With regard to the models in this thesis, additional neuroanatomical pathways could be added to the BG component, or the model could be extended to cover other cognitive phenomenon such as switch costs (see section 7.8.3) or error rates in the Stroop task. These inadequacies of the models could, I like to think, be corrected while being faithful to the tenor of the models – and are, hence, not an insurmountable problem.

Crick (Crick, 1989; Crick & Asanuma, 1986) makes criticisms on neural networks based on their lack of biologically plausible detail. This critique is a consequence of this complexity-clarity trade-off. Like most neural network modellers, I acknowledge that the model neurons used in the simulation are simplistic, but hope that the simplicity of the units is compensated for by allowing comprehensible models.

The large number of parameters in the models means analysis is required to understand them. Additionally, different predictions can be made by altering peripheral parameters. Although this removes the clarity that hard falsification provides (i.e. a situation where data contrary to the predictions proves the model wrong), it draws out the necessity of model exploration; something insufficiently emphasised in the literature.

7.7.2. Specific problems of these models

The Cohen model uses an architecture in which word reading and colour naming are processes in qualitatively identical architectures. Kanne et al (1998) question this, and, although I don't agree with the grounds for their criticism - as stated above simplifications are a unavoidable part of modelling and not necessarily obstructive to progress – the modifications to the front-end involved in Model 2 (chapter 5) go some way to answering this challenge.

Another questionable assumption of my models is the localist representation used through out. One school of thought (e.g. Plaut et al., 1996) would insist on the use of distributed representation to capture the complexities of word reading. I would agree with them, *for models which address word reading at that level of detail*. As it is, although augmenting the models of this thesis with distributed representations might be an interesting project, the effort would be prohibitive for the gain in the understanding of reaction times in the Stroop task. We are not attempting to model word learning, large vocabularies, differences between irregular and regular words, and/or word reading errors. The results of word processing in our models are compressed to two scalar values, the competing saliences. At this level of modelling the choice between localist and distributed representations is theory-incidental.

Our models fail to address the nature of early visual attention. Evidence suggests that the earliest visual selection may be spatial (Styles, 1997). The input representation used in the models doesn't allow for any distinctions to be made about the spatial nature of the stimuli. This may in fact be important (see MacLeod, 1991 for a discussion of integrated versus separate Stroop stimuli). For example, Glaser & Glaser (1982), who provide the empirical data for the SOA experiments, used Stroop stimuli in which the word component appears at a different location from the colour component. These models cannot account for this in their current form.

Our model of the basal ganglia omits reference to its possible role in learning, and role in working memory (Brown et al., 1997; Doya, 2000; Frank et al., 2001; Graybiel, 1998; Middleton & Strick, 2000; Wise, 1996; Wise et al., 1996); two things for which it is most widely investigated. Learning in the BG was not one of the original purposes of our model, and so this remains a possible future extension of the model programme, rather than a current feature.

It might also seem perverse that these models, which amongst other things, claim to adumbrate issues concerning automaticity do not include explicit reference to the cerebellum – the region most popularly associated with the acquisition of automatic motor responses. Modelling the functional anatomy of the cerebellum is beyond the scope of this thesis. We suppose that any role played by cerebellum in facilitating automaticity in the task is included at the functional level by the relative 'strength of processing' in each pathway, so that anatomical grounding need not be explicitly reckoned with at this stage.

It might seem incongruous that the front end of the models represent a connectionist-psychological level and the back-end, the basal ganglia, is based on a systems level analysis of the neurobiology of that region. I have taken the view that psychology must deal with all levels of inquiry, rather than being reducible to just one, and as a consequence it is productive to combine levels in this way. The front end could be extended to be neuroanatomically grounded – i.e. to locate processing loci in such regions as prefrontal cortex and anterior cingulate cortex.

If could be argued that the various ways premature selection is restrained in these models demonstrates a lack of theoretical commitment to any of them. A minimal threshold is used (section 3.2), as well as dynamic attentional inhibition (chapter 6) and the use of an independent ‘no go signal’ (section 7.2.3) is also discussed. The minimal threshold is a consequence of basing our BG model on the actual upstate/downstate functionality of striatal neurons. It is possible that this feature would be sufficient to prevent erroneous front-end selection in most circumstances, but given the constraints of the Stroop task (see section 6.4), I have demonstrated

that dynamic attentional inhibition is a sufficient and useful feature. The first solution, a no-go signal, could perhaps be accused of being a post-hoc fix for the erroneous selection problem, but in this case the question should be asked: is this a psychologically plausible fix? Is it feasible that imminent selection is actively inhibited by a 'do nothing' signal? I think the answer is yes, at least to an extent that justifies the investigation of the phenomenon in these models.

Kanne et al (1998) criticized the Cohen et al (1990) model because it failed to account for set size effects (i.e. that reaction times slow as the set of possible responses increases). These criticisms have been addressed (Cohen et al., 1998) but it remains the case that this type of result is not dealt with by my models. Kanne et al (1998) note that the experiment simulated by Cohen et al (1990) and here, in this thesis – with 2 response options – is not, strictly, the same as the original experiment (Dunbar & MacLeod, 1984) - which had 3 response options.

It may be that the BG model can account for set size effects. Increasing the number of channels which are active at sub-selection threshold levels decreases the average selection time, which is in line with the general effect of increasing set size. Unfortunately there has not been time to go into this issue further.

Mewhort et al (1992) raise a concern about the model of Cohen et al (1990), namely that it fails to properly account for the distribution of response latencies in the different conditions. Although response latency distributions were not within the remit of the model, there is a case for saying that they should be. Mewhort et al (1992) provide an example of RT distributions being used to more fully test a model, and claim that the Cohen model generates the right mean reaction times, but for the wrong reasons. This is pertinent to the discussion of data fitting by models (section 1.7.1 and section 7.6.3). The considerable effort invested in refining mathematical models of RT distributions (e.g. Ratcliff & Rouder, 1998; Ratcliff et al., 1999) suggest that this is an area that should be followed up.

Both Models 1 and 2, even with dynamical attentional inhibition, do not fit the SOA data perfectly (sections 4.4. and 5.3.1). Given the inherent variability in RT data, it would perhaps be foolish to hope for an exact fit, but the current discrepancy, such as it is, indicates that there are pertinent facets of human information processing not captured by any extant models of the Stroop task.

It is not clear in these models how simulation time relates to real time in human experiments. In the model of Cohen et al (1990) it is possible to take the reaction times in the 6 basic stroop conditions and perform a regression to provide a metric for conversion of simulation time into milliseconds. However, this scheme does not allow the separate investigation of the roles of the separate components of the model in producing reaction times. For example, it is not clear how to assign responsibility for RT effects between the front-end and the back-end. In the model, the RT *differences* are mainly due to the back-end (the BG model), but it is not established that such late processing is equally important for human RTs. It could be that BG influence on RT is marginal and all the Stroop RT differences are due to processing in the early / frontal areas of the brain. Lacking a metric to convert simulated time for the components individually thus obscures the investigation of their relative contributions to RT effects. This highlights another advantage of anatomically grounded models. Within the BG model we can assess the relative importance of the different regions with respect to selection time. It is not clear how the architecture of the front end in Model 1 and in Model 2 relates to the number of anatomical stages involved in processing.

Designing a Stroop experiment to test predictions (section 5.3.3) highlighted many nuances that the models do not encompass. In preparing the experimental protocol, decisions had to be made with regard to things such as which finger the subject should use to indicate their response or what stimuli should be used for the colour-naming control condition. The models do not account for such elements, and so, in some sense, can be considered incomplete.

7.7.3. Problems with the theoretical and empirical background

In light of the observation by Kanne et al (1998) that the experiment simulated by Cohen et al (1990, simulation 1) was never run in that exact form, the empirical data we have aimed to simulate is called into question. This is particularly pertinent given the ephemeral nature of the facilitation effect that is not unusually absent in Stroop experiments, and is particularly weak in experiments with low response set sizes. Given this, it may be misplaced to require a model of Stroop processing with a response set of two to simulate the facilitation and interference effects produced from experiments using a response set of four.

This is further complicated by the inadvertent reading hypothesis of facilitation (MacLeod, 1998; MacLeod & MacDonald, 2000). If, as MacLeod suggests, facilitation is entirely separate from interference and results from the inadvertent reading of the word (which produces a response which, obviously in the congruent condition, is indistinguishable from the response due to the colour), then the models in this thesis do not account for this and instead simulate (erroneously) facilitation endogenous to the same processes that produce interference.

The construction of models which claim to shed some light on the concept of automaticity could be considered futile given recent critiques of that very concept (see in particular Pashler (1998)). Are we attempting to model something that doesn't exist? Concerns about the unity and coherence of the concept of automaticity are valid, but modelling work such as this is entirely appropriate for dealing with the components of a deconstructed notion of automatic processing (as discussed in section 7.3).

These models, in common with the vast majority of PDP models, fail to properly address the nature of 'controlled processing', the diametric opposite to automatic processing. Controlled processing is another way of trying to tie down that recurring conundrum of psychologists – consciousness. It has been suggested that connectionist models, while an appropriate level for modelling pre-conscious

processes, are unsuitable to the modelling the pseudo-rational deliberate processing of conscious thought.

7.8. Model Development

These models represents an attempt to unify or reconcile models and theories from different fields or topics: the neuroscience of action selection; classical choice models; information processing models of word reading, PDP models of attention and stimulus and response conflict. This unification, as well as being an aim in itself, is an expression of a theoretical position, viz. the rejection of atomistic psychological models and the move towards functional wholes.

7.8.1. Embodiment

This unification project requires, for its conclusion, the inclusion of reference to the embodied, adaptive, function of the human cognitive system. Although it is entirely valid to develop models without reference to the purpose of the cognitive system as a whole, such a focus allows a more comprehensive test of the suitability of the model and its integration into a wider context (Clark, 1999).

7.8.2. Modelling pathologies

If these models were to be developed, one tactic would be to locate existing modules of the model in specific brain regions such as the prefrontal cortex, anterior cingulate cortex and the other brain regions involved in the Stroop task (for a review see MacLeod & MacDonald, 2000). The fully neuroanatomically grounded model could then be tested by simulating results from imaging, neurophysiology and neuropsychology experiments. An additional application is the investigation of dysfunction in pathologies such as schizophrenia. There already exists a large body of work on the performance of the Stroop task by schizophrenics (Grapperon & Delage, 1999) and Cohen and colleagues have already approached the problem

(Cohen & Usher, 1996). Basal ganglia dysfunction has already been linked to schizophrenia (Bogerts, Meertz, & Schonfeldt-Bausch, 1985; Braff & Swerdlow, 1997; Calabresi et al., 1997; Heckers, 1997; Lidsky, 1997) and a fruitful line of enquiry could be the degree that impairments of basal ganglia-based switching capacity underlie the physical and cognitive problems seen in schizophrenia. Computational modelling would provide a way of disentangling the involvement of weakened attentional control in schizophrenia from any putative disorders of switching capacity.

7.8.3. Switch costs

For the selection of responses in the absence of competition, the BG model predicts that response selection will be slowed if it involves deselection of a previous response (see Figure 33, comparing the control condition response time with and without a pre-competition task).

This finding suggests that the BG model may be useful in the investigation of switch costs (Wylie & Allport, 2000) and similar phenomena, such as negative priming. Although we would not suggest that the BG is the prime locus of such phenomena, there is sufficient reason – based on the current investigations – to think that the role of the response mechanism is important in their genesis and that a computational model which could include both attentional and response functions would be a good starting point for investigations.

7.8.4. Application to the study of attention

Our model informs the study of attention using computational methods. The model illustrates a case where preparatory attention is insufficient, and dynamic (stimulus evoked) inhibition is required. There are opportunities to explore this further, especially in relation to negative priming. The persistence of inhibition also bears upon the blossoming task switching literature (e.g. Dreher, Kohn, & Berman, 2001; Wylie & Allport, 2000).

Attentional selection interacts with subsequent decision processing. Attention modulates the stimuli that affect decision processing. Decision processing (via behaviour) affects which stimuli are perceived. Throughout this thesis I have argued the importance of explicitly acknowledging response mechanisms within psychological models. Models like this, which include both attentional (front) and response (back) control mechanisms, can provide a starting point for the exploration of the co-operative action of the two.

Our models also provide an example of computational cognitive neuroscience models being used to provide explanations of cognitive phenomena, and in turn being assessed by cognitive level experiments. Modelling, if it is to be of value must be explanatory – and so by definition must cross representational levels. In this way modelling efforts are an essential part of the project to unify cognitive and neuroscientific theories.

APPENDIX I

A.1. Other potential Sources of interference-facilitation asymmetry in the Cohen et al (1990) model

The response mechanism of the model calculates reaction times based on the relative strength of evidence, that is, the difference between the target output and the competing output. The inputs to the response mechanism, calculated from the outputs of the model with the weights and parameters given in Cohen et al (1990), are shown in table A1. For the piecewise linear activation function the increase in relative evidence for the conflict condition is exactly equal to the decrease in relative evidence in the congruent condition. Yet the simulation still produces the correct reaction times (see Figure 8). Therefore it must be the response mechanism that converts symmetrical differences in evidence into asymmetrical differences in response time.

Table A1: The strength of evidence for target response in the three colour-naming conditions for models using the two activation functions

	Logistic function		Piecewise linear function	
Control	0.48		0.46	
Conflict	0.27	(-0.21)	0.24	(-0.22)
Congruent	0.64	(+0.17)	0.68	(+0.22)

Note: The change in strength of evidence relative to the control condition is shown in brackets for the conflict and congruent conditions.

For the logistic activation function the change in relative evidence is greater in the conflict condition by 0.04. So, while there is a change in the response-outputs due to the non-linearity of the logistic function, it is too small to explain the magnitude of the interference-facilitation asymmetry.

The outputs for the word-reading conditions are not shown since they do not vary significantly between conditions or models and are not relevant to the discussion of the point at hand.

It was also suggested by Cohen et al (1990) that the time constant for leaky integration in the artificial neurons of the model played a role in producing the asymmetry of interference and facilitation. However, I have found that removing the neuron temporal dynamics altogether, so that their outputs change instantaneously, has a negligible effect on the simulation results.

A.2. Procedure for deriving log-log plots

Figures showing the fit of a response mechanism to an analogue of Pieron's Law were produced using the following method:

A Pieron's Law like curve, as defined in equation (4) was fitted to the data using the *fminsearch* function from MATLAB version 12. This is an unconstrained non-linear optimisation procedure which uses the simplex search method (Lagarias, Reeds, Wright, & Wright, 1998). The asymptotic value obtained was used to plot the log of reaction time less the asymptote on the y-axis, while the log of the input to the response mechanism was plotted along the x-axis. Functions which fit Pieron's Law exactly produce straight lines when plotted like this. The line shown is the best-fit line that results from using the full set of parameters derived from the optimisation procedure. The transformation to log-log coordinates exaggerates the discrepancy between the data and the best-fit line at lower RTs.

For some plots the range of data used to derive the asymptote was longer than the range of data shown. This was done in order to more accurately derive the asymptotic value, which is the major influence on the straightness of the line. When this was the case, a second simplex search was done for the range of data shown on the graph with the asymptotic value fixed but the other two parameters, k and β, unconstrained.

APPENDIX II

This supplement is intended to assist in the replication of the models presented in this thesis.

Chapter 2 - Replication of Cohen et al (1990)

Full details are given in Cohen et al (1990), however a number of small errors were made in the presentation of the model in that paper. The errors are corrected below, and presented alongside a number of caveats which need to be borne in mind.

Kanne et al (1998) note that the experiment upon which simulations of Cohen et al (1990) are based have never been done in exactly the form which Cohen et al purportedly simulate. (Dunbar & MacLeod, 1984; Glaser & Glaser, 1982) use response sets of 4 and 5 colours respectively. This variation in the response set may be significant, given work on how context and response set can affect Stroop effects (Kanne et al., 1998).

Cohen et al (1990, p.336) contains a misleading diagram of the network architecture. The task demand units shown in figure 1 are not connected to the hidden units in the opposite pathway – only to the hidden units in the corresponding pathway (as Cohen et al, 1990, figure 3 shows).

Although Cohen et al (1990, p.340) report that word patterns were trained 10 times as often as colour patterns, in fact the ratio used in the simulations was 5:1 (Kanne et al, 1998)

Cohen et al (1990) figure 5 is referenced as 'after Dunbar & MacLeod (1984), p.62'. There is no p.62 in Dunbar & MacLeod (1984). Most likely the data is from Dunbar & MacLeod (1984), experiment 2, p.630. The data shown is from is two conditions

from an experiment that was part of a series from the original paper. The same conditions could, in fact, have been taken from a number of experiments in the same paper. The variability between this data, from identical conditions, but recorded during different experiments, shows that it is unnecessary to try to get a simulation to match exactly any particular set of empirical reaction times for this experiment.

Several parameters of the model not specified or were incorrectly specified. (Mewhort et al., 1992) report the correct parameters[10] and explore the parameter space for simulation 1 using the model.

a) standard deviation of the Gaussian noise added to evidence accumulators = 0.01
b) standard deviation of the Gaussian noise added to the unit activation = 0.5
c) the cascade rate, $\tau = 0.1$
d) the rate of evidence accumulation, $\alpha = 0.1$

In the discussion of Simulation 2 (Cohen et al, 1990, p.344) mention is made that the maximum amount of interference from colours on words in the simulation 'appears to be…asymptotic'. As discussed (section 2.4.2) consideration of the cause of this interference at negative SOAs shows that this amount of interference *cannot* reach an asymptotic limit.

No mention is made of error catching in the original model. Since gaussian noise was added to the response mechanism and the unit activations it is a theoretically necessity that response errors could have been made by the model. Given the noise parameters used error responses were, in fact, extremely unlikely. It is presumed that Cohen et al checked that the correct responses were obtained from the model,

[10] Mewhort et al (1992) themselves make an error reporting the parameter values. The standard deviation of the noise added to the unit activation is reported both as 0.1 and as 0.5 in the paper (p.874). The correct value is assumed to be 0.5.

and discarded any data from trials on which the model provided the incorrect response.

Cohen et al (1990) modified their model parameters for their simulation 2 (SOA simulations). The resting net input for the hidden units was decreased from –4.0 to –4.9. This was claimed to be 'to simulate the reduced interference and facilitation effects observed at the 0-ms SOA in this experiment, in comparison with the standard experiment using integral stimuli.' In fact this change in the bias was necessary to prevent erroneous responses at negative SOAs (see section 2.4.2, this thesis). Our replication simulation shows that altering the bias to –4.9 actually *increases* the discrepancy between the 0ms SOA simulation data and the 0ms SOA empirical data.

Chapter 3 – The Basal Ganglia Model

The basic basal ganglia model is presented in two companion papers by Gurney et al (2001a; 2001b). The models in this thesis use this basic ('intrinsic') basal ganglia model but only as part of a wider model of thalamo-cortical connections, as presented by Humphries & Gurney (2002).

For the simulations discussed in Chapter 3 of this thesis the inputs (saliences) were presented to the BG model as step function inputs on 'motor cortex'. As presented in Figure 19, motor cortex is where thalamic feedback and cortical inputs meet. In this case it consists of six simple neurons, as used in the rest of the BG model, with one output each (to the striatum of the basal ganglia intrinsic model) and two inputs (saliences and thalamic input) which are weighted equally.

The outputs of the model are recorded from the GPi of the intrinsic model. When activity of units in this region drops to 0 selection is indicated. All the parameters of the BG model are as reported by Humphries & Gurney (2002).

Chapter 4 – Model 1

Model 1 is the concatenation of the Cohen model, with modifications, and the BG model. The modifications required of the Cohen model by this concatenation are discussed in section 4.2. No other changes were required.

The BG model weights are fixed for all simulations discussed in this thesis. Training of the network weights was still required for the front-end. To train the front-end weights the standard back-propagation algorithm was used, as implemented by Cohen et al (1990). The outputs were taken from the front-end (i.e. before entering the response mechanism) and compared to the targets, as with the original Cohen model.

Chapter 5 – Model 2

The front-end of Model 2 was also not trained. The architecture is shown in Figure 25. The activation function, the Weibul function is defined on p.103. Note that only the front-end network units used the Weibul function. The BG model continued to use the standard piecewise linear activation function, as in the other simulations.

Figure 25 should be interpreted like this. Each module represents two units, one for each of the two responses, 'green' and 'red'. There are no cross connections between units. Each unit receives input from the corresponding unit(s) in modules upstream, and projects to the corresponding unit(s) downstream. The inputs are represented in the same way as for the Cohen model.

REFERENCES

Alexander, G. E., Delong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience, 9*, 357-381.

Allport, A. (1987). Selection for Action: Some Behaviorial and Neurophysiological Considerations of Attention and Action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on Perception and Action* . London: Lawrence Erlbaum Associates.

Allport, A. (1993). Attention and Control: Have we been asking the wrong questions? A critical review of twenty-five years, *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience.* . Cambridge, MA.: MIT Press.

Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience, 12*(3), 505-519.

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z. P., Barad, V., Gullett, D., Shah, C., & Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMIRI. *Cognitive Brain Research, 10*(1-2), 1-9.

Bargh, J. A. (1989). Conditional Automaticity: Varieties of Automatic Influence in Social Perception and Cognition. In J. A. Bargh & J. Ullman (Eds.), *Unintended Thought* (pp. 3-51). London: Guilford Press.

Bechtel, W. (1994). Levels of Description and Explanation in Cognitive Science. *Minds and Machines, 4*(1), 1-25.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences, 4*(3), 91-99.

Beiser, D. G., & Houk, J. C. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology, 79*(6), 3168-3188.

Besner, D., & Stolz, J. A. (1999). Unconsciously controlled processing: The stroop effect reconsidered. *Psychonomic Bulletin & Review, 6*(3), 449-455.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4*(2), 221-225.

Bogerts, B., Meertz, E., & Schonfeldt-Bausch, R. (1985). Basal ganglia and limbic system pathology in schizophrenia. *Archiv Gen Psychiat, 42*, 784-791.

Bonnet, C., Zamora, M. C., Buratti, F., & Guirao, M. (1999). Group and individual gustatory reaction times and Pieron's law. *Physiology & Behavior, 66*(4), 549-558.

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*(6758), 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*(3), 624-652.

Braff, D. L., & Swerdlow, N. R. (1997). Neuroanatomy of schizophrenia. *Schizophrenia Bull, 23*, 509-512.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nature Neuroscience, 2*, 370-374.

Brito, G. N. O. (1997). A neurobiological model for Tourette syndrome centered on the nucleus accumbens. *Med. Hypotheses, 49*, 133-142.

Broadbent, D. (1985). A Question of Levels - Comment. *Journal of Experimental Psychology-General, 114*(2), 189-192.

Brooks, R. A. (1991). New Approaches to Robotics. *Science, 253*(5025), 1227-1232.

Brown, L. L., Schneider, J. S., & Lidsky, T. I. (1997). Sensory and cognitive functions of the basal ganglia. *Current Opinion in Neurobiology, 7*(2), 157-163.

Brown, T. L. (1996). Attentional selection and word processing in Stroop and word search tasks: The role of selection for action. *American Journal of Psychology, 109*(2), 265-286.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in stroop color naming: Is word recognition "automatic"? *Journal of Experimental Psychology-General, 131*(2), 220-240.

Brunia, C. H. M. (1999). Neural Aspects of anticipatory behaviour. *Acta Psychologia, 101*, 213-242.

Buchel, C., & Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex, 7*(8), 768-778.

Bullinaria, J. A. (1997). Modeling reading, spelling, and past tense learning with artificial neural networks. *Brain and Language, 59*(2), 236-266.

Bundesen, C. (1990). A Theory of Visual-Attention. *Psychological Review, 97*(4), 523-547.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*(6), 215-222.

Calabresi, P., DeMurtas, M., & Bernardi, G. (1997). The neostriatum beyond the motor function: Experimental and clinical evidence. *Neuroscience, 78*, 39-60.

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*(5364), 747-749.

Casey, B. J., Thomas, K. M., Welsh, T. F., Badgaiyan, R. D., Eccard, C. H., Jennings, J. R., & Crone, E. A. (2000). Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America, 97*(15), 8728-8733.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2*(7), 671-676.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory- guided visual search. *Journal of Neurophysiology, 80*(6), 2918-2940.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences, 20*(12), 553-557.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: The MIT Press.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*(9), 345-351.

Clark, A., & Thornton, C. (1997). Trading spaces: Computation, representation, and the limits of uninformed learning. *Behavioral and Brain Sciences, 20*(1), 57-&.

Cleeremans, A., & French, R. M. (1996). From chicken squawking to cognition: Levels of description and the computational approach in psychology. *Psychologica Belgica, 36*(1-2), 5-29.

Cohen, G. (2000). Overview. In G. Cohen, R. A. Johnston, & K. Plunkett (Eds.), *Exploring Cognition: Damaged Brains and Neural Networks* . Hove, UK: Psychology Press.

Cohen, J. D., Botvinick, M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: who's in control? *Nature Neuroscience, 3*(5), 421-423.

Cohen, J. D., Braver, T. S., & Oreilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society of London Series B- Biological Sciences, 351*(1346), 1515-1527.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes - a parallel distributed-processing account of the Stroop effect. *Psychological Review, 97*(3), 332-361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the Use of Interactive Models For Understanding Attention and Performance, *Attention and Performance XV* (Vol. 15, pp. 453-476).

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine - a connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99*(1), 45-77.

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed-processing approach to automaticity. *American Journal of Psychology, 105*(2), 239-269.

Cohen, J. D., & Usher, M. (1996). A neural network model of stroop interference and facilitation effects in schizophrenia. *Biological Psychiatry, 39*(7), 237.

Cohen, J. D., Usher, M., & McClelland, J. L. (1998). A PDP approach to set size effects within the Stroop task: Reply to Kanne, Balota, Spieler, and Faust (1998). *Psychological Review, 105*(1), 188-194.

Coltheart, M. (1993). Drc - a New Computational Model of Reading and Its Simulations of Normal Reading and Patterns of Acquired Dyslexia. *Australian Journal of Psychology, 45*(2), 110-111.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3*(3), 115-120.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of Reading Aloud - Dual-Route and Parallel-Distributed- Processing Approaches. *Psychological Review, 100*(4), 589-608.

Coltheart, M., & Langdon, R. (1998). Autism, modularity and levels of explanation in cognitive science. *Mind & Language, 13*(1), 138-152.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6*(3), 456-463.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J Neurosci, 18*, 7426-7435.

Cowan, G. (1998). *Statistical Data Analysis*. Oxford: Clarendon Press.

Crick, F. (1989). The Recent Excitement About Neural Networks. *Nature, 337*(6203), 129-132.

Crick, F., & Asanuma, C. (1986). Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 333-371). Cambridge, MA: MIT Press.

Crossman, A. R. (1987). Primate models of dyskinesia: The experimental approach to the study of basal ganglia-related involuntary movement disorders. *Neuroscience, 21*, 1-40.

Davis, K. D., Hutchison, W. D., Lozano, A. M., Tasker, R. R., & Dostrovsky, J. O. (2000). Human anterior cingulate cortex neurons modulated by attention-demanding tasks. *Journal of Neurophysiology, 83*(6), 3575-3577.

Dawson, M. R. W., Medler, D. A., & Berkeley, I. S. N. (1997). PDP networks can provide models that are not mere implementations of classical theories. *Philosophical Psychology, 10*(1), 25-40.

Dejong, R., Liang, C. C., & Lauber, E. (1994). Conditional and Unconditional Automaticity - a Dual-Process Model of Effects of Spatial Stimulus - Response Correspondence. *Journal of Experimental Psychology-Human Perception and Performance, 20*(4), 731-750.

Desoto, M. C., Fabiani, M., Geary, D. C., & Gratton, G. (2001). When in doubt, do it both ways: Brain evidence of the simultaneous activation of conflicting motor responses in a spatial Stroop task. *Journal of Cognitive Neuroscience, 13*(4), 523-536.

Di Paolo, E. A., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments, *Artificial Life Vii* (pp. 497-506). Cambridge, MA.: MIT Press.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition, 28*(8), 1437-1449.

Dixon, M., Brunet, A., & Laurence, J. R. (1990). Hypnotizability and Automaticity - Toward a Parallel Distributed-Processing Model of Hypnotic Responding. *Journal of Abnormal Psychology, 99*(4), 336-343.

Djurfeldt, M., Ekeberg, O., & Graybiel, A. M. (2001). Cortex-basal ganglia interaction and attractor states. *Neurocomputing, 38*, 573-579.

Dominey, P. F., & Boussaoud, D. (1997). Encoding behavioral context in recurrent networks of the fronto-striatal system: a simulation study. *Cognitive Brain Research, 6*(1), 53-65.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology, 10*(6), 732-739.

Doyon, J., Laforce, R., Bouchard, G., Gaudreau, D., Roy, J., Poirier, M., Bedard, P. J., Bedard, F., & Bouchard, J. P. (1998). Role of the striatum, cerebellum and

frontal lobes in the automatization of a repeated visuomotor sequence of movements. *Neuropsychologia, 36*(7), 625-641.

Dreher, J. C., Kohn, P. D., & Berman, K. (2001). Neural basis of backward inhibition during task switching. *Neuroimage, 13*(6), S311-S311.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color - Stroop interference patterns with transformed words. *Journal of Experimental Psychology-Human Perception and Performance, 10*(5), 622-639.

Duncan, J. (1986). Consistent and Varied Training in the Theory of Automatic and Controlled Information-Processing. *Cognition, 23*(3), 279-284.

Duncan, J. (1998). Converging levels of analysis in the cognitive neuroscience of visual attention. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 353*(1373), 1307-1317.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7*(2), 255-261.

Durgin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review, 7*(1), 121-125.

Dutton, J. M., & Starbuck, W. H. (Eds.). (1971). *Computer simulation of human behavior*. New York: Wiley.

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Ellis, R., & Humphreys, G. (1999). *Connectionist Psychology : a text with readings*. Hove, UK: Psychology Press Ltd.

Elman, J. L. (1994). Implicit Learning in Neural Networks - the Importance of Starting Small, *Attention and Performance Xv* (Vol. 15, pp. 861-888).

Elman, J. L., Bates, E. A., Johnston, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connection Perspective on Development*. Cambridge, MA: The MIT Press.

Eysenck, M. W., & Keane, M. T. (1995). *Cognitive Psychology: 3rd Edition*. Hove (UK): Lawrence Erlbaum Associates.

Fiebach, C. J., Friederici, A. D., Muller, K., & von Cramon, D. Y. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience, 14*(1), 11-23.

Fodor, J. A. (1983). *The modularity of mind : an essay on faculty psychology*. Cambridge, MA.: MIT Press.

Fox, E. (1995). Negative Priming From Ignored Distractors in Visual Selection - a Review. *Psychonomic Bulletin & Review, 2*(2), 145-173.

Frank, M., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective and Behavioural Neuroscience, 1*, 137-160.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 3314-3319.

Ghilardi, M. F., Ghez, C., Dhawan, V., Moeller, J., Mentis, M., Nakamura, T., Antonini, A., & Eidelberg, D. (2000). Patterns of regional brain activation associated with different forms of motor learning. *Brain Research, 871*(1), 127-145.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., Kim, Y. H., Meyer, J. R., & Mesulam, M. M. (1999). A large-scale distributed network for covert spatial attention - Further anatomical delineation based on stringent behavioural and cognitive controls. *Brain, 122*, 1093-1106.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the stroop phenomenon. *Journal of Experimental Psychology-Human Perception and Performance, 8*(6), 875-894.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences, 5*(1), 10-16.

Goodale, M. A., & Humphrey, G. K. (1998). The objects of action and perception. *Cognition, 67*(1-2), 181-207.

Goodale, M. A., & Milner, A. D. (1992). Separate Visual Pathways For Perception and Action. *Trends in Neurosciences, 15*(1), 20-25.

Grapperon, J., & Delage, M. (1999). Stroop test and schizophrenia. *Encephale-Revue De Psychiatrie Clinique Biologique Et Therapeutique, 25*(1), 50-58.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory, 70*(1-2), 119-136.

Gurney, K. (1997). *An Introduction to Neural Networks*. London: UCL Press.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia I: A new functional anatomy. *Biological cybernetics, 85*(6), 401-410.

Gurney, K., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia: II: Analysis and simulation of behaviour. *Biological cybernetics, 85*(6), 411-423.

Hayes, A. E., Davidson, M. C., Keele, S. W., & Rafal, R. D. (1998). Toward a functional analysis of the basal ganglia. *Journal of Cognitive Neuroscience, 10*(2), 178-198.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185-207.

Heckers, S. (1997). Neuropathology of schizophrenia: Cortex, thalamus, basal ganglia, and neurotransmitter-specific projection systems. *Schizophrenia Bull, 23*, 403-421.

Heinze, H. J., Manfun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Munte, T. F., Gos, G., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M. S., & Hillyard, S. A. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature, 372*, 543-546.

Henik, A., Ro, T., Merrill, D., Rafal, R., & Safadi, Z. (1999). Interactions between color and word processing in a flanker task. *Journal of Experimental Psychology-Human Perception and Performance, 25*(1), 198-209.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences, 353*, 1257-1270.

Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems, 1*, 495-502.

Hofstader, D. R. (1979). *Godel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

Holland, J. H. (1998). *Emergence: From Chaos to Order*. Oxford: Oxford University Press.

Holthoff-Detto, V. A., Kessler, J., Herholz, K., Bonner, H., Pietrzyk, U., Wurker, M., Ghaemi, M., Wienhard, K., Wagner, R., & Heiss, W. D. (1997). Functional effects of striatal dysfunction in Parkinson disease. *Arch. Neurol., 54*, 145-150.

Hommel, B. (1997). Interactions between stimulus-stimulus congruence and stimulus- response compatibility. *Psychological Research-Psychologische Forschung, 59*(4), 248-260.

Horwitz, B., Friston, K. J., & Taylor, J. G. (2000). Neural modeling and functional brain imaging: an overview. *Neural Networks, 13*(8-9), 829-846.

Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences, 3*(3), 91-98.

Houghton, G., Tipper, S. P., Weaver, B., & Shore, D. I. (1996). Inhibition and interference in selective attention: Some tests of a neural network model. *Visual Cognition, 3*(2), 119-164.

Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology, 77*(5), 1011-1025.

Humphries, M. D., & Gurney, K. N. (2002). The role of intra-thalamic and thalamocortical circuits in action selection. *Network-Computation in Neural Systems, 13*(1), 131-156.

Inzelberg, R., Plotnik, M., Flash, T., Schechtman, E., Shahar, I., & Korczyn, A. D. (1996). Switching abilities in Parkinson's disease. In L. Battistin, G. Scarlato, T. Caraceni, & S. Ruggieri (Eds.), *Advances in Neurology* (Vol. 69, pp. 361-369). Philadelphia, PA: Lippincott-Raven Publ.

Jackson, S., & Houghton, G. (1994). Sensorimotor selection and the basal ganglia: A neural network model. In J.C.Houk & J. Davis (Eds.), *Information processing models of basal ganglia* . Cambridge, MA.: MIT Press.

Jacobs, R. A., & Jordan, M. I. (1992). Computational Consequences of a Bias Toward Short Connections. *Journal of Cognitive Neuroscience, 4*(4), 323-336.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task Decomposition Through Competition in a Modular Connectionist Architecture - the What and Where Vision Tasks. *Cognitive Science, 15*(2), 219-250.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). Surround inhibition among projection neurones is weak or nonexistent in the rat neostriatum. *J. Neurophysiol., 72*, 2555-2558.

Jancke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266*, 125-128.

Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience, 13*(6), 730-743.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science, 286*(5445), 1745-1749.

Kanne, S. M., Balota, D. A., Spieler, D. H., & Faust, M. E. (1998). Explorations of Cohen, Dunbar, and McClelland's (1990) connectionist model of Stroop performance. *Psychological Review, 105*(1), 174-187.

Kaplan, W. (1952). *Advanced Calculus*. Reading, Mass.: Addison-Wesley.

Kastner, S., DeWeerd, P., Desimone, R., & Ungerleider, L. C. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282*, 108-111.

Kastner, S., Pinsk, M. A., DeWeerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*, 751-761.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*(5280), 1399-1402.

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. New York: Oxford University Press.

Koos, T., Tepper, J., Goldman-Rakic, P., & Wilson, C. J. (2002). *Electrophysiological properties and dopaminergic modulation of GABAergic*

inhibition among neostriatal projection neurons. Paper presented at the Society for Neuroscience Annual Meeting, Orlando, USA.

Kornblum, S., Stevens, G. T., Whipple, A., & Requin, J. (1999). The effects of irrelevant stimuli: 1. The time course of stimulus-stimulus and stimulus-response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *Journal of Experimental Psychology-Human Perception and Performance, 25*(3), 688-714.

Koski, L., Paus, T., Hofle, N., & Petrides, M. (1999). Increased blood flow in the basal ganglia when using cues to direct attention. *Experimental Brain Research, 129*(2), 241-246.

LaBerge, D., Auclair, L., & Sieroff, E. (2000). Preparatory attention: Experiment and theory. *Consciousness and Cognition, 9*(3), 396-434.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization, 9*(1), 112-147.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. New York: Cambridge University Press.

Lanthorn, T., Storm, J., & Andersen, P. (1984). Current-to-Frequency Transduction in Ca1 Hippocampal Pyramidal Cells - Slow Prepotentials Dominate the Primary Range Firing. *Experimental Brain Research, 53*(2), 431-443.

Levy, R., Friedman, H. R., Davachi, L., & GoldmanRakic, P. S. (1997). Differential activation of the caudate nucleus in primates performing spatial and nonspatial working memory tasks. *Journal of Neuroscience, 17*(10), 3870-3882.

Lewandowsky, S. (1993). The Rewards and Hazards of Computer-Simulations. *Psychological Science, 4*(4), 236-243.

Lidsky, T. I. (1997). Neuropsychiatric implications of basal ganglia dysfunction. *Biol Psychiatry, 41*, 383-385.

Logan, G. D. (1987). Toward an Instance Theory of Automatization. *Bulletin of the Psychonomic Society, 25*(5), 342-342.

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review, 95*(4), 492-527.

Lu, C. H., & Proctor, R. W. (2001). Influence of irrelevant information on human performance: Effects of S-R association strength and relative timing. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 54*(1), 95-136.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organisation*. New York: Clarendon Press.

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect - an integrative review. *Psychological Bulletin, 109*(2), 163-203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition, 26*(2), 201-211.

MacLeod, C. M., & Dunbar, K. (1988). Training and stroop-like interference - evidence for a continuum of automaticity. *Journal of Experimental Psychology-Learning Memory and Cognition, 14*(1), 126-135.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383-391.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology, 37*(3), 243-282.

Marr, D. (1982). *Vision*. New York: W.H. Freeman and Company.

Marton, F., Fensham, P., & Chaiklin, S. (1994). A Nobels Eye View of Scientific Intuition - Discussions With the Nobel Prizewinners in Physics, Chemistry and Medicine (1970-86). *International Journal of Science Education, 16*(4), 457-473.

Masterman, D. L., & Cummings, J. L. (1997). Frontal-subcortical circuits: The anatomic basis of executive, social and motivated behaviors. *Journal of Psychopharmacology, 11*(2), 107-114.

Matthews, G., & Harley, T. A. (1996). Connectionist models of emotional distress and attentional bias. *Cognition & Emotion, 10*(6), 561-600.

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of Negative Priming. *Psychological Bulletin, 118*(1), 35-54.

Maynard Smith, J. (1987). When Learning Guides Evolution. *Nature, 329*(6142), 761-762.

McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

McClelland, J. L. (1988). Connectionist Models and Psychological Evidence. *Journal of Memory and Language, 27*(2), 107-123.

McClelland, J. L. (1993). Toward a Theory of Information-Processing in Graded, Random, and Interactive Networks. *Attention and Performance*(14), 655-688.

McCloskey, M. (1991). Networks and Theories - the Place of Connectionism in Cognitive Science. *Psychological Science, 2*(6), 387-395.

McNellis, M. G., & Blumstein, S. E. (2001). Self-organizing dynamics of lexical access in normals and aphasics. *Journal of Cognitive Neuroscience, 13*(2), 151-170.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to stroop dimensions - effects of base-line discriminability, response-mode, and practice. *Memory & Cognition, 21*(5), 627-645.

Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response-time distributions and the stroop task - a test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 872-882.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. *Brain Cognition, 42*, 183-200.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50*(4), 381-425.

Mink, J. W. (2001). Basal ganglia dysfunction in Tourette's syndrome: A new hypothesis. *Pediatric Neurology, 25*(3), 190-198.

Monchi, O., Taylor, J. G., & Dagher, A. (2000). A neural model of working memory processes in normal subjects, Parkinson's disease and schizophrenia for fMRI design and predictions. *Neural Networks, 13*(8-9), 953-973.

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition, 29*(1), 137-151.

Montes-Gonzalez, F. M., Prescott, T. J., Gurney, K., & Redgrave, P. (2000). The robot basal ganglia: Control of robot action selection by an embodied model of the mammalian basal ganglia. *European Journal of Neuroscience, 12*, 134-134.

Morton, J., & Chambers, S. M. (1973). Selective Attention to Words and Colours. *Quarterly Journal of Experimental Psychology, 25*(387-397).

Nakahara, H., Doya, K., & Hikosaka, O. (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences - A computational approach. *Journal of Cognitive Neuroscience, 13*(5), 626-647.

Neill, W. T., Valdes, L., A., & Terry, K. M. (1995). Selective Attention and Inhibitory Control of Cognition. In F. N. Dempster (Ed.), *Interference and Inhibition in Cognition* . London: Academic Press.

Newman, D. V. (1996). Emergence and strange attractors. *Philosophy of Science, 63*(2), 245-261.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences, 2*(11), 455-462.

O'Reilly, R. C., & Farah, M. J. (1999). Simulation and explanation in neuropsychology and beyond. *Cognitive Neuropsychology, 16*(1), 49-72.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23*(4), 443-+.

Palmeri, T. J. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(2), 543-551.

Pashler, H. E. (1998). *The Psychology of Attention*. Cambridge, MA.: The MIT Press.

Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience, 2*(6), 417-424.

Phaf, R. H., Vanderheijden, A. H. C., & Hudson, P. T. W. (1990). SLAM - a connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*(3), 273-341.

Pieron, H. (1914). Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensites excitatrices. *L'Annee Psychologique, 20*, 17-96.

Pieron, H. (1920). Nouvelles recherches sur l'analyse du temps de latence senorialle et sur la loi qui relie de temp a l'intensite d'excitation. *Annee Psychologique, 22*, 58-142.

Pieron, H. (1952). *The Sensations: Their Functions, Processes and Mechanisms*. London: Frederick Muller Ltd.

Pinker, S., & Prince, A. (1988). On Language and Connectionism - Analysis of a Parallel Distributed-Processing Model of Language-Acquisition. *Cognition, 28*(1-2), 73-193.

Pins, D., & Bonnet, C. (1996). On the relation between stimulus intensity and processing time: Pieron's law and choice reaction time. *Perception & Psychophysics, 58*(3), 390-400.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15*(4-5), 445-485.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56-115.

Plaut, D. C., & Shallice, T. (1993). Deep Dyslexia - a Case-Study of Connectionist Neuropsychology. *Cognitive Neuropsychology, 10*(5), 377-500.

Popper, K. R. (1963). *Conjectures and refutations: the growth of scienfic knowledge*. London: Routledge and Kegan Paul.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* . Hillsdale, NJ: Erlbaum.

Prescott, T. J., Redgrave, P., & Gurney, K. (1999). Layered control architectures in robots and vertebrates. *Adaptive Behavior, 7*(1), 99-127.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59-108.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*(2), 261-300.

Ravizza, S. M., & Ivry, R. B. (2001). Comparison of the basal ganglia and cerebellum in shifting attention. *Journal of Cognitive Neuroscience, 13*(3), 285-297.

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience, 3*(8), 827-830.

Redgrave, P. (1998). The cognitive neuroscience of action. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 51*(4), 379-380.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience, 89*(4), 1009-1023.

Rees, G., & Frith, C. D. (1998). How do we select perceptions and actions? Human brain imaging studies. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences, 353*, 1283-1293.

Riolo, R. L., Cohen, M. D., & Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature, 414*(6862), 441-443.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358-367.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318-362). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986b). *Parallel Distributed Processing: Explorations in the microstructure of cognition*. Cambridge, MA: The MIT Press.

Ryan, C. (1983). Reassessing the Automaticity Control Distinction - Item Recognition As a Paradigm Case. *Psychological Review, 90*(2), 171-178.

Salmon, D. P., & Butters, N. (1995). Neurobiology of Skill and Habit Learning. *Current Opinion in Neurobiology, 5*(2), 184-190.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience, 2*(1), 33-42.

Schooler, C., Neumann, E., Caplan, L. J., & Roberts, B. R. (1997). A time course analysis of stroop interference and facilitation: Comparing normal individuals and individuals with schizophrenia. *Journal of Experimental Psychology-General, 126*(1), 19-36.

Seidenberg, M. S. (1993). Connectionist Models and Cognitive Theory. *Psychological Science, 4*(4), 228-235.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, UK.: CUP.

Sharkey, A. J. C., & Sharkey, N. E. (1995). Cognitive Modeling: psychology and Connectionism. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 200-203). Cambridge, MA.: The MIT Press.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition, 26*(5), 1033-1040.

Smolensky, P. (1988). On the Proper Treatment of Connectionism. *Behavioral and Brain Sciences, 11*(1), 1-23.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology-Human Perception and Performance, 26*(2), 506-526.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Styles, E. A. (1997). *The Psychology of Attention*. Hove: Psychology Press Ltd.

Sugg, M. J., & McDonald, J. E. (1994). Time-course of inhibition in color-response and word-response versions of the stroop task. *Journal of Experimental Psychology-Human Perception and Performance, 20*(3), 647-675.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 54*(2), 321-343.

Treisman, A., & Fearnley, S. (1969). The Stroop Test: Selective Attention to Colours and Words. *Nature, 222*, 437-439.

Tunstall, M. J., Oorschot, D. E., Kean, A., & Wickens, J. R. (2002). Inhibitory interactions between spiny projection neurons in the rat striatum. *Journal of Neurophysiology, 88*(3), 1263-1269.

Turken, A. U., & Swick, D. (1999). Response selection in the human anterior cingulate cortex. *Nature Neuroscience, 2*(10), 920-924.

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop Effects By Manipulating Expectations For Color Words. *Memory & Cognition, 20*(6), 727-735.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550-592.

Vallacher, R. R., & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry, 8*(2), 73-99.

Weekes, N. Y., & Zaidel, E. (1996). The effects of procedural variations on lateralized stroop effects. *Brain and Cognition, 31*(3), 308-330.

White, N. M. (1997). Mnemonic functions of the basal ganglia. *Current Opinion in Neurobiology, 7*(2), 164-169.

Wiles, J., Chenery, H. J., Hallinan, J., Blair, A., & Naumann, D. (2000). Stroop performance in Alzheimer's disease: A preliminary test of theories of damage using a connectionist simulation. *Brain and Language, 74*(3), 341-344.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychological Bulletin, 120*(1), 3-24.

Wilson, C. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 29-50). Cambridge, MA.: MIT Press.

Wise, S. P. (1996). The role of the basal ganglia in procedural memeory. *Seminars in Neurosci., 8*, 39-46.

Wise, S. P., Murray, E. A., & Gerfen, C. R. (1996). The frontal cortex-basal ganglia system in primates. *Critical Reviews in Neurobiology, 10*(3-4), 317-356.

Wood, T. J., & Milliken, B. (1998). Negative priming without ignoring. *Psychonomic Bulletin & Review, 5*(3), 470-475.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research-Psychologische Forschung, 63*(3-4), 212-233.

Young, A. W., & Burton, A. M. (1999). Simulating face recognition: Implications for modelling cognition. *Cognitive Neuropsychology, 16*(1), 1-48.

Zhang, H. Z., & Kornblum, S. (1998). The effects of stimulus-response mapping and irrelevant stimulus- response and stimulus-stimulus overlap in four-choice stroop tasks with single-carrier stimuli. *Journal of Experimental Psychology-Human Perception and Performance, 24*(1), 3-19.

Zhang, H. Z. H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology, 38*(3), 386-432.

Zorzi, M., & Umilta, C. (1995). A Computational Model of the Simon Effect. *Psychological Research-Psychologische Forschung, 58*(3), 193-205.