Human task switching and the role of inhibitory processes: A computational modelling and empirical approach

Thesis submitted for the degree of Doctor of Philosophy in the
Department of Psychological Sciences

Birkbeck, University of London

Nicholas J. Sexton

2018

Declaration

I declare that this thesis and the research contained herein is entirely my own work. All work and ideas on which this thesis has drawn which are not my own have been clearly attributed.

Nicholas J. Sexton

For Hazel.

“What would it have looked like,” he said, “if this TV set projected all channels onto the cathode ray screen *at the same time?* Could we have distinguished anything, in the mixture?”

“I don’t think so.”

“Maybe we could learn to. Learn to be selective; do our own job of perceiving what we wanted to and what we didn’t. Think of the possibilities, if our brain could handle twenty images at once; think of the amount of knowledge which could be stored during a given period. I wonder if the brain, the human brain —” He broke off. “The human brain couldn’t do it,” he said, presently, reflecting to himself.

(Dick, 1969)

Acknowledgements

It is said that it takes a village to finish a PhD thesis. I'd like to express thanks to some of them specifically. To Hazel, for her ongoing support, patience and encouragement throughout this whole project and continuing to believe in it when I didn't. There have been times when I haven't been easy to live with, so thank you for putting up with me. To my parents, for their kind and unconditional support when I've needed it, and particularly to my mum, for nurturing an interest in psychology and encouraging me through my studies. To my supervisor, Rick, for his continued inspiration, constructive criticism, patience, trust, and gentle encouragement. I hope you all share the satisfaction and pride that I do at seeing it finally completed. To my big sister Rosi, for giving such a shining example of how to get the most out of yourself. To all the friends and fellow PhD travellers I've met along the way, but particularly Olivia, Suzanne, Aji, Anna, Nicole, Chris, and Eva. To my undergraduate tutors, particularly Karen, who encouraged and inspired me to think that science was something I could do. To the members of Islington Cycling Club for keeping me going, and patiently letting me explain ideas and think out loud, on and off the road. But most particularly, for helping me to feel that this project — like a good bike ride — has an end to it, after which you can lean the bike up against a wall, contemplate the road you've travelled, and look forward to all the roads yet to come.

This work was supported by the Economic and Social Research Council [grant

number ES/J500021/1].

Abstract

Task switching is a behavioural paradigm within cognitive psychology that has been claimed to reflect the activity of high-level cognitive control processes. However, classic behavioural markers such as the (n-1) switch cost have also been shown to reflect a multitude of other cognitive processes. The n-2 repetition paradigm has proven more successful, with a behavioural measure (the n-2 repetition cost) agreed to be reflective of a cognitive inhibition mechanism ('backward inhibition'). The present thesis develops computational models of task switching, including a backward inhibition model. The models are developed within the interactive-activation and competition (IAC) framework, as a development of an existing task switching model. Modelling is constrained by the general computational principles of the IAC framework and default parameter settings where these are shared with earlier models. The effect of specific novel parameter settings on behaviour is explored systematically. The backward inhibition model predicts a range of empirically observed behavioural phenomena including both n-1 switch and n-2 repetition costs, and the modulation of the n-2 repetition cost under certain circumstances, including the manipulation of intertrial intervals. A specific prediction of the model, the modulation of n-2 repetition costs according to switch direction when tasks are of different difficulties, is tested empirically, with results confirming and providing validation of the model. Finally, consideration is given to how such a backward inhibition model could be adapted to maximise performance benefits in different task switching contexts, via a process of parameter tuning.

Contents

List of Tables

Chapter 1

Overview

This chapter introduces the notion of task switching and the key theoretical constructs of switch costs, n-2 repetition costs, and backward inhibition. The chapter ends with an overview of the computational methodology and the structure of the thesis as a whole.

1.1 Introduction

In everyday situations, people frequently switch between two or more tasks. For example, when preparing a manuscript on a computer, a researcher might switch between manipulating features of the word processor (e.g., formatting text) and typing. Additional goal-directed diversions may present themselves, such as engaging in conversation or checking e-mail. One component of this ubiquitous 'multitasking' is the ability to switch between tasks. It is frequently assumed that cognitive control processes are required to manage cognitive resources in order to ensure that current behaviour is appropriate to the task at hand.

Within cognitive psychology, it has long been known that alternating between two different tasks is slower than when performing the same tasks in single-task blocks (Jersild, 1927). More recently, task-switching paradigms have been developed which compare reaction times when a given task is a repetition of the previous task, to when it is a switch from the previous task (e.g., Allport, Styles, & Hsieh, 1994; Rogers & Monsell, 1995; Meiran, 1996). Such experimental procedures robustly reveal 'switch costs' — a slower reaction time and/or increased error rate for switch trials than repeat trials. In the last two decades, a wealth of empirical data has been collected suggesting that switch costs are affected by a range of task parameters. For example, preparation time, task cues, experience on the task, response selection and execution processes, etc. (see Kiesel et al., 2010, for a review). However, the theoretical interpretation of switch costs has been controversial, and the extent to which switch costs directly reflect cognitive control processes remains unclear.

While task switching is in itself a theoretically interesting domain of research, with potential applications such as in assessing inhibitory function in clinical neuropsychology (e.g., Mayr, Diedrichsen, Ivry, & Keele, 2006, but see Grange, Juvina, & Houghton, 2013), this thesis regards task switching as a vehicle for the more general study of cognitive control in complex cognition. Within the executive function literature, task switching is commonly cited as one, of many, computationally heterogeneous executive functions in a fractionated model (Cooper, 2010). One strand of evidence comes from individual differences studies that have suggested that multiple candidate executive functions, such as task switching, working memory updating, and response inhibition, have a common basis, while remaining computationally distinct (Miyake et al., 2000; Miyake & Friedman, 2012). These studies have identified executive functions as dimensions of individual difference, based on factor analysis of a wide range of putative executive tasks, including various task switching procedures.

A limitation of this approach is the absence of a comprehensive account of the cognitive processes involved. While computational models of putative executive tasks or mechanisms exist (see Cooper, 2010, for a review), typically these remain poorly integrated. For example, models of task switching (Gilbert & Shallice, 2002) and conflict monitoring (Botvinick, Braver, Barch, Carter, & Cohen, 2001) explain different control processes using the same computational mechanism, namely attentional biasing of processing pathways (cf. Cohen, Dunbar, & McClelland, 1990; Cohen & Huston, 1994). This thesis extends existing computational approaches to task switching in a theoretically and empirically grounded fashion by incorporating a conflict monitoring mechanism.

1.2 Theoretical overview

It has frequently been assumed that switch costs primarily reflect the duration of cognitive control processes which are specific to task switching (Rogers & Monsell, 1995) or time required to reconfigure the cognitive system appropriately to perform the alternative task (e.g., Vandierendonck, Liefooghe, & Verbruggen, 2010) (henceforth, 'reconfiguration' accounts, e.g., Rogers & Monsell, 1995; Monsell, 2003). However, some have argued that switch costs primarily reflect interference caused by competing tasks ('interference' accounts, e.g., Allport et al., 1994; Allport & Wylie, 2000; Gilbert & Shallice, 2002). Importantly, interference accounts are agnostic on the subject of extra cognitive processes specific to switching task or reconfiguring the cognitive system. Thus, theoretically, the interference and reconfiguration hypotheses are not incompatible. Indeed, integrative theories attempt to combine both hypotheses, suggesting that task switching involves both overcoming interference from a previously performed task, and reconfiguring the cognitive system (Monsell, 2003; Vandierendonck et al., 2010). However, reconfiguration and interference accounts give conflicting accounts on the source of switch costs. Reconfiguration accounts implicitly assume that cognitive processing occurs in linear, sequential stages, and thus suggest that switch costs can be partitioned into components which directly reflect components of cognitive processing. In contrast, interference accounts argue that residual activation of a previously performed task interacts with nascent activation of the current task, and that the resulting interference is sufficient to produce the empirically observed pattern of switch costs. These accounts suggest that the reaction time of processing is not additive; reaction times of repeat trials cannot simply be subtracted from those of switch trials to give a meaningful index of a specific cognitive process. Thus, while reconfiguration accounts accept that a component of switch costs may be attributable to interference, the interference view holds that inferences cannot be drawn about cognitive control processes directly on the basis of switch costs. It is therefore important to establish what quantitative predictions the interference view makes about behaviour in task switching experiments. Indeed, a computational model of the interference account, implemented in a parallel distributed processing (PDP) framework, was able to simulate a range of empirical task switching effects that had previously been taken as primary evidence for discrete stages of cognitive processing proposed by the reconfiguration account (Gilbert & Shallice, 2002). Empirical and computational work on task switching is reviewed and evaluated in chapter 2.

Given the limitations of paradigms involving switches between two tasks to provide insight into cognitive control processes, instead this thesis turns to phenomena within a related paradigm. In a paradigm involving three tasks, (A, B and C), Mayr and Keele (2000) established that ‘n-2 repetitions’ (A-B-A) are both slower, and more error prone, than ‘n-2 repeats’ (C-B-A). It has been argued that this reflects ‘backward inhibition’, or the operation of a process of cognitive inhibition, acting on the representation of the n-2 task-set. To-date, the phenomenon has proven robust to non-inhibitory explanations. The extant literature on the backward inhibition hypothesis is reviewed in chapter 3.

Given the similarity between the two- and three-task switching paradigms, psychological theories and computational models of task switching should readily generalise between them. Thus, additional insight in the cognitive processes underlying task switching may be obtained by applying models from one paradigm to the other. To date, only Grange and colleagues (Grange et al., 2013; Grange & Juvina, 2015) have developed computational model of the three-task switching paradigm, in this case, based within the ACT-R architecture. However, it is a limitation that their model does not generalise between the two- and three-task paradigms without significant modification, or additional theoretical assumptions (e.g., the strategic adaptation hypothesis, explored in chapter 10).

This thesis therefore explores the computational cognitive mechanisms required by the phenomenon of backward inhibition, additional to the processes of cross-task interference that give a good account of empirical phenomena within the two-task switching literature.

1.3 Methodology and thesis structure

Theory development within cognitive psychology inevitably involves specification of numerous entities, including processes (e.g., inhibition, active maintenance, selection, etc.) resources (e.g., stores or buffers) representations (e.g., schemas, stimulus-response pairings), and their relationships. Verbal (or 'box-and-arrow') models have usefully been used to specify theories, and offer a ready means of conceptualising cognitive processes. However, the scope of this utility is limited in a number of ways. Firstly, such models may not specify cognitive entities in precise enough terms to ensure a theory fully explains a phenomenon. Secondly, it is not clear from such models whether a theory is both necessary and sufficient to explain the phenomenon. Finally, such theories have limited ability to make detailed quantitative predictions that can be empirically falsified.

The use of computational cognitive models, therefore, provides a means of specifying a theory in a detailed way. By having to specify all of the aspects of a model, in a way that can be fully implemented in a computer program, the process can draw attention to aspects of the theory that had not been fully worked out in verbal models (as an example, Grange et al., 2013, established that the lack of a 'backward inhibition' effect did not imply the absence of cognitive inhibition, as had been assumed in the empirical literature, but merely reduced inhibition). The need to fully specify all aspects of a working computational model may include additional details which are not fully contained within the theory, or on which the theory is agnostic. For example, the specific choice activation function for the units in connectionist networks is not generally regarded as an important part of a psychological theory. The amount of implementational detail may be considerable for a computational model of any great complexity. If modeling is to contribute to psychological theory, it is therefore important the predictions of models are shown to be independent of any specific implementational details that are not a part of the theory (Cooper, Fox, Farringdon, & Shallice, 1996). Alternatively, for those details that are important to the behaviour of a model, the theory must be expanded. In the present thesis, the potential role of the extra-theoretical concept of 'negative conflict' in the simulation of the empirically observed behaviour pattern is explored by presenting three alternative versions of the model (chapters 4 and 6).

A related issue is the degree of freedom afforded to the modeler. With a potentially vast array of implementational details that are left unspecified by psychological theory, the specification of models may potentially be left unconstrained, giving modelers great scope for simulating arbitrary patterns of behaviour. One approach to constraining the range of implementational detail available to the modeler working within a specific domain is to create their model within a cognitive architecture (Newell, 1990). The architecture itself, by providing the general-purpose cognitive aspects of the model common to all tasks (e.g., inputs and outputs, such as sensory and motor systems; generalised cognitive processing characteristics such as activation functions) reduces the details that must be specified in a particular model to those specific to the theory in question. While the most prolific example of this approach has been the ACT-R architecture (Anderson, 2007), it has been argued that an alternative is the development of interactive-activation and competition (IAC) networks, under the constraints of certain common principles (McClelland, 1993). This is the approach adopted by this thesis, in the continued development of specific IAC models (i.e., Cohen et al., 1990; Cohen & Huston, 1994; Gilbert & Shallice, 2002).

A final methodological issue concerns the falsifiability of computational models. As the implementation of a specific psychological theory, it is important that a model can be potentially falsified. While it is important that a putative model of a cognitive process fits the available empirical data, at least equally important is the range of data (that is, behavioural patterns) that the model does not fit, or excludes (Roberts & Pashler, 2000). Thus, if model fitting is used exclusively as a criterion for the success of a model, a careful approach to fitting is essential, as is exploring the dependence of a particular behaviour pattern on specific parameter settings. One alternative approach to model validation is to use the model to generate specific, testable behavioural hypotheses and to show that these predictions hold over a broad range of theoretically justified parameter values. If a model makes a clear, falsifiable prediction in a novel behavioural paradigm, and this prediction is empirically verified, support for the model is provided in a similar way to non-computational theories. This approach is adopted in this thesis. Chapters 4 and 5 describe a model of three-task switching and present initial simulations, while chapter 6 conducts systematic studies of the model's behaviour across parameter space and chapter 7 generates empirical predictions on manipulation of the n-2 repetition cost when switching between three-tasks of asymmetric difficulty. These predictions are subsequently tested in chapter 8. Further simulation of empirical effects are reported in chapter 9, while chapter 10 addresses the computational mechanisms underlying participants' strategic adaptation to their current task context. Finally, chapter 11 presents a general discussion and questions for future research.

Chapter 2

Task Switching: A Review of Empirical and Computational Evidence

This chapter presents an overview of the main procedural paradigms used in task switching experiments and the main empirical effects. This is followed by discussion of the implications of these results for theoretical accounts of task switching. Within the field of task switching, one strand of research has focused on switching between tasks of unequal difficulty. This sub-paradigm presents a range of its own empirical effects, such as asymmetric switch costs. The second half of the chapter reports a re-implementation of two computational models of task switching between tasks of unequal difficulty.

2.1 Empirical paradigms and effects

Within cognitive psychology, the domain of task switching aims to study the cognitive processes underlying the human ability to switch between different tasks. Nearly two decades of research has revealed a large range of empirical effects, of which the most well known is the switch cost — performing a new task is both slower and more error prone than when repeating that task. A number of different empirical procedures have been used to examine different aspects of task switching behaviour, and a number of the empirical effects are therefore procedure-specific. This section presents a non-exhaustive overview of the main empirical procedures that have been used historically, and the typically observed empirical effects (see Kiesel et al., 2010, for a full review).

A further major division in empirical task switching research has been between studies using tasks which are relatively equally well-learned tasks, and those which use tasks with different levels of practice. The former studies include classification tasks, such as classifying letters as consonants or vowels, or digits as odd or even (e.g., Rogers & Monsell, 1995). The latter include tasks such as word reading and colour naming of incongruent Stroop stimuli (e.g., Allport & Wylie, 2000). Importantly, a body of empirical effects are specific to switching between tasks of different difficulties, such as asymmetric switch costs.

2.1.1 The list procedure

In laboratory situations, it has long been known that alternating between two different tasks is associated with a performance cost ('switch costs', usually measured in terms of a slower response time, but also, frequently, response accuracy) when compared with performing the same tasks in single-task blocks. For example, completing a list comprised of two simple arithmetic tasks, such as adding and subtracting 3, is slower when participants must complete a mixed list, alternating between addition and subtraction, than a list consisting of two single task blocks, completing all the addition tasks followed by all the subtraction tasks (Jersild, 1927). While similar list-based paradigms are still used to assess task switching (e.g., Miyake et al., 2000, used a list procedure alongside other switching tasks) and remain attractive for their simplicity, they have a number of limitations. The time-cost of switching is attributable to the overall difference between blocks of trials, and is difficult to attribute to any specific components of behaviour. For this reason, they are usually referred to as ‘global’ switch costs or ‘mixing costs’, Additionally, such procedures may be confounded by the additional requirement to hold two tasks in memory when completing the mixed list, whereas single-task blocks may be completed holding only a single task in memory for most of the procedure.

2.1.2 The alternating runs paradigm

To address the limitations of the list procedure, Rogers and Monsell (1995) developed a procedure in which participants complete alternating runs of tasks. While the run length is variable, it is typically fixed such that the task sequence is entirely predictable. For example, two tasks A and B, might be sequenced in runs of length 2, resulting in the task sequence $AABBAA$ etc., (Rogers & Monsell, 1995), although studies with longer sequences have been conducted (Monsell, Sumner, & Waters, 2003). Typically, the stimuli appear in a rotating sequence around a four-quadrant grid. Participants are instructed to perform task A when stimuli appear in one half of the grid, and task B in the other half. Thus, ‘switch’ and ‘repeat’ trials regularly alternate within the same block of trials, and the task to perform is implicitly visually cued, removing the requirement to maintain the current task in working memory.

Switch costs

The main emprical effect in this paradigm is robust switch costs, calculated as the average RT difference between ‘switch’ and ‘repeat’ instances of each task. Assessed in this way, switch costs reflect the additional time required to complete a task switch.

The theoretical interpretation of switch costs has been controversial. One frequent suggestion has been that switch costs reflect the operation of cognitive control processes required by switch trials but not by repeat trials (Rogers & Monsell, 1995), or the time to reconfigure the cognitive system and set task parameters appropriately for a new task (Vandierendonck et al., 2010). Alternatively, it has been argued that switch costs primarily reflect an interference effect of having recently performed an alternative, competing task. This concept has been expressed intuitively in terms of 'task-set inertia' (Allport et al., 1994; Allport & Wylie, 2000) and expressed formally as residual activation of the competing task in an interactive activation model.

However, the utility of the alternating runs paradigm for assessing the costs of a task switch was called into question by Altmann (2007a). Within this paradigm, every task switch is also the first trial in a run. This measure of switch costs, therefore, is fundamentally confounded with any effects of trial position. Indeed, Altmann's experiments provide evidence for such run-position effects on RT when runs are separated by a longer RSI (see restart costs, section 2.1.2) but also when the RSI is uniform within and between runs. This criticism provides a rationale for the inclusion of runs with no switch of task (i.e., repeat runs) to separate the effects of a task-switch from those of run position, as in what Vandierendonck et al. (2010) refer to as the "intermittent instruction procedure". Additionally, it implies caution when comparing switch costs between alternating runs and explicit cueing paradigms. As each trial in the latter paradigm is cued separately, it is effectively a series of runs of length 1, implying switch costs may be confounded with restart costs.

Task serial position effect

A frequently reported property is that an elevated RT following a switch of task is restricted to the first trial after a task switch (henceforth, the 'switch trial'), while RTs are usually approximately equal for subsequent trials in which the task repeats ('repeat trials') (e.g., Rogers & Monsell, 1995; Allport & Wylie, 2000; Rubinstein, Meyer, & Evans, 2001). While it has been argued that this effect is not consistent with interference accounts of switch costs (e.g., Rogers & Monsell, 1995; Rubinstein et al., 2001), the computational model of interference in task switching of Gilbert and Shallice (2002) reproduced this effect, suggesting that the presence of this effect alone does not discriminate between competing theories. While it has been debated whether this effect constitutes evidence against an interference-based account of switch costs (Rogers & Monsell, 1995; Rubinstein et al., 2001, but see Gilbert & Shallice, 2002), direct comparison with the explicit cueing paradigm suggests that this effect may be specific to switching procedures with predictable runs (see section 2.1.4).

Preparation effects

The effect of preparation on switch costs has been of major interest. According to the logic of the reconfiguration account, when the task required on the next trial is predictable and/or explicitly cued, a greater interval in which to prepare for the upcoming task should allow participants to complete any control processes required by the task switch, and thus should reduce switch costs (e.g., Rogers & Monsell, 1995). A number of studies have shown that in the alternating runs paradigm, switch costs are indeed reduced by an increased response-stimulus interval (RSI) (e.g., Meiran, 1996; Rogers & Monsell, 1995, experiment 3).

This robust finding has often been taken as primary evidence for a reconfiguration account of task switching. According to this perspective, participants use preparation intervals to complete 'advance reconfiguration', leading to a reduction in the reconfiguration required after stimulus onset. As this paradigm typically involves predictable sequences and obvious indication of the participant's current position in the sequence, the preparation interval (i.e., the time interval between knowing which task will be required next and the stimulus required to actually initiate performing the task) cannot be directly manipulated. As a result, it is difficult to definitively demonstrate using this paradigm whether preparation effects reflect factors relating to the task switch (i.e., reconfiguration) or factors relating to the previous trial (i.e., passive task-set decay). Thus, most studies on preparation in task switching have preferred the explicit cueing paradigm, which allows these two variables to be manipulated independently.

Residual switch costs

While switch costs are reduced with preparation, a number of studies have shown that they have a residual component which is not entirely eliminated by preparation, even for preparation intervals as long as 8s (Kimberg, Aguirre, & D'Esposito, 2000), while others do not show such a residual effect (Meiran, 1996; Meiran, Chorev, & Sapir, 2000; Tornay & Milán, 2001; De Jong, 2000). Within reconfiguration accounts of the preparation effect, some theorists have argued that preparation for a task consists of two stages, only the first of which may be completed in advance of the target stimulus (so-called two stage models, e.g., Rogers & Monsell, 1995; Meiran et al., 2000). According to these models, it is the second stage of reconfiguration that requires a target stimulus, and this is the cause of the residual switch cost. This claim will be discussed further in section 2.2, below.

Alternative accounts of the residual switch cost propose that it does not reflect reconfiguration processes but is the result of interference. These accounts may attribute the entirety of the switch cost to interference (e.g., Allport & Wylie, 2000). Alternatively, 'hybrid' accounts of switching assume an active reconfiguration process, but that the residual switch cost reflects an interference component (Monsell, 2003).

Restart costs

In procedures in which runs are separated by a longer intertrial interval (RSI), the first trial of the next run typically has a longer RT than subsequent trials, even when the task does not switch (e.g., Allport & Wylie, 2000; Waszak, Hommel, & Allport, 2003). This effect is further discussed in section 2.1.3, below.

Asymmetric task difficulties in the alternating runs paradigm

A range of studies in the alternating runs paradigm have been conducted on switching between pairs of trials with asymmetric difficulties. Most typically, in a modified task-switching version of the Stroop paradigm, participants switch between word reading and colour naming of Stroop stimuli (Allport et al., 1994; Allport & Wylie, 2000; Bryck & Mayr, 2008; Rogers & Monsell, 1995; Yeung & Monsell, 2003), although similar effects have been found with simple congruent or incongruent spatial compatibility rules (Bryck & Mayr, 2008), picture naming and word reading tasks (Waszak et al., 2003), arithmetic tasks (Schneider & Anderson, 2010), and switching between languages (Meuter & Allport, 1999; Philipp, Gade, & Koch, 2007)[1].

However, the findings discussed in this section are not universal. Some studies have failed to find asymmetric switch costs (e.g., Monsell, Yeung, & Azuma, 2000), and switch cost asymmetries have successfully been reversed by experimental manipulations, such as by introducing a stimulus onset asynchrony between the word and colour elements of Stroop stimuli (Yeung & Monsell, 2003).

Asymmetric switch costs Empirically, it has been found that when completing two tasks with unequal levels of practice, in response to stimuli which afford both tasks — such as word reading and colour naming in response to Stroop stimuli — that switch costs are highly asymmetric. Counter-intuitively, switch-costs are higher for switching to the *more* practiced task (i.e., word reading), than for switching to the *less* practiced task. Given that reaction times are generally slower for the less practiced tasks, this effect has been described as ‘paradoxical’ (Allport et

---

[1] Additionally, similar effects have been found when switching between pro-saccades and anti-saccades, although generalising from these studies may be problematic due to inhibition mechanisms specific to the oculomotor system. See Kiesel et al. (2010) for a review.

al., 1994). Switch cost asymmetries have been demonstrated using Stroop stimuli (Allport et al., 1994) and first and second language reading in bilinguals (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999; Philipp, Gade, & Koch, 2007). While the switch cost asymmetry is a robust finding, some studies have shown that the asymmetry may be eliminated or even reversed under some circumstances (e.g., Yeung & Monsell, 2003).

Reverse Stroop interference It has been observed that when participants switch between word reading and colour naming of Stroop stimuli, the RT of word reading trials is more affected by switching than colour naming trials (Allport et al., 1994; Allport & Wylie, 2000). Word reading is slower for incongruent than neutral stimuli on switch trials, but not repeat trials; this effect is referred to as a 'reverse-Stroop' effect. In contrast, colour naming is slower for incongruent stimuli in both switch and repeat trials. This effect has been interpreted as evidence in favour of interference-based accounts of switching.

Limitations of the alternating runs paradigm

The alternating runs paradigm has a number of limitations. Firstly, the procedure only investigates switching where task sequences are predictable. The procedure tends not to be used to study the effect of predictability on task switch effects (although, see Monsell et al., 2003). Secondly, studies using this procedure have typically used short run lengths, of two or four. According to the reconfiguration account, it is typically assumed that once switching processes have taken place, there should be no residual effects of a task switch on subsequent repeat trials. This is reasonable, given that switch costs are typically restricted only to the first trial in a run. However, Allport and Wylie (2000) reported experiments in which persisting effects of interference were found after 24 trials, and even after several blocks of exclusively performing repeat trials. These findings are reviewed in section 2.1.3, below. Thirdly, while preparation effects suggest that longer intertrial intervals may reduce switch costs, the paradigm does not offer a means of distinguishing the passive decay of residual activation from the previous trial (Meiran et al., 2000), from active processes which prepare for the upcoming task (e.g., Rogers & Monsell, 1995; Rubinstein et al., 2001). This limitation is addressed by the explicit cueing paradigm, discussed in section 2.1.4. Finally, in most instances of the alternating runs paradigm, there are no runs in which the task repeats – each run is a switch run. Thus, Altmann (2007a) persuasively argued that the alternating runs paradigm structurally conflates switch costs with first-trial effects.

2.1.3 The before-and-after paradigm of Allport and Wylie (2000)

Allport and Wylie (2000) used a variation of the blocked, predictable switching paradigm to examine the longer-lasting effects of prior performance of a conflicting task. In a series of experiments using colour naming and word reading tasks in response to Stroop stimuli, participants either performed 30 colour naming trials, or rested, in between baseline and 'postcolour' phases of exclusively word reading blocks, each consisting of 10 trials and alternating between neutral and Stroop stimuli. Each block of trials was separated by a longer intertrial interval, and the task was explicitly cued.

Typically, switch costs and a characteristic asymmetry (i.e., greater for switches into the dominant task than the non-dominant task) are observed in this paradigm. Additional effects are reviewed below.

Restart costs

Following a short break the first trial in a block is typically slower than subsequent trials, even with no switch of task. For example, Allport and Wylie (2000, experiment 3) report restart costs following a 2s break between blocks of 10 trials. Restart effects on the first trial of each block were observed for all participants, including control participants who performed no colour naming, but rested between baseline and 'postcolour' phases. This effect robustly appears in blocked paradigms (Allport & Wylie, 2000; Bryck & Mayr, 2008) and the intermittent instructions paradigm (Altmann, 2007a).

Rebound effects

Allport and Wylie (2000) report a striking interaction between switching and restart costs. In their experiments 3–5, participants performed consecutive blocks of word reading (termed 'postcolour' blocks), with no further task switches, after either an unrelated RT task (control condition) or a block of colour naming (colour naming condition). On the first trial of postcolour block 2, control participants were slower than on subsequent trials (a standard restart effect), while those who performed colour naming showed an additional RT cost above that of the control participants. Additionally, participants who performed colour naming showed a large increase in RT (i.e., a switch cost) on the first postcolour word reading block, but also a smaller, but statistically significant, increase in RT on the first trial of the second block. Allport and Wylie (2000) interpreted this effect as suggesting that, not only was interference persistent over a large number of trials, but that it interacts with restart costs ('rebounds') on the first trial of a run.

While this effect has been relatively overlooked, Bryck and Mayr (2008) replicated the finding, in both simple congruent or incongruent spatial compatibility tasks, and Stroop colour naming and word reading. In a modified alternating runs paradigm with runs of 4 trials, interrupted by a longer RSI between trials 2 and 3 (i.e., $AA - AA - BB - BB - AA\ldots$) the restart trial (i.e., trial 3) exhibited the RT asymmetry characteristic of switch costs (i.e., a greater switch cost when switching to the dominant task) even with no switch of task.

Rebound effects are problematic for a number of reasons. Firstly, they assume that switch costs directly reflect switch-related control processes (although, they do not, in themselves, preclude an additional time cost of such control processes.) Secondly, they assume that repeat trials reflect a stable, fully configured cognitive system. Thirdly, they assume that interference is restricted to switch trials, and is overcome or eliminated on repeat trials. Fourthly, it suggests that interference is latent, and may persist during repeat trials despite their stable RT. Rebound effects are challenging for reconfiguration accounts of task switching to explain, and provide compelling evidence that switch costs are a problematic measure of any cognitive reconfiguration processes. While they are more consistent with an interference account of task switching, interference resulting from the carryover of activation from the previous trial does not provide a sufficient explanation.

Instead, both Allport and Wylie (2000) and Bryck and Mayr (2008) interpret these findings as suggesting an associative-learning mechanism whereby task sets become associated with specific stimulus items. Additionally, Bryck and Mayr (2008) propose that associative learning is further modulated by cognitive control: the higher level of control required to perform the non-dominant task facilitates associative learning, thereby strengthening associations between the stimulus item and the non-dominant task, more than the dominant task. Such long-term effects of interference were not addressed by Gilbert and Shallice (2002) in their computational model of the interference account, and without further tests it is unclear whether the model can simulate such effects.

Item-specific switch costs

Allport and Wylie (2000, experiment 5) reported compelling evidence that a component of switch costs could be attributed to negative priming. Specifically, stimuli which had previously appeared in a colour-naming block of trials resulted in slower RTs when they appeared on word reading trials, than when they had not appeared for a colour-naming task. This effect was greater for switch than for repeat trials.

Neutral stimuli in the destination (WR) task

Firstly, when neutral stimuli are used to perform the dominant task, costs of a switch into this task are typically reduced, although not eliminated. For example, Allport and Wylie (2000, in experiment 4), the effects of switching from colour naming of incongruent stimuli to word reading, contrasted incongruent against neutral word reading stimuli. The cost of switching to word reading was significantly reduced for neutral stimuli, although the switch cost was still substantial (see also Rogers & Monsell, 1995, experiment 1).

Neutral stimuli in the prior (CN) task

Allport and Wylie (2000, experiment 1) examined the effect of varying the degree of interference between colour naming and word reading tasks, by using neutral stimuli only on colour naming trials, compared to the typical incongruent stimuli. The cost of switching to word reading was much greater when incongruent stimuli were used in the colour-naming task, rather than neutral stimuli, even though all word-reading stimuli were incongruent. This was interpreted as suggesting that switch costs depend primarily on characteristics of the previous, rather than the current trial, and is key evidence that interference is a key cause of the switch cost.

2.1.4 The explicit cueing paradigm

We recall that the reduction of switch costs with greater time to prepare for the forthcoming trial is taken as primary evidence of reconfiguration accounts of task switching. However, within the alternating runs paradigm, where only the inter-trial interval (the response-stimulus interval, or RSI) is manipulated, it is difficult to discriminate between effects of preparation from what might be passive decay-effects of the previous trial. For this purpose, the explicit cueing paradigm offers better control of task timing parameters.

In an explicitly cued switching task, the order of tasks is randomised, hence both the task sequence and the need to switch tasks are unpredictable. Participants are instructed which task to perform by an on-screen cue, which appears before the task stimulus. The RSI in the alternating runs paradigm is partitioned into response-cue interval (RCI) and cue-stimulus interval (CSI).

Within this paradigm, the RSI and CSI parameters may be varied independently, and are assumed to primarily influence the residual effects of the previous trial and preparation for the forthcoming trial, respectively.

Preparation effects

1. **Preparation effects from variable foreperiod (CSI)** Typically, the effect of preparation for an upcoming trial is assessed by varying the CSI between trials. In this manipulation, longer CSI's robustly reduce switch costs, and also reaction times generally (e.g., Meiran, 1996). However, some caution is required in interpreting this finding, and two main lines of research suggest that preparation may not be specific to task switching, making questionable theoretical accounts taking preparation effects as primary evidence for reconfiguration (see Kiesel et al., 2010, pp. 856-858, for a detailed review).

   Firstly, when CSI was varied between-subjects, longer CSIs produced a general reduction in RTs, but no significant reduction in switch costs (Koch, 2005; Koch & Allport, 2006), and two studies which specifically compared within-subjects and between-subjects manipulations of CSI found that switch costs were only significantly affected in the within-subjects design (Koch, 2001; Altmann, 2004). Altmann and Gray (2008) present a computational model accounting for these preparation effects, although the model may not be compatible with evidence from Monsell and Mizon (2006). This model is reviewed in section **??**.

2. **Preparation effects from sequence predictability** The second line of evidence comes from studies which used a constant CSI, but manipulated the

predictability of the task sequence. For example, presentation of an additional cue on each trial, indicating the probability of a task repeat or a switch, produced faster overall reaction times for high compared to low probability cues, but did not produce a significant difference in switch costs, with preparation benefits due to predictability seeming to affect both switch and repeat trials (Dreisbach, Haider, & Kluwe, 2002). Similarly, in a procedure in which the sequence was explicitly cued, but the predictability of the sequence itself was manipulated (varying between the $AABB$ design used in the alternating runs paradigm, and an unpredictable sequence), RTs were generally slower for an unpredictable sequence but switch costs did not differ between predictable and unpredictable sequences (Koch, 2005).

Summary - preparation effects Taken together, the previous two effects suggest that the benefits of preparation resulting, firstly, from a preparation foreperiod after a task cue (i.e., CSI) and secondly, from sequence predictability, may not specifically benefit task switching. Thus, it cannot be assumed that preparation effects necessarily reflect advance reconfiguration, as is sometimes assumed by discrete stage-based reconfiguration accounts (e.g., Schmitz & Voss, 2012). The computational model of Gilbert and Shallice (2002) may also be theoretically under-specified with respect to preparation effects. The simulation of preparation by the prior activation of task demand units assumes that preparation also makes the above assumption, and probably is not sufficient to account for a wider range of empirical effects.

In response, alternative theoretical proposals have been made that cognitive processes such as task-updating (cf. Rubinstein et al., 2001) need to be performed on both switch and repeat trials, or that participants may switch between task-specific or generic preparation strategies, depending on low- or high-stress contexts (Steinhauser, Maier, & Hübner, 2007).

Task serial position effects

In contrast to the alternating runs paradigm, RTs on the trials consequent to a switch of task do not tend to show a sharp discontinuity between the first (switch) trial, and subsequent (repeat) trials. Instead, in EC procedures with longer runs and less frequent switches, a graded reduction in reaction times is typically observed (e.g., Tornay & Milán, 2001; Monsell, 2003; Milán, Sanabria, Tornay, & González, 2005; Koch, 2005). This difference has been attributed to the differences in processing predictable and unpredictable sequences (see Altmann, 2007a, for a review and one theoretical account).

2.1.5 Effects of cue switching

The interpretation of explicitly cued task switching has recently been challenged by a line of research which argues that apparent switch costs actually reflect the effects of cue-encoding processes, rather than switches of task-set (see Jost, De Baene, Koch, & Brass, 2013; Kiesel et al., 2010, for recent reviews). Logan and Bundesen (2003) and Mayr and Kliegl (2003) both argued that a major limitation of the explicit cueing paradigm is that in the usual procedure task cues indicate tasks with a 1:1 mapping, meaning that every task switch also implies a cue switch. Hence, the effects of a task switch are confounded with those of a cue switch. Both sets of researchers independently devised a procedure which partially deconfounded these factors by using a 2:1 cue:task mapping. For example, in a procedure involving colour naming and shape classification tasks, the colour task may be identified by the cues “colour” or “hue”, while the shape classification task may be identified by either “shape” or “form” (Logan & Bundesen, 2003). Thus, transitions are possible in which both the cue and task switch, the cue switches but the task repeats, or where both repeat. These experiments, typically, have found an RT cost (‘cue-switch cost’) on trials where the cue switches but the task does not. Based on subtracting the cue-switch cost from the overall switch cost, a ‘corrected task-switch cost' may be derived, putatively fractionating the cost of the switch into cue-switch and task-switch components (Altmann, 2006; Logan & Bundesen, 2004; Mayr & Kliegl, 2003; Monsell & Mizon, 2006).

Problematically, Logan and Bundesen (2003) found only a small remaining corrected task-switch cost, and argued that the main cause of the switch cost was the switch of cue, rather than of the task (see also Mayr & Kliegl, 2003). However, other studies have reported a larger corrected task-switch cost in some circumstances, suggesting that cue-switch costs cannot entirely account for switch costs (Logan & Bundesen, 2004; Monsell & Mizon, 2006)

Cue-switch costs may therefore reflect the benefits of priming to cue-encoding when the cue repeats (Logan & Bundesen, 2003; Logan, Schneider, & Bundesen, 2007). Logan and colleagues used mathematical modelling to test three competing models of cue-switch costs (cue priming, reconfiguration, and priming-and-reconfiguration) by fitting them to empirical data (see also Logan & Bundesen, 2004; Schneider & Logan, 2005). The priming model was found to be sufficient to account for the empirical data, and was preferred as the most parsimonious theory (although, see Altmann, 2006, for a less adequate fit of the model). Schneider and Logan (2005) argued that participants may perform the explicitly cued task switch task switching paradigm without any actual switches of task. According to this view, by using a *stimulus-compound strategy* in which the combination of cue and stimulus may be used to uniquely identify a single correct response, participants may appear to perform switching between two, two-choice tasks, but actually be performing a single four-choice stimulus-response mapping task. Corrected switch costs, according to this view, are actually due to cues mapping onto the same task priming each other either directly, or indirectly via a common task representation.

2.1.6 Switch costs and first-trial effects in the alternating runs and explicit cueing paradigms

While first-trial effects (restart costs) are frequently reported by a number of studies in which runs are separated by a longer RSI (Waszak et al., 2003; Allport & Wylie, 2000), Altmann (2007a) reported an elegant experiment which found first-trial effects even with a uniform RSI between and within runs. In a variant of the alternating runs paradigm, participants performed runs of 2 trials, however as the task was randomised for each run, runs were either switch or repeat. Thus, the factors of run position, and task type (switch vs. repeat) could be separated.

Altmann reasoned that certain simple effects provided an analog of switch costs as assessed in both the alternating runs, and explicit cueing paradigms. Thus, the simple effect of position (1st vs. 2nd trial) on switch runs only is an analog of the alternating runs switch cost, and the simple effect of run type (switch vs. repeat) on position 1 only is an analog of the explicit cueing switch cost. Furthermore, the design allowed an assessment of the relative contribution of first-trial effects to switch costs — the simple effect of position (1st vs. 2nd trial) on repeat runs only. The results supported the suggestion that the alternating runs switch cost conflates effects of switching and trial position; in this experiment, the first-trial cost was almost exactly the difference between the analogs of the alternating runs and explicit cueing switch costs.

2.2 Theoretical accounts of task switching 1: Reconfiguration models

Reconfiguration accounts of task switching propose that the source of the switch cost is the additional time required to reconfigure the cognitive system appropriately to perform an upcoming task, which is longer for switch trials than for repeat trials. A number of reconfiguration-based cognitive models of task switching have been proposed, and various distinctions between them will be discussed in the following section.

2.2.1 Two-stage models

Two-stage models, in order to explain the preparatory reduction in switch costs and the residual switch cost, typically distinguish between *endogenous* cognitive operations, which may be performed as part of preparation for the forthcoming task, prior to stimulus onset, and *exogenous* processes, which can only be performed after the stimulus has been presented (Rogers & Monsell, 1995). According to this distinction, advance endogenous reconfiguration of the task sets during a preparation interval results in the observed reduction in switch costs, while exogenous reconfiguration processes cannot take place until after stimulus onset, accounting for the residual switch cost (Rogers & Monsell, 1995).

Various subsequent authors have accommodated this distinction within a wider theory of executive processing. For example, Mayr and Kliegl (2003) propose two discrete serial stages. The *retrieval stage* is a memory retrieval operation of task rules from long term memory (LTM) into working memory (WM). This stage may be triggered either by an externally provided cue, or through an internal act of preparation. The second stage, the *application stage*, involves a relatively automatic processing of the stimulus according to the task rules in WM, and requires the presentation of the stimulus in order to be carried out. Critically, this model assumes that only a single set of task rules may be contained in WM at any time. Therefore, a LTM retrieval operation is essential for every switch of task, even for the simple choice tasks typically used in task-switching experiments. This shared use of a limited resource is also the source for interference between the tasks (also see Meiran et al., 2000, for a similar two-process model).

Similarly, Rubinstein et al. (2001) proposed that task switching involves a goal shifting (executive) process, which loads a new abstract goal — to complete the specified task — into declarative memory. This process may be either endogenous, or exogenous, depending on knowledge of the forthcoming task and an appropriate preparation interval. Following stimulus onset, a chain of cognitive processes are executed, including stimulus identification, rule activation, response selection, and movement production. The executive process of rule activation involves activating the correct rule (conceptualised as a production rule) in procedural memory, to correspond with the current goal. According to this formulation, it is necessary to keep the rules for only a single task activated in procedural memory at a time, to minimise conflict during response selection. Rubinstein et al. (2001) explain the effects of previous-trial interference in terms of next-trial stimulus features partially activating previous-trial rules, before rule activation has been completed for the next trial, thus making them harder to deactivate, although this explanation of interference appears to violate the principle of strictly successive stages.

2.2.2 Extra process models

Somewhat orthogonally to the previous distinction, 'extra process' models assume that additional cognitive processes are required for a task switch, which are not required on a repeat trial. Once a task-set has been activated, it is usually assumed that it remains active to some extent, until an alternative task-set is activated.

Reconfiguration models which do not share this assumption, in contrast, suggest that the same processes take place in switch and repeat trials, but proceed faster on a task repeat due to mechanisms such as repetition priming (see Vandierendonck et al., 2010, p. 605, for a review).

Various cognitive operations have been proposed as part of the reconfiguration process, including the activation/chaining of task-set parameters (Logan & Gordon, 2001; Monsell & Mizon, 2006; Rogers & Monsell, 1995), or retrieval of the task-set from memory (Altmann & Gray, 2008; Rubinstein et al., 2001; Sohn &

Anderson, 2001).

2.2.3 Discrete processing stage models, and the assumption of additivity

A majority of reconfiguration models share the strong assumption that cognitive processing takes place in discrete, sequential stages (Mayr & Kliegl, 2003; Rubinstein et al., 2001; Altmann & Gray, 2008). According to these models, task switching comprises a chain of discrete subprocesses. For example, Rubinstein et al. (2001) proposes that task switching involves executive processing stages (goal shifing, rule activation) and task execution stages (stimulus identification, response selection, and movement production), which must be completed in a specific order.

Rubinstein et al. (2001) formalise these assumptions using an additive factors logic (Sternberg, 1969, 1998). According to this logic, it is assumed that the overall RT of a cognitive process is a simple sum of processing times for each stage of processing. Each stage of processing is strictly successive. That is, processing may only begin once preceding processes have been completed. Given such an arrangement of cognitive processing, multiple factors may affect each stage of processing, and some factors may affect multiple stages. However, some factors may be selective. That is, given two stages of processing, it should be possible to find factors that affect one but not the other, and vice versa. The operational assumption is that if two factors each have a main effect on the overall RT, but with no interaction, they affect different stages of processing; if they interact, they affect the same stage of processing. On this basis, it is reasoned that the structure of processing stages may be ascertained through finding a set of factors whose effects are purely additive.

The additive factors method has been influential in cognitive psychology, accounting for a wide range of RT data (Sternberg, 1969, 1998). Assuming discrete stage-based cognitive processing is, therefore, methodologically convenient, allowing the decomposition of RT differences between experimental conditions in a theoretically meaningful way. However, the assumptions of the approach are questionable, especially from the perspective of connectionist models of human cognition, which propose that rather than information processing being separated into strictly successive, discrete stages, processing instead takes place in parallel and may be interactive. Indeed, connectionist models have been shown to produce RTs which closely approximate factor additivity observed in human data (Ashby, 1982).

In summary, while stages-of-processing models of task-set reconfiguration and task execution processes remain influential (e.g., Schmitz & Voss, 2012), the demonstration of factor additivity alone cannot be taken as evidence for a discrete separation of stages, as connectionist models have demonstrated that processing with a different architecture can, in principle, approximate the same behavioural data. Empirical verification is therefore required to conclusively demonstrate that processing takes place in discrete stages. In addition, as Allport and Wylie (2000, p. 65) argued, models based on discrete, strictly successive stages in which reconfiguration must occur prior to stimulus processing cannot account for interference effects such as item-specific priming of stimuli. Thus, such models must be incomplete.

2.2.4 Do preparatory effects imply reconfiguration processes?

A number of explanations for preparatory reductions in switch costs which do not assume advance reconfiguration are possible. These are briefly reviewed here. (See Kiesel et al., 2010, pp. 854–855, for a more complete review.)

The first possibility is that performance on a trial is facilitated by a non-specific readiness to respond at a particular time (‘temporal preparation’). For example, studies using single tasks have shown that temporal preparation (e.g., afforded by an uninformative warning signal) may speed reaction times (see Kiesel et al., 2010, p. 854). However, the use of warning signals in predictable (Rogers & Monsell, 1995) or unpredictable (Meiran et al., 2000) task-switching has effects which are weak and not specific to switch trials. Hence, temporal preparation is not a satisfactory explanation for preparatory reduction in switch costs.

A second possibility is that task-set activation is subject to passive decay with longer preparation intervals. (Note this is essentially the explanation for restart costs in the model of Gilbert & Shallice, 2002.) According to this 'passive task-set decay' hypothesis, preparatory reduction in switch costs may be, to some extent, attributable to the passive decay of activation relating to the previous trial.[2] Addressing this possibility, Rogers and Monsell (1995) reported findings that varying RSI randomly within-blocks (experiment 2) did not produce the reduction in switch costs that was observed when RSI varied only from block-to-block (experiment 3). They argued that a passive task-set decay process should be independent of such experimental manipulations, and found it more plausible to assume that reduction in switch costs with longer RSI in a blocked design resulted from an advance reconfiguration process, which was disturbed when the RSI varied randomly between trials.

Most empirical studies of preparation effects have used the explicit cuing paradigm. Within this paradigm, participants are assumed to use the CSI to prepare for the upcoming task, while the RCI affects passive decay of the previous task. Thus, if the passive decay hypothesis is correct, longer RCIs should produce a reduction in switch costs. This prediction has been confirmed in a number of experimental studies (e.g., Koch, 2001; Meiran et al., 2000). However, studies which varied the CSI independently within-subjects have also found that switch costs are reduced with longer CSIs, suggesting that switch costs are, indeed, reduced by an active

---

[2] In this respect, passive task-set decay sits more naturally with an interference-based account of switch costs. From that perspective, passive task-set decay is one mechanism for the reduction of previous-trial interference.

preparation process (e.g., Koch, 2001, Logan & Bundesen, 2003, Logan & Schneider, 2006, Meiran, 1996, Meiran et al., 2000; see Monsell, 2003, and Kiesel et al., 2010, p. 855, for reviews).

In summary, empirical evidence suggest that passive task-set decay is one contributor to a decrease in switch costs. However, separately to this effect, there is also evidence that preparation for the forthcoming trial also reduces switch costs. Behavioural evidence further suggests that preparation is not switch-specific. Both preparation and sequence predictability facilitates RTs in both switch and repeat trials. Preparation effects of switch costs can be explained by assuming that preparation has more of an effect on switch trials because switches are more affected by interference.

2.3 Theoretical accounts of task switching 2: Interference models

Interference accounts of task switching propose that task performance is impaired by the prior performance of an alternative, conflicting task. For example, in the 'Task Set Inertia' hypothesis, Allport et al. (1994) suggest that performance of a task requires the activation of an appropriate task set, and the inhibition of competing task sets. Once a task has been performed, activation and inhibition of task sets (positive and negative priming) persists into the next trial. Repeat trials are thus facilitated by the carryover of activation and inhibition from the previous trial, while task switches are impaired. According to this account, interference is caused by residual activation or inhibition at the level of the task-set. In the light of evidence for some associative priming when specific stimulus items are exclusively associated with a particular task set (Allport & Wylie, 2000, experiment 5), in a modified version of the theory, task-sets may be directly primed by individual stimulus items with which they have become associated. According to this perspective, then, positive and negative priming of task sets is sufficient to account for switch costs. Slower RTs on switch trials are caused by prolonging of the same cognitive processes which take place on repeat trials.

A number of computational models of interference in task switching have been presented, demonstrating mechanistically how task switch phenomena such as the switch cost can be caused by interference between competing task representations. While these models are typically agnostic on the existence of reconfiguration processes, they present an argument that the switch cost is not, in itself, evidence of such processes. Three models will be discussed, the models of Altmann and Gray (2008), Gilbert and Shallice (2002), and Yeung and Monsell (2003). The model of Brown et al. (2007), while relevant, deals with a slightly separate set of task switching phenemona, and is a much more complex model incorporating a wider range of cognitive mechanisms.

2.3.1 The model of Altmann and Gray (2008)

?? In the model of Altmann and Gray (2008), based on the ACT-R cognitive architecture, task switching is conceptualised as a memory retrieval process. Task cues are encoded as an episodic memory trace, used as the basis for retrieving the relevant stimulus dimension (e.g., odd/even), identifying the correct category (odd) and making a response (left). As is standard in all ACT-R models, the time taken to access representations ('chunks') from memory is based on their activation. ACT-R's base-level-learning (BLL) equation, which governs chunk activation, ensures that the activation of a chunk is increased each time it is accessed, after which it decays according to a power law. Given this formulation, task repeats are faster than task switches because the relevant task representations have been accessed more recently — a form of priming.

As with all ACT-R models, the activation of chunks in declarative memory flows associatively. Thus, activation of a task dimension (parity) spreads to associated dimensions (odd/even), and then to the associated responses (left/right). Conceptually, then, the explanation for switch costs in the model is similar to that of the model of Gilbert and Shallice (2002), being due to residual activation/priming of recently performed tasks and the related representations, although in the Gilbert and Shallice model there is a greater emphasis on between-task interference, occurring due to lateral inhibition from residually active irrelevant representations.

2.3.2 The IAC model of Gilbert and Shallice (2002)

Gilbert and Shallice (2002) implemented the modified Allport and Wylie (2000) interference account, as a connectionist, interactive activation and competition (IAC) model (McClelland & Rumelhart, 1981). In an IAC model, inputs, outputs, and intermediate processing layers are represented as a network of simple processing units, each with a bounded, real-valued activation value. Units may be connected by excitatory or inhibitory connections. The strength of connections between units are governed by weights, which are either hand-set or set via a learning algorithm and training procedure. In this class of models, processing takes place as a series of iterations, in which the activation values of each unit are updated in parallel, based on the excitatory and inhibitory inputs to the unit and its previous activation value.

The Gilbert and Shallice (2002) task-switching model was based on an earlier IAC implementation of the Stroop task (Cohen et al., 1990). Common to both models, stimulus dimensions and possible responses are represented as single units for simplicity (i.e., both the 'ink colour' and 'word' input layers consist of units representing specific colours, such as red and green). The alternative tasks of word reading and colour naming are implemented as segregated processing pathways. Control of the current task is provided by a control 'task demand' layer. This consists of two units, analogous with task-sets, and reflects the currently active task. Task demand units bias processing by facilitating activation in the processing pathway of the corresponding task. In the Gilbert and Shallice (2002) model, this is implemented in terms of excitatory connections between task demand units and the output units of the corresponding task, and inhibitory connections to the output units of the competing task.

On a typical incongruent Stroop trial, different input units are activated for each processing pathway (e.g., the red colour unit and the green word unit) and activation propagated to the outputs of both pathways. Simultaneously, a single task-demand unit becomes active, biasing processing in favour of the current task, while suppressing processing in the competing task. Crucially, the weights between input and output units are greater for the word reading pathway than the colour naming pathway, assumed to reflect the greater strength of the word reading pathway.[3] Additionally, the colour-naming task demand unit receives greater top-down control activation than the word-reading unit, assumed to reflect the greater cognitive control required to carry out a non-dominant task. Activation is updated iteratively in cycles, with each cycle corresponding to a unit of time, simulating the temporal dynamics of cognitive processing. The model stops when the activation of an output unit is greater than that of its nearest competitor unit by more than a threshold value.

Processing in the model is interactive between bottom-up and top-down processes, and occurs in parallel. The model has no stage-like reconfiguration mechanisms. The effect of control processes is restricted to modulating the strength of top-down input into the task-demand units.

The behaviour of the model is entirely dependent on this style of interactive, parallel processing, and by the following two mechanisms of interference. First, the residual activation and inhibition of task sets from the previous trial, is implemented by carrying over a certain proportion of the final activation of task-demand

---

[3] While these weights were hand-set, Cohen et al. (1990) demonstrated that similar weights could be learned via the delta-rule algorithm and a training set consisting of a large number of word reading trials and a small number of colour naming trials.

units at the end of the previous trial. Second, item-specific negative priming is implemented by excitatory connections directly from stimulus inputs to task-demand units. These connections are initialised with a weight of zero. However, long-term associative priming is simulated by altering the weights after each trial using a Hebbian learning rule: if a task demand unit and a stimulus input unit are both active, the weight is strengthened, while if one is active and the other inactive, the weight is weakened.[4] The model can therefore be seen as implementing the interference-based account of task switching proposed by Allport and Wylie (2000).

Simulation of empirical task switching effects

The model was used to simulate a number of variations of the alternating runs paradigm, and succeeded in modeling a range of the empirical effects in the human data. To the extent this is successful, it validates interference-based accounts of switch costs, by demonstrating that principles of task interference are sufficient to produce the empirical effects. However, the reason why the model produces empirical effects are not always consistent with the theoretical explanation given by Allport and colleagues. Additionally, there are a number of empirical effects which the model does not simulate. This section evaluates the performance of the model on simulating the empirical data.

Cost of switching restricted to the switch trial The model was used to simulate the alternating runs paradigm, with a run length of 4 (Rogers & Monsell, 1995, experiment 6). Consistent with human participant behaviour, RTs were generally slower for colour naming trials than for word reading. The switch cost was also largely confined to the first trial in a run (i.e., elevated RTs were restricted to the switch trial, apart from a negligible increase on

---

[4] Unless otherwise stated, weights were zeroed before recalculating weight changes, hence only the effects of negative priming from the immediately preceding trial were modelled.

the first repeat trial). The restriction of switch costs to the switch trial on longer runs had previously been argued as evidence in favour of an explanation based on additional cognitive processes which must be completed on switch trials (e.g., Rogers & Monsell, 1995). In the model, however, this was due to inappropriate residual task-demand activation; the current task begins switch trials with negative residual activation, while the alternative task begins the trial with positive residual activation. Switch costs thus reflect the additional processing cycles required for the model to overcome these inappropriate starting values. Thus, the model demonstrates that the TSI hypothesis accounts for these basic findings of switch costs.

Switch cost asymmetry Switch costs also exhibited an asymmetry characteristic of human participants — greater for switches into the dominant task than the non-dominant task. However, there is a difference between previous interference-based theoretical explanations for this effect and the reason why it occurs in the model. In the model, greater switch costs for switches into the dominant task is due to weaker top-down control activation of the dominant, compared to the non-dominant task (i.e., weaker for word reading than colour-naming). In contrast, the TSI theory of Allport et al. (1994) suggests that switch cost asymmetries occur because performance of the non-dominant colour naming task requires greater suppression of the dominant word-reading task than vice versa. The switch cost asymmetry in the model is not rigid, however, but is sensitive to certain parameter manipulations. For example, when the top-down control strength for both tasks is the same, but the two tasks differ in stimulus input strength, the switch cost asymmetry is reversed — i.e., it is larger for switches to the weaker task.

Reverse Stroop interference When human participants perform task switching for Stroop-like stimuli, an interaction is typically observed for reaction times between switch condition (switch vs. non-switch) and item congruence. On switch trials, RTs for word reading are slower for incongruent than neutral trials, whereas on repeat trials, RTs are unaffected by target congruence. For colour naming, however, RTs are slower for incongruent trials for both switch and repeat conditions. This effect is also an intrinsic property of the model.[5] In the model, the connection weights between input and output units in the colour naming pathway are too weak to interfere much with the word reading pathway. However, on switch trials, residual activation in the colour-naming task demand unit boosts activation in this pathway sufficiently to produce interference. On the other hand, the word reading pathway is strong enough to interfere with colour naming, even in the absence of greater task-demand activation. Because of a non-linear activation function, residual activation in the word-reading task demand unit has less of an activation-boosting effect on an already active pathway. Thus, reverse Stroop interference is explained as an effect of relatively stronger processing for word reading than colour naming pathways, combined with a non-linear activation function which 'squashes' activation, such that a given amount of additional input produces a larger activation boost when a unit is in the middle of its activation range compared to an extreme (cf. Cohen et al., 1990).

An additional mechanism for reverse-Stroop effects was item-specific priming, discussed in detail below. As neutral stimuli cannot be repeated between word-reading and colour-naming tasks, any effect of associative priming of task-demand units would be absent. Indeed, this was demonstrated in an additional simulation. When item-specific priming effects were removed, reverse-Stroop effects persisted, but were reduced.

---

[5] However, it is dependent on the use of a non-linear activation function, and also specific parameters for word-reading and colour-naming input-output weights. Parameter studies conducted by Gilbert and Shallice (2002) provide confidence that the model is not overly dependent on specific parameters.

Item-specific switch costs Stroop stimuli result in slower RTs on word reading trials when they have previously appeared in colour naming trials, then when they have not appeared previously, and this effect is greater for switch trials. Allport and Wylie (2000) interpreted this in terms of their interference account, suggesting that a substantial component of interference was item-specific and caused by the long-term priming of task-sets by stimuli which they had been exclusively associated with. In the case of performing a conflicting task, this would imply negative priming, in which the competing task-set is activated while the current task-set is inhibited.

In one simulation, Gilbert and Shallice (2002) implemented direct connections between stimulus input units and task demand units. These connections were initialised with a weight of zero, but at the end of each trial a Hebbian learning rule was used to update weights, such that when task demand units and stimulus input units were both active, their weights were strengthened, but when one was active and the other inactive, their weights were weakened. The model simulated the interaction between whether the task switched or repeated, and whether the stimulus was primed or unprimed; with the RT being more affected by primed stimuli on switch, compared to repeat, trials. Primed stimuli directly activated the competing (colour-naming) task demand unit, which in turn had an excitatory effect on the competing, colour-naming pathway, and an inhibitory effect on the word reading pathway. This activation was either counteracted by residual task-demand activation from the previous trial (on repeat trials) or combined with it (on switch trials), thus producing a disproportionate effect on switch compared to repeat trials.

Effects of neutral trials The model simulated two reported empirical effects of responding to neutral rather than incongruent stimuli on switch costs. Firstly, there is a reduction in switch costs when switching to a task using neutral stimuli. Gilbert and Shallice (2002) explained this effect in terms of the previous explanation of reverse-Stroop interference. RTs for word reading switch trials are slower for incongruent than neutral stimuli, while there is a smaller difference between corresponding repeat trials.

Secondly, Allport and Wylie (2000, experiment 1) examined the effect of stimulus congruity in the colour naming task on the cost of a switch into the word reading task. Typically, much greater switch costs are observed on the word reading (switch) trial when incongruent stimuli are used on the preceding colour naming trials, than neutral stimuli. This is interpreted as key evidence for an interference-based account of the switch cost. The effect was also produced by the model. In the model, two mechanisms explain the reduction in switch costs. One is the absence of item-specific priming effects with neutral stimuli, as discussed above. The second results from the fact that responses to neutral trials tend to be generated quickly, and thus task demand activation is substantially less at the end of a neutral trial than an incongruent trial. As a result, less residual activation of the competing task is carried over onto the next trial.

Restart costs The model produces restart costs observed in the before-and-after paradigm, by assuming that residual task-demand unit activation dissipates during a break between blocks. Thus, the absence of residual (facilitatory, for repeat trials) activation from previous trials is sufficient to account for restart costs.

Rebound effects In the before-and-after paradigm, Allport and Wylie (2000) reported larger restart costs in participants who had previously performed a conflicting colour-naming task than controls who had performed an unrelated task. Allport and Wylie (2000) interpreted these ‘rebound’ effects as demonstrating a persistent, long-lasting form of interference. As this interference appeared to be latent, producing effects only on the first trial of a block, and not subsequent repeat trials, it cannot be explained in terms of item-specific costs. While Gilbert and Shallice (2002) explain restart costs in terms of the absence of residual task-demand activation, this does not explain why restart costs should interact with latent interference, to produce rebound effects. Thus, rebound effects are not explained by the model, and the latent characteristics of interference appear resistant to explanation in terms of residual task demand activation.

Preparation effects/Residual switch costs A key strand of empirical evidence is the effects of preparation on switch costs. According to reconfiguration accounts, a preparation interval should allow participants to complete task reconfiguration prior to stimulus onset, and thus should reduce switch costs. While a reduction in switch costs is a robust finding, a number of studies have found a residual component of the switch cost that is not eliminated even by very long preparation intervals.

In the model, task preparation is simulated by allowing the task-demand units to become active for a number of cycles before the activation of input units. In simulations, switch costs were indeed, reduced by such preparation; switch costs were entirely eliminated after 150 cycles, corresponding to 1200 ms based on regressing model cycles on empirical data. Additionally, preparation also speeded repeat trials, although to a lesser extent than switch trials. Thus, while the model simulates the effect of preparation to an extent, it does not produce residual switch costs as are sometimes reported in the experimental literature.

In summary, the IAC task-switching model of Gilbert and Shallice (2002), instantiating the interference account of switch costs, has parsimoniously accounted for a wide variety of empirical findings, not all of which could be intuitively predicted from an informal statement of the theory. Many of these empirical phenomena, such as the restriction of switch costs to switch trials, have been taken as evidence for the operation of cognitive control processes, yet the model suggests that such processes are not necessary to account for many of the effects seen in the empirical data. A number of phenomena, such as reverse-Stroop interference, are intrinsic to the activation and competition-based parallel processing implemented by the model.

Despite its success in accounting for a wide range of empirical phenomena, the model has a number of limitations. Firstly, some robust empirical effects, including residual switch costs, are not well simulated by the model without additional assumptions or mechanisms. Thus, this instantiation of the interference account does not provide a good explanation for why participants are unable to fully prepare for a task switch given a long period of time. Secondly, the characterisation of interference (as residual task-set activation), while successfully accounting for a wide range of empirical phenomena, does not explain why interference appears to be so persisting, or explain how interference can be latent, in order to cause rebound effects. Thirdly, the model has not been extended to a range of more recently discovered task switching phenomena, such as the effects of sequence predictability or cue-switching phenomena.

Finally, in theoretical terms the model implements an interference-based account of switch cost phenomena. Residual interference from previous tasks is modelled as residually active task demand units. However, some have argued that the carryover of activation for top-level units, but not lower level input or output units, could be seen as arbitrary (Grange & Houghton, 2014). It is true that without simulating residual effects at an input or output level (i.e., priming), as a model of the human cognitive system, it must be regarded as incomplete. In addition, the model's mechanism for effecting a switch of task, top-down-control units as a source of activation to task demand units, arguably does implement a form of reconfiguration. This process of reconfiguration could be seen as merely being explained away by invoking higher level cognitive processes that are not addressed by the model.

As Gilbert and Shallice (2002) argued, however, the model was published in response to an ongoing debate at the time, regarding whether task switching was a reconfiguration-based process, or due to residual task activation. More deeply, this debate reflected deeper assumptions about whether cognitive processing was stage-based, and could therefore be structurally decomposed through additive-factors logic (i.e., Sternberg, 1969, see Rubinstein et al., 2001, as an example of this approach) or whether parallel, interactive processing, of the type implemented by artificial neural networks, poses challenges for this methodology — a debate which is ongoing (e.g., Stafford & Gurney, 2011; Sternberg, 2013). Gilbert and Shallice (2002) argued not that reconfiguration models of task switching were wrong, but that the extant data held to support reconfiguration models were consistent with interference between competing tasks. Rather than viewing top-down control inputs as an irreducible homunculus, Gilbert and Shallice (2002) advocated a research strategy of trying to understand the action and structure of higher-level processes by examining their output — that is, the top-down control input into this model — and seeing the task-switching model as a form of filter. For example, they describe a number of empirical dual-tasking studies where participants perform task switching concurrently with another task designed to load a specific higher-level cognitive resource, such as working memory, and modelling performance in terms of a reduced top-down control input. In this context, the model should not be regarded as a complete ‘reference-model’ of a complete cognitive system, but as a relatively simple tool for factoring the outputs of higher-level cognitive processes into behaviour which can be observed empirically.

2.3.3 The IAC model of Brown et al. (2007)

The model of Brown et al. (2007) is conceptually similar, although substantially more complex, than that of Gilbert and Shallice (2002). It incorporates two additional mechanisms for managing cognitive conflict within the model. Firstly, an incongruency detector signals conflict between co-active incompatible responses, and in response sends additional excitation to the currently active task unit. Secondly, a change detector responds to trial-by-trial changes, in either the task units or the response units, by removing a 'tonic arousal signal' — an effect which slows all responses and lasts for a number of trials. Through slowing of responses, this latter mechanism also leads to increased accuracy. Thus, the model reacts to between-task interference (sequential conflict) by slowing all responses, while within-trial conflict (between simultaneously active responses) is resolved by reinforcing the activation of the currently active task unit.

Importantly, this model integrates conflict-monitoring mechanisms within a task switching model, and is comparable to the proposal of Goschke (2000) that the conflict generated between incongruent responses is a trigger for inhibition at the task level. While this proposal shares commonalities with certain accounts of backward inhibition (e.g., Koch, Gade, Schuch, & Philipp, 2010), the authors do not explore the model's performance in the n-2 repetition paradigm, and instead focus on sequential stimulus congruency and response repetition effects.

2.3.4 The mathematical model of Yeung and Monsell (2003)

Yeung and Monsell (2003) also implemented a model of task switching to explain and predict empirical effects in asymmetric task switching (word reading and colour naming). The authors explain the asymmetry of switch costs by proposing a mathematical model that describes the contribution of various component processes to switch costs. The model is implemented as a set of equations that model the duration of various aspects of cognitive processing in terms of a response time distribution. These equations are presented in section 2.8. The activation of each task set is calculated from the sum of task strength, priming (i.e., both interference on switch trials and facilitation on repeat trials), endogenous control, and a noise term. As in the model of Gilbert and Shallice (2002), the shape of the activation function ensures that the same incremental input resulting from priming effects has a greater effect on weaker tasks than stronger tasks. Task strength is greater for word reading than colour naming, while priming reflects a positive, transient increase in activation for the previously performed task (e.g., word reading for a colour switch trial, or for a word repeat trial). It is assumed that exerting endogenous control is effortful, and that participants minimise the level of endogenous control required on a given task, in order to trade-off processing speed against accuracy. Implementing this assumption, a training phase is used to calibrate the strength of endogenous control input according to a staircase procedure, in which the level of control is dynamically adjusted in response to correct or error responses on preceding trials. In their simulations, Yeung and Monsell (2003) calibrated the control input separately for each of four conditions: switch and repeat trials for word reading and colour naming tasks. Thus, their model assumes that a different level of endogenous control is applied for the same task, depending on whether it is a switch or a repeat trial.

In the model, competing task-sets generate responses asynchronously but in parallel, with the rate given by the ratio of the task-set's activation to total task-set activation. These responses then converge onto a single resolution stage, in which the time to produce a response is modeled by scaling r, a variable randomly drawn from an ex-Gaussian distribution, by a function of the difference in the two generation rates, which varies between r (when interference is minimised, such as for neutral trials or for competing tasks with non-overlapping response sets), or when one task is much stronger than the other, such that the difference in generation times is large (greater than or equal to r), and $1.5r$ for incongruent trials, when the difference between generation times approaches zero. The ex-Gaussian distribution provides the characteristic distribution of human response times.

Using their model, Yeung and Monsell (2003) simulated a number of empirical effects.

Asymmetric Switch Costs Due to the non-linear activation function used by the model, a priming input that is constant for both word-reading and colour-naming tasks has a greater effect on the resulting activation of the weaker task (i.e., colour naming). Therefore interference occurring on switch trials has a greater effect on the colour naming task.

Reversal of asymmetric switch costs In an empirical study, Yeung and Monsell (2003) were able to reverse the direction of the switch cost asymmetry by de-synchronising aspects of stimulus presentation. Specifically, the names of colours were presented over a coloured rectangle, where the colour names and the colour of the rectangle were incongruent. However, the presentation of the word could be either simultaneous with the stimulus onset (i.e., as in the experiments of Allport et al., 1994), or delayed by 160 or 320 ms. While the synchronous condition replicated previous findings by Allport and colleagues (i.e., greater switch costs when switching into the easier task), the delayed conditions reversed the direction of the asymmetry, instead producing larger costs for switches into the harder task.

Yeung and Monsell (2003) simulated the delayed-stimulus onset manipulation of asymmetric switching by re-setting the endogenous control inputs to the same value (0.15) for both colour reading and word naming tasks, on the assumption that less control was required for colour naming given the reduced interference. Due to the lowered control strength, colour switch trials became harder to perform when the low control strength input to the colour task was combined with the constant priming input for the word task.

While the role of the model is descriptive, rather than explanatory, one weakness is that the distribution of human response times is not modeled, merely reproduced by incorporating a random component drawn from a distribution selected for its similarity to the desired empirical distribution.

2.3.5 Discussion

A number of differences exist between the models in their aims and theoretical interpretation. One distinction may be drawn between mathematical models of the computational level, which aim to describe the mathematical function being computed, and process models, which aim to replicate within the model itself something of the nature of the cognitive processing which is occurring.

In these terms, while Gilbert and Shallice (2002) present a process model, Yeung and Monsell (2003) is situated mid-way between a process model and a mathematical model of the overall computation. The algorithmic content of the latter model proposes that task-set activations are calculated in parallel, as a non-linear function of their inputs. This function 'squashes' input activation, such that the variable inputs (e.g., priming and endogenous control) change overall activation more, when the level of other inputs is low. As in Gilbert and Shallice (2002), this is why task priming has a greater effect on colour naming than on word reading. Additionally, the model proposes that competing responses are prepared in parallel, according to equations describing the time-course of response generation. Ultimately, competing responses converge on a response-resolution stage, where they interact to either facilitate or interfere with processing. In these terms, the model of Yeung and Monsell (2003) describes a cognitive process, conceived in linear, sequential stages in contrast to the parallel, interactive processing in Gilbert and Shallice (2002). However, the equations governing response production and response resolution are not clearly theoretically grounded. In particular, the final calculation of the time taken to resolve responses is deterministically calculated as a difference between the activation value of task-sets, and scaled according to a distribution specifically chosen to reflect the RT distributions characteristic of human participants performing RT tasks. While such scaling is not methodologically problematic, it limits the explanatory power of the model to explain why RTs on switching tasks might have such a distribution.

To the extent that the model of Yeung and Monsell (2003) makes theoretical claims about cognitive processes, there are points of both agreement and disagreement with the model of Gilbert and Shallice (2002). Task-set priming in the Yeung and Monsell (2003) model is conceived in terms of positive priming. In Gilbert and Shallice (2002), the priming of task-sets may be both positive or negative. While this issue has been the subject of theoretical debate, it remains to be seen whether it produces substantial behavioural differences in the models, or is merely an implementational detail. In particular, longer-lasting inhibitory effects (e.g., mechanisms responsible for n-2 inhibition effects) are not explicitly implemented by either model, and are argued to not be responsible for asymmetric switch cost effects.

Nevertheless, the overall theoretical interpretation of the Yeung and Monsell (2003) model is similar to that of Gilbert and Shallice (2002). The models agree that processing on a given trial is a function of constant-valued top-down endogenous control inputs, previous-trial task priming, and the asymmetric strengths of the task-relevant processing pathways. However, while the top-down (endogenous) control strengths do not change between switch and repeat trials in the model of Gilbert and Shallice (2002), in the model of Yeung and Monsell (2003) control strengths are calibrated separately for switch and repeat conditions, based on a training phase. The latter model therefore assumes that practice on the task switching paradigm is a key part of setting cognitive parameters appropriate to performing the task, and that human participants are able to set cognitive parameters in a near-optimal way. This assumption is also central to the simulation of the reversed switch cost asymmetry due to delayed onset of the word dimension of the stimulus. In this simulation, endogenous control strengths for each condition are re-set on the basis of reduced interference in the colour naming condition.

In summary, both models simulate empirical effects of switching between asymmetric tasks, characterising switch costs as an interaction between interference from previous tasks, and endogenous control, although the model of Gilbert and Shallice (2002) also incorporates a bottom-up source of interference — long term, item-specific priming of task sets. Both models give an explanation of switch costs in asymmetric switching, without recourse to time-consuming control processes. However, both models also invoke an element of cognitive reconfiguration — in the model of Gilbert and Shallice (2002) this is implemented by the switching of top-down control inputs according to the desired task, which is itself held to be the output of a higher-level executive system. The model of Yeung and Monsell (2003) involves calibrating control parameters during a training phase, and assumes that the modelled agent is capable of setting appropriate levels of control for switch and repeat instances of each task in a near-optimal way. The mechanistic detail of this parameter setting, however, is left unspecified.

2.4 'Hybrid models' and the reconciliation of interference and reconfiguration accounts

Hybrid accounts (e.g., Monsell, 2003) attempt to reconcile reconfiguration and interference-based accounts of task switching, by suggesting that task switches reflect both task carryover and additional cognitive processes. However, they analyse switch costs in terms of discrete components: interference and a component attributable to control processes. According to this logic, the cost attributable to interference can be subtracted from the overall switch cost.

Importantly, with specific respect to switch costs, even hybrid accounts are not compatible with interference-based accounts such as that of Allport and Wylie (2000) or Gilbert and Shallice (2002). While these accounts do not deny the existence of control processes, they suggest that their effect on the chronometrics of task performance is indirect, rather than direct. The computational model of Gilbert and Shallice (2002) gives a good account of basic empirical findings on switch costs, in terms of principles of interactive activation and competition between task-processing pathways and task demand units. The temporal dynamics of processing in the model cast doubt on attempts to characterise control processes underlying task switching in terms of discrete, linear stages with additive time costs. From a connectionist perspective, control processes are likely to interact dynamically with processing in the network, thus their effects on reaction times are likely to be nonlinear, and interact with a range of other processing factors. Therefore, attempting to separate out the effects of interference through a simple subtractive logic is unlikely to be successful.

A more fruitful approach to unpicking the effects of control processes may lie in further use of computational models. On the assumption that the model of Gilbert and Shallice (2002) gives a good account of task-set interference, it may be used as a baseline for comparison. Rather than seeking to subtract away the effects of interference from empirical data, the model may be used as a basis for modeling putative control processes. An advantage of a model-based approach is that it provides a means of deriving specific quantitative predictions from theoretical hypotheses. If we accept the assumption that control processes and interference both have effects on quantitative measures such as response time, one approach is to systematically compare the predictions of a purely interference-based account (i.e., the Gilbert & Shallice, 2002, model 'out of the box') and the model augmented with putative control processes, and see which best fits the empirical data.

2.5 Interference and preparation in task switching: Interim summary

Much of the existing literature on switching has focused on the concept of switch costs, either in terms of RT or error rate. However, much of the debate has concerned the question of what the switch cost really reflects, in cognitive terms. The switch cost is often thought to reflect some combination of between-task interference and preparation (or reconfiguration) for a new task. However, switch costs themselves may be problematic for a number of reasons. Firstly, depending on the precise experimental paradigm, switch costs may be difficult to isolate and are potentially conflated with subsidiary task switching phenomena, such as restart costs (e.g., Altmann, 2007a). Secondly, the costs attributable to switching interference may be long-lasting and re-evoked in a non-switching context, such as rebound RT effects (e.g, Allport & Wylie, 2000, experiments 3-5). Finally, and again, depending on experimental paradigm, there are compelling arguments that so-called switch-costs do not reflect the cognitive components of switching task at all, but merely reflect the use of a compound cue-retrieval strategy (e.g., Logan & Schneider, 2010).

Kiesel et al. (2010), in her review, arranges the conceptual landscape of task switching around the two key landmarks of preparation (reconfiguration) and interference. Of course, it is difficult to clearly separate the two concepts. Vandierendonck et al. (2010) suggests an ‘interplay of reconfiguration and interference control’, casting cognitive control processes in terms of proactive (i.e., preparatory) and reactive control, with both potentially involving control of interference, such as through the setting of appropriate attentional bias.

This somewhat complicated relationship between interference and putative control processes makes their study through behavioural approaches alone complex, if not problematic. The use of computational models has proven a fruitful approach to specifying the effects of interference and its relationship with cognitive control processes in more quantifiable terms.

Two computational models of task switching explicitly deal with cross-task interference in a comparable task switching paradigm, that of Gilbert and Shallice (2002) and Yeung and Monsell (2003). The remainder of this chapter reports on the replication of these two models, and in the case of the former, further simulations conducted using the replicated model.

2.6 Re-implementation of the Gilbert and Shallice (2002) model

The task-switching model of Gilbert and Shallice (2002) was successfully re-implemented, using the published model description and default parameter settings. Programming of the model did not re-use code from the original authors, but proceeded from scratch, on the basis of the published description, equations, and parameter settings. This provided some confidence in the model's reproducibility.

As discussed above, the model implements the theory of Allport and Wylie (2000), whereby (n-1) switch costs are the result of residual, interfering activation from the competing task and associative (stimulus-task) priming effects. The authors simulated a range of behavioural effects typically observed when human participants switch between the two tasks of word reading and colour naming of Stroop stimuli (i.e., tasks of asymmetric difficulty).

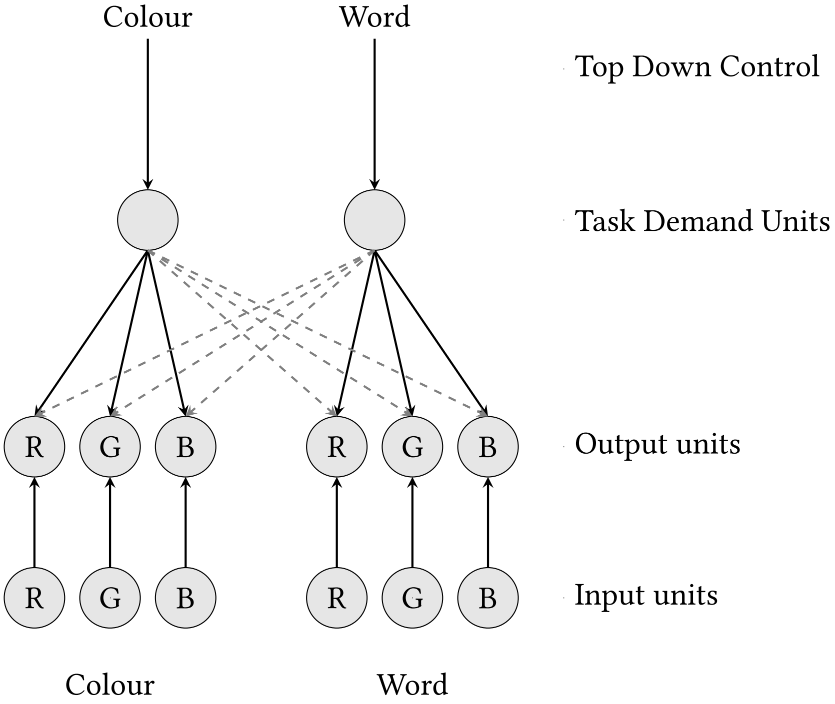
The re-implemented model successfully replicated a number of the most critical simulations of behavioural task switching effects, including simulations of switch cost asymmetries (Rogers & Monsell, 1995), and item-specific priming (Allport & Wylie, 2000, Experiment 5). This section details the results of this replication, and compares them to the claims for the model of Gilbert and Shallice (2002).

2.6.1 Model description

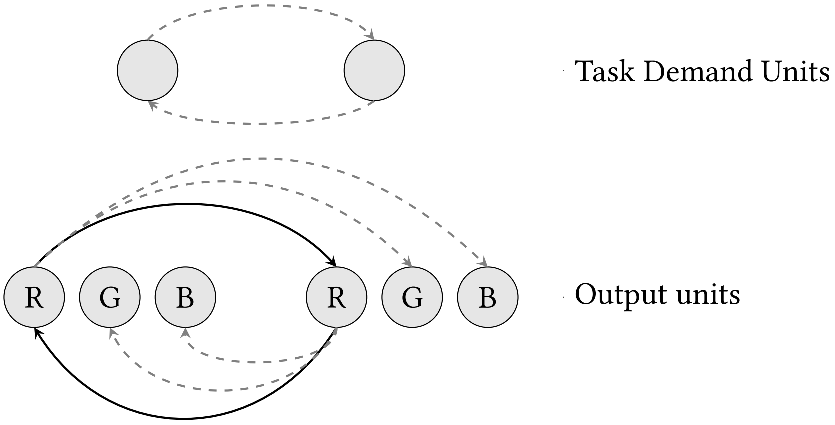
The architecture of the model used by Gilbert and Shallice (2002) is illustrated in figure 2.1. The main architecture of the model showing connections between different levels is illustrated in figure 2.1a. For simplicity, only between-module connections are shown. The model consists of two processing pathways, for the colour and the word task respectively. For each pathway, individual units at input level correspond with three available stimulus dimensions. For the colour task, these represent the screen colours red, green and blue, while for the word task, they represent the words red, green and blue. At output level, three units in each pathway each correspond to one of the three possible responses (red, green or blue).

On a simulated congruent trial, two corresponding input level units would be activated, e.g., both ‘R’ units. For an incongruent trial, non-corresponding units would be activated, e.g., one ‘R’ and one ‘G’ unit. Neutral trials are represented by the activation of an input unit for one pathway only. On each trial, an additional top-down control input is activated, indicating the desired task on the basis of a task cue. This sends input to one of the two task demand units, which may be thought of as a cognitive representation of the currently active task-set. Task demand units have a general effect on the output level. The word task-demand unit excites all word output units and inhibits all colour output units, while the colour task-demand unit has the opposite effect. Task demand units also retain a proportion (by default, 20%) of their activation at the end of the previous trial, this represents a certain degree of task carryover, which both facilitates repeat trials, and causes interference on switch trials. Although 20% of final activation only represents a small value, in the early stages of a trial it is enough to bias processing in the output units enough to substantially facilitate, or interfere with, the competition process by which an output unit is selected.

Lateral connections between units at the same level are also present, as illus-



(a) Overall architecture of the model



(b) Detail of lateral inhibition between units at output and task demand level. For simplicity, only the connections from the first unit in each layer are shown. For output units, lateral connections to corresponding units are excitatory, and inhibitory to non-corresponding units. Task demand lateral connections are mutually inhibitory.

Figure 2.1: The Task Switching model of Gilbert and Shallice (2002). Excitatory connections are shown with black solid arrows, inhibitory connections with grey dashed arrows. The model consists of two pathways, corresponding to the word task and the colour task. Individual units at input and output level represent three stimulus dimensions and the three available responses, respectively.

trated in figure 2.1b. For output units, corresponding units are mutually excitatory, while non-corresponding units are inhibitory. Task demand units are mutually inhibitory.

Additionally, input units are directly connected to task-demand units. While these connections initially have zero weight, following each trial a Hebbian learning mechanism modifies the connection weight from the previously active input and task demand units, to simulate a direct priming effect of task sets by the previously active stimulus dimension. These connection weights are reset after every trial, so only a single bottom-up connection is active at any one time.

In the output layer of the model, corresponding to a response selection level of processing, three available responses are each represented twice, for the colour and word pathway, however it is assumed that the corresponding colour and word output units both correspond to the same physical response, presumably implemented by a response execution stage, which is not present in the model.

Activation in the model is therefore fully interactive, and propagates both bottom up, from input units directly to output units — and top down, from the top-down control inputs to the task demand units, which bias processing in the output units, either exciting or inhibiting activation on the basis of task.

To take a concrete example, on an incongruent trial, activation propagates from the input units (e.g., colour unit 'R' and word unit 'G') up to the output level, directly activating colour output unit 'R' and word output unit 'G'. However, due to lateral connections, the colour unit 'R' sends an excitatory input to the word unit 'R', and an inhibitory input to the word unit 'G', causing the activation of the output units to compete. Eventually, one unit 'wins' this competition, its activation becoming sufficient to suppress activation in the competing output units, and a response is generated. The time taken for this to occur, in model cycles, is the simulated reaction time.

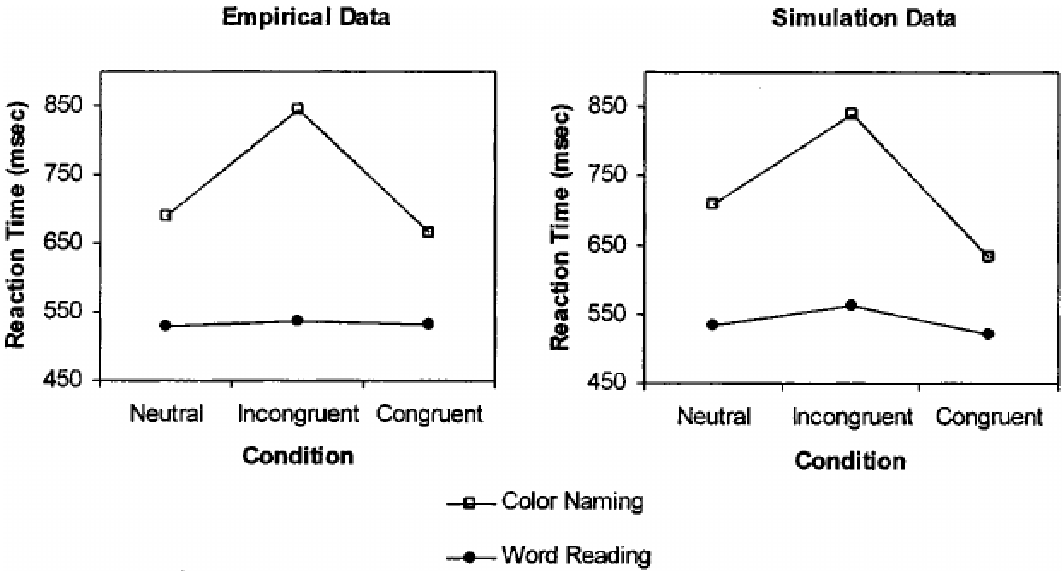
2.6.2 Simulations of empirical phenomena

Stroop interference and facilitation

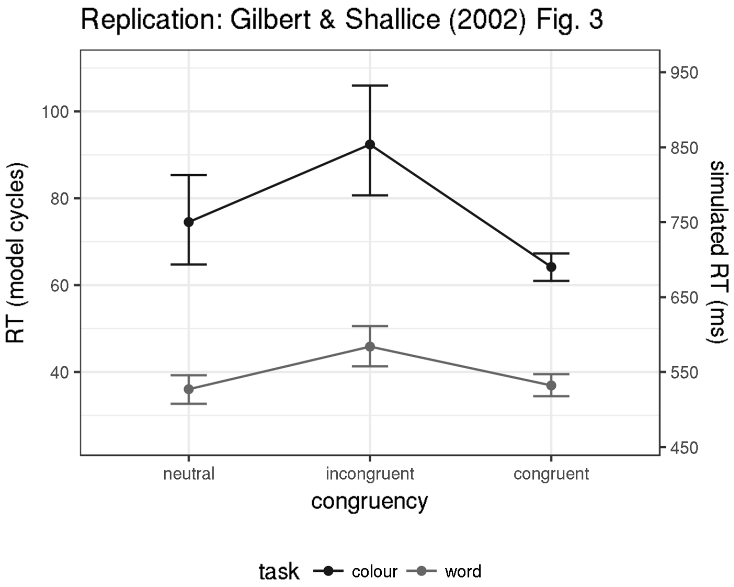
Gilbert and Shallice (2002) demonstrated that their model accurately simulated the facilitation and interference effects empirically observed when human participants perform pure blocks of colour naming and word reading in response to Stroop stimuli. While this property of the model is unsurprising, given its similarity with previous Stroop models of Cohen et al. (1990) and Cohen and Huston (1994), it was important to demonstrate, especially given the subsequent modifications to the model. This simulation was illustrated in figure 3 of the original paper (p. 311), which is reproduced below as figure 2.2a. The simulated RT in model cycles was also converted to milliseconds using the regression equation:

$$RT_{ms} = 5.8RT_{cycles} + 318$$

Error bars or an indication of model variability were not provided in the original simulation, although the original authors performed an analysis on '50,000 simulated reaction times'. (It is unclear whether this refers to 50,000 trials per condition, or a total of 50,000 trials, but given 50,000 does not divide exactly between six conditions, this interpretation seems unlikely.) In the replication error bars represent bootstrapped 95% confidence intervals. These suggest that the variability of model reaction time, especially in the colour naming condition where there is more interference, is considerable, although it does not cast any doubt on the validity of the original simulation. Specifically, the neutral, incongruent and congruent conditions for colour naming all have means which lie outside the 95% confidence interval of the other conditions, suggesting statistically significant differences according to a conventional $alpha$ of .05. For colour naming, incongruent trials are significantly slower than congruent or neutral trials, which do not differ from each other.



(a) Figure 3 from Gilbert and Shallice (2002), showing performance of the Gilbert and Shallice (2002) model in pure blocks of the standard Stroop task, with empirical data based on Dunbar and MacLeod (1984, Experiment 1B). Reproduced with permission.



(b) The corresponding simulation in the replicated model. The secondary axis represents response time (milliseconds) as simulated by Gilbert and Shallice (2002) using the equation $RT_{ms} = 5.8RT_{cycles} + 318$. Error bars represent bootstrapped 95% confidence intervals.

Figure 2.2: Replication of Gilbert and Shallice (2002) figure 3: Stroop interference and facilitation in pure blocks.

The size of the confidence intervals is largely, although not entirely, dependent on the number of trials in each condition, and may be reduced somewhat by merely running more trials, therefore the previous observation of ‘significant’ differences between conditions is somewhat arbitrary. However, it should be noted that even with only 100 trials per block, ‘significant’ differences were observed between incongruent and congruent/neutral conditions for word reading, which were not observed in the empirical data (Dunbar & MacLeod, 1984).

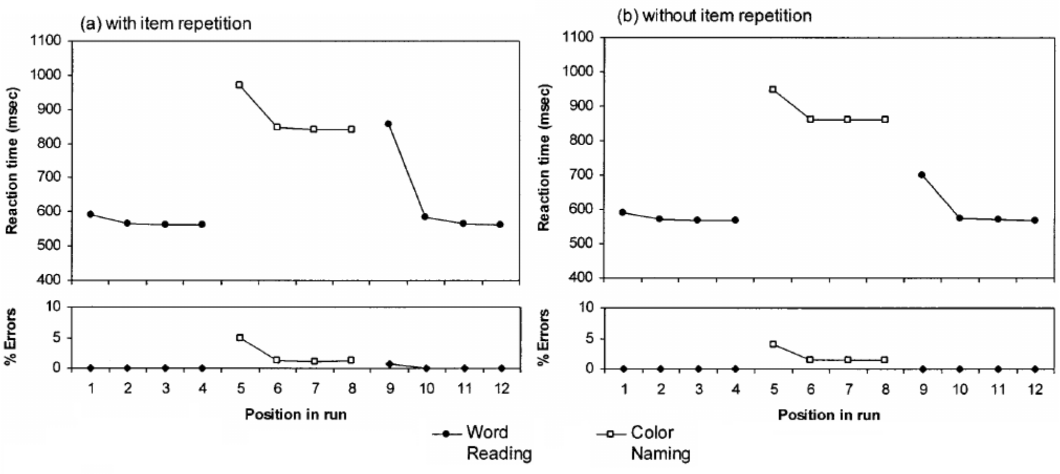
As observed by Gilbert and Shallice (2002), errors in the model occurred extremely infrequently in pure blocks. When the model was run for 1000 trials in each block, errors only occurred in the colour naming incongruent condition, at around 0.6% (with a 95% confidence interval of [0% - 1.5%]).

Task switching in mixed blocks

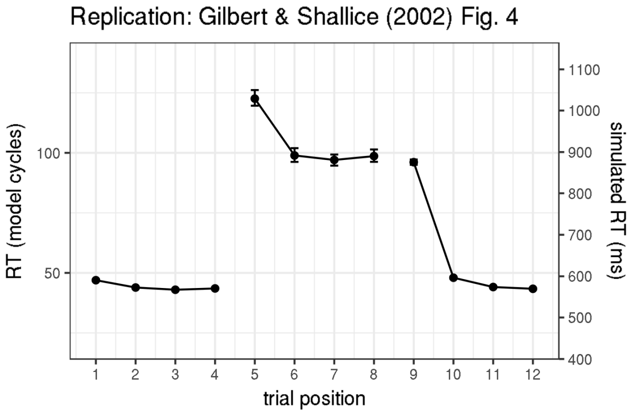
Next, Gilbert and Shallice (2002) adapted the model to blocked task switching of predictable length runs. As an analog of the paradigm used by Rogers and Monsell (1995, experiment 6), the model simulated three blocks, each consisting of four trials of the same task, with switches between tasks occurring on the first trial of each block. This simulation demonstrated three key effects of task switching. Firstly, Rogers and Monsell (1995) found that the cost of a switch was confined to the first trial following a switch of task, an effect previously described as the task serial position effect (section 2.1.2). In early literature on task switching, this effect was used to argue for a reconfiguration model of task switching, where the switch cost reflected the time taken to ‘reconfigure’ the cognitive system appropriate to each task. If the switch cost reflected interference, it was argued, there should be a graded reduction in switch cost for a number of trials following the switch. Subsequent research, including the computational evidence of Gilbert and Shallice (2002), implementing an interference-based account of switching, demonstrated that the confining of the switch cost to switch-trials only is indeed consistent with interference based accounts. Thus, replication of this finding is critical for the model and interference-based theories. Secondly, in paradigms involving switching between two tasks of asymmetric difficulty, two characteristic behavioural effects are typically observed: asymmetric switch costs, and so-called reverse-Stroop interference (see section 2.1.2 for a complete review). In the former effect, switches from the weaker to the stronger task are associated with a greater cost, relative to baseline repeat performance of that task, than switches from the stronger to the weaker task, despite performance on repeat trials of the weaker task being slower and more error prone. Thirdly, RTs are typically slower on the first trial of a run, even when that trial does not involve a task switch, an effect known as 'restart costs' (section 2.1.3).

These effects were demonstrated in a simulation, reproduced here as figure 2.3a. A small elevation in RT is observed on the very first trial, compared to the subsequent three trials; this reflects the effect of restart costs. The switch cost observed is much smaller for the switch to the more difficult task, the difference in RT and errors between trials 5 and 6, than that for the easier task, between trials 9 and 10, even though this asymmetry is reduced when item repetition does not occur.

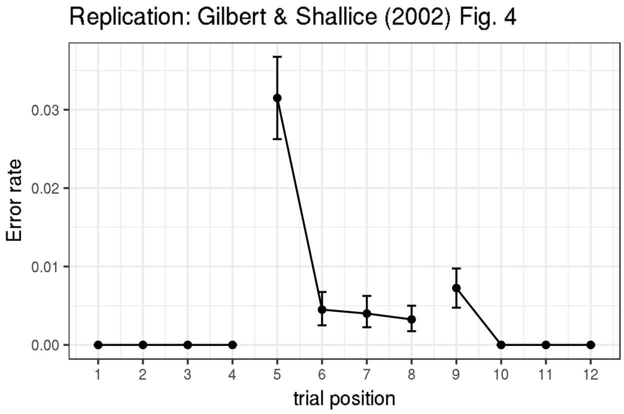
This simulation was replicated, with results presented in a similar format in 2.3b for RTs, and 2.3c for error rates. A comparison of the simulation of Gilbert and Shallice (2002) and the corresponding replication shows a very close correspondence. The key features of the simulation are all replicated. Specifically, switch costs are largely confined to the switch trial, although as found by Gilbert and Shallice (2002), RTs continue to decline a very small amount between the second and third trials. The switch costs are asymmetric: that is, the cost of switching from the harder to the easier task (trials 9-10) is much greater than the cost of switching from the easier to harder task (trials 5-6). However, this is only the case for reaction times: for error rates, the asymmetry is reversed, with a much larger



(a) Figure 4 from Gilbert and Shallice (2002) demonstrating switch costs between alternating-task blocks of 4 trials. Reproduced with permission.



(b) Corresponding simulated RTs in the replicated model. The complete run of 12 trials was run 10,000 times. Error bars represent 95% confidence intervals.



(c) Error rates in the replicated model. Error bars represent 95% confidence intervals.

Figure 2.3: Replication of Gilbert and Shallice (2002) figure 4: Switch costs in alternating single-task blocks of trials.

cost for switching from the easier to the harder task, than vice versa. This gives the appearance of a speed/accuracy trade-off, although it is purely due to the structure of the task, as no mechanism for prioritising either speed or accuracy is present in the model. While this feature is the same for both the original simulation and the present replication, Gilbert and Shallice (2002) did not comment on this feature of their results. Finally, a very small restart cost is observed for RTs only.

Item-specific switch costs

Empirical evidence has shown that for the stronger task of word reading, a certain component of the switch cost is related to the repetition of specific stimulus items that have previously been associated with the alternative task. Word reading switch trials are slower when the stimulus item has previously been associated with colour naming (see section 2.1.3).

Gilbert and Shallice (2002) simulated this effect by employing a form of Hebbian learning, applied to the weights of connections between input units and task demand units, strengthening connections between units which were both active following each trial. Thus, input units which had previously evoked a particular task-set on the previous trial had a small positive connection with that task-set on the subsequent trial.

Specific stimulus sequences were used to model the effects of repeat or unrepeated stimulus items, occurring either on switch or repeat trials. Let us represent cued stimuli with a digit and a letter, such as $W1$ or $C2$, where the letter refers to the task (word reading or colour naming) and the digit refers to a specific stimulus item.

$$C1, C2, W3, W4 \tag{2.1}$$

$$C1, C2, W2, W3 \tag{2.2}$$

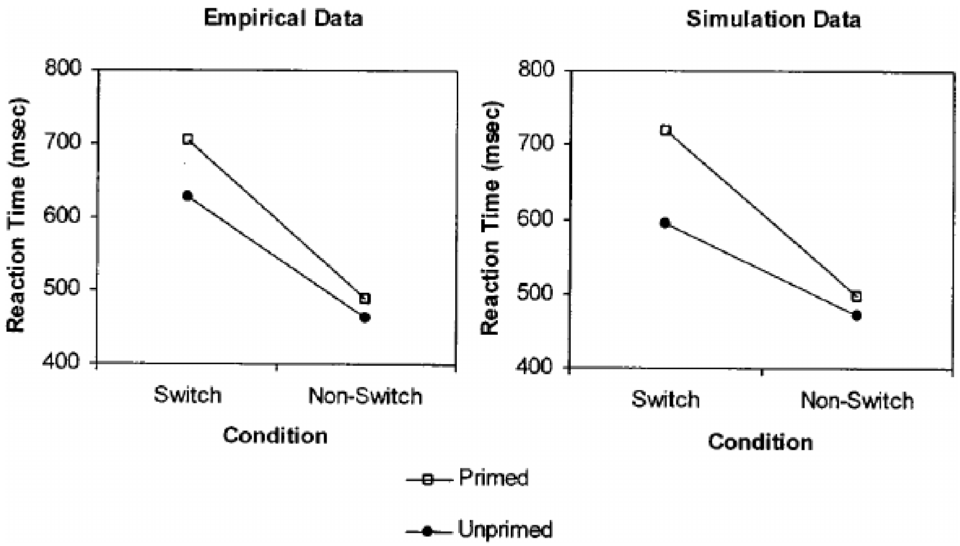
$$C1, C2, W3, W2 \tag{2.3}$$

Sequence (2.1) was used to obtain baseline RTs for non-repeating stimuli. As the model only implemented three available stimuli, for the baseline measure, Hebbian learning was only allowed to persist for one trial. After re-initialisation, Sequence (2.2) was run; comparison between (2.1) and (2.2) contrasted primed and unprimed switch trials, occurring on position 3 of sequences (2.1) and (2.2). Finally, comparison of position 4 in sequences (2.2) and (2.3) gave a contrast between primed and unprimed repeat trials. In their simulation, Gilbert and Shallice (2002) ran each sequence 50,000 times.

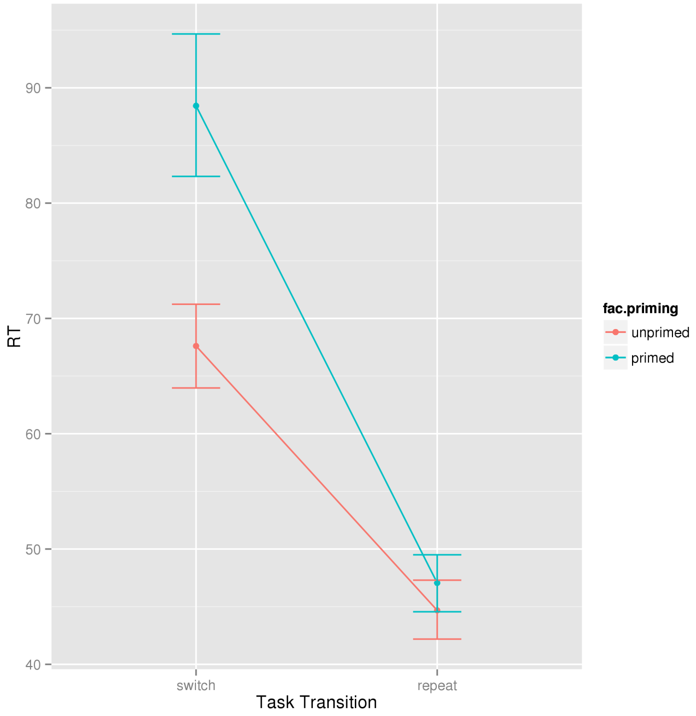
The simulation of item-specific priming effects was replicated. The interaction between priming and task transition (see Figure 2.4b), replicates the simulation of item-specific switch costs (Gilbert & Shallice, 2002, p. 319), which in turn simulated an empirical effect (Allport & Wylie, 2000, experiment 5). Thus, item-specific priming effects were effectively simulated by direct connections from stimulus input units to task demand units, using a Hebbian learning mechanism.

Preparation effects

Gilbert and Shallice (2002) also used their model to address the empirically observed effects of a preparation interval. Typically, participants are presented with a task cue for a specific interval, before stimuli are presented, and instructed to use the interval to prepare for the forthcoming task (see section 2.3.2 for a review of this paradigm). Typically, preparation intervals have an advantageous effect on the size of switch costs, although in most studies it is not eliminated altogether, leaving a reduced, residual switch cost (Rogers & Monsell, 1995, experiment 3; Meiran, 1996; although, see De Jong, 2000). The effect of preparation, and the residual switch cost, was originally a central plank of reconfiguration accounts of the switch cost. According to this view, while most of a task-switch reconfiguration could occur in response to a task cue, a final stage could only take place in the presence of a specific stimulus, accounting for the residual switch cost.



(a) Figure 7 from Gilbert and Shallice (2002) demonstrating the effects of item-specific priming on switch costs. Empirical data are based on Allport and Wylie (2000). Reproduced with permission.



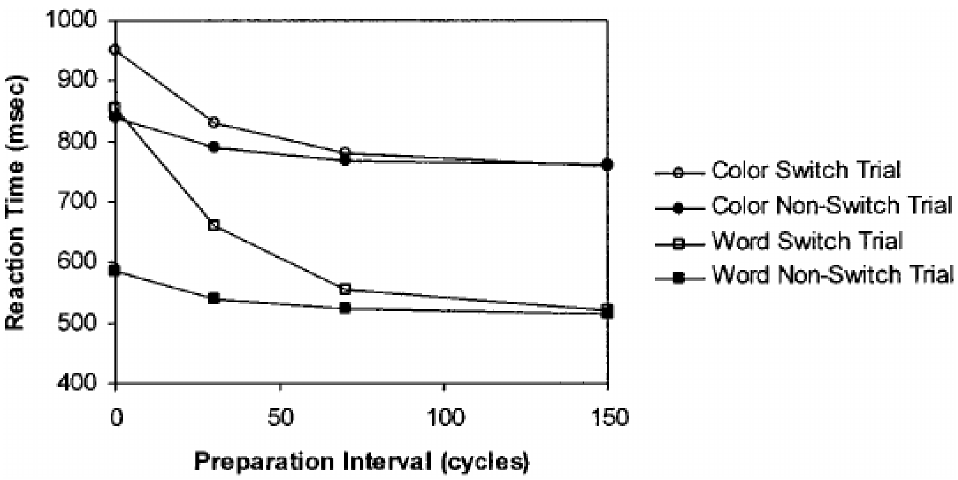
(b) The effects of item repetition for word reading switch and repeat trials in the reimplementation of the Gilbert and Shallice (2002) model.

Figure 2.4: Replication of Gilbert and Shallice (2002) figure 7: Item-specific priming

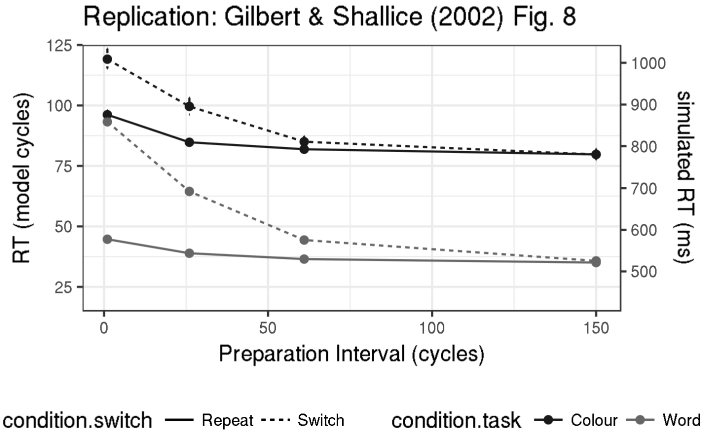
Gilbert and Shallice (2002) argued that the effect of a preparation interval is entirely consistent with an interference-based account of switching. In their model, top-down control units are activated and allowed to cycle for a certain number of cycles, prior to the activation of the input units that represent stimuli. It could therefore be argued that the effect of top-down control and the processing that takes place in the task-demand layer is, indeed, implementing a form of reconfiguration. However, a crucial difference is that this is not a stage-based reconfiguration process, except to the extent that is imposed by an interval in which only preparation is allowed to take place. Instead, any reconfiguration is interactive with activation in the rest of the model. However, the model of Gilbert and Shallice (2002) did not simulate residual switch costs. In their model, with the longest preparation interval (150 cycles) the switch cost was entirely eliminated for both the simulated colour naming and word reading tasks.

In the replicated model, the simulation of preparation interval was accomplished as in the original paper, by allowing the task demand units to cycle based on residual activation from previous trials, inputs from top-down control units, and lateral inhibition between themselves, only. Here, four different preparation intervals were used: 0, 25, 60, and 149 cycles. Gilbert and Shallice (2002) do not specify the exact task sequence used, but for simplicity the same sequence as for the original demonstration of switch costs, in section 2.6.2, was used. The replicated results, together with the original figure, are presented below in figure 2.5.

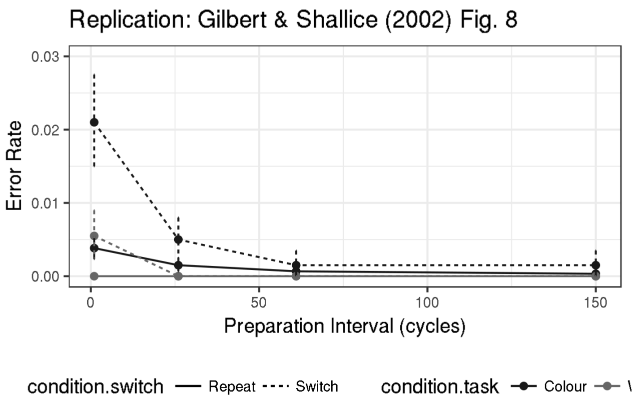
Gilbert and Shallice (2002) present only a simulation of reaction times (figure 2.5a). The replication of the RT analysis corresponds very closely with the original (figure 2.5b), showing a reduction and eventual abolition of switch costs for the longest preparation intervals. The analysis of error rates is also presented (figure 2.5c). Interestingly, error rates do show a small residual switch cost for the colour task only.



(a) Figure 8 from Gilbert and Shallice (2002) demonstrating the effects of preparation on switch costs. Reproduced with permission.



(b) Replication of the effects of variable preparation interval on simulated reaction times



(c) Effects of preparation on error rates

Figure 2.5: Replication of Gilbert and Shallice (2002) figure 8: The effects of preparation. As in the original article, simulated RTs were estimated using the same regression equation used in the replication of figure 3 (reproduced here as figure 2.2a).

2.7 Further simulations using the Gilbert and Shallice (2002) model

The model of Gilbert and Shallice (2002) effectively implements task-set interference within a two-task switching paradigm. As the authors demonstrated with their account of asymmetric switch costs, interference effects have often proven counter-intuitive. This section reports two novel simulations in which the model was used as a vehicle to explore what may prove to be similarly counter-intuitive effects of interference in a wider range of task-switching scenarios.

2.7.1 Simulation 1: Individual Differences in Executive Functions

This simulation addresses reported correlations in individual differences studies, on what Miyake et al. (2000) argued are basic executive tasks. 'Response inhibition', putatively required in the Stroop task, was operationalised as RT differences between Stroop colour naming and word reading. 'Task switching' tasks included the list procedure Jersild (1927), measuring performance as the difference in total time to complete two sets of tasks, when tasks alternated, compared to occurring in single-task blocks; and the alternating runs procedure of Rogers and Monsell (1995). The correlation coefficients reported by Miyake et al. (2000) for the executive tasks pertinent to this simulation are presented in table 2.1.

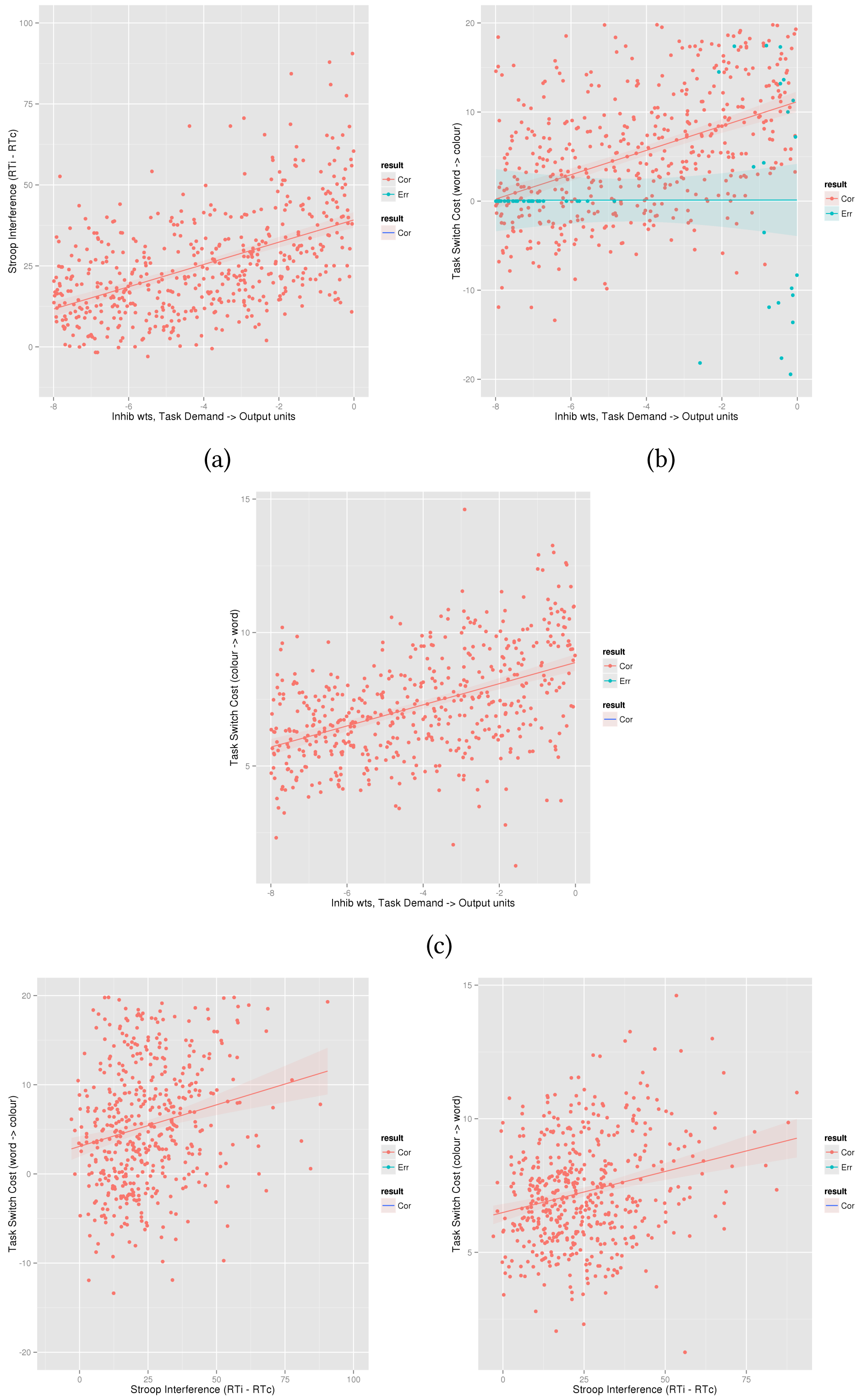
Table 2.1: Pearson correlations between dependent measures on executive tasks Miyake et al. (2000)

	Stroop	Switch cost
Switch cost	.09	
Global switch cost	.07	0.32*

* significant at 0.05 level

Miyake and colleagues reasoned that performance correlations on tasks putatively requiring different executive functions (the 'unity' of executive function) may be due to a shared executive resource (Friedman et al., 2008; Miyake & Friedman, 2012), while individual executive functions also had a function-specific component (the 'diversity' of executive function). However, it remains to be explored whether both unity and diversity can be accounted for by variation of a single factor, or parameter, in a cognitive model which can perform both tasks (see Cooper & Davelaar, 2010, for an example of this approach). In the model of Gilbert and Shallice (2002), one candidate parameter was the weight of connections between task-demand units and task-specific processing pathways, controlling attentional biasing in the model. Simulation 1 was exploratory, examining the effects of varying the parameter on both 'response inhibition' and 'task switching' performance. The simulation varied this parameter randomly between 0 and -8 (compared to a default value of -2.5).

Scatter plots of simulated individual differences in model performance are presented in figure 2.6. Panels 2.6a, 2.6b and 2.6c show the relationship between weights of inhibitory connections between task demand and output units (i.e., the strength of inhibitory attentional biasing) and dependent measures on executive tasks: Stroop interference, a measure of 'inhibition' (2.6a), and two measures of task switch cost (2.6b, 2.6c). Panels 2.6d and 2.6e show the resulting relationships between dependent 'executive' measures. Table 2.2 presents the resulting Pearson correlation coefficients, which were all highly significant for a population of 500 models. The magnitudinal difference in correlations suggests that while both Stroop and switch cost performance are modulated by a single parameter (reflected by a strong correlation between these dependent measures and inhibitory bias weight) the intercorrelation between the two dependent measures is somewhat weaker. This finding implies that strong correlations between tasks putatively requiring the same executive function, and weak correlations between tasks requiring different executive functions, might in principle be explained in terms of a single parameter or factor.



(a)

(b)

(c)

(d)

(e)

Figure 2.6: Scatter plots showing simulated individual differences in model performance. (2.6a) Inhibitory bias weight vs. ‘Response inhibition’; (2.6b) Inhibitory bias weight vs. ‘Task Switching’ (colour naming); (2.6c) Inhibitory bias weight vs. ‘Task Switching’ (word reading); (2.6d) ‘Response inhibition’ vs. ‘Task Switching’ (colour naming); and (2.6e) ‘Response inhibition’ vs. ‘Task Switching’ (word reading). Correct trials (red) and errors (blue) have been plotted separately, although on some trials (e.g., Stroop word reading) no errors were made.

Table 2.2: Stimulated individual differences in executive functions. Pearson correlation coefficients for performance on Stroop (‘response inhibition’) and Switch costs

	Stroop	Switch cost (colour)	Switch cost (word)
Inhibitory bias weight	0.51***	0.51***	0.47***
Stroop		0.23***	0.24***
Switch cost (colour)			0.23***

*** significant at 0.001 level

While these findings are interesting, caution is required in the comparison of simulated correlations with published individual differences data. The specific ‘task switching’ tasks used by Miyake et al. (2000) did not involve switching between Stroop stimuli, as in Allport and Wylie (2000) and the model, and this renders direct comparison problematic. Future systematic studies of individual differences must either simulate performance on a wider range of switching paradigms using a more developed model of task switching, or obtain empirical individual differences data including switching between asymmetric tasks.

The proposal that variation of attentional bias parameters might produce a specific pattern of correlations in different dependent measures was largely exploratory. However, the simulation of individual differences in various executive tasks remains an important methodology for future simulations, which may adopt a more theoretically motivated approach. Specifically, the modelling of theoretical proposals such as task inhibition, goal encoding and conflict monitoring, provides scope for simulating individual differences in putative executive functions.

2.7.2 Simulation 2: ‘Rebound effects’ in task switching

In switching between two tasks of asymmetric difficulty (colour-naming and word-reading in response to Stroop stimuli), Allport and Wylie (2000, experiments 3, 4) found effects of interference which persisted beyond the subsequent trial. Specifically, following a block of colour-naming trials, task-set interference was re-evoked on the first trial after a break, with no switch of task (a restart trial). This effect was

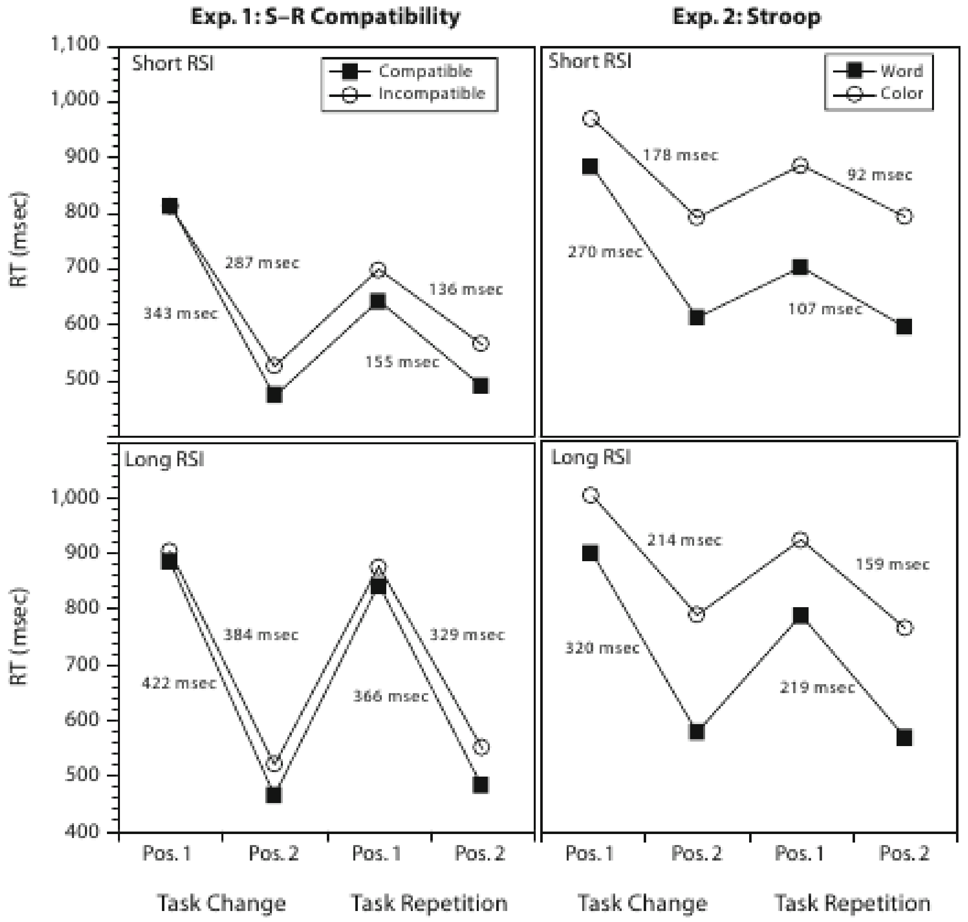


Figure 2.7: Asymmetric restart costs: Results from experiments 1 and 2 of Bryck and Mayr (2008). Restart costs, although lower than switch costs, exhibited the same pattern of asymmetry, i.e., greater in switching from the non-dominant to the dominant task. Reproduced with permission.

larger for word reading than colour naming trials, even though RTs were generally faster for word reading, mirroring the pattern of switch cost for asymmetric tasks (replicated in Bryck & Mayr, 2008, see figure 2.7).

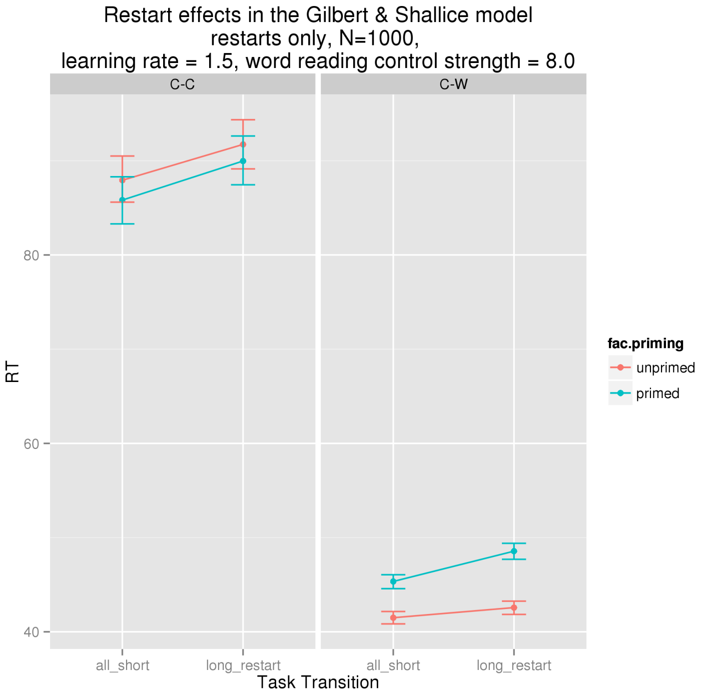
Recent authors have argued that longer-term interference effects such as this cannot readily be accounted for in terms of residual task activation, and instead favoured explanations in terms of variable memory encoding of item-specific associations (Bryck & Mayr, 2008) or task difficulty effects (Schneider & Anderson, 2010). However, although Gilbert and Shallice (2002) did not extend their model to account for longer-term interference, certain effects may be produced in interactive activation models as the dynamic interaction between different levels of processing, such as item-specific priming effects (i.e., affecting stimulus encoding stage) and task preparation (affecting task reconfiguration). Importantly, this distinguishes interactive models (Allport & Wylie, 2000; Gilbert & Shallice, 2002) from discrete-stage models of task processing (Rubinstein et al., 2001; Mayr & Kliegl, 2003) in which sequential stages, such as stimulus encoding and response selection, do not interact. Hence, if this effect can be parsimoniously accounted for by existing mechanisms in an interactive model, it would present problems for discrete stage-based accounts of task switching, which must invoke additional mechanisms.

Simulation 2 takes as point of departure the simulation of item-specific priming effects using the replicated model of Gilbert and Shallice (2002), presented previously (section 2.6.2). In this model, item-specific priming was implemented with weights acquired through Hebbian learning, which persist onto the subsequent trial. Thus, we already have available an implementation of item-specific priming with some empirical validation. To test the hypothesis that rebound effects (asymmetric restart costs) are caused by an interaction between item-specific priming and residual task-demand activation, the simulation varied two factors: the inter-trial interval (RSI) and whether the stimulus was primed or unprimed (i.e., primed stimuli directly activated the colour-naming task-demand unit). It was predicted that longer RSIs would produce restart costs (i.e., longer RTs on restart trials) and that this effect would be greater for word reading than colour naming. For this simulation, two parameters were adjusted from their default settings in Gilbert and Shallice (2002). Firstly, the learning rate was increased from 1.0 to 1.5 to increase the size of the priming effect in the model. (In the model, prior task performance is based on only a single trial, compared to multiple blocks for human participants (Allport & Wylie, 2000), hence the exact learning rate used in the model to simulate participant data is arbitrary.) Secondly, two values of word reading top-down control strength were tested, the default of 8.0, and a lower value of 4.5. In the model, more automatised tasks are reflected by greater bottom-up activation, and thus a lower level of control. Although these strengths were hand-set by Gilbert and Shallice (2002), in a related model Cohen et al. (1990) demonstrated that asymmetric control strengths for two tasks could be learned through different numbers of training runs.

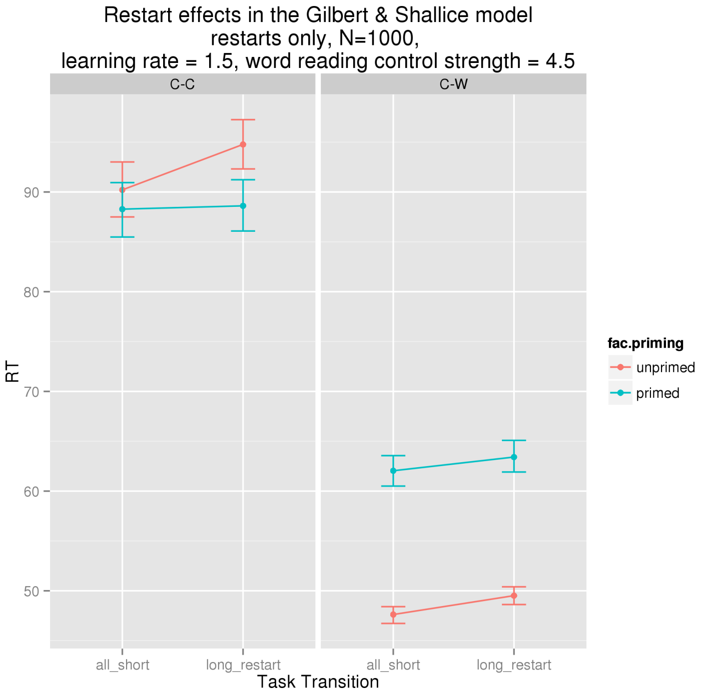
These results are presented in figures 2.8a and 2.8b. Analysis of the results suggested that intertrial intervals could interact with priming to cause additional interference (an item-specific component of restart costs), but that this only occurred for one task. Importantly, which pathway was affected by item-specific restart costs depended on the top-down control strength of the word reading pathway. This parameter was adjusted as a means of controlling interference between colour naming and word reading: lower top-down control strength should produce greater interference on word reading trials. When the word reading control strength was high (i.e., a default value of 8.0) the item-specific restart cost (an interaction between priming and intertrial interval) was highly significant for word reading, $F(1, 3956) = 7.74, p < .01$, but not colour naming trials, $F(1, 3866) = 0.02$, figure 2.8a. When word reading control strength was low (4.5) the item-specific restart cost was not significant for word-reading, but approached significance for colour-naming $F(1, 3852) = 2.46, p = .12$, figure 2.8b.

These findings do not support the hypothesis that asymmetric restart costs reflect an interaction between item-specific priming and the decay of residual activation during an intertrial interval — the predicted 3-way interaction between intertrial interval, priming and task did not approach significance. However, the data suggests that the model may produce a task-specific interaction between priming and interval, modulated by top-down control strength.

In the model of Gilbert and Shallice (2002), top-down control strength was fixed. However, in the original model of Cohen et al. (1990), top-down control strengths were learned, such that the top down control in the two pathways was



(a) Item-specific restart costs with word reading control strength of 8.0



(b) Item-specific restart costs with word reading control strength of 4.5

Figure 2.8: Item-specific restart costs.

variable. Although not predicted, this observation fits with findings from Bryck and Mayr (2008) that restart cost asymmetries are affected by the relative strengths of the two tasks. This observation may warrant further systematic study. In addition to theoretical interest in applying the task-switching model to empirical phenomena for which no computational explanation exists, this simulation touches on a wider theoretical issue of whether discrete stage-based models or interactive models of cognitive processing are compatible, or make diverging behavioural predictions.

2.8 The mathematical task-switching model of Yeung and Monsell (2003)

In this section, we report an attempt to replicate the model of Yeung and Monsell (2003), described above in section 2.3.4.

2.8.1 Model description

The model consists as a series of equations describing the inputs, and the resulting activation, of two task-sets. For each task, *i*:

$$input_i = strength_i + priming_i + control_i + noise \tag{2.4}$$

Strength reflects relative differences in the of tasks as a result of different levels of practice. In the reported simulations, this was set to 0.1 for colour reading and 0.5 for word naming. *Priming* is modelled as an additional input to the most recently activated task-set, taking a constant value of 0.3 for both tasks. *Control* reflects endogenous control input. It is assumed that this input is effortful, and is minimised in order to regulate performance to an arbitrary error threshold. In the model, the values for control inputs were determined during a training phase to keep the error rates below 5% in each condition.[6] The final values used were reported as 0.20 for word switch, 0.15 for word repeat, 0.97 for colour switch, and 0.38 for colour repeat trials. Finally, a *noise* term is drawn from a Gaussian distribution with a mean of 0 and standard deviation of 0.1.

The activation of the task-set is given as a function of its input, by the following activation equation:

$$activation_i = 1 - e^{(-c \times input_i)} \tag{2.5}$$

where c is a constant equal to 1.5 in the reported simulations. Importantly, due to the shape of the activation function, a constant priming input has a greater effect on the task-set's activation when the input is relatively small, such as for the colour-naming task, than when the input is larger, as for the word reading task.

Once the activation of the two task-sets is determined by the above equations, response times are calculated through the following equations describing a response selection process. Firstly, it is assumed that the duration of a response selection process is affected by the activation of each task-set, and that similar activation levels produce greater response times due to interference.

$$generation\ rate_i = \frac{activation_i}{\Sigma activation} \tag{2.6}$$

$$generation\ time_i = \frac{threshold}{generation\ rate_i} \tag{2.7}$$

where *threshold* is a time constant with a value of 100. Secondly, it is assumed that response codes are generated for each task-set individually, and that resolving these competing responses requires a duration given by the following equation:

$$resolution\ time = r + f[r - (generation\ time_j - generation\ time_i)] \tag{2.8}$$

---

[6] In the training phase, control inputs for each task were initialized at 0.15, and 'several trials' were performed. Control inputs were incremented by 0.05 following erroneous responses and 0.001 following correct responses.

where r is a term drawn on each trial from an ex-Gaussian distribution, generated by convolving a normal distribution (mean of 140, standard deviation of 10) and an exponential distribution (mean of 40). This equation implements the assumption that the magnitude of interference or facilitation depends on the proximity in time with which competing responses are generated — competing responses generated within a short space of time will produce a large amount of interference. The coefficient of this time difference, f, takes a value of 0 for neutral stimuli and 0.5 for incongruent stimuli. Thus for neutral stimuli response interference does not contribute to the resolution time.

Finally, the output of the model gives the simulated RT in milliseconds according to the following equation:

$$RT = P + generation\ time + resolution\ time + R \tag{2.9}$$

where P and R represent the time taken by perceptual and response processes respectively, and take a combined value of 150 in the reported simulations. Yeung and Monsell (2003) report conducting 50 simulations of 600 trials each.

2.8.2 Model replication

Unfortunately, the model could not be fully implemented based on the description in Yeung and Monsell (2003). Some additional assumptions were required in order for the model to produce the described results. These are detailed below:

Negative generation rates as a result of equation 2.6 Because a noise term is added to the input for each task pathway (equation 1), with mean 0 and SD 0.1 (parameters specified in the paper) it was possible, on occasion, for the net input to be negative, especially with low levels of control (e.g., the 160 ms SOA condition). A small, negative activation produces a small, negatively signed generation rate. In equation 2.7, this produces a large, negatively signed generation time. The end result is a large, negative RT for that particular task. Additionally, a zero activation value is problematic, as it produces a generation rate of 0 (equation 2.6) which leads to division by zero in calculating generation time (equation 2.7). In order to prevent these results from occurring, in the present replication a lower bound of 0.0001 was used for activation values.

Implementation of delayed onset of stimulus dimensions The paper is vague on what was varied in the model to simulate delayed onset of stimulus dimensions. The authors write:

> To implement the second critical assumption — that top down control is effortful and, hence, minimised where possible — control inputs are determined by iteration during a training phase to be at the minimum level required to keep error rates low in each condition (5%). In this training phase, control input to each task set was initialized at a minimum value (0.15), and then the performance of the model was assessed over several trials for each task and trial type (nonswitch and switch). The control input for the relevant task set and trial type was incremented by 0.05 units each time the model made an error and was reduced slightly (by 0.001 units) for each correct response. In this way, control input was set to the minimum level required to produced (*sic*) generally accurate performance, capturing our assumption that levels of control input reflect a trade-off between accuracy and effort. Performance quickly stabilized to a level at which responding was accurate on most trials, and the output of the model was assessed once this stable level of performance was reached.
>
> (Yeung & Monsell, 2003, p. 465)

> Our [empirical] data show that introducing a delay reduced Stroop interference without creating any reverse Stroop effect. The model naturally reproduces this basic finding, as interference depends on the relative rate at which response tendencies are generated. Of interest is the impact of these changes on simulated switch costs, and this is shown in Table 5 and Figure 6 (simulation of 160-ms delay condition). Evidently, the primary effect of delaying word onset is to increase the cost of switching to the color naming task.
>
> [. . .]
>
> The model behaves in this way because the reduced interference from the delayed word stimulus means that a reduced level of control input is required to perform the color naming task. [. . .] The model therefore explains the results of Experiment 1 through the effect of delaying word onset on the control inputs required to perform color naming: As control input is decreased, task priming effects come to have a large effect on color naming performance. . .
>
> (Yeung & Monsell, 2003, pp. 466-467)

The above quotations clearly imply that the independent variable in the model was the introduction of a delay into some aspect of model processing, with the changes in control levels dependent on this. However, the paper does not specify implementation details. The tabulated settings (Table 5 in the quotation above, reproduced here as Table 2.3) is inconsistent with this interpretation, however, suggesting that a fixed value of 0.15 is used as the control strength for all trial types, consistent with the presentation of control parameters used for the no delay condition, presented in their Table 4. In the re-implementation of the model, the fixed value of 0.15 was input as the

Tabulated RTs

Trial type and task set	Strength	Control	Priming	Total input	Activation
Word Switch					
Color	0.10	0.00	0.30	0.40	0.45
Word	0.50	0.15	0.00	0.70 [*sic*]	0.60
Word nonswitch					
Color	0.10	0.00	0.00	0.10	0.14
Word	0.50	0.15	0.30	0.95	0.76
Color Switch					
Color	0.10	0.15	0.00	0.25	0.31
Word	0.50	0.00	0.30	0.80	0.70
Color nonswitch					
Color	0.10	0.15	0.30	0.55	0.56
Word	0.50	0.00	0.00	0.50	0.53

Table 2.3: Simulation of the 160ms delay condition, Yeung and Monsell (2003) Table 5.

control parameter for all trials. However, while the pattern of RTs produced by the model did fit the description and the qualitative pattern in the text (and presented in their Figure 7), the exact values were not accurate, and the model produced extremely high error rates (above 95%) in the colour-switch condition. This was taken to suggest that the values printed in Table 2.3 were only the starting values, before the training phase in the model. Thus, the training regime was also replicated.

Implementation of the training regime The training regime was also vaguely described. No details were provided of how many trials the model was trained for, only that 'Performance quickly stabilized to a level at which responding was accurate on most trials'. However, in the replication, control weights did not stabilize, and with the given parameter settings, continued to increase even after a relatively large number of trials. Training simulations for up to 10,000 trials revealed that control strength for colour naming on switch trials continued to increase to over 30, while control strength for word reading on nonswitch conditions became negative.

Use of trial-and-improvement found that 50 trials with the stated training parameters produced approximately the right control strengths for the control (synchronous) condition. This number of trials also produced control strength settings which stabilised in the delayed onset condition: Word switch = 0.10; Word Nonswitch = 0.10; Colour Switch = 0.56; Colour nonswitch = 0.15. While these control settings control the error rate of trials, the model produces the same direction of switch cost asymmetry observed in the control condition.

Implementation of delayed stimulus-onset in the model In the model, sequential stages of cognitive processing are modelled using equations. As a result, the state of the model at each stage of processing is represented as some psychological quantity, such as (task) activation, generation rates, response times, etc. Time, in the model, is calculated proportionately (in ms) for each stage of processing. This may be regarded as a strength of the model, and is in contrast to models (e.g., IAC models) where the state of a model at a particular stage of processing is somewhat opaque in comparison. Given this, then, the implementation of the different experimental conditions is unintuitive. For example, the effect of delaying the onset of the word aspects of the stimulus by 160ms (in human participants) is modelled by lowering the control strength required to perform the colour naming task. Model control strengths are then re-established through a training phase, with the result that control values of 0.15 are used for both word reading and colour naming. While this produces the desired reversed-asymmetry RT switch cost (i.e., greater switch cost for colour naming than word reading) the model also produces extremely high error rates (over 90%) in one condition (colour naming switches).

Consider, however, that, if the two stimulus dimensions are initially processed in parallel, as the model assumes, one might equally choose to model the effects of SOA by adding the SOA (160ms) to the generation time for the irrelevant stimulus, based on the assumption that both dimensions are processed as before, but with different start times. However, under this manipulation, the model does not produce the main empirical effect (inverted switch cost asymmetries).[7] Therefore, the production of reversed-asymmetry switch costs in the model is dependent on specific settings for control input.

In summary, replication of the model faced certain issues which required additional assumptions to duplicate the results as published. In particular, the key model prediction of reversed-asymmetry switch costs was dependent on somewhat arbitrary setting of control values, and not the adjustment of some other parameter (SOA) that might reasonably be expected to correspond to the behavioural condition in question. In the light of the limited replicability of the model, no additional simulations were attempted, and subsequent simulation work focused on developing the model of Gilbert and Shallice (2002).

2.9 Summary

In the past two decades since its emergence as a domain of cognitive psychology, the task switching paradigm has matured and diversified, and now encompasses a wide range of empirical phenomena, cognitive theorising, and methodological approaches (Koch & Brass, 2013). The switch cost, as a putative indicator of task switching ability (Miyake et al., 2000) has proven particularly problematic. On a theoretical level, there has been extensive debate over whether the switch cost primarily reflects the reconfiguration of the cognitive system appropriate to a new task, or to interference from competing task representations. It has been argued

---

[7] Of note, when both manipulations are combined (i.e., equal low control settings for both tasks and introducing SOA into the resolution time equation (equation 2.8), the reversal of the switch cost asymmetry is more pronounced. However, it seems it is the lower control settings that are both necessary and sufficient for the effect.

that of these two alternatives, switch costs may more plausibly be understood in terms of interference models, rather than reconfiguration-based models. However, certain alternative behavioural paradigms have raised the question of whether the switch cost really reflects a switch of task at all, with compound-cue retrieval models (e.g., Logan & Schneider, 2010) providing a good fit to behavioural data, and parsimoniously suggesting that switch costs merely reflect a switch of cue, rather than a task-switch per se. In the light of this development, the prospect of the task-switching paradigms discussed here offering insight into executive processes has become increasingly distant. Meanwhile, a number of theoretical accounts offer accounts of interference-based processes and the cognitive challenges facing any executive reconfiguration process.

The following chapter discusses an alternative task switching paradigm to the two-task switching ones presented in this chapter. It will be argued that in contrast to the phenomena discussed in this chapter, the n-2 repetition paradigm provides good evidence for at least one cognitive task switching mechanism, the inhibition of task-sets.

Chapter 3

Task-Set Inhibition: The n-2 Repetition Paradigm

Classic effects of task switching, such as switch costs, have proven to be of limited empirical value in revealing the action of cognitive control processes. Instead, the n-2 repetition cost, observed in switching paradigms with three tasks, may reflect a cognitive inhibition process. This chapter reviews the evidence for the n-2 repetition cost as an index for the inhibition of task-sets ('Backward Inhibition'). In particular, n-2 repetition costs appear resistant to manipulations of preparation intervals or cueing, that reduce the observed effects of switch costs in two-task switching, although the magnitude of n-2 repetition costs may be reduced (but not eliminated) by controlling for episodic priming. N-2 repetition costs also appear to be driven by conflict between competing task representations, although there is ambiguity in the literature as to whether they depend on conflict at the level of cue-task retrieval or of response processes, or both. This chapter also reviews mechanistic and computational accounts of task-set inhibition.

3.1 Introduction

A number of the theoretical accounts of task-switching phenomena propose a central role for inhibitory mechanisms in task switching. For example, in their task-set interference (TSI) account of switch costs, Allport and colleagues proposed that interference was composed of "continued priming of the previous task (competitor priming), and suppression (negative priming) of the currently intended task" (Allport & Wylie, 1999). According to this view, active suppression or inhibition of task-sets is a central mechanism in the control of task-sets, and the persistence of such inhibition over time is a contributor to the empirical phenomenon of switch costs (Allport et al., 1994). Similarly, in the computational model of Gilbert and Shallice (2002), lateral inhibition between units at the same level ensures that only a single unit is active at any time. Current task-sets are therefore mutually inhibitory, and inhibition at the end of each trial persists into the next trial, producing switch costs. Lateral inhibition between task-set units is similarly implemented in the model of Brown et al. (2007). Numerous theories proposing similar roles for inhibition in task switching have been proposed (e.g., Arbuthnott, 2005 Goschke, 2000, Mayr & Keele, 2000, Schuch & Koch, 2003, see Koch et al., 2010 for a review). For example, Goschke (2000) proposed that inhibition directed at task sets is dynamically adjusted based on the level of response conflict evoked by incongruent stimuli.

It has been argued, however, that many cognitive phenomena which have been explained in terms of inhibition may more parsimoniously be explained without invoking specific inhibitory mechanisms (MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). For example, task-switching accounts exist which explain switch costs in terms of persisting activation (Altmann & Gray, 2008). In general, without direct empirical evidence, theories based on the competing activation of task-sets could equally propose that control operates via either excitatory or inhibitory mechanisms (e.g., Yeung & Monsell, 2003). It is now generally accepted that evidence from procedures in which participants switch between two tasks fails to conclusively demonstrate the existence of inhibitory mechanisms in task switching Koch et al. (2010). Instead, the most compelling evidence is provided by a different measure, n-2 repetition costs.

In the n-2 repetition procedure (e.g., Mayr & Keele, 2000), participants switch between three tasks. In most versions of this procedure, tasks do not repeat, hence each trial (n) represents a switch compared to the previous (n-1) trial. The main dependent measure in such a procedure is the contrast between 'n-2 repeats' — trials in which the participant returns to a task after a single intervening trial (i.e., given switching between three tasks, A, B and C, the third trial in the task sequence ABA) — and 'n-2 switches', in which the task has not been performed for two or more trials (such as the third trial in CBA sequences). Based on the assumption that persisting activation and inhibition decay over time, such that the residual effect on trial n of trial n-2 is greater than that of trial n-3+, it has been reasoned that the persisting-activation and persisting-inhibition hypotheses make opposite predictions for this contrast (Koch et al., 2010). In the event of persisting activation of task sets, n-2 repeats should be facilitated by having performed the task more recently compared to n-2 switches, and therefore reaction times should be shorter. In contrast, the persisting inhibition of task sets hypothesis suggests that n-2 repeats should be slowed compared to n-2 switches (see Mayr & Keele, 2000, for a one-tailed statement of this hypothesis).

Typically, studies using an n-2 repetition procedure find a robust n-2 repetition cost, that is, n-2 repeats are significantly slower than n-2 switches. This effect has been observed for a range of different tasks (see section 3.2.5).[1]

---

[1] In the literature on the n-2 repetition procedure, various terms ('backward inhibition', 'task/task-set inhibition', 'alternating switch cost') have been used somewhat interchangeably. Here, I adopt the practice of Koch et al. (2010) of using the theoretically neutral term 'n-2 repetition costs' to refer to the empirical phenomenon, while the terms 'backward inhibition' or 'task-set inhibition' are reserved for putative theoretical mechanisms.

3.2 Empirical and theoretical issues

3.2.1 N-2 repetition costs as task-set inhibition

It has frequently been suggested that n-2 repetition costs reflect the operation of a mechanism for sequential task control. In situations that afford a number of competing tasks, rapidly transitioning from one task to another would require the suppression of a highly active task-set in favour of a less active task-set. Mayr and Keele (2000) proposed that a specific cognitive mechanism which they termed 'backward inhibition' facilitates task selection by inhibiting no-longer relevant task-sets.

Assuming such an inhibitory mechanism raises a theoretical question — at what level of the cognitive system is it situated? One possibility is that it is a high-level executive process, perhaps belonging to a supervisory attentional system (SAS; Norman & Shallice, 1986). However, Mayr and Keele (2000) discounted this possibility based on evidence that the n-2 repetition cost is resistant even to long preparation intervals (reviewed in section 3.2.9) and when participants know that the inhibited task is likely to be useful in the immediate future, such as when sequences are fully predictable (see section 3.2.2). Mayr and Keele (2000) preferred an account in which task-set inhibition is a relatively automatic process, resistant to high-level intervention. According to this view, "inhibition may be contributed by a process that is insensitive to to the current context, but instead, once triggered, obeys its own temporal dynamics. To be more specific, disengagement from a task set [. . .] may turn on an inhibitory node associated with the task-set representation. Once initiated, inhibition is fed into the task-set node until activity of the inhibitory node has waned according to its inherent decay function. As long as the inhibitory node is active, selection of the associated task-set node will be impaired" (Mayr & Keele, 2000, p. 23). Subsequent research on this point has not been definitive, and the question remains open (see section 3.2.9).

3.2.2 Non-inhibitory accounts of n-2 repetition costs

A number of possible explanations for the n-2 repetition cost are possible which do not involve inhibitory mechanisms. While these accounts fail to provide a complete account of the data, they are considered briefly below.

N-2 repetition costs due to violation of sequential expectancies

N-2 repetition costs observed in ABA compared to CBA sequences may be due to violations of an implicit expectation that tasks are equally distributed in short runs. However, it has been argued that this explanation is ruled out by the evidence. For example, n-2 repetition costs remain even when sequences are fully predictable (Koch, Philipp, & Gade, 2006; Mayr & Keele, 2000; Mayr, 2009; Schneider, 2007). See Koch et al. (2010) and Mayr and Keele (2000, p. 5-6) for additional arguments.

N-2 repetition costs due to cue switching

Within the two-task-switching literature, extensive research has been conducted to determine whether the (n-1) switch cost on explicitly cued tasks can be attributed to cue-related processing, rather than control of task-sets. Such studies have used procedures in which two alternative cues are used to indicate each task, such that the effects of a cue switch can be partially separated from those of a task switch (e.g., Logan & Bundesen, 2003; Mayr & Kliegl, 2003). In a recent review of the two-task switching literature, Jost et al. (2013) concluded that some, but not all, of the switch cost in the explicit cueing procedure was attributable to cue switches, rather than task switches (see section 2.1.5 for a review).

Importantly, using a similar 2:1 cue:task mappings procedure in the n-2 repetition procedure produces a dissociation between (n-1) switch costs and n-2 repetition costs; while the former were affected by cue switches, the latter were not (Altmann, 2007b; Gade & Koch, 2008; Mayr & Kliegl, 2003). Based on this evidence, it has been argued that task switching cannot be reduced down to cue switching (e.g., Jost et al., 2013; Koch et al., 2010). Mayr and Kliegl (2003) reasoned that this pattern of findings supports a two-stage model of task switching. According to this model, a first stage involves cue-driven retrieval of task rules from long-term memory into working memory, and is affected by cue-switching. A second stage is assumed to reflect task-set reconfiguration, which involves using active task rules to produce a task-appropriate attentional configuration, and is indicated by 'pure task-switch costs' (i.e., the difference between the overall switch costs and the cue switch cost). Whether this second 'application' stage can be affected by preparation is considered by Jost et al. (2013, p. 8). It was proposed that the application stage could only take place after stimulus presentation (Mayr & Keele, 2000). However, later studies revealed that true task-switch costs were affected by manipulations of the preparation interval (Monsell & Mizon, 2006; Jost, Mayr, & Roesler, 2008).

Set-specific episodic priming

Mayr and Keele (2000, p. 20) suggested that n-2 repetition costs may alternatively be explained in terms of associative priming of the task-set and response, a similar explanation to that of item-specific priming effects in the two-task switching literature (Allport & Wylie, 2000). For example, in a visual search task, a previous instance of the target dimension ('colour') may form an association with a specific response ('upper right'). Such a mechanism would lead to both benefits and costs, depending on whether the current target was congruent or incongruent with the associated response, respectively. If incongruencies occurred with a higher probability than congruencies (e.g., the visual search procedure requires identifying a target among four possibilities) or if the RT cost of incongruencies was quantitatively greater than the RT benefit of congruencies, then the net effect might appear as an n-2 repetition cost.

Mayr and Keele (2000) ruled this possibility out on the basis that it makes specific predictions for the effect of similarity between instances of the same task-set, which were not supported by their data. The set-specific episodic priming account would predict that highly similar trials should produce RT benefits, while low similarity trials should produce repetition costs. In contrast, a task-set inhibition account would predict similar repetition costs for both high and low similarity trials. In five out of six experiments, n-2 feature repetitions produced greater RT costs than non-repetitions, the opposite of what would be predicted by a set-specific priming account. However, n-2 repetition costs were reduced slightly for n-2 response repetitions than n-2 response switches, indicating a small subsidiary effect of episodic priming. However, the fact that n-2 repetition costs were persistent in such trials strongly indicates that set-specific episodic priming is inadequate as an overall explanation for n-2 repetition costs.

Because the procedure used by Mayr and Keele (2000) randomised the generation of task parameters trial-by-trial, n-2 response repetitions very rarely also coincided with cue and stimulus repetitions. These experiments therefore cannot rule out the possibility of an episodic priming component to the n-2 repetition cost. Mayr (2002) tested this possibility more directly, developing a paradigm in which stimuli (a dot appearing in one of four possible locations) had to be translated along one of three dimensions: vertically, horizontally, or diagonally, according to a rule given by the task cue. For an n-2 repetition, therefore, the stimulus and correct response either matched or mismatched those that appeared on the corresponding n-2 trial. Mayr (2002) reasoned that according to an inhibition-based account, n-2 repetition costs should be associated with both n-2 response repetitions and non-repetitions. If n-2 response repetitions were associated with a smaller n-2 repetition cost than n-2 response non-repetitions, however, this would be suggestive of a response priming effect, implying that at least a component of the n-2 reptition cost could be ascribed to an episodic priming effect. Mayr (2002) reported an experiment in which 39 participants completed 3 blocks of 120 trials. While both n-2 repetition costs and n-2 response-repetition facilitation were observed, crucially there was no significant interaction between the two factors. It was therefore concluded that while episodic priming affected reaction times on n-2 task alternations, it was inadequate as a sole explanation for the n-2 repetition cost.

More recently, however, Grange, Kowalczyk, and O'Loughlin (2017) replicated the earlier experiment of Mayr (2002) in 3 separate experiments with a large number of participants (76, 66, and 25) each performing 4 blocks of 120 trials. On this occasion, in addition to a significant n-2 repetition cost and n-2 response repetition facilitation, a significant interaction between these two factors was observed for reaction times. Although the effect could not fully account for the n-2 repetition cost, Grange et al. (2017) argued that the n-2 repetition cost is potentially contaminated with episodic priming effects, and cannot be viewed as a 'pure' measure of inhibition. Analysing data from all three of their experiments, they found evidence for a small 'residual n-2 repetition cost' which may be attributable to inhibition. The authors acknowledged that the extent to which n-2 response-repetition produces a facilitation, and n-2 response-nonrepetition produces interference, is difficult to ascertain experimentally, and therefore the residual n-2 repetition cost may be underestimated in their data.

3.2.3 The role of conflict

In an alternative account of task switching in general, Goschke (2000) proposed that task inhibition was dynamically regulated in response to the degree of conflict (cf. Botvinick et al., 2001) occurring during task. More recently, the concept of conflict has been applied as a specific mechanism for the recruitment of task-set inhibition. In a review of the n-2 repetition paradigm, Koch et al. (2010) argued that a wide range of evidence suggests that task-set inhibition is a somewhat flexible mechanism, which is directed toward any elements of a task-set that produce intertrial conflict (e.g., Arbuthnott, 2009; Gade & Koch, 2005, 2007; Houghton, Pritchard, & Grange, 2009; Koch et al., 2010; Schuch & Koch, 2003). Specific evidence supporting this view is discussed throughout this chapter.

3.2.4 Direct evidence for task-set inhibition: The n-1 flanker facilitation effect

The n-2 repetition cost is usually interpreted as an indirect measure of the residual inhibition of the task on trial n (Koch et al., 2010; Mayr & Keele, 2000). According to theoretical accounts, task-set inhibition is a means of facilitating the performance of the current task by reducing the interference of previously performed tasks (Mayr & Keele, 2000, p. 4). While it has been proposed that task-set inhibition facilitates task switching, the n-2 repetition cost does not provide direct evidence of such facilitation. Instead it measures the residual interference when resuming a task which was previously switched-away from. Thus, it is important to demonstrate that a task-set inhibition mechanism actually facilitates switching performance. Hübner, Dreisbach, Haider, and Kluwe (2003) provided some evidence using a different procedure. In their experiments, participants switched between three univalent tasks (i.e., stimuli were associated with only a single task-set), with 'flankers' (distractors) presented either side of the target stimulus. The distractors were target stimuli incongruent with the current task. Hence each flanker produced task interference, uniquely affording the previous *n-1* or *n-2* tasks. Hübner et al. (2003) reasoned that if, when switching to the current (*n*) task, a task-set inhibition mechanism was applied to the *n-1* task, acting to reduce the interference from the previous task, then interference from *n-1* flankers should be lower than *n-2* flankers, when measured as an increase in reaction time or error rates. It was further hypothesised that if task-set inhibition was a top-down, endogenous mechanism, such an effect of task-set inhibition should only be observed in conditions where each trial was cued in advance.

The authors found that this was in fact the case for reaction times: trials with

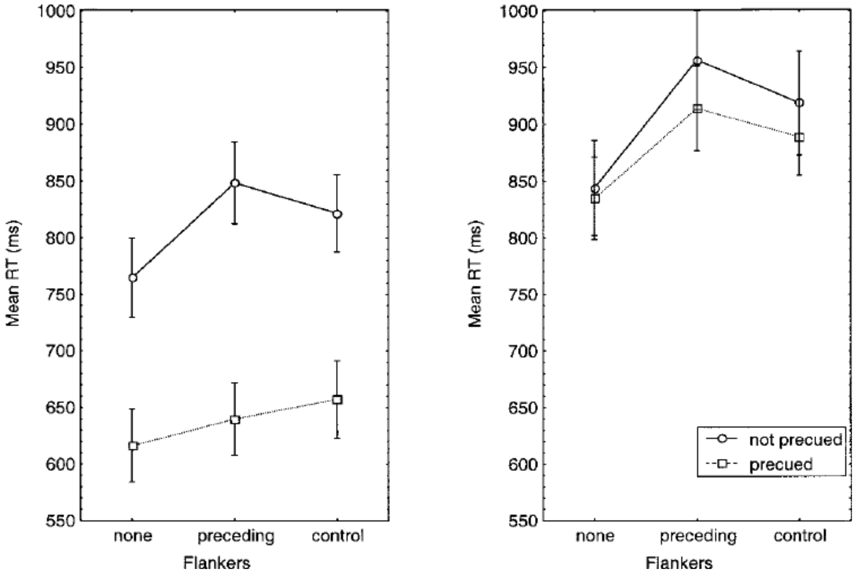


Figure 3.1: Geometric means of reaction times (RTs) as a function of cue condition (not precued or precued) and flanking conditon (no flankers, flankers from the preceding task, or flankers from the control task) for Experiments 1 (left) and 2 (right) from Hübner et al. (2003). Error bars represent standard errors of the means. Reproduced with permission from Hübner et al. (2003).

n-1 flankers had shorter RTs than trials with *n-2* flankers in the pre-cued condition, while trials with *n-1* flankers had longer RTs than those with *n-2* flankers in the non-cued condition (fig. 3.1 left panel). This pattern of results was found to depend on advance cues which specified the identity of the upcoming task. Thus, in a second experiment, advance cues which only indicated when the task would occur did not achieve the same reduction in interference for $n-1$ flankers (fig. 3.1 right panel). Importantly, no repeat trials occurred during this procedure, implying that although participants were aware that the $n-1$ task was to be abandoned when preparing for trial n, knowledge of which task was next to be performed was essential for producing the $n-1$ flanker facilitation effect. Taken together, this pattern of results was interpreted as supporting the proposal that task-set inhibition acts to facilitate processing on the current task by inhibiting task-sets associated with a previous task. Additionally, it provides evidence that task-set inhibition is an endogenously-triggered mechanism that depends on advance preparation specific to the forthcoming task.

However, the study has a number of limitations. Based on the studies of Mayr and Keele (2000), one might expect to observe an n-2 repetition cost in addition to the flanker n-1 facilitation effect. However, this was not observed. Kuhns, Lien, and Ruthruff (2007), in an attempt to directly replicate the findings of Hübner et al. (2003) using a slightly simplified procedure, did find n-2 repetition costs, but not the n-1 flanker facilitation effect. In a second experiment, in which the procedure was modified to maximise conflict between tasks, Kuhns et al. (2007) did find both both behavioural effects of interest. Kuhns et al. (2007) speculated that task-set inhibition may be modulated by task difficulty, with greater inhibition recruited in more difficult conditions. However, this proposal awaits experimental investigation.

In summary, the experiments of Hübner et al. (2003) and Kuhns et al. (2007) provide an important additional behavioural measure of task-set inhibition, independent from the n-2 repetition costs observable in the procedure of Mayr and Keele (2000). However, the results they produce appear somewhat brittle. Additional studies are required to determine whether the effects of task-set inhibition are robust, and if so, to establish their boundary conditions.

3.2.5 The target of task-set inhibition

The n-2 repetition cost has been observed in a range of different tasks. For example, Mayr and Keele (2000) used a visual search paradigm, in which participants were required to identify which of four stimuli differed from the others, on a specified visual dimension, out of colour, orientation and movement. For this type of task, the task-set must presumably specify the visual dimension on which to maintain visual attention. Thus in this case task-set inhibition comprises inhibition of target stimulus dimensions. (See also Arbuthnott, 2005; Arbuthnott & Frank, 2000; Arbuthnott & Woodward, 2002; Mayr & Kliegl, 2003, in which participants switched between attending digits, letters and symbols which were combined in a display.)

More frequently, symbolic classification tasks have been used. For example, Schuch and Koch (2003) had participants classify single digits as greater or less than 5, even or odd, or located centrally (i.e., 3, 4, 6, 7) or peripherally (i.e., 1, 2, 8, 9) (see also Arbuthnott & Frank, 2000; Arbuthnott & Woodward, 2002; Gade & Koch, 2005, 2007, 2008; Hübner et al., 2003; Kuhns et al., 2007; Philipp & Koch, 2006; Sdoia & Ferlazzo, 2008). In this task domain, task-sets must include abstract symbolic categories as a target for inhibition.

Additionally, n-2 repetition costs have been observed in trilingual language switching (Philipp, Gade, & Koch, 2007; Philipp & Koch, 2009), and switching between response modalities (Philipp & Koch, 2005). In this latter procedure, participants performed a single symbolic classification task but switched between different response modes. In these examples, task-sets must include representations of the response modality, which may also be the target of inhibition. This is highlighted in a study by Koch, Gade, and Philipp (2004), in which one of the three tasks was a simple response, requiring participants to simply press both response keys. N-2 repetition costs were observed even for this simple response task, suggesting that task-set inhibition could be specifically directed within-modality at a particular means of using responses.

In summary, it appears that task-set inhibition may be applied to multiple different aspects of the task-set: including visual stimulus dimensions, abstract stimulus categories, and response modalities.

3.2.6 N-2 repetition costs and cue-encoding

N-2 repetition costs are affected by certain experimental manipulations of task cueing. Firstly, there is evidence for an influence of spatial position on task cues. Arbuthnott and Woodward (2002) found significantly smaller n-2 repetition costs when task cues were presented in different spatial locations, compared to when locations of task cues spatially overlapped. (See also Arbuthnott, 2005, 2008b, 2009;

Druey & Hübner, 2007, for replications.) Arbuthnott (2009) suggested that such non-overlapping cues are encoded as 'location-task representations', which do not interfere with each other sufficiently during response generation to recruit task-set inhibition. According to this view, task inhibition is a flexible mechanism that is sensitive to the degree of overlap in task-set components, which Koch et al. (2010) argue is consistent with a dynamic task-set inhibition mechanism recruited by task conflict (an argument discussed in detail in section 3.3.3).

Secondly, cue transparency has been found to affect n-2 repetition costs. In a series of experiments using a visual search procedure, Houghton et al. (2009) reported larger n-2 repetition costs when the cue for the target stimulus dimension was more abstract. Specifically, cues that were identical with the stimulus target produced no n-2 repetition costs. Iconic cues that directly represented the relevant dimension produced small costs, verbal cues produced an intermediate cost and arbitrary cues, either verbal or iconic, produced the largest costs. Additional evidence that cue-encoding processes may be the target of backward inhibition was reported by Grange and Houghton (2010b), who showed that n-2 repetition costs were increased by switching cue-task assignments after several blocks of trials. Houghton et al. (2009) interpreted their findings in the context of the two-process model of Mayr and Keele (2000), as affecting the first stage whereby cues are used to retrieve task-sets from long-term memory. According to their account, this first stage (cue-target translation) is more difficult for less transparent cues. Additionally, the cue-target translation process may generate conflict with recently performed translations (i.e., on the n-1 trial). In this experimental context, then, task-set inhibition affects the cue encoding stage, rather than the later response stage. Houghton et al. (2009) suggest that this account of task-set inhibition is consistent with the proposal that its functional role is as a means of clearing working memory, and thereby facilitating the establishment of a new task-set on the following trial (Mayr & Keele, 2000). While the verbal specification of this model is lacking in detail, the same authors have more recently implemented similar ideas as a computational model (see section 3.3.4). However, this account of task-set inhibition would not appear to explain evidence that task-set inhibition depends on response processes (section 3.2.7).

A further cue-releated influence on n-2 repetition costs that has been suggested is temporal cue-target overlap. Druey and Hübner (2007) reported significantly smaller n-2 repetition costs when task cues disappeared before the stimulus was presented. However, Grange and Houghton (2009) failed to replicate this effect in three experiments. Therefore, while there is evidence that cue characteristics (spatial relationships and cue-task transparency) influence n-2 repetition costs, there is no clear evidence for an influence of the temporal relation between cues and stimulus targets.

More recent support for the involvement of cue encoding processes to the n-2 repetition cost comes from a study by Scheil (2016). In two experiments on the effects of practice on n-2 repetition costs, one changed stimulus-response associations of one task halfway through the procedure, while the other changed the cue-task association of one task. In the first, n-2 repetition costs gradually declined as a result of practice. Switching the stimulus-response associations had no effect on n-2 repetition costs for that task. In the second, however, switching cue-task associations increased n-2 repetition costs substantially for the switched task, while n-2 repetition costs decreased for the other two tasks where cue-task associations remained constant. While these results do not rule out response-related processes as a source of the n-2 repetition cost, they do provide additional weight to the involvement of cue-task associations.

Recently, Grange et al. (2017) found that the n-2 repetition cost was modulated by n-2 response repetitions, thought to reflect an effect of episodic priming (see section 3.2.2). In addition to varying task sequence (n-2 repetition vs. n-2 switch) and episodic prime response (n-2 response repetition vs. non-repetition) one of their experiments also varied cue type, from fully transparent (arrows indicating the direction of the required spatial translation) to non-transparent (polygons arbitrarily indicating which translation to perform). A significant 3-way interaction was found between cue type, n-2 task repetition and n-2 response repetition. The modulation of the n-2 repetition cost by n-2 response repetition was larger for non-transparent cues than transparent ones. The authors argued that this was likely due to the additional working memory demands of non-transparent cues, and that it localised the episodic retrieval effect on n-2 repetition costs to a stage of cue-based task retrieval. This finding led Grange et al. (2017) to question the findings of recent studies on cue-encoding and the n-2 repetition cost. They argued that putative modulation of the n-2 repetition cost by cue-encoding processes may in fact be an episodic priming effect. This might equally apply to modulations of the n-2 repetition cost associated with response-led processes (section 3.2.7), a possibility which remains to be explored.

3.2.7 N-2 repetition costs and response processes

A range of empirical evidence suggests n-2 repetition costs are modulated through the manipulation of response processes. The main empirical findings are discussed below, in terms of manipulations of response set overlap and response execution. However, the sensitivity of the n-2 repetition cost to response processes is by no means universal. For example, in the study by Scheil (2016) examining the effects of practice on n-2 repetition costs, in which n-2 repetition costs were observed to gradually decline with practice, switching response mappings mid-way through the procedure did not disrupt the general trend. Switching cue-task mappings, on the other hand, temporarily caused a large increase in n-2 repetition costs. This was interpreted by the author as suggesting that the cue-based process of retrieving the task from working memory was an important cause of backward inhibition.

Response set overlap

Gade and Koch (2007) operationalised the conflict between different tasks as the degree of overlap between different response sets. Their participants performed four symbol classification tasks in response to trivalent stimuli, responding to each task with a vocal response (cf. manipulations of response modality, see section 3.2.5). For three tasks (classifying according to shape, colour and size), responses were mapped to overlapping responses (“left”' and “right”). The fourth task (responding whether the stimulus was filled or empty) was univalent. Across three experiments, N-2 repetition costs were found for the trivalent, overlapping tasks. In contrast, n-2 repetition costs were not reported when the “fill” task occurred on trial n-1, and the responses (“up” and “down”) did not overlap with the other tasks. N-2 repetition costs did occur, however, when responses to the “fill” task were mapped to overlap with the other tasks “left” and “right”. These findings were interpreted as supporting the suggestion that task-inhibition is recruited by response conflict, and that task-set inhibition acts on the level of response mappings (Gade & Koch, 2007; Koch et al., 2010; see also Schneider & Verbruggen, 2008).

Response execution

If the function of task-set inhibition is to facilitate task processing by inhibiting recently abandoned task-sets, one question is when this process takes place and how it is triggered. While a number of studies have examined the role that actually producing a response has on task switching (see Koch et al., 2010, p. 10 for a review), to date only two have focused on n-2 repetition costs (Philipp, Jolicoeur, Falkenstein, & Koch, 2007; Schuch & Koch, 2003).

Schuch and Koch (2003) devised a task which combined switching between three digit classification tasks, with an auditory go/no-go signal presented simultaneously with stimulus onset. 25% of all trials were no-go trials. Tasks were cued in advance, and the preparation interval (CSI) was manipulated. A main effect of CSI indicated that participants indeed prepared for the upcoming task. Although n-2 repetition costs were reported when trial n-1 was a go trial, they were significantly reduced or eliminated following no-go trials. This finding suggests that the process of generating a response, in addition to preparing for the task (i.e., encoding of cues and encoding task rules), is necessary for triggering task-set inhibition.

One potential criticism of this interpretation is that the act of withholding a response is not the same as simply not generating a response. For example, it might be argued that stopping a response on no-go trials involves the use of a dedicated response inhibition mechanism (Verbruggen & Logan, 2009) which interferes with any task-set inhibition processes. However, the effect was replicated when the correct response to a no-go trial was to press both response keys (Koch & Philipp, 2005; Schuch & Koch, 2003). Hence, the reduction of n-2 repetition costs occurs when the execution of a task-specific response is withheld (i.e., execution following response selection), rather than merely occurring following non-specific motor stopping. It thus appears that response selection is a critical trigger for task-set inhibition. In the study of Schuch and Koch (2003), the degree of response selection which actually occurs on no-go trials is not controlled, as the no-go signal is presented at the same time as the stimulus.

This issue is addressed by Philipp, Jolicoeur, et al. (2007), who used a modified procedure, the “go-signal paradigm”, to separate response selection and response execution processes. In this procedure, participants prepare a response after a stimulus is presented, but do not execute it until an auditory go signal is presented, after either a short or a long delay (go-signal delay, or GSD). In 25% of trials, a no-go signal is presented instead, and the response must be withheld. The authors reasoned that for no-go trials, a short GSD on trial n would tend to produce neither response selection nor response execution, while a long GSD on trial n would tend to produce response selection but not response execution. Go-trials with a long

GSD on trial n would tend to produce both response selection and execution.

For trials with a short GSD on trial n, a similar pattern of results to Schuch and Koch (2003) was reported, with significantly reduced n-2 repetition costs following no-go trials compared to go-trials.[2] To examine the effects of response execution in trial n-1, sequences were compared in which go trials with short GSD on trial n followed either no-go trials with long GSD (i.e., response selection but not execution in trial n-1) or go-trials with long GSD (i.e., response selection and execution in trial n). In this contrast, n-2 repetition costs were significantly larger in go-trials following go-trials than following no-go trials. This pattern of results thus indicates that response execution, as well as response selection in trial n-1, is important for the recruiting of task-set inhibition.

The findings from Schuch and Koch (2003) and Philipp, Jolicoeur, et al. (2007) have been interpreted as consistent with the more general hypothesis that a task-set inhibition mechanism is triggered by conflict (Koch et al., 2010), specifically response conflict in this case. According to this view, response conflict generated from the simultaneous activation of conflicting responses (according to the current task-set and an irrelevant task-set) signals the need for task-set inhibition (see section 3.3.3).

However, such a model is difficult to reconcile with evidence which suggests that an earlier cue-encoding stage may be affected by task-set inhibition (Houghton et al., 2009, see section 3.2.6). If, as Houghton et al. (2009) propose, task-set inhibition occurs in response to conflict generated at a cue-encoding stage, it is not obvious why it should later be triggered or reset by go- or no-go trials, respectively. Addressing this discrepancy, Houghton et al. (2009) argue that a complete account of the cognitive processes involved in stopping a response in the context of a task-switching trial has not yet been presented, and that such an account

---

[2] In contrast, on trials with a long GSD on trial n, n-2 repetition costs were substantially reduced, and the interaction with trial type (go vs. no-go) in trial n-1 was not significant, indicating that with sufficient time to select a response on trial n, the effects of task-set inhibition were overcome.

might include a general 'cognitive resetting' which disrupts ongoing WM processes. Alternatively, no-go trials might be better characterised as interleaving an additional (fourth) task, which involves its own memory (cue-target translation) processes (Houghton et al., 2009, p. 475). To date, no empirical manipulation of cue transparency in combination with a go- or a no-go signal has been conducted.

3.2.8 Dissipation of task-set inhibition: Manipulations of intertrial intervals

A frequent assumption of the n-2 repetition paradigm is that task-set inhibition dissipates over time. That is, the residual task-set inhibition for the final A trial on the sequence ABA, when the A task was last performed two trials previously, is greater than the persisting inhibition in the sequence CBA, where the A trial has not repeated for at least three trials.[3] In support of this assumption, it has frequently been reported that n-2 repetition costs are smaller in conditions where the intertrial interval, specifically the interval between the previous trial's response and the next cue (Response-Cue Interval, RCI), was long compared to when it was short (e.g., Grange & Houghton, 2009; Koch et al., 2004; Mayr & Keele, 2000). This finding has been interpreted to suggest that task-set inhibition does indeed dissipate over time.

One problem with the methodology used in these studies, however, is that RCI was manipulated in a blocked design, such that the interval after trial n-1 was the same as that after trial n-2, etc. However, blockwise manipulations of RCI conflate potentially different functional effects of RCI. In particular, Gade and Koch (2005) reasoned that in a design where RCIs following trials n-2 and n-1 may be varied independently, if task-set inhibition dissipates over time, then a reduction in n-2 repetition costs should be observed for shorter RCIs following trial n-1. Or-

---

[3] However, Vandierendonck (2013) has questioned this assumption, challenging the empirical evidence that backward inhibition declines with increasing numbers of intervening trials (e.g., ABA vs. ABBA, ABBBA, etc.).

thogonal to this prediction, if task-set inhibition is recruited by response conflict, then the critical interval determining the size of the n-2 repetition cost should be the RCI following trial n-2, as a shorter RCI would maximise conflict. In fact, in two experiments which used such a within-block manipulation of RCIs, the RCI following trial n-1 had no systematic effect on n-2 repetition costs. In contrast, RCI following trial n-2 did significantly affect the size of n-2 repetition costs, with larger n-2 repetition costs produced by a shorter RCI. This finding was interpreted to suggest that task-set inhibition was partly determined by the residual activation of the n-2 task (i.e., the inhibited task) at the onset of trial n-1.

While these findings are, to date, based on a single study, if they prove robust, they are important for a theory of task-set inhibition in two ways. Firstly, they support a conflict-driven account of task-set inhibition. Secondly, they provide evidence that task-set inhibition does not passively dissipate with time,[4] but is instead relatively persistent (see also Mayr & Keele, 2000). This raises the question of what, if not passive decay, causes task-set inhibition to dissipate?[5] Koch et al. (2010) suggest that inhibition is released as a result of processing new tasks. For example, in performance of the sequence ABC, switching from task B to C reactively releases the previous inhibition of task A, making it easier to switch back to task A compared with task B. However, this account has not yet been specified in mechanistic terms.

3.2.9 Task-set inhibition as an endogenously triggered process

One line of inquiry in the n-2 repetition literature is whether task-set inhibition operates as a top-down or a bottom-up process. Within this paradigm, a "top-down"

---

[4] At least, over the RCIs used in the experiment, which were 1,400 ms and 1,900 ms in experiments 1 and 2, respectively (Gade & Koch, 2005).

[5] The explanation favoured by Gade and Koch (2005), that inhibition was directly linked to residual task activation, seems functionally incomplete, although compatible with an account that relates task-set inhibition to conflict.

process has been operationalised as one in which the identity of the task is revealed by a specific task cue, while for a "bottom-up" process, the task is inferred from the stimulus (Koch et al., 2010; Mayr & Keele, 2000). This distinction maps onto that between endogenous/exogenous stages of processing in two-process models, proposed within the 2-task switching literature (e.g., Rogers & Monsell, 1995). Top-down/endogenous processes, by definition, may take place in response to the task cue, in the absence of the stimulus, whereas bottom-up/exogenous processes, by definition, may only begin after onset of the target stimulus (Koch et al., 2010; Mayr & Keele, 2000). On this basis, two lines of research have examined whether, firstly, task-set inhibition is dependent on advance cueing of the specific task, and secondly, whether it is affected by variation of the preparation interval.

Does backward inhibition depend on advance, task-specific cueing?

In their model of task switching, Mayr and Keele (2000) argued that inhibition only occurred during endogenous task-switches, and not during exogenous task-switches triggered by the target stimulus. In their Experiment 2, Mayr and Keele (2000) adapted their visual search paradigm such that distractors only differed from the target on dimensions which were never used to identify targets. Thus, target stimuli could be identified purely from the stimulus display. The authors contrasted two conditions, a top-down condition in which a cue indicated the relevant stimulus dimension, and a bottom-up condition, with no cue. They reported significant n-2 repetition costs only in the top-down condition, supporting their hypothesis that top-down task selection was necessary for the triggering of task-set inhibition. However, they noted that in their experiment, a bottom-up approach for identifying the target was always a viable strategy in both conditions. Indeed, the n-2 repetition cost was only significant for the fastest group of participants, which was argued to reflect those who had used the cue to initiate a task-switch. It has been argued that the small n-2 repetition cost (7 ms), significant for only this group, does not provide strong evidence that task-set inhibition is exclusively a top-down process (Koch et al., 2010, p. 7).

Converging evidence that task-set inhibition is a top-down process is provided from the flanker compatability procedure (Hübner et al., 2003; Kuhns et al., 2007, see section 3.2.4 for a review of this procedure). Contrasting results from two experiments, Hübner et al. (2003) found that the n-1 flanker facilitation effect was only observed when the cues used were task-specific (i.e., they directly indicated which task to perform next), and not when cues were non-specific (i.e., they merely indicated a change of task, but did not specify the new task). This was interpreted to suggest that not only was preparation required for the recruitment of task-set inhibition, but preparation must be task-specific; unspecific preparation is not sufficient.

While these interpretations are consistent with the hypothesis that task-set inhibition only occurs with top-down task-specific preparation, a number of criticisms have been raised. Firstly, Koch et al. (2010) highlighted that a flanker facilitation effect was indeed observed following task-unspecific cues in the error rates, rather than the RTs, although Hübner et al. (2003) do not report a test for significance.[6] Secondly, the strength of the evidence for task-set inhibition has been questioned (e.g., by Kuhns et al., 2007, p. 978). While the n-1 flanker facilitation effect was significant when the specific task was cued in advance, the n-2 repetition cost in this condition failed to reach significance. Kuhns et al. (2007) reasoned, however, that procedural details used by Hübner et al. (2003) discouraged participants from inhibiting tasks. Firstly, target stimuli were univalent, which facilitated bottom-up activation of the task (cf. Mayr & Keele, 2000). Secondly, non-overlapping response sets were used, leading to a lower level of response conflict, and thus task-set inhibition, than might be expected from overlapping response sets (Gade & Koch, 2007).

---

[6] Generally, this paradigm predominantly used RTs as dependent measures with relatively little consideration given to error rates, or complications due to a speed/accuracy trade offs.

Kuhns et al. (2007) addressed these problems in an attempt to replicate the earlier findings. Additionally, they reasoned that top-down or bottom-up accounts of task-set inhibition (or, as they characterised it, proactive or reactive) could directly be distinguished as they made different predictions. They argued that the crucial comparison was sequences in which the current flanker task matched the preceding target task (e.g., $Cb - Ac$, with upper-case representing the target task, and lower-case, the flanker) compared with sequences where the current flanker matched the preceding flanker (e.g., $Bc - Ac$). In this comparison, Kuhns et al. (2007) reasoned that a proactive account of inhibition implies that the most strongly inhibited task should be the preceding target task, hence flanker effects should be smaller for trials with flankers matching the preceding target task ($Cb - Ac$). In contrast, if inhibition is reactive, the most strongly inhibited task should be the preceding flanker task, hence flanker effects should be smaller for the trials with flankers matching the preceding flanker task ($Bc - Ac$).

In fact, Kuhns et al. (2007) had mixed results replicating the original effects of Hübner et al. (2003) (see section 3.2.4 for a detailed review). While significant n-2 repetition costs were reported for their Experiment 1, no significant n-1 flanker facilitation effect was observed, while Experiment 2, which increased task conflict, produced both a significant n-2 repetition cost and n-1 flanker facilitation effect. In Experiment 1, RTs were significantly faster when the flanker on the current trial matched the preceding flanker, a finding argued to support the reactive inhibition hypothesis. However, in Experiment 2, the flanker compatibility effect was (nonsignificantly) smaller when the current flanker matched the immediately preceding target, and larger when it matched the immediately preceding flanker. This finding was more consistent with the proactive inhibition hypothesis. To account for these seemingly incompatible findings, Kuhns et al. (2007) speculated that inhibition may be either reactive or proactive, depending on task difficulty.

In their account of proactive (top-down) task-set inhibition, Kuhns et al. (2007)

make two assumptions. Firstly, they assume that proactive inhibition is triggered by endogenous task selection processes (consistent with Mayr & Keele, 2000). Secondly, they make the stronger assumption that proactive inhibition is necessarily directed toward the previous (endogenously activated) target. By the same logic, their reactive account of task-set inhibition assumes, firstly, that inhibition occurs equally when engaging in endogenous task preparation (i.e., for the current task) as it does when preparation occurs exogenously. Secondly, reactive inhibition is assumed to be directed toward the task activated exogenously on the preceding trial (i.e., the task cued by the preceding flanker). This logic thus involves a somewhat stronger set of assumptions than other characterisations of the top-down/bottom-up distinction (e.g., Koch et al., 2010; Mayr & Keele, 2000). It is not clear whether such assumptions necessarily follow from the basic assumption that task-set inhibition is top-down (proactive) or bottom-up (reactive).

In summary, empirical evidence based on advance cueing only weakly supports the assertion that task-set inhibition is a top-down process. While the visual search paradigm of Mayr and Keele (2000) is methodologically flawed for distinguishing whether a top-down or bottom-up task-selection strategy has been used, the flanker-facilitation procedure of Hübner et al. (2003) and Kuhns et al. (2007) appears equivocal on the issue. The reasoning about empirical effects in task switching, in this instance, suffers from vaguely specified implicit models (e.g., Kuhns et al., 2007). As a result, theoretical interpretations (i.e., whether n-2 inhibition is reactive or proactive) remain open to question. This is one area which would be clarified by the use of computational models, which make processing architectures explicit, and allow quantitative predictions to be directly simulated and compared with predictions obtained through verbal reasoning.

Is backward inhibition affected by preparation?

A second group of studies have addressed whether task-set inhibition is a top-down process by manipulating the cuing interval (CSI). According to this logic, if task-set inhibition is a top-down process, longer preparation intervals should produce more inhibition, thus producing larger n-2 repetition costs. Most studies have found no effect of varied CSI on n-2 repetition costs (Bao, Li, Chen, & Zhang, 2006; Gade & Koch, 2008; Mayr & Keele, 2000; Schuch & Koch, 2003), although some studies reported mixed results (Koch et al., 2004; Philipp, Gade, & Koch, 2007). Following this pattern of findings, some have concluded that n-2 repetition costs reflect passive inhibition which decays over time, but is not affected by endogenous preparation (e.g., Mayr, 2002).

However, two problems have been raised with respect to this reasoning (Koch et al., 2010). Firstly, the reasoning that longer preparation intervals would produce a greater amount of task-set inhibition is only valid if one conceptualises the influence of inhibition as constant or increasing over time. Koch et al. (2010) have suggested that if, alternatively, inhibition is 'ballistic' — recruited over a short triggering period but taking a longer length of time to take full effect — then longer CSIs would not necessarily produce greater n-2 repetition costs. The second criticism is methodological: similarly to the criticisms of blockwise manipulation of intertrial intervals (section 3.2.8), generally CSIs have been manipulated blockwise. This approach confounds processes which take place in preparation for trial n-1 (namely, inhibition of the n-2 task) and processes which take place during preparation for trial n (such as overcoming any residual task-set inhibition). Thus, effects of task-set inhibition and task preparation may counteract each other (Koch et al., 2010).

Addressing this latter issue, Scheil and Kleinsorge (2014) varied CSI trial-by-trial, such that the preparation interval for trials n-2, n-1 and n could be either short (200ms) or long (1000ms). In two experiments, the authors found that while n-2 repetition costs were independent of the CSI in the current trial (n), they were indeed affected by CSIs in trials n-2 and n-1. Specifically, n-2 repetition costs were significantly larger when CSI was long in both trials n-2 and n-1. In three conditions involving shorter CSIs, (short-short, short-long or long-short) n-2 repetition costs were smaller, with no significant difference between the three conditions. Scheil and Kleinsorge (2014) argued that these findings were consistent with the view that task-set inhibition is required when the activation of the previous task-set is high (i.e., long n-2 CSI) and is recruited during the preparation period for the switched-to task (n-1).

Although this interpretation is intuitively plausible, it assumes that the final activation level of a task-set following response execution is related to the preparation interval. Yet it also seems possible that residual task-set activation may actually be greater for exogenously cued tasks. For example, in the connectionist model of task switching of Gilbert and Shallice (2002), an advantage in starting activation of the relevant task-demand unit at the start of a trial, counterintuitively, leads to lower absolute levels of activation at the moment of response execution, as task-demand activation continues to increase over the duration of trials, which are slower when starting activations are similar (i.e., on uncued trials). Thus, whether the assumption that final task activation is related to the preparation interval is consistent with extant models of task switching is an empirical question.

3.2.10 Distributional analyses of response times

While most analyses of RT data in the n-2 repetition paradigm have dealt exclusively with mean RTs, an alternative approach looks at potential effects of interest over the entire RT distribution for each participant. For example, if switch or n-2 repetition costs had differential effects on a participant's faster responses compared to their slower responses, this may provide insight that is not captured by mean RTs. In two-task switching, De Jong (2000) argued that for each participant, faster responses reflected a greater state of preparation than slower responses. A number of authors have applied this logic to n-2 repetition costs: if n-2 repetition costs were reduced by preparation, they should be smaller for faster trials than for slower trials.

Schneider and Verbruggen (2008) assigned data for each particpant into quantiles ordered by RT, with data for each quantile then being combined. N-2 repetition costs were then calculated separately for each quantile. These authors found consistent n-2 repetition costs throughout the RT distribution. In contrast, Grange and Houghton (2011), using the same approach, re-analysed data from two previous studies. The first was where arbitrary cue-task mappings were switched mid-way through the procedure (Grange & Houghton, 2010b). In the original paper, the authors argued that this cue reassignment would create conflict during the cue-target translation process and would therefore increase n-2 repetition costs, and this hypothesis was confirmed. In the second study, the original authors manipulated cue-task transparency, reasoning that less transparent cues would again increase cue-target translation conflict, producing greater n-2 repetition costs. Again, this hypothesis was confirmed (Grange & Houghton, 2010a).

Grange and Houghton (2011) used the same procedure as Schneider and Verbruggen (2008), constructing cumulative distributions by assigning RTs for each participant into deciles from slowest to fastest. For data from Grange and Houghton (2010b), n-2 repetition costs were found to gradually increase from faster to slower trials, with the trend being significantly elevated for switched cues (i.e., the larger n-2 repetition cost condition). For data from Grange and Houghton (2010a), no n-2 repetition costs were detected throughout the distribution for the transparent cue condition, while costs increased throughout the distribution for non-transparent cues. Importantly, the absence of any increase in n-2 repetition costs throughout the distribution in the transparent cue condition suggests that the distributional effect is not just a general effect of slower trials, but is related to the condition that causes n-2 repetition costs. Grange and Houghton (2011) argued that as faster RTs quantiles represent trials where participants have completed relatively more cue-based preparation than in slower quantiles, this distributional approach is sensitive to experimental manipulations of cue-processing conflict.

Subsequent studies have reported analogous findings. Thus, Grange and Juvina (2015), in examining the effect of extensive practice on n-2 repetition costs, conducted an ex-Gaussian analysis of RT distributions. In this approach, a bootstrap-resampling method was used to fit ex-Gaussian parameters to RT distributions. In this case, significant n-2 repetition costs were reported only for the τ parameter — representing the tail of the distribution (i.e., slower responses) — which decreased over the course of the experiment. While Grange and Juvina (2015) interpreted the reduction in n-2 repetition costs with practice in terms of increasing automatisation in the cue-task translation process, cue-task translation conflict was not explicitly manipulated in other ways.

Overall, sophisticated analyses of response time distributions offers future additional insights into the n-2 repetition cost. Although there is some evidence that in some circumstances n-2 repetition costs are elevated for slower responses, the theoretical grounding for this effect is currently weak. Grange and Houghton (2011) link this finding to an effect of preparation on n-2 repetition costs. However, this account rests on the assumptions, firstly, that faster and slower responses represent lesser and greater degrees of task preparedness respectively (and not a more general alertness or readiness) and secondly, that the n-2 repetition cost is decreased by preparation. In terms of the second assumption, as has previously been discussed (section 3.2.9), alternative evidence for an effect of preparation on the n-2 repetition cost is currently equivocal. The suggestion of Grange and Houghton (2011) that this effect is sensitive to cue-based conflict and not response-based conflict is plausible but has yet to receive solid empirical support. Distributional effects, at present, lack theoretical grounding.

3.3 Mechanistic accounts of Backward Inhibition

As discussed above, theoretical accounts of task inhibition differ on the source of the inhibition. This section will discuss three mechanistic proposals: self-inhibition, lateral inhibition, and inhibition triggered by conflict.

3.3.1 Self inhibition

One suggestion is that tasks-sets are self-inhibiting, after they have been used to generate a response. Similar mechanisms have frequently been used as a means of counteracting residual activation in sequential behaviour, and thus preventing perseveration. Such explanations have been proposed in task-switching, specifically by Grange et al. (2013) and Mayr and Keele (2000), and the cognitive psychology literature in general (reviewed in Grange et al., 2013).

However, Koch et al. (2010) have argued that a self-inhibition account of n-2 repetition costs is inadequate on emprical grounds. If task-inhibition was self-directed, one would expect immediate task repetitions (e.g., repeat trials in 2-task switching contexts) to be associated with a cost, rather than the benefit that is routinely observed (reviewed in Kiesel et al., 2010). In response, Grange et al. (2013) suggested that the degree of self-inhibition exerted might be under the control of the participant, and adjusted strategically depending on the task context. For example, in a 2-task switching paradigm with unpredictable but infrequent task switches, self-inhibition would be counterproductive as task repetitions are more common. However, in a 3-task switching paradigm with no task repetitions, such as the n-2 repetition procedure, self-inhibition would be advantageous. In support, Grange et al. (2013) argue that 2-task switching paradigms have been successfully simulated using an activation-only model (Altmann & Gray, 2008), whereas n-2 repetition paradigms appear to require an additional inhibitory mechanism (Grange et al., 2013). From this account, one might predict that n-2 repetition costs would not be observed in 3-task contexts where task repeats were possible. In fact, using such a procedure, Arbuthnott (2005) did find n-2 repetition costs, as did Philipp and Koch (2006), although in the latter study n-2 repetition costs were significantly reduced compared to a condition with no immediate repetitions. Based on this pattern of data, strategic modulation of self-inhibition based on task context does not give an adequate account of n-2 repetition costs (but see chapter 10 for additional discussion).

Additionally, Koch et al. (2010) argued that if task inhibition was self-directed, n-2 repetition costs should not be sensitive to characteristics of the n-1 trial. For example, no-go trials (Schuch & Koch, 2003) or go-signal trials (Philipp, Jolicoeur, et al., 2007) should not affect n-2 repetition costs. Once again, Grange et al. (2013) argued that in task contexts where no-go trials were a possibility, self-inhibition of tasks would not be an adaptive strategy, as performance would be impaired specifically on trials where the n-1 task was a no-go trial (i.e., one experimental condition in Schuch & Koch, 2003, for which no n-2 repetition costs were found). However, if this is the case, it does not explain why n-2 repetition costs *were* found in sequences in which the n-1 task was *not* a no-go trial. One possibility not considered by Grange et al. (2013), however, is that self-inhibition of task-sets may only be triggered by response execution. This would be compatible with the experiments of Schuch and Koch (2003) and Philipp, Jolicoeur, et al. (2007), which found that response execution was required for the triggering of task inhibition.

Nevertheless, overall self-inhibition does not give a satisfactory account of the range of behavioural findings on the n-2 repetition cost. Furthermore, suggesting that self-inhibition might be strategically modulated, although plausible, is less parsimonious than alternative explanations. Such accounts do not explain on what basis self-inhibition is itself modulated, or what this mechanism might be, thus reintroducing an homunculus into an overall account of task control.

3.3.2 Lateral inhibition

Various authors have proposed lateral inhibition as a mechanism that implements task inhibition (e.g., Koch et al., 2010; Mayr & Keele, 2000; Philipp & Koch, 2006). However, Grange et al. (2013) argued that lateral inhibition is an insufficient mechanism. Specifically, they suggested that according to a lateral inhibition account, the final task A should be inhibited *more* in the sequence CBA than in the sequence ABA, as two of its competitor tasks are active. To support this argument, they produced a simple simulation demonstrating activation levels in a simple connectionist network architecture with lateral inhibition.

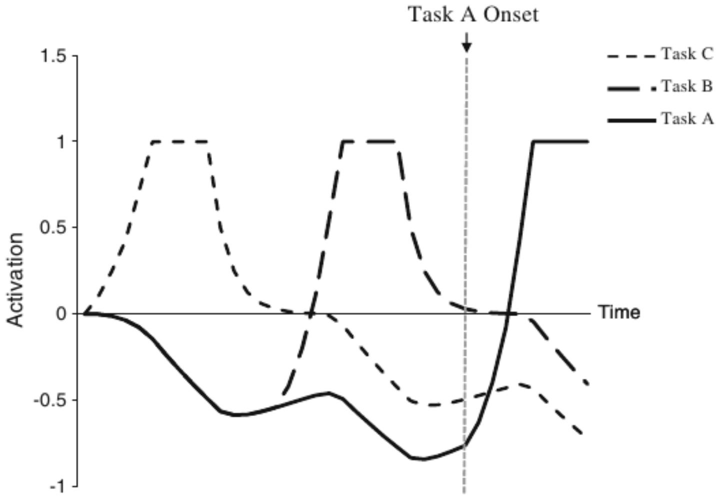


Figure 3.2: Activation in a simple connectionist model with lateral inhibition, as simulated by Grange et al. (2013). According to this simulation, n-2 repetition costs should be greater in sequences where a task has not been repeated for two or more trials (CBA), than a single trial (ABA). Reproduced with permission from Grange et al. (2013).

While this argument is plausible, their simulation Grange et al. (2013) calculated activation merely as the sum of a unit's inputs hard clipped to a value between -1 and 1 (see fig. 3.2). However, this does not accurately reflect the way lateral inhibition is actually implemented in most connectionist networks (e.g., Gilbert & Shallice, 2002; cf. McClelland, 1993). In such networks, a unit's activation is usually calculated as some function of the sum of its inputs. Activation functions usually take the form of a logistic (sigmoid) function, which ‘squashes’ activation such that a difference in one input to a unit has a much larger effect on its overall activation when the input to the other units is small, than when it is large. It has been demonstrated that the use of such a function produces radically different overall behaviour from a network without a squashing function (McClelland, 1993)[7]. In the current instance, lateral inhibition may well produce much smaller differences between inhibited units than the network of Grange et al. (2013). Thus, while their argument that lateral inhibition alone is insufficient to cause n-2 repetition costs is plausible, a more complete simulation using a network implementing lateral inhibition and a squashing activation function would further support their argument.

3.3.3 Inhibition triggered by conflict

Koch et al. (2010) argued that the empirical phenomena attributed to task-set inhibition, including sensitivity to intertrial intervals between trials n-2 and n-1 (discussed in section 3.2.8), dependencies on response selection/execution processes in the preceding trial (section 3.2.7), and manipulations of response set overlap (section 3.2.7), can all be parsimoniously explained in terms of a general task-set inhibition mechanism. According to this view, task-set inhibition is a somewhat general mechanism, which can be targeted at different aspects of the task-set (e.g., stimulus dimensions or response processing: Koch et al., 2010) or even more generally, ‘wherever the locus of conflict appears’ including cue-encoding processes (Grange et al., 2017; Houghton et al., 2009; Scheil, 2016).

It has been proposed that the processing of conflict occurs by way of a mechanism similar to that of Botvinick et al. (2001), proposed to explain conflict monitoring effects in a single-task context. This mechanism dynamically inhibits any task-relevant representation (e.g., competing response alternatives, or task-sets)

---

[7] Indeed, in the model of Gilbert and Shallice (2002), this property is important to the simulation of a number of empirical effects, cf. section 2.3.2.

that generates conflict during task-relevant processing. Koch et al. (2010, p. 11) propose that this explains the role of response selection as a trigger for task inhibition (Schuch & Koch, 2003). In the models of Botvinick et al., conflict is typically monitored at the response level (e.g., Botvinick et al., 2001; Botvinick, Cohen, & Carter, 2004). In these implementations, conflict is defined as simultaneous activation of competing response alternatives, and is calculated as the product of the activation of response units. Koch et al. (2010) argue that the current empirical evidence most strongly supports a critical role for conflict at the response level for the inhibition of task sets. Some data (e.g., Philipp, Jolicoeur, et al., 2007; Schuch & Koch, 2003) implicates the role of response selection and execution processes in the recruitment of task inhibition. However, a mechanistic account of these processes has not been specified. Two possibilities have been suggested (Koch et al., 2010). Inhibition may be triggered by post-response evaluation (Botvinick et al., 2004). Alternatively, performing the response could provide performance feedback, driving reinforcement learning (Holroyd & Coles, 2002) which assures a positive bias toward the recently performed task, thus indirectly inhibiting the the competing task.

Alternatively, task-set inhibition has been characterised as a flexible process which can selectively be directed at different elements of the task-set depending on which elements cause conflict in a given task context. Yet such characterisations leave many algorithmic details unspecified. One possibility is that there are multiple parallel conflict monitoring/task-set inhibition loops each sensitive to different elements of the task-set. Another possibility is that there is a single task-set inhibition mechanism which is triggered by detection of conflict at multiple levels.

Computational models of conflict monitoring mechanisms have, to date, been somewhat ad-hoc, with different models bearing a family resemblence rather than any consensus emerging on a single reference implementation (e.g., Blais, Robidoux, Risko, & Besner, 2007; Botvinick et al., 2001; Botvinick, 2007; Yeung, Botvinick,

& Cohen, 2004). While there have been steps made toward integrating conflict monitoring and task switching, both empirically (e.g., evidence that committing errors leads to strengthening of the wrong task: Steinhauser & Hübner, 2006, 2008) and theoretically (e.g., the task-switching model of Brown et al., 2007), to date the task-switching and conflict monitoring paradigms remain somewhat separate.

3.3.4 The computational models of Grange and colleagues

Altmann and Gray (2008) present a model of cognitive control in which task-switching is conceptualised as a serial memory retrieval process which proceeds according to the following stages. Firstly, on presentation of a task cue, the cue is used to encode a goal representation. Next, the goal representation is used to retrieve a representation of the target from memory. Finally, the target representation is used as the basis for a visual search for a matching target in the stimulus display. Thus, the model conceptualises task switching as goal-driven retrieval of task-relevant representations from memory (see Mayr & Keele, 2000, for a similar proposal). Memory representations (‘chunks’) have an associated activation value, which determines the time taken to retrieve the chunk from memory. The activation of a chunk is calculated according to a standard ACT-R equation (the base-level learning equation, BLL (Anderson, 2007), by which the activation of a chunk is increased when it is accessed, after which it decays according to a power law. More recently accessed memory representations are therefore quicker to retrieve from memory, and this is indeed the case for task repeats, compared to task switches.

Given the absence of inhibitory processes within the ACT-R architecture, Grange et al. (2013) argued that the model of Altmann and Gray (2008) cannot account for n-2 repetition costs, and would instead predict n-2 repetition facilitation, for similar reasons as for n-1 repeats. Grange et al. (2013) therefore implemented a model within ACT-R by modifying the BLL equation to include an early, inhibitory component (i.e., self-inhibition) which is large immediately after chunk retrieval, making chunks harder to retrieve immediately, but dissipates passively over time, leading to an inverted-U shaped activation curve (see fig. 3.3). In this model, then, the locus of the n-2 repetition cost is the process of retrieving target representations from memory — for a certain period this process is slower for more recently retrieved target representations, as in the n-2 repetition condition.

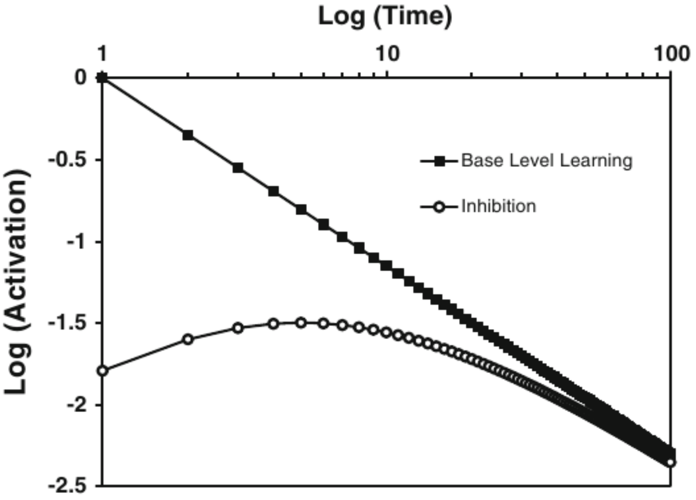


Figure 3.3: Inhibitory and default versions of the ACT-R Base Level Learning Equation, plotted in log-log space, hence BLL decay is linear. Reproduced with permission from Grange et al. (2013).

This conceptualisation of task-set inhibition is based on a number of implicit assumptions. Firstly, it is assumed that inhibition decays passively over time. While the rate is controlled by an inhibition-decay parameter, this is fixed for a particular simulation. However, this assumption is challenged by empirical evidence that suggests inhibition does not passively decay but is persistent (e.g., Gade & Koch, 2005, reviewed in section 3.2.8), and is incompatible with theories proposing that the release of inhibition is event-based (Koch et al., 2010).

Secondly, it is assumed that task-sets are self-inhibitory. Indeed, this is the strong theoretical claim of Grange et al. (2013). This claim is discussed in detail in section 3.3.1. In short, self-inhibitory models of task inhibition fail to give an adequate account of a number of robust empirical findings in the n-2 repetition paradigm. It is therefore doubtful that such a model would generalise to a more complete model, simulating a wider range of empirical effects. While modelling task inhibition as a form of self-inhibition provides a good fit to the data, whether it represents a viable theoretical proposal remains unresolved (see Grange et al., 2013; Koch et al., 2010, for a full range of arguments).

Through the use of simulations, Grange et al. (2013) demonstrate that their model cannot produce n-2 inhibition effects in the absence of inhibitory mechanisms. Furthermore, while the absence of n-2 inhibition effects have, in the past, been interpreted as the absence of inhibition (see Grange et al., 2013, for a review), the absence of inhibitory mechanisms in their model actually predicts an opposite effect, an n-2 repetition benefit, rather than a cost. Their model also made additional behavioural predictions: Specifically, in conditions with reduced n-2 repetition costs, RTs should be generally faster. This prediction is partially supported by the data (e.g., Arbuthnott, 2008b; Schneider & Verbruggen, 2008, but see Arbuthnott, 2005). Thus, the absence of n-2 repetition costs was argued to imply a reduced level of inhibition, rather than its absence.

Using their previous model, Grange and Juvina (2015) predicted that n-2 repetition costs should be specific to relatively novel tasks: when tasks become relatively automatised after a large amount of practice, n-2 repetition costs should be reduced or eliminated, due to the increase in the base-level activation with learning for each task representation, making well-learned tasks easier to retrieve. In an experiment with 10 participants completing 6100 trials over five days, the average n-2 repetition cost indeed significantly decreased with practice. While the model of Grange et al. (2013) predicted a reduction in costs slower than was observed empirically, the empirical data could be fit with the additional assumption that the cue-target association also becomes stronger as a result of practice.

Despite these successes, the model has a number of limitations. Grange and colleagues themselves concede that a limitation of their model is that a single version cannot simultaneously account for both empirically observed n-1 and n-2 effects of task switching; the model with the inhibitory version of the BLL would predict n-1 repetition costs rather than switch costs. Dealing with this issue, Grange et al. (2013) argued that the task self-inhibition may be strategically recruited by the human participant in three-task switching paradigms, as it provides an advantage in situations where task repeats are not possible. This view has some empirical support. While some studies have shown that n-2 repetition costs occur when n-1 repeats are possible (Arbuthnott, 2005), others have shown that the n-2 repetition cost is significantly reduced in such circumstances (e.g., Philipp & Koch, 2006). Yet other studies suggest that consistent n-1 switch and n-2 repetition costs tend not to occur in the same block of trials (Philipp & Koch, 2006, for review). While this is therefore an intriguing suggestion, the cognitive system responsible for such strategic (re)configuration remains to be specified in theoretical terms.

Additionally, the model does not easily accommodate the range of behavioural findings which demonstrate that the n-2 repetition cost is modulated by manipulations of conflict generated on the n-1 trial.

The model predictions and the assumption that practice strengthens cue-task associations received further recent support from a study by Scheil (2016). Two experiments were conducted to determine whether the reduction in the n-2 repetition cost was due to practice on the paradigm overall, or whether the practice was specific to a particular task. In each experiment, 24 participants completed 10 blocks, each of 120 trials. In the first experiment, the stimulus-response associations of one of the three tasks was reversed after six blocks. Similar to the overall findings of Grange and Juvina (2015), n-2 repetition costs declined over the course of 12 blocks, however while basic RTs were affected by the switch of stimulus-response associations, n-2 repetition costs were not. In the second experiment, the cue-task association changed after six blocks. For three blocks following the change, n-2 repetition costs increased significantly for the changed task, while n-2 repetition costs were much reduced for the unchanged tasks. For the final three blocks, n-2 repetition costs were unchanged compared to those prior to the change. Scheil (2016) interpreted her results as consistent with the model of Grange and Juvina (2015), but not Grange et al. (2013). That is, changing the cue-task association could be implemented in the model as a decrease in base-level-learning, and also a reset of the cue-task association parameter, which in turn would lead to an increase in influence of the inhibitory component, temporarily increasing n-2 repetition costs before cue-task associations and with base-level-learning activation once again increasing with practice. However, changing the stimulus-response mappings in experiment 1 would also be expected to affect base-level learning activation, hence the difference between the two experiments could only be modelled with the additional assumption in the model of Grange and Juvina (2015) — that cue-task associations become stronger over time.

3.4 Summary

N-2 repetition costs are a widely replicated empirical phenomenon, thought to reflect the inhibition of task-sets (i.e., backward inhibition). They represent a definitive instance of an inhibitory cognitive phenomenon. Unlike the switch cost, initially thought to represent evidence of a cognitive reconfiguration process but subsequently partitioned into lower-level effects (e.g., interference, cue switching), the n-2 repetition cost has to-date proven resistant to non-inhibitory explanations.

Nevertheless, substantial theoretical questions remain about the n-2 repetition cost. What is inhibited? Is inhibition endogenously or exogenously triggered? What is the mechanism for inhibition? The strongest empirical evidence links backward inhibition strongly to both response-based processes and cue-based processes. This division may reflect two different behavioural procedures which emphasise, respectively, response-based or cue-based difficulty for the participant.

The most complete theory is that of Koch and colleagues, suggesting that backward inhibition is triggered by the conflict between task representations that occurs on the n-1 trial; and this may occur at multiple locations including cue-encoding and response-generation processes.

Despite obvious theoretical interest in the phenomenon, computational accounts have, to-date, been scarce. The models of Grange and colleagues explain backward inhibition in terms of self-inhibition of task representations at the cue-task encoding level, and these models have had successfully predicted the effect of extensive practice. However, while this model was developed from a two-task switching model, its modifications mean that it does not generalise to two-task switching — in a two-task context, the backward inhibition model predicts switch facilitation, rather than switch costs.

Clearly, for human participants, both switch costs and n-2 repetition costs are robust behavioural effects, hence a complete model of human task switching should be able to capture both effects with external modification of its mechanisms or parameters. However, this remains an open empirical question, with some suggestion that in studies which permit both measures within the same block, switch costs and n-2 repetition costs may not co-occur or may be significantly reduced, with the suggestion that participants may employ some form of strategic adaptation to their current task context, leading to switch costs when task repeats occur (such as in the two-task paradigms) or when they do not (typically, in the three-task paradigms). If this is the case, the computational mechanisms driving such adaptation are of theoretical interest and have yet to be specified.

The remainder of this thesis develops a computational model of generalised task switching effects, applicable to both two or three-task paradigms. The following chapter begins this development by generalising the two-task switching model of Gilbert and Shallice (2002) to three tasks, and exploring the model's behaviour in the n-2 paradigm. Critically, this model does not incorporate any inhibitory processes beyond lateral inhibition. The work therefore explores from a computational perspective whether a specific inhibitory mechanism is indeed required to account for the empirical effects observed in three-task switching studies.

Chapter 4

A Simple Model of Three-Task Switching

This chapter moves toward a model of backward inhibition by adapting the two-task switching model of Gilbert and Shallice (2002) to the three-task n-2 repetition paradigm. It is demonstrated (Simulations 3 and 4) that the mechanisms of the adapted model (i.e., lateral inhibition) are not sufficient to produce n-2 repetition costs, but instead yield n-2 repetition facilitation. This facilitation is not affected by stimulus congruency effects (Simulation 3) or asymmetric task difficulty (Simulation 4). These simulations provide a baseline against which to compare the performance of a backward inhibition model.

4.1 Introduction

As argued in the previous chapter, the n-2 repetition paradigm has been the most theoretically productive area within the task switching paradigm in terms of evidencing specific cognitive processes for the control or regulation of task sets, specifically, backward inhibition. To-date, behavioural evidence suggests that task-set inhibition may be flexibly directed toward relevant aspects of task-set. It has been proposed that this is achieved on the basis of conflict (Koch et al., 2010). However, this idea has not been formally specified in a computational model. Additionally, some theoretical predictions of this hypothesis (e.g., congruence effects: Goschke, 2000; response repetition effects: Rogers & Monsell, 1995, Meiran et al., 2000, Koch et al., 2010, p. 10) have not yet been empirically tested. Additionally, some theoretical issues remain unresolved by the behavioural evidence. Firstly, it is unclear whether task-set inhibition is a top-down (endogenous) mechanism, or a more automatic process, with behavioural evidence remaining somewhat equivocal (Koch et al., 2010). Secondly, dissipation of inhibition is a key assumption of the n-2 repetition paradigm — yet this issue remains unresolved. Some evidence (Vandierendonck, 2013) suggests there may be problems with time-based decay of inhibition. Koch et al. (2010) suggested event-related release of inhibition, but this hypothesis remains theoretically under-specified. Finally, there is debate over whether backward inhibition takes place at the cue-encoding/retrieval stage, as envisaged by Mayr and Keele (2000), the response selection stage, or even later in the response execution stage (Philipp, Jolicoeur, et al., 2007). Alternatively, if backward inhibition is flexible, as suggested by Koch et al. (2010), it may be directed at any or all of these stages of processing. However, this proposal remains vague and it is difficult to see how it could easily be implemented computationally.

Of the existing models of task switching, none cover all stages of the response process. For example, the ACT-R model of Altmann and Gray (2008) provides interesting insights in cue encoding and the retrieval of task rules, in a similar manner to the verbal model of Mayr and Keele (2000), but does not model response processes. Therefore, accounting for the range of empirically effects implicating response processes within this approach remains problematic. Related to this model, the models of Grange et al. (2013) and Grange and Juvina (2015) simulate certain n-2 repetition cost effects within the four-choice visual search paradigm in terms of cue-driven memory retrieval (cf. Mayr & Keele, 2000). However, these models require modifying the base-level learning equation governing the activation of items in working memory, which is core to the wider ACT-R architecture (Anderson, 2007). This modification incorporates an early inhibitory component to the activation of recently accessed items — a form of self-inhibition, which is not easily reconciled with empirical evidence that n-2 repetition costs are not a self-inhibitory phenomenon. The model of Brown et al. (2007) incorporates sequential congruency and conflict monitoring mechanisms, and is the closest to a complete end-to-end model of task switching. However the complexity of the model is a disadvantage. Due to the number of mechanisms present, its operation is somewhat opaque, and it is unclear which mechanisms are responsible for its success. The model of Gilbert and Shallice (2002) is somewhat abstract, and does not model cue encoding or response generation processes. However, it is presented within the GRAIN framework (McClelland, 1993), which has productively generalised to a range of executive phenomena (cf. Cohen & Huston, 1994) and is thus compatible with other related models, such as the conflict monitoring model of Botvinick et al. (2001). Additionally, the close relation of these models suggests corresponding computational similarities between the concepts of task switching and conflict monitoring, which are likely to impinge on any conflict-based model of task switching phenomena. The simplicity of these models may prove a virtue, in allowing easy modification to incorporate additional theoretical mechanisms, and allowing easy comparison. Finally, the model inherits additional theoretical constraints from the GRAIN framework, in an analogous way to the inheritance of constraints by models developed within a cognitive architecture (cf. Newell, 1990).

In basing a model of a backward inhibition mechanism on the task switching model of Gilbert and Shallice (2002), the first objective is to adapt the existing model of switching between two tasks to the paradigm of three task switching. While this change is relatively trivial in cognitive terms, it is important to establish the behaviour of the existing model in the new paradigm, in order to properly assess the effects of any additional mechanisms. For example, Grange et al. (2013) argued that lateral inhibition alone was insufficient to produce lasting inhibition effects such as n-2 repetition costs. While this argument was supported by a simulation of lateral inhibition, that simulation differed from lateral inhibition as implemented in most complete interactive activation models, in that it did not incorporate a non-linear activation (squashing) function (cf. section 3.3.2), which is an important feature in the simulation of certain empirically observed behaviors (e.g., 'Reverse' Stroop interference, or larger switch costs for the dominant task when switching between tasks of asymmetric difficulty, cf. section 2.3.2). This was because the harder task produced greater activation of the irrelevant task demand unit. At the end of a colour naming trial, activation in both task demand units was higher than following a word reading trial, which was carried over (and interfered with) the subsequent trial. This effect is one potential mechanism by which sequential RT effects may be produced from more conflicting trials, even with no specific conflict monitoring mechanisms.

Although the four-choice visual detection paradigm of Mayr and Keele (2000) has has been popularly used in backward inhibition studies (such as by Grange and colleagues, e.g., Grange & Houghton, 2009, Grange & Houghton, 2010a, Grange & Juvina, 2015), a majority of studies of the n-2 repetition cost have used the symbolic classification paradigm (e.g., Arbuthnott & Frank, 2000; Arbuthnott & Woodward, 2002; Gade & Koch, 2005, 2007, 2008; Hübner et al., 2003; Kuhns et al., 2007; Philipp & Koch, 2006; Schuch & Koch, 2003; Sdoia & Ferlazzo, 2008). Symbolic classification also represents the paradigm modelled by Gilbert and Shallice (2002), of which the colour naming and word reading tasks they simulated, represent a specific case. Therefore, the present model retains the cued-task two-choice reaction time format.

4.2 The lateral-inhibition-only model: a 3-task variant of the Gilbert and Shallice (2002) model

We have already observed that the interactive activation model of Gilbert and Shallice (2002) produces behaviour that is difficult to anticipate purely from verbal descriptions of a phenomenon. For these reasons, fully establishing behaviour of the Gilbert and Shallice (2002) model in the three-task paradigm through simulation, is an important first step.

The adapted model architecture is illustrated in figure 4.1. The model is based closely on the Gilbert and Shallice model of 2-task switching, for which the successful re-implementation was described in section 2.6. That model was modified to include three task pathways, representing arbitrary classification tasks, here designated A, B and C. Following the original model, lateral inhibitory connections are also present at output and task demand level, with excitatory lateral connections to congruent units at the response level, and inhibitory connections to incongruent units at both the response and task demand levels (figure 4.2)

4.2.1 Input/output units

The three-task switching model processes each stimulus dimension (A, B or C) according to a different pathway. These three pathways are each mapped to a pair of output units (0, 1), which are assumed to reflect two response keys (e.g., L and R). Incongruent output units are mutually inhibitory, such that activating the 0 response according to dimension A inhibits the 1 response corresponding with dimensions A, B and C. Thus, the digit ‘4’ (i.e., even, smaller, central) would be encoded in the model as 1 on dimension A (odd vs. even), 0 on dimension B (smaller vs. larger) and 0 on dimension C (central vs. peripheral).

Input units have feed-forward connections to the corresponding output units, of which there are six, representing two possible responses to each of the three

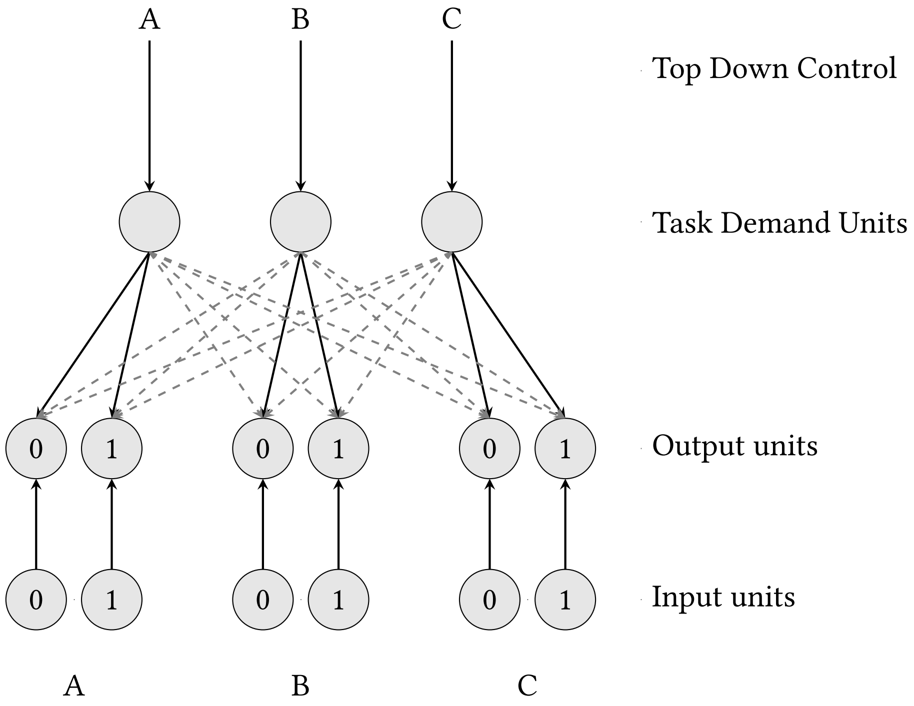


Figure 4.1: 3-task switching variant of the model of Gilbert and Shallice (2002), simulating three arbitrary classification tasks (A, B, C). Each task has two possible responses (0, 1). Solid black arrows represent excitatory connections, while grey broken arrows represent inhibitory connections. Not shown are lateral inhibitory connections, which are illustrated in figure 4.2.

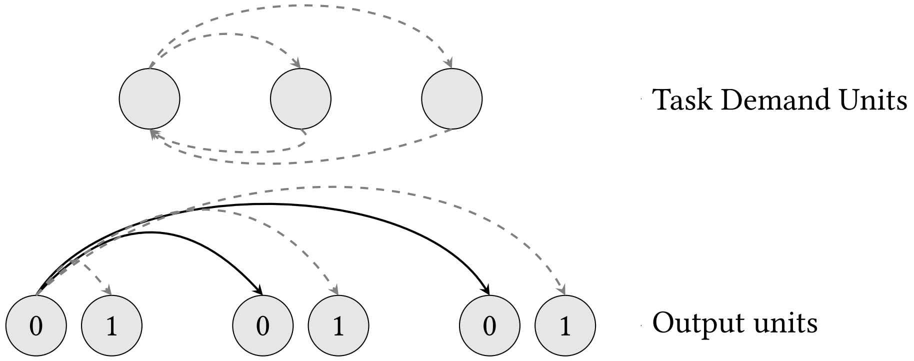


Figure 4.2: 3-task switching version of the Gilbert and Shallice model, showing detail of lateral inhibitory connections between units at output and task demand level. For simplicity, only the connections from the first unit in each layer are shown. For output units, lateral connections to corresponding units are excitatory (black solid arrows), and inhibitory (grey, dashed arrows) to non-corresponding units. Task demand lateral connections are all mutually inhibitory.

tasks. Within the set of output units, units that correspond to the same response are mutually excitatory, while units that correspond to alternative responses are mutually inhibitory (figure 4.2). The model therefore implements an experimental procedure in which responses for all tasks are mapped to the same set of response keys (cf. Gade & Koch, 2007).

4.2.2 Task demand units

As in the model of Gilbert and Shallice (2002), task processing is biased by task-demand units — one per task — which have excitatory (i.e., positively weighted) connections to their respective output units and inhibitory (i.e., negatively weighted) connections to the output units associated with the other tasks. These connections are bi-directional, so a response activated bottom-up by a strong input connection will tend to activate the task with which the response is associated. The currently relevant task unit also receives a 'top-down control' input, simulating the level of deliberate control required to perform each task to a reasonable level.

As a simplification, the model omits the modifiable connections from input to task demand units, which Gilbert and Shallice (2002) used to simulate the influence of item-specific priming.

4.2.3 Activation calculation

The equation for calculating unit activation is unchanged from the model of Gilbert and Shallice (2002). For each unit i, the change in activation value on each model cycle, $\Delta\alpha_i$, is calculated as follows:

$$\Delta\alpha_i = \begin{cases} \sigma I_i(\alpha_{max} - \alpha_i) + \mu & \text{if } I_i \geq 0 \\ \sigma I_i(\alpha_i - \alpha_{min}) + \mu & \text{if } I_i < 0 \end{cases}$$

where α_i is the unit's current activation, I_i is its net input, and σ, α_{max} and α_{min} are parameters affecting step size (0.0015 for all simulations reported here), and maximum (1.0) and minimum (-1.0) unit activation values respectively. μ is a noise term, drawn from a Gaussian distribution, with a mean of 0 and standard deviation of 0.006.[1]

4.2.4 Processing within the model

On the first trial of each block, all units are initialised with zero activation. On subsequent trials, Task Demand (TD) units, which carry over 20% of their activation from the final step of the previous trial (as in the original model of Gilbert & Shallice, 2002), and conflict units, which carry over a certain proportion of their activation (set to 50% for all simulations), model the effects of residual task inhibition. All other units are initialised as for the first trial.

In a simulated trial, one input unit in each task pathway (representing a trivalent stimulus), and a top-down control unit (representing the currently cued task) are set to 1. Activation then iteratively propagates throughout the model. As response units suppress their non-congruent competitors via lateral inhibition, they exhibit a 'winner-takes-all' property, in which small differences in activation between the 'winning' and 'losing' units become decisive as the losing units are actively suppressed. A response is therefore taken as the most active output unit, when its activation exceeds that of the next most active, non-congruent output unit by a response threshold of 0.15. The number of cycles taken for this to occur is the simulated response time (RT).

4.2.5 Parameter settings

Where possible, parameter settings were kept identical to those used by Gilbert and Shallice (2002), to maximise implementational consistency with that model.

---

[1] These parameter values were the defaults used by Gilbert and Shallice (2002).

However, due to the presence of three task pathways, some changes were made to stimulus input strength and top-down-control strength parameters to stabilise the overall level of activation in the model. In the following simulations, for symmetric tasks, a stimulus input strength of 2.0 and top-down control strength of 14.0 was used for all tasks. Once again, trials ran for a maximum of 500 cycles.

4.3 Simulation 3: Response congruency in the symbolic classification paradigm

Primarily, this simulation aimed to determine the predictions of the lateral inhibition-only model on n-2 repetition costs in a 3-task switching paradigm. While it has been reasoned that lateral inhibition alone is an unlikely source for the n-2 repetition cost (Grange et al., 2013), as argued above it remains important to demonstrate clearly the behaviour of the model with the lateral inhibitory connections that are a standard part of the architecture of IAC models (Cohen & Huston, 1994). This simulation additionally provided a baseline for understanding the behaviour of any additional mechanisms.

Specifying the task sequence performed by the model necessitates specifying the congruence or incongruence of the irrelevant stimulus dimensions. Given three stimulus dimensions, for any given stimulus, one dimension would correspond to the currently relevant task, with the other two being irrelevant. Those two irrelevant dimensions could: both be congruent with the correct response to the currently relevant task (fully congruent, or CC); be congruent and incongruent (semi-congruent, IC); or both be incongruent (fully-incongruent, or II). In the two-task switching literature, the stimulus congruency has been shown to affect response times on the current trial (Meiran & Kessler, 2008; Rogers & Monsell, 1995). Additionally, response times are affected by stimulus congruence on the preceding trial, known as the n-1 congruence effect (Brown et al., 2007; Goschke, 2000; Monsell et al., 2003). This has been explained in terms of an increased level of response conflict on the previous trial, where conflict is linked to the on-line recruitment of cognitive control. This mechanism has been simulated in the model of Brown et al. (2007). A secondary aim of this simulation, therefore, is to determine the predictions of the lateral-inhibition-only model for the effect of response congruency on reaction times, n-1 switch costs, and n-2 repetition costs.

In the lateral inhibition model, response conflict evoked by stimulus dimension incongruence is simply operationalised as the simultaneous activation of competing target stimuli, each of which passively activates a task-demand node through spreading activation. Unlike the model of Brown et al. (2007), there is no specific conflict detection mechanism in this model. Any sequential effects of stimulus congruence on response times are therefore purely due to the carryover of task-demand activation between subsequent trials.

4.3.1 Procedure

The procedure was simulated as a series of runs of three trials. Each run of trials could be an n-2 repeat, (i.e., ABA) or an n-2 switch (i.e., CBA). As the task strengths used in this simulation were identical (i.e., symmetrical tasks), no permutations of tasks (e.g., BAB, CAB) were considered.

The congruence or incongruence of the irrelevant stimulus dimensions with the correct response was considered when generating the task sequence performed by the model. Irrelevant dimensions could possibly be congruent/congruent, incongruent/congruent or incongruent/incongruent. Even without considering fully congruent trials (i.e., congruent/congruent), each of the three trials in a block could be either semi-incongruent (IC) or fully incongruent (II), giving a total of eight congruency conditions for each type of run[2], which were fully counterbalanced. Selection of stimuli and tasks was randomised for each task. Item-specific priming

---

[2] IC IC IC, IC IC II, IC II IC, II IC IC, II II IC, II IC II, IC II II, II II II.

effects (i.e., the direct connections from input to task-demand units, with limited Hebbian learning, as in the model of Gilbert and Shallice (2002), were disabled for the purposes of this simulation. For each sequence type (ABA, CBA), each stimulus congruency combination (II/II/II, II/II/IC, II/IC/IC etc.) was run 500 times, for a total of 8,000 three-task blocks, or 24,000 trials.

4.3.2 Results

When collapsed across sequence congruency (figure 4.3), the effect of sequence position was highly significant, $F(1, 23237) = 63.71,\ p < .001$. Sequence alone was not statistically significant, $F(1, 23237) = 1.09,\ p = .297$, suggesting that overall, ABA sequences did not differ significantly from CBA sequences. The n-2 repetition cost is assessed specifically looking at the RT on the third trial, and is suggested by the interaction between sequence type and trial position. As hypothesized, the lateral inhibition-only model did not produce an n-2 repetition cost, with comparison of means suggesting a slight n-2 repetition facilitation, although this effect was not statistically significant, $F(1, 23237) = 1.96,\ p = .161$. As the significance of the trial position × sequence type interaction is affected by the first two trials, which would not be expected to differ in RTs, the n-2 repetition cost (i.e., difference in RT on the third trial alone) was also assessed with an independent samples t-test on RT_3. This was significant, $t(7732) = 2.39$, $p = .017$, with n-2 switch trials being slower (mean of 82.10 cycles) than n-2 repeats (80.03 cycles).

Next, we consider differences in the effects of stimulus congruency on reaction times. This analysis was conducted as a 4-way ANOVA ($Sequence \times Congruency_1 \times Congruency_2 \times Congruency_3$) on the reaction time on trial 3, RT_3 (figure 4.4). Firstly, we consider the main effects. The main effect of sequence was significant, $F(1, 7719) = 6.10,\ p = .014$, suggesting that residual activation from preceding trials did affect RTs on the final trial (i.e., an n-2 effect). The main effects of trial congruency were not significant for $Congruency_1$, $F(1, 7719) = 0.16,\ p =$

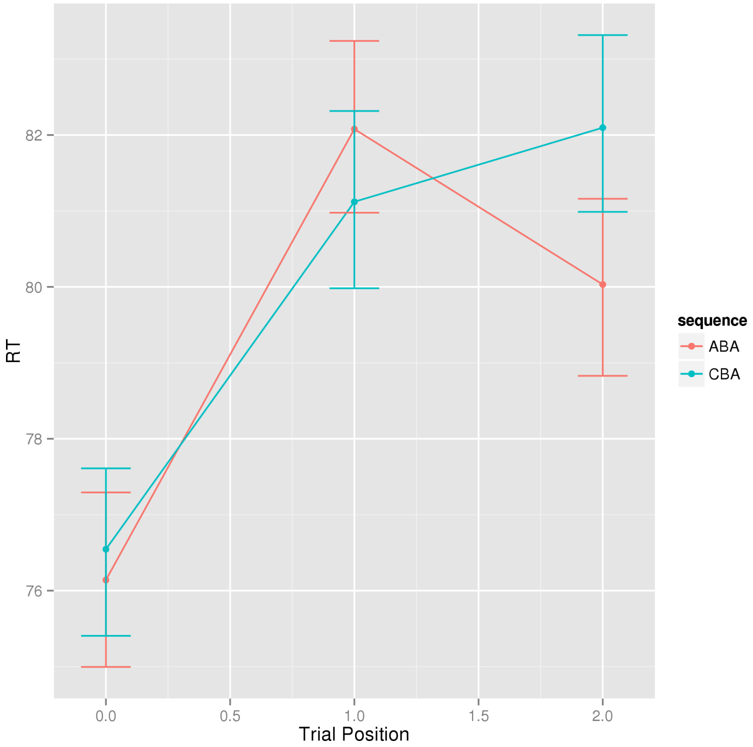


Figure 4.3: Simulation 3: Basic simulation of the n-2 repetition procedure. When collapsed across stimulus congruency, the model produces a small n-2 facilitation effect, not n-2 costs as in the empirical literature. Error bars represent 95% confidence intervals

.68; marginally significant for $Congruency_2$, $F(1, 7719) = 3.82,\ p = .051$, and highly significant for the current trial, $Congruency_3$, $F(1, 7719) = 505.21,\ p < .001$. Perhaps surprisingly, none of the interactions were significant ($F(1, 7719) < 1.21,\ p = .272$) except the marginally significant $Congruency_1 \times Congruency_2$ interaction, $F(1, 7719) = 3.27,\ p = .070$.

4.3.3 Discussion

The results confirm that while $Sequence$ has no statistically significant effect on trial 3 RT (i.e., no reliable n-2 repetition cost or facilitation), when the error was partitioned by including the three stimulus congruency factors in a 4-way ANOVA, the effect of n-2 repetition became significant, in this case an n-2 repetition facilitation rather than a cost. This finding confirms recent arguments that in the absence of backward inhibition, recent performance of a particular task (i.e., of the A task

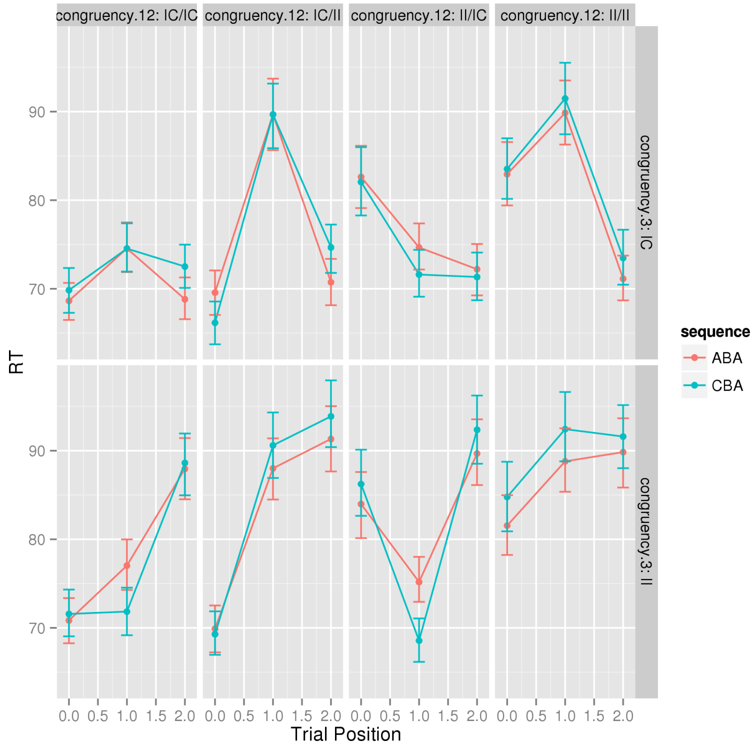


Figure 4.4: Simulation 3: Basic simulation of the n-2 repetition procedure. Sequential congruency effects on trial 3 RT. The $Congruency_3$ factor is represented across upper and lower panels, the combination of $Congruency_1$ and $Congruency_2$ is represented across horizontal panels.

in ABA sequences) would lead to reduced RTs, as a result of residual activation of the task — effectively, a form of priming (Grange et al., 2013).

The marginally significant main effect of $Congruency_2$ suggests the model does predict an n-1 congruency effect, as found empirically by Goschke (2000) and Monsell et al. (2003). In the absence of a specific incongruency detection mechanism, this effect is purely due to the residual activation in the model caused by an incongruent trial. However, given the large number of trials in the simulation, and the resulting marginally statistically significant result, it is unlikely that this simulated effect would be detected empirically, and so the mechanism within the model is correspondingly unlikely to fully account for the observed behaviour.

The lack of higher-order interaction effects, especially interactions with $Sequence$ suggests that while the congruency of trial 3 (and to a lesser extent trial 2) affected final trial RTs, it did not affect n-2 repetition costs. Similarly, the lateral inhibition-only model does not predict higher-order sequential congruency effects on the n-2 repetition cost.

4.4 Simulation 4: Asymmetric tasks in the n-2 repetition paradigm

Most empirical three-task switching studies have used tasks of equal, or approximately equal difficulty. However, equal task difficulty is frequently an assumption: to-date, most empirical studies have not compared the tasks in terms of systematic RT or error rate differences. In one of the few studies that examined asymmetric task difficulty systematically, Arbuthnott (2008a) combined three tasks of varying difficulty. All three tasks involved classifying certain single digits (2, 3, 4, 6, 7, 9). The easiest, magnitude task, involved stating whether a digit was high (6, 7, 9) or low (2, 3, 4). The next most difficult task, parity, involved classifying whether a digit was odd (3, 7, 9) or even (2, 4, 6). Finally, the prime task, on which participants were slowest and most error prone, involved classifying whether a digit was prime (2, 3, 7) or non-prime (4, 6, 9).[3] Arbuthnott derived a directional hypothesis for asymmetric n-2 repetition costs, reasoning from a verbal model of backward inhibition that contrasted a residual-activation based explanation with a residual-inhibition based explanation. She reasoned that if the n-2 repetition cost was primarily due to residual inhibition (as compared to residual activation), conditions with a larger switch cost (i.e., Easy-Hard-Easy sequences) should also exhibit larger n-2 repetition costs than those with a smaller switch cost (i.e., Hard-Easy-Hard). Alternatively, if the asymmetric switch cost was largely due to residual activation, n-2 repetition costs should be equal across conditions.

---

[3] See section 8.1.1 for criticism of this paradigm.

4.4.1 Target data

Arbuthnott (2008a) conducted two experiments. The first used non-overlapping sets of response keys for the three tasks, enabling differentiation between errors due to the participant performing the incorrect task, and those due to the correct task but the incorrect response. The second used overlapping sets of response keys.[4] Experiment 2, therefore, corresponds with the current model, which implements overlapping response sets in terms of lateral inhibition/excitation at the output/response level (recall figure 4.2).

Arbuthnott assessed relative task difficulties by comparing RTs and error rates (wrong-task errors and decision-errors) in the no-switch condition. In experiment 1, these were as follows: Magnitude task: 663ms, 0.33%, 2.15%; Parity task: 764ms, 1.44%, 2.21%; Prime task: 821ms, 0.38%, 3.93%. For magnitude and prime switches (i.e., easiest and hardest tasks), the direction of the typical switch cost asymmetry was reversed, with greater costs for switches toward the harder, prime task. N-2 repetition costs were significantly different for the two alternation directions, with greater costs for the magnitude task (i.e., EHE) and a slight n-2 repetition facilitation for the prime task (i.e., HEH). The analysis of error rates was marginally significant for decision errors, in the same direction as the RT analysis, but not wrong-task errors. For parity and prime switches (i.e., intermediate and hardest tasks), unlike for the magnitude/prime pairing, the typical switch cost asymmetry was observed, with greater costs for switches to the parity task, for both RTs and error rates. N-2 repetition costs were not significantly different for the two task pairings, for RT, although mean n-2 repetition costs were higher for the easier (EHE) task. Wrong-task error rate n-2 repetition costs were significantly different, with more errors for the prime (HEH) task than the parity task.

In experiment 2, different error types could not be distinguished due to the use

---

[4] The distinction between overlapping and non-overlapping responses is important, as there is empirical suggestion that response set overlap is one parameter increasing the magnitude of the n-2 repetition cost (Gade & Koch, 2007).

of overlapping response keys. RTs and overall error rates were as follows: Magnitude task: 635ms, 1.67%; Parity task: 741ms, 4.52%; Prime task: 759ms, 4.61%. For magnitude and prime switches, the asymmetric switch cost was reversed (i.e., greater switch cost for switches to the harder, prime task) for RTs, but not error rates. The difference in n-2 repetition cost did not approach significance, although the mean n-2 repetition cost was higher for the easier task (i.e., EHE) than the harder task (HEH). For parity vs. prime switches, no significant difference in difficulty between these tasks was reported in experiment 2, and the difference in switch costs and n-2 repetition costs was not analysed.

Although Arbuthnott (2008a) argued that the results supported her hypothesis, one substantial empirical limitation is the lack of robustness of the reported effect. Asymmetric n-2 repetition costs are only reported in experiment 1, and only for one pair of tasks. For the other pairing, no significant n-2 repetition costs occurred for RTs, with a significant effect in the opposite direction occurring for wrong-task errors. In experiment 2 (overlapping response sets), only one pair of tasks had significant differences in difficulty, and for these tasks, the n-2 repetition costs were not significantly different between alternations. Switch cost asymmetries were also unusually non-robust in this experiment, with a reversal in the typical direction of the effect in one task pairing in experiment 1, and in experiment 2. In neither experiment were the direction of switch or n-2 repetition costs compared for the easy and intermediate task pairing. Additionally, deriving a directional hypothesis by reasoning from poorly specified verbal models of backward inhibition is clearly fraught with difficulty. An opposite directioned hypothesis, derived from alternative reasoning from the same two verbal models, might be similarly plausible.

In summary, although the empirical study of Arbuthnott (2008a) did not produce robust empirical results, or falsify a clearly stated hypothesis, the empirical paradigm has subsequently proven under-utilised. The use of computational models can readily be used to derive clear directional hypotheses through simulation, rather than needing to derive them verbally. As in the previous simulation, prior to simulating asymmetric switching in a backward inhibition model, it is first necessary to establish the predictions of a lateral-inhibition-only model.

4.4.2 Procedure

The simulation consisted of multiple runs of 3-trial blocks. The model was re-initialised before each block. Therefore, comparison of performance on the 3rd trial in each task triplet gives the cleanest indication of the behaviour resulting from the residual effects of the preceding two trials, only. As in the analysis conducted by Arbuthnott (2008a), four different types of block were defined by sequence type, as follows:

1. No switch (e.g., BAA, CBB, BCC etc)

2. One-switch (e.g., BBA, CCB)

3. Two-switch (e.g., ABC, CBA)

4. Alternating switch (e.g., ABA, BCB)

The simulation explicitly controlled for sequence congruency — fully incongruent stimuli were used for all 3 trials in a block (e.g., II/II/II, as in section 4.3). The simulated sequences consisted of all switch conditions, each repeated for all permutations of three tasks (i.e., six possible sequence permutations), and each occurring five hundred times, for a total of 12,000 blocks of 3 trials.

The model parameters governing the strength of tasks A, B and C respectively were hand-set to 4.4, 3.5 and 2.2 for stimulus input strength, and 5.8, 9.0 and 13.0 for top-down control strength. The hand setting of parameters was done such that the model produced asymmetric switch costs for all three task pairings, in the typically observed direction (i.e., greater costs for switches to the easy task).

4.4.3 Results

The basic results of the simulation are presented in tables 4.1, for basic RT data (see also figure 4.5), and 4.2, for n-1 switch and n-2 repetition costs.

Tasks	Switch condition	RT (easier)	RT (harder)
Between easy (0) and hard (2)	No-switch	54.7 (18.8)	129.2 (124.9)
	1-switch	106.2 (33.1)	145.5 (89.4)
	2-switch	103.8 (32.0)	155.3 (89.3)
	Alternating switch	93.6 (32.5)	170.4 (125.7)
Between intermediate (1) and hard (2)	No-switch	72.9 (41.3)	129.1 (124.9)
	1-switch	122.3 (49.8)	143.2 (84.5)
	2-switch	116.4 (38.5)	171.6 (115.4)
	Alternating switch	111.7 (47.1)	158.9 (121.3)

Table 4.1: Simulation 4 - Simulated RTs resulting from switching between three asymmetric tasks. Reported RTs are mean (SD) model cycles. cf. Arbuthnott (2008a), tables 2 & 3

Tasks	Switch Direction	Switch cost	n-2 repetition cost
Between easy (0) and hard (2)	Easy to hard	16.3	15.1
	Hard to easy	51.5	-10.2
Between intermediate (1) and hard (2)	Intermed to hard	14.1	-12.7
	Hard to intermed	49.4	-4.7

Table 4.2: Simulation 4 - Simulated RT switch costs and n-2 repetition costs resulting from switching between three asymmetric tasks. RTs are model cycles.

Switch costs

The analysis proceeded parallel to that of Arbuthnott (2008a). Switch cost and switch cost asymmetry was assessed with a 2×2 ANOVA, with factors of switch condition (0-switch vs. 1-switch) and Task (i.e., Switch direction: Easy-Hard-Easy vs. Hard-Easy-Hard). Anticipated effects were a main effect of switch condition (reflecting a general switch cost), a main effect of task (reflecting relative task difficulty) and a $Task \times SwitchCondition$ interaction (reflecting differences in switch costs for different task direction — potentially, a switch cost asymmetry).[5]

---

[5] This class of model has already been demonstrated to produce asymmetric switch costs for two-task switching, for a variety of parameter settings (Gilbert & Shallice, 2002). Thus, it was

Arbuthnott (2008a) compared each easier task (magnitude and parity judgments) with the harder task (prime judgments). Here, each combination of two tasks was tested against each other.

All expected effects and interactions reached significance for chosen parameter settings. For tasks 0 (easiest) and 1 (intermediate), highly significant main effects were obtained for $Sequence$, $F(1, 2883) = 1343.2$, $p < .001$, and $Task$, $F(1, 2883) = 212.5$, $p < .001$. The interaction was also significant, $F(1, 2883) = 4.21$, $p = .040$. For tasks 0 (easiest) and 2 (hardest), highly significant main effects were obtained for $Sequence$, $F(1, 2094) = 108.3$, $p < .001$, and $Task$, $F(1, 2094) = 414.6$, $p < .001$. The interaction was also highly significant, $F(1, 2094) = 19.7$, $p < .001$. For tasks 1 (intermediate) and 2 (hardest), highly significant effects were obtained for $Sequence, F(1, 2084) = 92.0$, $p < .001$, and $Task, F(1, 2084) = 181.1$, $p < .001$, with a highly significant interaction, $F(1, 2084) = 16.9$, $p < .001$. Comparison of overall switch cost (table 4.2) suggests the switch cost asymmetry to be in the typical direction, rather than the reversal observed by Arbuthnott (2008a).

N-2 repetition costs

As established previously (section 4.3), the lateral-inhibition only model did not produce n-2 repetition costs, hence they are not anticipated here. Arbuthnott (2008a), addressing the question of whether the asymmetric switch cost was due to residual task inhibition, or residual task activation, reasoned that if it were primarily due to residual inhibition, conditions with a larger switch cost (i.e., Easy-Hard-Easy sequences) should also exhibit larger n-2 repetition costs than Hard-Easy-Hard sequences. Alternatively, if the asymmetric switch cost were largely due to residual inhibition, n-2 repetition costs would be equal across conditions. Although the present model implements residual activation in terms of task-demand

---

assumed that the model should produce asymmetric switch costs for two out of three possible tasks, and parameters were hand-set accordingly.

unit activation, and residual inhibition via lateral inhibitory task-demand unit connections, no directional effects are anticipated.

As with switch costs, a 2×2 ANOVA was used to assess the n-2 repetition cost, with factors of $Sequence$ (2SW vs. ALT) and $Task$ (HEH vs. EHE).

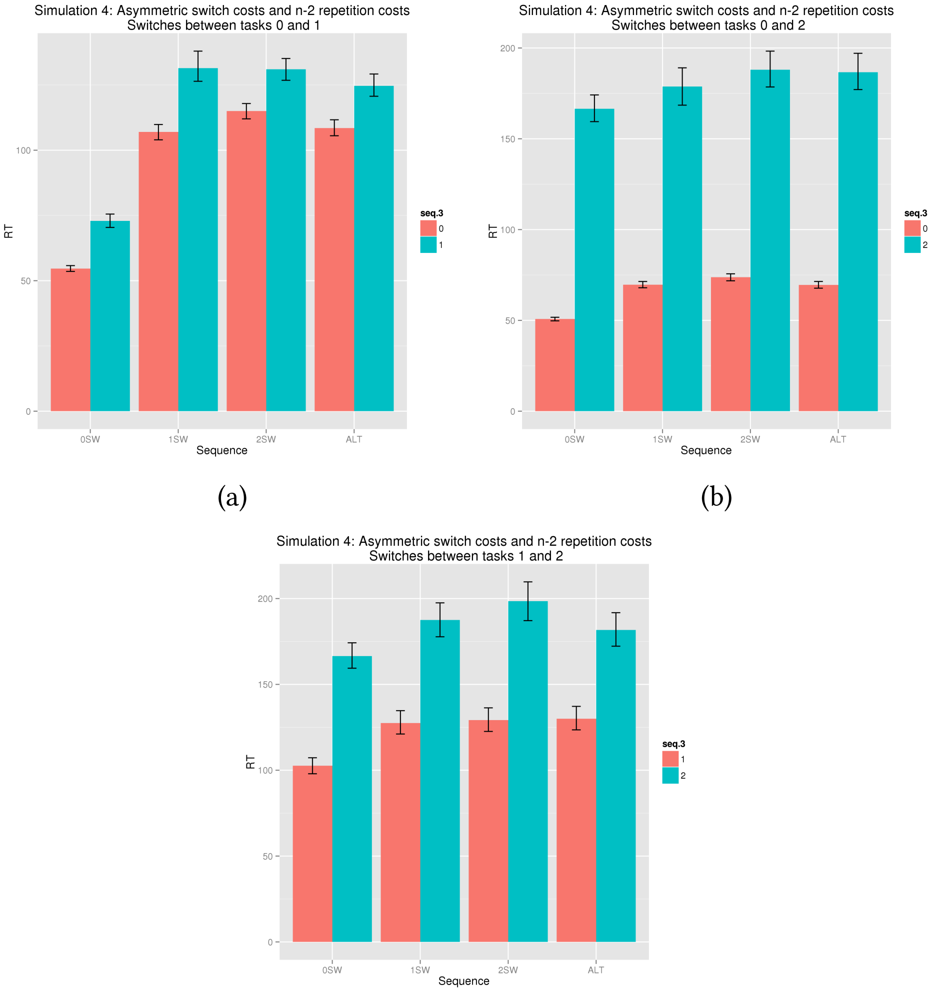
For tasks 0 (easiest) and 1 (intermediate), $Sequence$ was highly significant, $F(1, 1766) = 10.5,\ p = .001$, and $Task$ was highly significant, $F(1, 1766) = 71.3,\ p < .001$, but with no significant interaction, $F(1, 1766) = 0.003$. The significant main effect of Sequence reflects an n-2 repetition facilitation, as in the previous simulation (section 4.3).

For tasks 0 (easiest) and 2 (hardest), $Sequence$ was not significant, $F(1, 1218) = 0.40,\ p = .53$, $Task$ was highly significant, $F(1, 1218) = 253.7,\ p < .001$, and the interaction was highly significant, $F(1, 1218) = 9.13,\ p = .003$. The significant interaction reflects a difference in direction of the n-2 repetition effect: For 0-2-0 (EHE) switches it reflects a facilitation, while for 2-0-2 (HEH) switches it reflects a cost. The significance of these effects was tested individually using post-hoc t-tests (Welch two-sample t-test). The n-2 repetition cost for 2-0-2 sequences was marginally significant, $t(236.56) = -1.6,\ p = .01$ (uncorrected). Note that this is in the opposite direction to that predicted by Arbuthnott (2008a). The n-2 repetition facilitation for 0-2-0 sequences was also significant, $t(621.2) = 31.6,\ p < .001$ (uncorrected).

For tasks 1 (intermediate) and 2 (hardest), $Sequence$ was not significant, $F(1, 1181) = 1.24,\ p = .27$), $Task$ was significant, $F(1, 1181) = 113.5,\ p < .001$, while the interaction was not significant, $F(1, 1181) = 0.67,\ p = .41$.

4.4.4 Discussion

Asymmetric switch costs were obtained for all pairs of tasks after hand-setting of parameters. Moreover, consistent with the simulation reported in section 4.3, this simulation produces a small but statistically significant n-2 repetition facilitation



(a)

(b)

(c)

Figure 4.5: RTs for each pair of tasks (4.5a) Simulation 4 - switches between tasks 0 (easy) and 1 (intermediate); (4.5b) Simulation 4 - switches between tasks 0 (easy) and 2 (hard); (4.5c) Simulation 4 - switches between tasks 1 (intermediate) and 2 (hard).

(i.e., a negative n-2 repetition cost) for all three combinations of tasks.

In the model, the topdown control strength (TDCS) parameter (or more specifically, the balance between topdown control and stimulus input strength, SIS) regulates performance, effectively decreasing the dispersion of RTs. One finding that emerged from the process of hand-setting parameters was that the asymmetry of the switch cost was somewhat sensitive to specific parameter settings for TDCS and SIS. In a real cognitive system, however, determining the appropriate equivalent of the TDCS parameter is a separate cognitive problem. Presumably, this cannot purely be due to conflict monitoring (e.g., Botvinick et al., 2001), as this is not specific to the task about to be performed. Setting of the appropriate parameter value would fall within the remit of a task reconfiguration system, yet this was intentionally outside the scope of the model of the model of Gilbert and Shallice (2002). Thus, within the existing model, this problem is difficult to resolve.

In her first experiment (i.e., non-overlapping response sets), Arbuthnott (2008a) found that for one pair of tasks, the switch cost asymmetry was reversed (i.e., greater switch costs for switches into the harder task), which she did not explain in theoretical terms. In the model presented here, while a variety of parameter settings were tried (for stimulus input strength and top-down control strength), a reversal of the switch cost asymmetry could not be obtained. However, this issue cannot be approached more formally without a more systematic approach to setting the parameters for stimulus input strength and top-down control strength, and is therefore returned to in a later chapter.

It might be argued that Arbuthnott (2008a) found a reversal of the switch cost asymmetry, thus potentially invalidating the assumption underlying the hand-setting of task difficulty parameters. However, the reversal of the switch cost asymmetry was not robustly demonstrated in Arbuthnott (2008a). The assumption of switch cost asymmetry was made in order to maintain consistency with previous findings from the two-task switching literature. It should be acknowledged, however, that when hand-setting model parameters, reversal of the switch cost asymmetry for one of the three possible task pairings did occasionally occur. Therefore, it is possible that the introduction of a third task into the model changes the behaviour of the model for two-task effects. One priority for future simulations, therefore, is a more systematic exploration of the model behaviour across varying parameters for task difficulty.

4.5 Summary

This chapter has demonstrated, through simulation, that a lateral-inhibition-only model that predicts a wide range of two-task switching effects, does not predict n-2 repetition costs. This finding is not affected by the modulation of either response congruency or variable task difficulty. This justifies the need for introducing additional cognitive mechanisms to explain these empirical phenomena. In the next chapter, we consider a modification of the task switching model where backward inhibition is triggered by inter-task conflict.

Chapter 5

A Model of Backward Inhibition

This chapter extends the lateral-inhibition-only model of the previous chapter with the addition of a backward inhibition mechanism, based on the empirical suggestion that backward inhibition is triggered by conflict (Koch et al., 2010) between aspects of task representations. In the model, backward inhibition is implemented by conflict units sensitive to co-activation of task demand units, with recurrent inhibitory connections. The chapter presents a series of simulations showing the effect of backward inhibition compared to the lateral inhibition only model (simulation 5, section 5.3), contrasts model performance for symmetric and asymmetric task strengths (simulations 6a, section 5.4.1, & 6b, 5.4.2), examines the activation dynamics of the model in the n-1 switch and n-2 repetition paradigms (section 5.4.1), and finally examines various treatments of an implementational issue, negative conflict (simulation 6c, section 5.5).

5.1 Introduction: Task conflict and backward inhibition

Based on a range of evidence linking backward inhibition to conflict between competing task representations, the model developed in this chapter is based on three theoretical assumptions:

1. Task inhibition is invoked by a conflict-monitoring mechanism (following the suggestion of Koch et al., 2010)
2. Between-task conflict causes active Backward Inhibition of the non-relevant task
3. Backward Inhibition is persistent (across trials)

Within the cognitive control literature, conflict is a related concept to interference in task switching. In theoretical terms, Botvinick et al. (2001) posited conflict as a means of signaling the need for (top-down) control within the cognitive system. Computationally, Botvinick et al. (2001) implemented conflict monitoring units by taking as their input the product of the activations of two task demand units (figure 5.1), in a modification of the Stroop model of Cohen and Huston (1994). The signal generated in conflict monitoring units on each trial was then used to adjust performance on subsequent trials, accounting for inter-trial effects such as sequential adjustments in performance in the Eriksen flanker task, the effects of manipulating trial-type frequency (i.e., the proportion of congruent, incongruent and neutral trials) in the Stroop task, and within-run slowing effects following errors.

Substantial commonalities exist between the models of Gilbert and Shallice (2002) and Botvinick et al. (2001). Both represent modifications of the same IAC architecture, and both use computational properties arising from the simultaneous activation of competing representations to simulate human performance. The

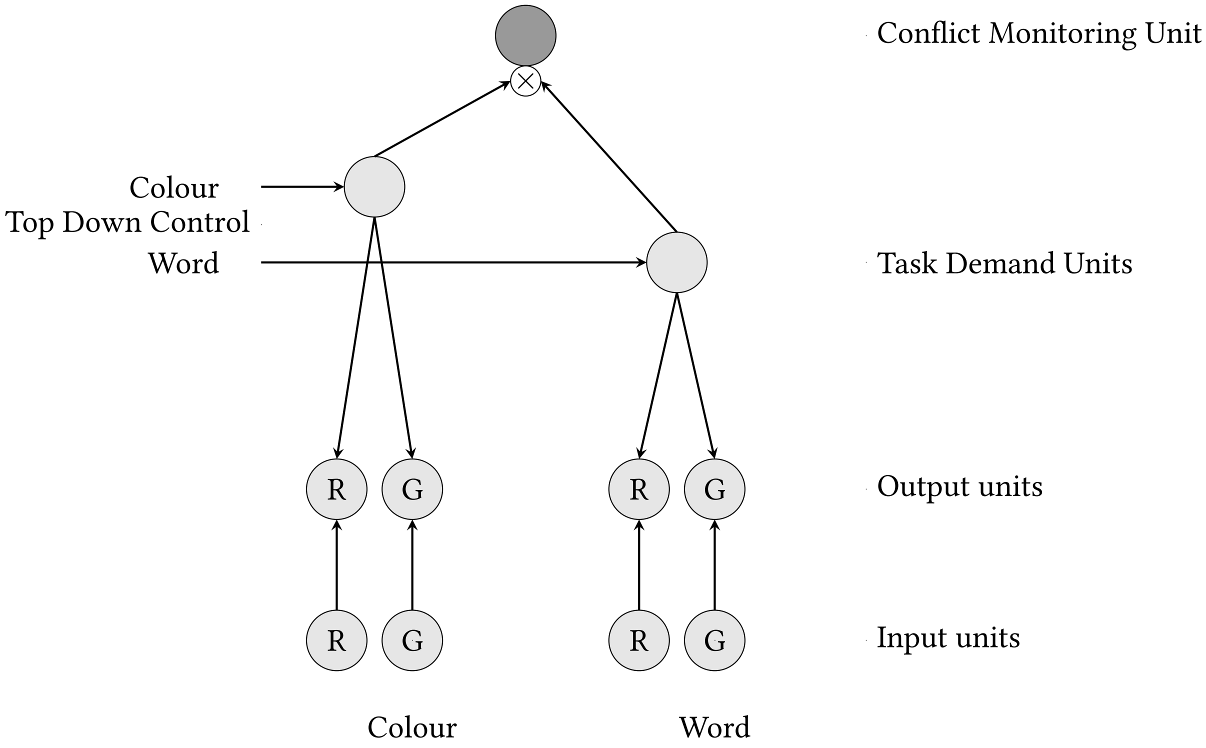


Figure 5.1: A simplified diagram of the conflict monitoring model (Stroop task) of Botvinick et al (2001). Output units are further connected to response units, and an additional (neutral) unit is added at input and output level.

conflict monitoring hypothesis of Botvinick et al. (2001) proposes, however, that specific units (in computational models) or regions of cortex (anterior cingulate, in humans) have the function of detecting such interference and using it as a signal for triggering cognitive control.

The proposal of Koch et al. (2010), that conflict signals the need for a form of higher-level control, is therefore entirely consistent with the role of conflict in the conflict monitoring hypothesis, and implementable within the current computational framework by introducing conflict monitoring units similar to those of Botvinick et al. (2001) into the 3-task version of the task switching model, described in the previous chapter.

Various possibilities were considered for adding conflict monitoring units to the model of Gilbert and Shallice (2002). For example, persistent inhibition might be modelled as changes in weights (i.e., a form of Hebbian learning) of inhibitory connections between conflict units and task-demand units. Alternatively, backward inhibition might result from residual activation (carryover) of a conflict signal from previous trials. Both of these mechanisms, to an extent, feature in the model of Gilbert and Shallice (2002). However, given that their model succeeded in explaining a range of switch-cost phenomena in terms of the dynamics of residual activation of task-demand units, it seems reasonable to adopt the same approach to explaining n-2 repetition costs, i.e., as a consequence of residual activation of units at a higher level in the model.

The present theoretical perspective conceptualises backward inhibition as a cognitive control mechanism that operates by biasing processing between multiple task pathways, on the basis of conflict at the level of task representations. The model extends the interactive activation model of Gilbert and Shallice (2002), which in itself is analogous to a lower-level contention scheduling system (Norman & Shallice, 1986; Cooper & Shallice, 2000), with an additional task inhibition mechanism.

5.2 Model description

5.2.1 Model architecture

The model architecture is illustrated in figure 5.2[1]. The portion of the figure below the dashed line is equivalent to the model of Gilbert and Shallice (2002) applied to three tasks. Specifically, it is implemented as an interactive activation model in which processing in each unit of the model is allowed to bias processing at other levels (McClelland, 1993). The model has three sets of input and output units, corresponding to each of three tasks (referred to in figure 5.2 as tasks A, B and C). For each task, two input units correspond to the two possible input values. For

---

[1] Unless otherwise specified, the weights of connection are fixed and take the default values used by Gilbert and Shallice (2002).

example, the input stimulus ‘9’, affording the three tasks *parity*, *magnitude* and *centrality*, might be represented as *odd* (left), *high* (right) and *peripheral* (right). Input units have feed-forward connections to the corresponding output unit. There are six output units, representing two possible responses to each of the three tasks. Within the set of output units, units that correspond to the same response are mutually excitatory, while units that correspond to alternative responses are mutually inhibitory. The model therefore implements an experimental procedure in which responses for all tasks are mapped to the same set of response keys (cf. Gade & Koch, 2007).

The strength of the connection weights between input and output units represents the strength of ‘bottom-up’ processing in the model: a well-learned or prepotent task (e.g., word reading of Stroop stimuli) has stronger connection weights than a weaker task (e.g., colour naming). While training of connection weights in the model did not occur, previous work has demonstrated that a similar architecture can learn this arrangement of connection weights when presented with training sets biased toward particular tasks (Cohen et al., 1990).

5.2.2 Task demand units

As in the model of Gilbert and Shallice (2002), task processing is biased by task-demand units — one per task — which have excitatory (i.e., positively weighted) connections to their respective output units and inhibitory (i.e., negatively weighted) connections to the output units associated with the other tasks. These connections are bi-directional, so a response activated bottom-up by a strong input connection will tend to activate the task with which the response is associated. The currently relevant task unit also receives a ‘top-down control’ input, from units represented on the left of figure 5.2, simulating the level of deliberate control required to perform each task to a reasonable level. As a simplification, the model omits the modifiable connections from input to task demand units, which Gilbert and Shallice

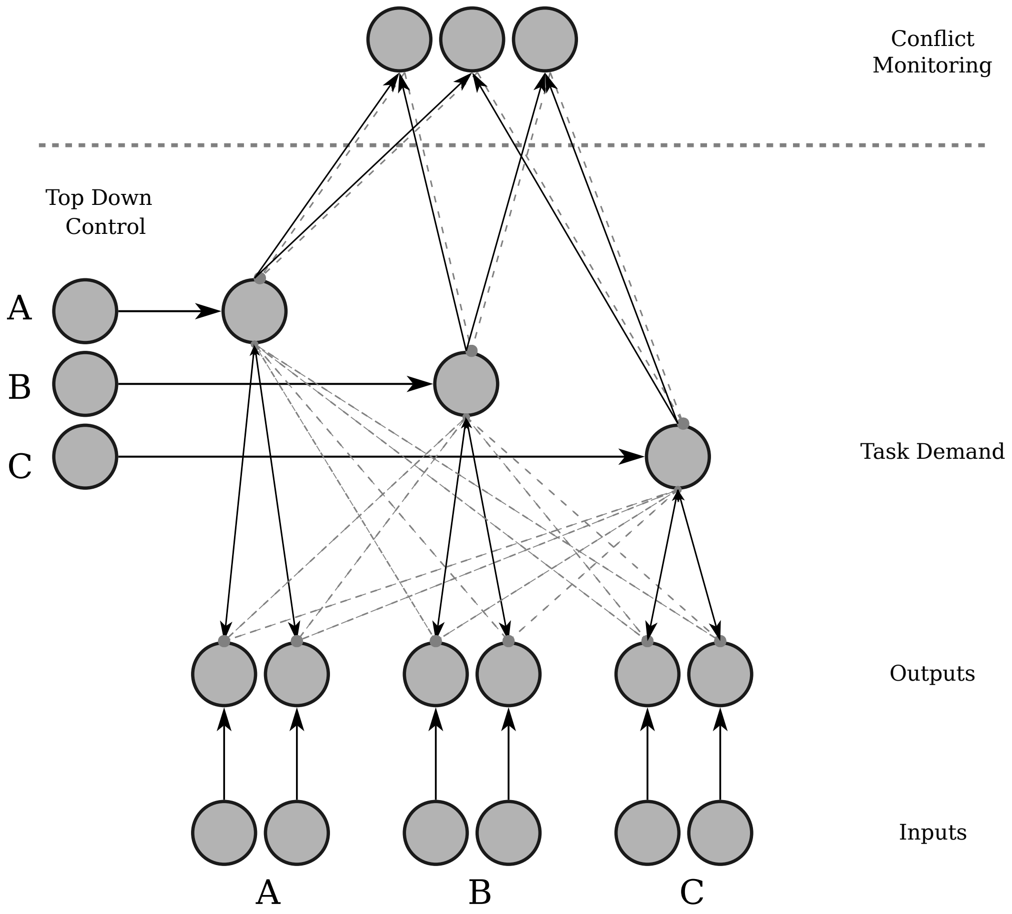


Figure 5.2: Architecture of the model capable of switching between three tasks (A, B & C). Excitatory connections are represented as lines in solid black with sharp arrows, inhibitory connections are dashed grey lines with circular arrows. Arrowheads show the direction of the connection. Not shown are within-module connections (e.g., lateral inhibition).

(2002) used to simulate the influence of item-specific priming.

5.2.3 Conflict monitoring units

The current model extends previous models by introducing an upper level of conflict monitoring units, operating similarly to those in the model of Botvinick et al. (2001). The input (I_c) to these units is somewhat different to elsewhere in the model. Each monitors the conflict (i.e., simultaneous activation) between two Task Demand (TD) units, according to equation 5.1:

$$I_c = \gamma_c(\frac{\alpha_1 + 1}{2})(\frac{\alpha_2 + 1}{2}) + \beta_c \tag{5.1}$$

First, the activations of each TD unit (α_1, α_2) are linearly rescaled from (-1,1) to (0,1), to ensure the minimum activation value is zero. The input to conflict units (I_c) on each cycle is then calculated as the product of these values, multiplied by a gain parameter (γ_c), and adding a constant negative bias (β_c). Thus, in the absence of conflict, the conflict units receive a constant negative baseline input ensuring their activation decreases. If the rescaled activation values of two TD units are both greater than zero, the mutually connected conflict unit will receive an input greater than this baseline.

Unlike the model of Botvinick et al. (2001), conflict units bias model processing interactively, via inhibitory connections to both respective task demand units multiplied by a weight parameter (ω). In sum, the model instantiates a theoretical position related to the proposal by Koch et al. (2010), i.e., that task inhibition is recruited by conflict generated during task processing. The theoretical position differs from that of Koch et al. (2010), in that those authors speculated that task inhibition was recruited by response conflict. In their account, incompatible response alternatives generate response conflict, which inhibit the task set responsible for exciting the (irrelevant) conflicting response. Here, conflict is generated at the task demand level, by simultaneous activation of two or more task units, which feeds an inhibitory signal back to the input of the task demand units. This has the advantage that the task conflict monitoring units do not require any additional information than that already available at the task level. Both tasks are inhibited, although in the case of the currently active task, this inhibition is counteracted by the top-down control signal.

5.2.4 Activation calculation

The calculation of activation in the model is unchanged from the 3-task version of the Gilbert and Shallice (2002) model, described in the previous chapter. However, while the overall behaviour of the model is not anticipated to be particularly dependent on any one activation function, the specific function is not well suited to modelling more complex activation dynamics, which introduce a number of implementational issues. Specifically, the absence of any residual decay in the activation of units implies steady-state unit activation in the absence of input. Instead, Gilbert and Shallice (2002) modelled decay in terms of bias units, which supply a constant negative input to each layer of units. This means a somewhat linear decay of activation in the model, with no level of rest activation.

Additionally, unit activations range from -1 to 1 (rather than, for example, 0 to 1). This, combined with the absence of a rest activation state, means that a unit with an activation of 0 (i.e., starting state) is effectively half-active. One implication of this is the possibility of 'negative conflict', where a conflict unit's input is calculated as the product of activations of a negatively active task demand unit with a positively active one. (This tends to occur in some longer trials, where the conflict signal decreases below zero.) In this situation, conflict units effectively provide excitatory input to TD units because of their negatively weighted connections (e.g., see figure 5.6a). This phenomenon is not a part of the conflict monitoring theory proposed by Koch et al. (2010). Neither is it an issue in the model of Botvinick et al. (2001), who scale the activation of units between 0 and 1. Instead, it is an unanticipated consequence of using a specific activation function and set of parameters specifying maximum and minimum activations.

A number of approaches to eliminating negative conflict from the model are considered here. While negative conflict is considered to be an undesirable property of a particular activation function, it is important to demonstrate the performance of the model is independent from a specific implementation, so three alternatives are considered. First, 'clipping' the conflict signal, by implementing a threshold of zero, with only above-zero activation being passed to task-demand units (as a negative input due to negative connection weights). Second, rescaling the output of the conflict monitoring units to 0 – 1 (similar to the rescaling of task demand activation before calculation of the conflict signal). Third, allowing negative conflict to determine its behavioural consequences and compare with the first two treatments.

Implementation of an alternative activation function with more desirable properties — such as squashing activation to between 0 and 1, with a resting activation (e.g., 0.1) to which activation extremes will decay in the absence of further input — would solve the negative conflict issue and also eliminate the need to balance input with static bias units. However, it would make a fundamental change to the activation dynamics (and, potentially, the behaviour) of the model of Gilbert and Shallice (2002), necessitating replication of previous simulations if the model is to progress toward the status of a consistent task-switching architecture that simulates a range of empirical behaviour across multiple paradigms.

5.3 Simulation 5: Basic effects in the backward inhibition model

This simulation aimed to explore the qualitative effect of the proposed inhibitory control mechanism. Two versions of the model were therefore compared: a no-inhibition version, where the weight of inhibitory conflict task-demand connections was zero (functionally, identical to the three-task lateral-inhibition only model from the previous chapter) and a model where the value of the weight parameter (ω) was -20.0. Conflict unit gain (γ) and bias (β_c) were 50.0 and -10.0, respectively, for both models. Other model parameters were identical with the preceding simulations (i.e., stimulus input strength of 3.0 and top-down control strength of 12.0).

5.3.1 Method

In order to evaluate model performance over a large block of consecutive trials, paralleling a typical behavioural paradigm, each version of the model was run on 100 blocks consisting of 600 trials with re-initialisation of the model occurring between blocks but not between trials. The sequence of tasks was generated randomly, with all three tasks having an equal probability of occurring on any given trial. As in previous empirical research, each trial was categorised according to the task sequence formed by it and its two preceding trials as described above.

5.3.2 Results

Mean response times and error rates for each sequence of switches were calculated for each block. Figure 5.3 plots the mean and 95% confidence intervals for all blocks, for both versions of the model (without and with inhibition) and triplet type. Confidence intervals allow direct inference of statistically significant differences from the figure.

Firstly, consider the no-inhibition version of the model. For mean reaction

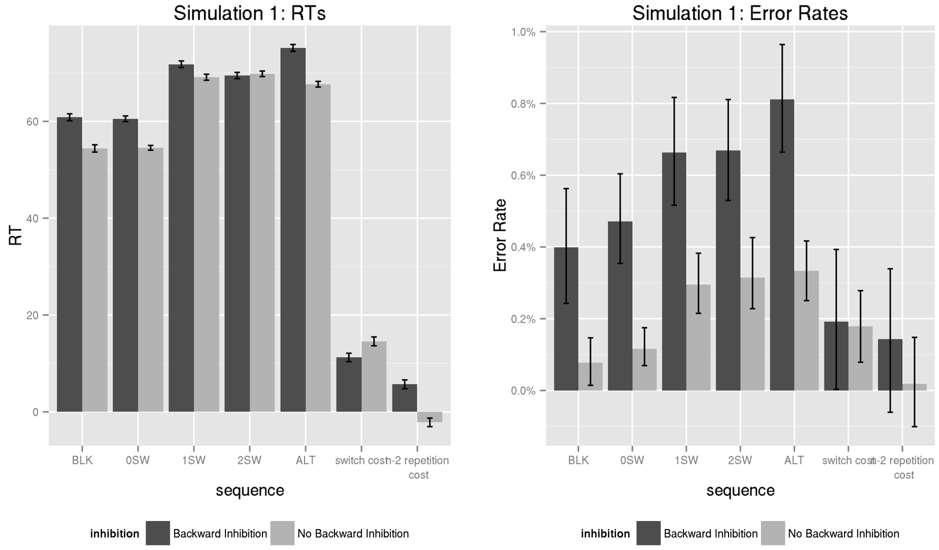


Figure 5.3: Simulation 5 results: RTs and error rates for sequences of task switches, and derived costs. Error bars represent bootstrapped 95% confidence intervals.

times, there is no statistically significant difference between repeats where the previous switch was recent (0SW) or more distant (BLK), inferred from 95% confidence intervals for each condition which include the other condition's mean. Similarly, 1SW is significantly slower than 0SW, an n-1 switch cost. These two findings replicate the results of Gilbert and Shallice (2002). There is also little difference between 1SW and 2SW conditions, while the ALT condition is faster than 2SW, i.e., a slight n-2 repetition facilitation (or equivalently, a negative n-2 repetition cost). Due to the stochastic nature of the data, more variance is present in the error rate data, reflected by larger confidence intervals. Nevertheless, a similar pattern is observed: No significant difference between BLK or 0SW repeats, a significant n-1 switch cost, but no significant difference between any of the switching conditions (1SW, 2SW or ALT).

Next, consider the backward inhibition model. For reaction times, again, no significant differences are observed between different repeat conditions (BLK or 0SW). As in the no-inhibition model, 1SW sequences are slower than 0SW, producing a significant n-1 switch cost, however the switch cost is reduced, due to larger 0SW RTs and smaller 1SW RTs than the no-inhibition model: Effectively, switches are facilitated, at the cost of slower repeat trials. However, ALT switches are significantly slower than 2SW sequences — a positive n-2 repetition cost, compared with n-2 facilitation in the no-inhibition model.

5.3.3 Discussion

Simulations 3 and 4 of the previous chapter demonstrate that the basic task switching model (analogous to the model of Gilbert & Shallice, 2002), which incorporates lateral inhibition at the response and task demand level but no backward inhibition, predicts n-1 switch costs but negative n-2 repetition costs, for both reaction times and error rates. As argued previously, this finding strengthens the conclusion of Grange et al. (2013), who argued against lateral inhibition as a plausible mechanism for backward inhibition, on the basis that lateral inhibition alone was incapable of producing lasting, residual inhibition of task demand units on the basis of activity on the n-2 trial. In contrast, the backward inhibition model, in the present simulation, predicts both n-1 switch costs (albeit reduced in magnitude compared with the no-inhibition model) and n-2 repetition costs for both RTs and error rates.

Turning to the error rate data, although the pattern of results is similar to the response time data, larger variability in the data leads to few statistically significant differences. In particular, the 95% confidence interval for the n-2 repetition cost overlaps zero, suggesting the model does not predict statistically significant effect for error rate data. Although the size of the confidence intervals is somewhat arbitrary, given the nature of simulation (i.e., a smaller CI could be obtained simply by running more blocks, or increasing the size of each block), the point remains that more statistical power is available for the response time data, paralleling the empirical literature.

It is notable that one effect of backward inhibition is a general increase in the error rate, irrespective of triplet type. On the face of it, this may raise concerns about the function of a backward inhibition mechanism: what benefits would such an error-causing mechanism bring to a cognitive system? This question is deferred until the general discussion. However, it is worth noting at this point that the model's error rate, even with backward inhibition, is generally below 1% — substantially below empirically observed error rates. Nevertheless, while the model accounts well for the empirically observed pattern of reaction times, as a general model of speeded response generation/execution it remains somewhat incomplete in comparison with other models which account for a speed/accuracy tradeoff (e.g., Brown et al., 2007).

5.4 Asymmetric tasks with the backward inhibition model

This section returns to the question of asymmetric n-2 repetition cost effects using the backward inhibition model. The remaining simulations in this chapter (6a, 6b, 6c) use the same procedure as simulation 4 (described in section 4.4). Two versions of the model were tested, with symmetric and asymmetric tasks (simulations 6a, section 5.4.1, and 6b, section 5.4.2, respectively).

5.4.1 Simulation 6a: Symmetric settings

This simulation served two purposes. The first was as a control simulation to test correct model implementation. With fully symmetric tasks (i.e., equal stimulus input strength and top-down control strength parameters) performance should be identical across all three tasks. The second was to establish model performance (especially n-2 repetition cost) with symmetric task difficulties, as a point of comparison for asymmetric tasks.

Tasks	Switch condition	RT (easier)
Symmetric switching (all tasks aggregated)	No-switch	72.0 (30.5)
	1-switch	78.2 (28.1)
	2-switch	79.1 (27.7)
	Alternating switch	87.3 (29.6)

Table 5.1: Simulation 6a - (Switching between three symmetric tasks, with conflict monitoring), RTs are mean (SD) cycles. cf. Arbuthnott (2008a), tables 2 & 3

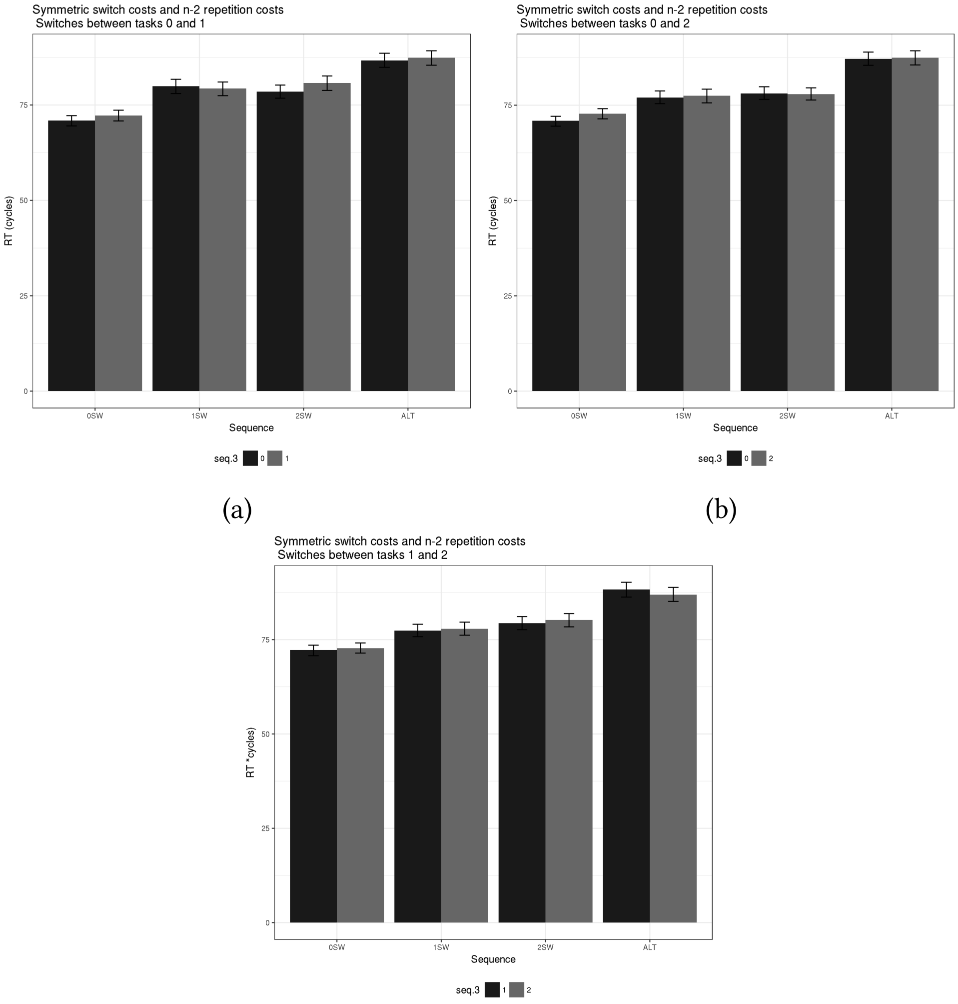
Method

Stimulus input strength of 3.0 and top-down control strength of 12.0 was used for each task. All other parameters were identical with the previous simulation (i.e., the weight parameter (ω) was -20.0, conflict unit gain (γ) and bias (β_c) were 50.0 and -10.0), with negative conflict set to 'clip'. Direct connections from input to task demand units, with Hebbian learning of weights (Gilbert & Shallice, 2002) were disabled. Each cell of the simulation (congruency condition × sequence × task permutation) was simulated 1000 times, for a total of 24,000 three-task blocks.

Results: Switch and repetition costs

The results are presented below, in figure 5.4 and table 5.1 for RT data. As previous simulations (section 5.3) have shown, error rate data from the model provide little additional information over RT data. These are not considered further in this simulation.

The reliability of the switch cost, and n-2 repetition cost, were assessed by comparing 0SW with 1SW RTs, and 2SW with ALT RTs, respectively, using independent-samples t-tests. The conflict monitoring model with symmetric tasks produces a switch cost of 6.2 cycles, $t(11347) = 11.31, p < .0001, d = 0.21$ 95% CI [0.17, 0.25], with an n-2 repetition cost of 8.1 cycles, $t(11391) = 15.25, p < .0001, d = 0.28$ 95% CI [0.25, 0.32].



(a)

(b)

(c)

Figure 5.4: RTs for each pair of tasks, for fully symmetric tasks. Error bars represent 95% confidence intervals. (5.4a) Switches between tasks 0 and 1; (5.4b) switches between tasks 0 and 2; (5.4c) switches between tasks 1 and 2.

Results: Activation dynamics

To explore how n-2 repetition costs arise as the result of differences in activation dynamics between different task sequences, activation within each trial was graphed, rather than merely considering final trial activation. Each cell of the simulation (congruency condition × sequence × task permutation) was simulated 500 times, for a total of 12,000 three-task blocks. As previously, a stimulus input strength of 3.0 and top-down control strength of 12.0 was used for each task. All other parameters were identical with the previous simulation (i.e., the weight parameter (ω) was -14.0, conflict unit gain (γ) and bias (β_c) were 39.0 and -7.5, with negative conflict set to 'clip'.

For each successfully completed trial, activation, in each unit in the model was averaged across trials for each cycle. This was done separately for each switch condition (figures 5.5 and 5.6) The activation of units at three different levels of output, task-demand, and conflict units, is represented in the lower, middle, and upper section of the plot respectively. Task demand units are colour-coded (red, green, and blue, representing tasks A, B and C, respectively). Conflict units are also colour-coded, with colours representing the combination of two task demand units (i.e., the AB conflict unit is yellow (red + green), the BC conflict unit is cyan (green + blue) and the AC conflict unit is fuchsia (red + blue). These graphs visualise activation for correct trials only, excluding the fastest and slowest decile of trials for each combination of trial position and condition, in order to produce more representative visualisation for the bulk of trials.[2] The vertical lines on each plot represent mean RT.

We begin by considering the n-1 switch cost in the three-task switching model (figure 5.5).

In the 0SW condition (i.e., BAA), the final trial (i.e., the right-most vertical

---

[2] Note that this is a different filtering criterion used elsewhere, such as that used for statistical tests which excluded RTs outside 3SDs of the mean (as in Gilbert & Shallice, 2002).

panels in figure 5.5a) is a repeat of the previous task (here represented by the red task demand unit). The unit has an initial advantage in starting activation over the B/C (green/blue) task demand units, due to the residual activation from the previous trial. Due to this residual activation, and the top-down control input, the unit quickly becomes highly active, as it laterally inhibits the other competing task demand units, in a competitive process.

In the 1SW condition (i.e., BBA), the final trial (‘2’ in figure 5.5b) is a switch trial. In this instance, the model must activate the relevant task demand unit to perform each trial (here represented as green, red, red at the task demand level, the horizontal central series of panels). On the middle trial, the B task demand unit (green) has a higher starting activation than the A or C (red/blue) units, due to residual activation from the previous trial. Therefore, the model must activate the red task demand unit via the active top-down control units. However, lateral inhibition at the task demand level means the A unit is simultaneously inhibited by the residually active B unit.

Thus far, this explanation is essentially the same as that for switch costs in Gilbert and Shallice (2002). However, the current model also includes a higher level of conflict units. As conflict units are responsive to the product of two TD units’ activations, when one TD unit is highly active, two conflict units tend to also be highly active, reflecting the conflict between the active TD unit and each of the two, less active TD units.

In the 0SW example, on the final (i.e., repeat) trial, the highly active (A/red) TD unit leads to a higher conflict signal for the yellow (A/B) and fuchsia (A/C) conflict units, while a lower signal for the cyan conflict unit reflects the decreasing signal from the two lower-activation B/C (green/blue) TD units. Here, the effect of the backward inhibition mechanism is to mildly inhibit TD units A/B (yellow) and A/C (fuchsia), with the effect that task A receives a greater level of backward inhibition than either the B or C task. Effectively, then, on repeat trials, backward inhibition has the effect of inhibiting the repeat task.

In the final trial of the 1SW example (figure 5.5b), initially, the cyan (B/C) and yellow (A/B) conflict units are more active, reflecting the residual activation of the green (B) TD unit at the and of the previous trial. The resulting inhibition from the conflict units suppresses the activation of the green (B) TD unit, which rapidly becomes even less active than the so-far unused blue (C) TD unit. As a result, throughout the final trial, activation of the cyan (B/C) conflict unit decreases, while activation of the fuchsia (A/C) unit increases, due to the increasing activation of the performed (red/A) task. An initial increase in yellow (A/B) conflict reflects the initial conflict between the residual activation of the B unit, and the activation of the A unit, at the start of the switch trial. This is resolved over the course of the trial. The effect, therefore, of the backward inhibition units on the switch trial is to facilitate performance, by inhibiting the recently performed but now irrelevant task representation.

To understand the role backward inhibition plays in the n-2 repetition cost, let us now consider the 2SW and ALT trials (figure 5.6). In the 2SW (CBA) sequence, activation on the middle trial proceeds similarly to the final trial of the 1SW sequence, discussed previously. By the end of the trial, the recently performed (blue/C) task is the most suppressed, compared to the red/A (not recently performed) or green/B (performed on this trial) tasks. As a result, the most active conflict units by the end of the middle trial are cyan (B/C) and yellow (A/B), which actively facilitate switching at the start of the final, switch trial, where the high level of inhibition of the green (B) TD unit helps to counteract its residual activation. This is why the final (switch) trial of the 2SW trial is slightly faster than the middle (also switch) trial, and slightly faster than the corresponding trial in the 1SW condition.

In the ALT (ABA) condition (figure 5.6b), activation proceeds similarly on the first two trials of the 2SW trials. By the end of the middle trial, the red/A TD unit is more suppressed than the green/B or blue/C units, again, due to the inhibition from the yellow (red/green, A/B) and fuchsia (red/blue, A/C) conflict units throughout the middle trial. Critically, at the start of the final trial, in contrast to the 2SW condition, the yellow conflict unit is more active than the cyan (B/C) unit, which in turn is more active than the fuchsia (A/C) unit. This is due to red/green (A/B) conflict which occurred on the middle trial. Thus, on the final trial where the red/A TD unit must become highly active, inhibition is being sent most strongly by the yellow (A/B) conflict unit, followed by the cyan (B/C) unit. While the strongest combined effect of backward inhibition is to suppress activation of the no-longer-relevant B unit, the A unit is also inhibited as a result.

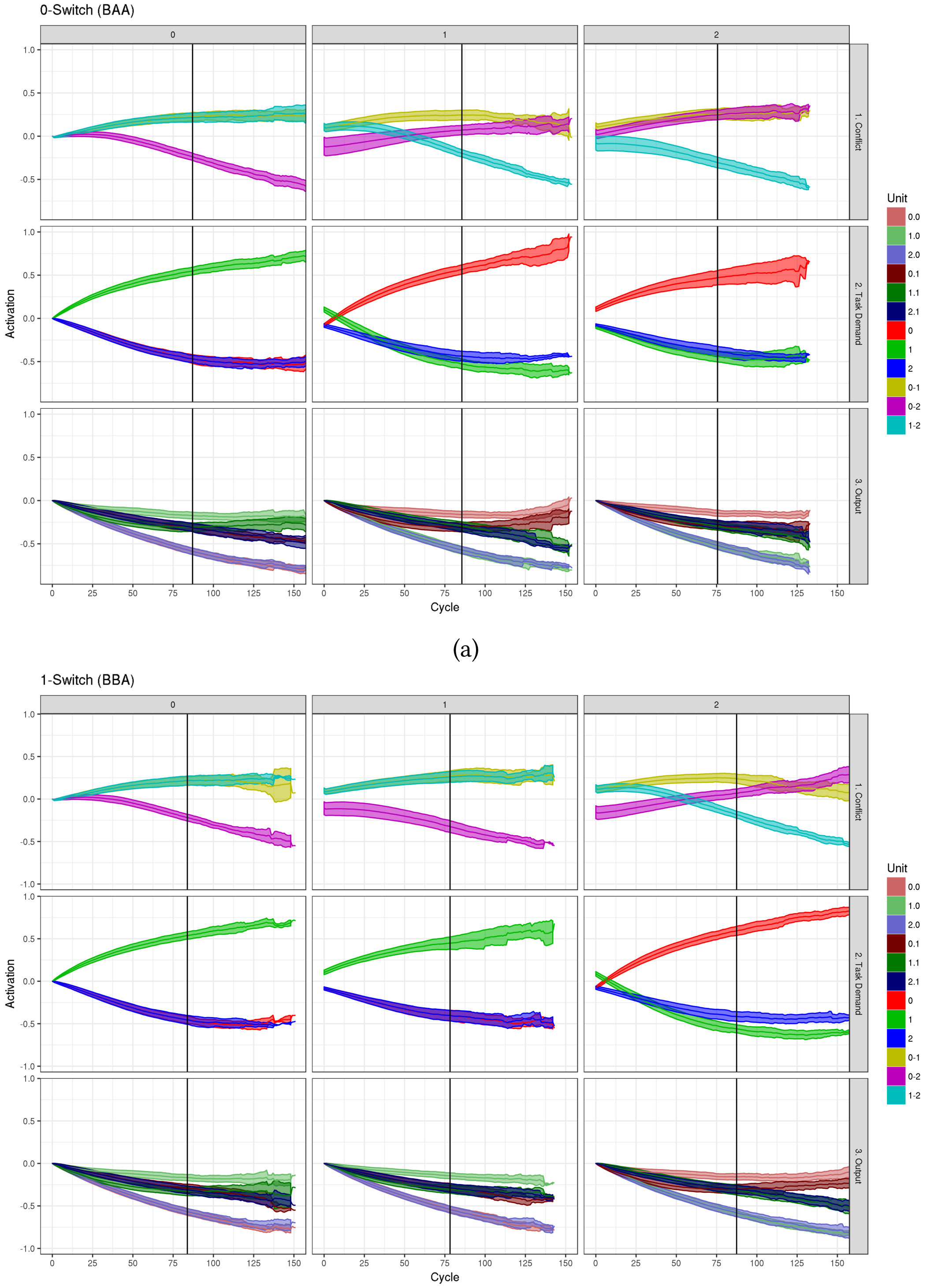
Essentially, then, this model represents a conflict-based theory of backward inhibition: Conflict occurring on the middle trial of ABA sequences results in longer-lasting suppression of the 'A' task. In the model, this is implemented by residual activation of conflict units, whose activation changes relatively slowly, provoked by recent co-activation of pairs of task-demand units.

5.4.2 Simulation 6b: Asymmetric switching

In a companion to the previous simulation (section 5.4.1), this simulation tested the backward inhibition model on asymmetric switching. The task control strengths were identical with those used in simulation 4 (section 4.4), specifically, stimulus input strengths of 4.4, 3.5 and 2.2, and top-down control strengths of 5.8, 9.0 and 13.0, for tasks A, B and C, respectively.

Method

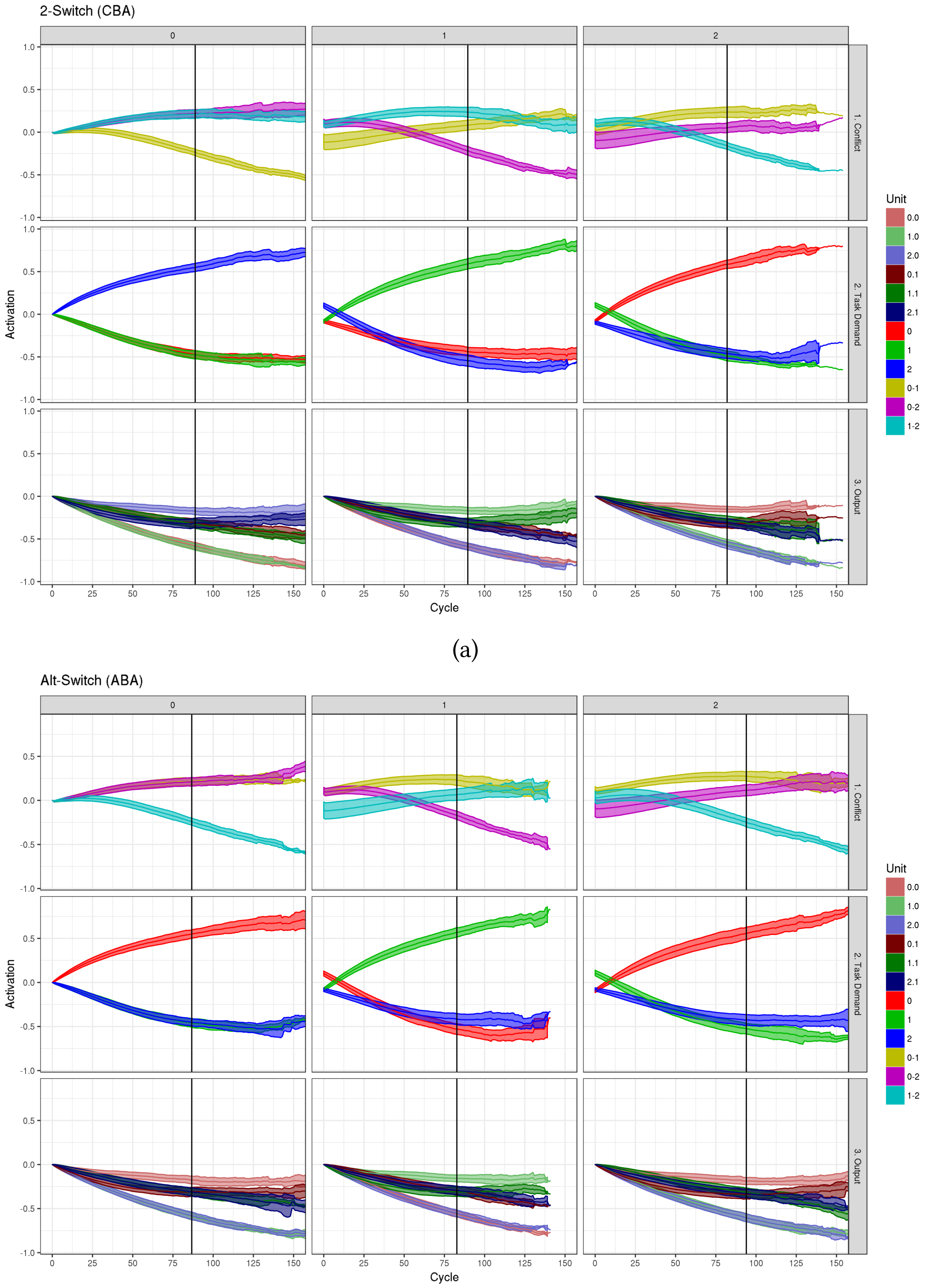
Once again, associative learning between inputs and task demand units was disabled. Negative conflict treatment was 'clip'. A larger number of trials was used, with 1000 occurrences of all switch conditions, each repeated for all permutations of three tasks (i.e., six possible sequence permutations), compared to 500 occur-



(a)

(b)

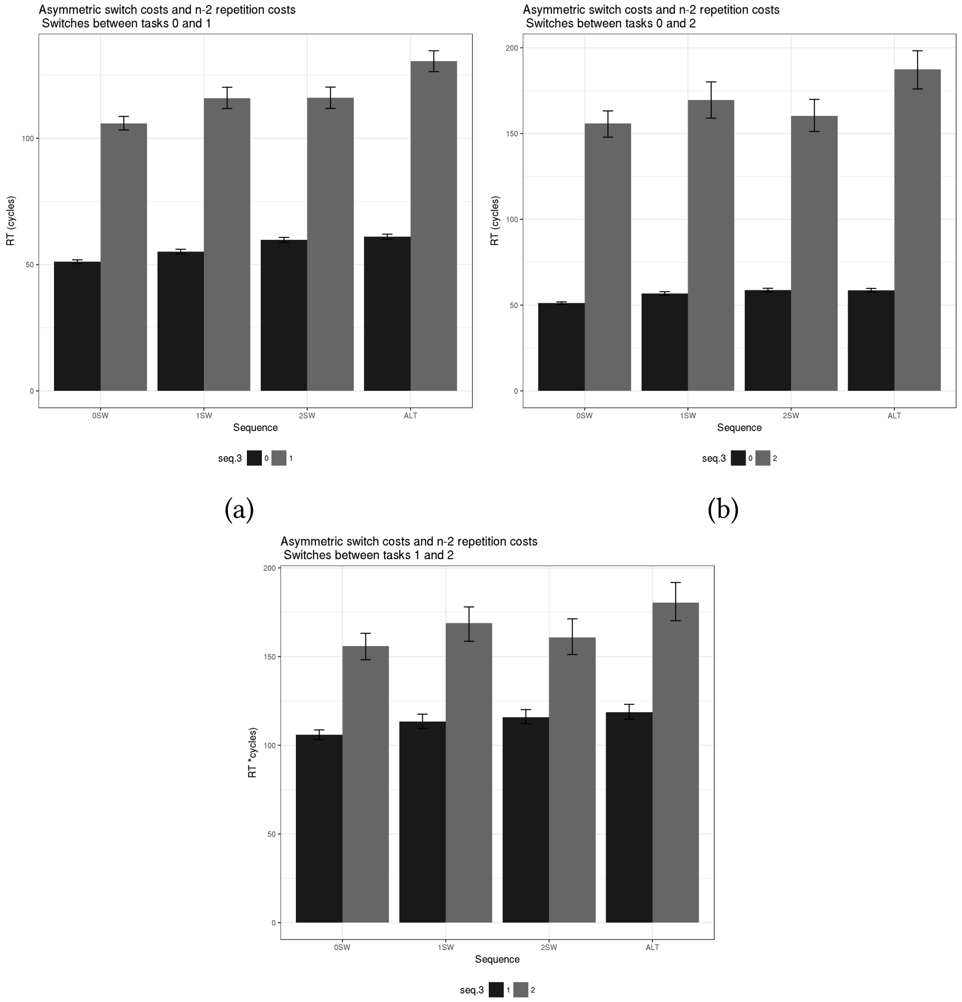
Figure 5.5: Activation dynamics in 3-task switching with symmetric task strengths: 0-Switch (BAA, panel 5.5a) and 1-Switch (ABA, panel 5.5b) conditions. The 3rd trial RT difference between conditions represents the switch cost. Vertical bars represent lower quartile, median, and upper quartile RTs.



(a)

(b)

Figure 5.6: Activation dynamics in 3-task switching with symmetric task strengths: 2-Switch (CBA, panel 5.6a) and Alt-Switch (ABA, panel 5.6b) conditions. The 3rd trial RT difference between conditions represents the n-2 repetition cost. The vertical bar represents the mean RT.



(a) (b)

(c)

Figure 5.7: RTs for switching between each pair of tasks, for asymmetric task strengths. (5.7a) Simulation 6b: Switches between tasks 0 (easy) and 1 (intermediate); (5.7b) Simulation 6b: Switches between tasks 0 (easy) and 2 (hard); (5.7c) Simulation 6b: Switches between tasks 1 (intermediate) and 2 (hard).

rences in simulation 4. Thus, the simulation totaled 24,000 blocks of 3 trials.

Results

The RTs for switching between different pairs of tasks are presented in figure 5.7, while the magnitude of n-1 switch costs and n-2 repetition costs is summarised in table 5.2.

Tasks 0 and 1

Switch costs: The main effect of task was significant, $F(1, 5499) = 2554.3, p < .0001, \eta_p^2 = .321$, reflecting the different RT resulting from tasks of different difficulties. The effect of sequence condition (0SW vs. 1SW) was also significant, $F(1, 5399) = 60.86, p < .0001, \eta_p^2 = .006$, indicating a switch cost, although with a small effect size. The interaction between these two factors, indicating a switch cost asymmetry, was also significant, but with a very small effect size: $F(1, 5399) = 6.54, p = 0.01, \eta_p^2 = 0.001$. Planned t-test comparisons revealed the switch cost was significant for both switches from 1 to 0, $t(2067) = 6.5, p < .0001, d = 0.25$ 95% CI [0.18, 0.33] and switches from 0 to 1, $t(1725.5) = 3.92, p < .0001, d = 0.16$ 95% CI [0.08, 0.25].

N-2 repetition costs: As for n-1 switch costs, the main effect of task/switch direction was significant, $F(1, 3694) = 1884.4, p < .0001, \eta_p^2 = .338$, indicating an RT performance difference between tasks. The effect of sequence condition (2SW vs. ALT) was significant, $F(1, 3694) = 18.22, p < .0001, \eta_p^2 = .007$, indicating an n-2 repetition cost, with effect size similar to that of the switch cost. The interaction was also significant, $F(1, 3694) = 20.88, p < .0001, \eta_p^2 = .006$, indicating an asymmetry in n-2 repetition costs depending on the direction of switch. Planned t-tests revealed that the n-2 repetition cost was only marginally significant for switches from 1 to 0, $t(1968) = 1.76, p = .07, d = 0.08$ 95% CI [0.01, 0.17], with magnitude of 1.3 cycles, but the effect was larger for switches from 0 to 1, $t(1721.7) = 4.88, p < .0001, d = 0.24$ 95% CI [0.14, 0.33], with magnitude 9.4 cycles. Thus, a larger effect size for the n-2 repetition cost was observed on task 1 (i.e., HEH) than on task 0, where it only reached marginal statistical significance. Note that this pattern is in the opposite direction to that observed by Arbuthnott (2008a).

Tasks 1 and 2

Switch costs: The effect of task/switch direction (i.e., RT difference from tasks of different difficulty) was significant, $F(1, 3967) = 320.8, p < .0001, \eta_p^2 = 0.075$, while the effect of sequence condition (i.e., n-1 switch cost) was also significant, $F(1, 3967) = 11.96, p = .0005, \eta_p^2 = .003$. The interaction was not significant, $F(1, 3967) = .09, p = .35$, indicating no switch cost asymmetry. Planned t-tests revealed the switch cost was significant for switches from task 2 to task 1, $t(1749.1) = 3.0, p = .003, d = 0.12$ 95% CI [0.04, 0.21], and also for switches from task 1 to task 2, $t(1192) = 2.0, p = .05, d = 1.0$ 95% CI [0, 0.21].

N-2 repetition costs: The effect of task/switch direction was highly significant, $F(1, 2965) = 233.5, p < .0001, \eta_p^2 = .073$, as was the effect of sequence condition (2SW vs. ALT), $F(1, 2965) = 6.86, p = .009, \eta_p^2 = .003$, indicating RT differences based on task difficulty, and an n-2 repetition cost. The interaction was also significant, albeit with very small effect size, $F(1, 2965) = 5.85, p = .016, \eta_p^2 = .002$. Planned t-tests revealed no significant n-2 repetition cost for from 2 to 1, $t(1807.9) = 0.95, p = .34$. However, the cost for switches for 212 (i.e., HEH alternations) was significant, $t(1118.2) = 2.66, p = .008, d = 0.16$ 95% CI [0.04, 0.27]. This finding, an n-2 repetition cost for the HEH alternation, but no significant cost for the EHE alternation, was consistent with the results for tasks 0 and 1, and also in the opposite direction to the effect observed by Arbuthnott (2008a).

Tasks 0 and 2

Switch costs: Once again, the effect of task was significant, $F(1, 4512) = 2204.66, p < .0001, \eta_p^2 = .328$, as was the effect of sequence condition (0SW vs. 1SW), $F(1, 4512) = 39.6, p < .0001, \eta_p^2 = .003$, indicating RT task asymmetry and switch costs, respectively. The interaction was only marginally significant, $F(1, 4512) = 2.91, p = .09, \eta_p^2 < .001$, indicating only a negligable switch cost asymmetry, if any. Planned t-tests revealed the switch cost was significant for switches from task 2 to 0, $t(1924.1) = 8.72, p < .0001, d = 0.34$ 95% CI [0.26, 0.42], and also for switches from 0 to 2, $t(1209.4) = 2.10, p = .04, d = 0.11$ 95% CI [0.01, 0.21].

N-2 repetition costs: The effect of task/switch direction was highly significant, $F(1, 3084) = 1685.9, p < .0001, \eta_p^2 = .353$. The effect of sequence condition (2SW vs. ALT) was also significant, $F(1, 3084) = 10.9, p = .0009$. The interaction was also highly significant, $F(1, 3084) = 23.7, p < .0001, \eta_p^2 = .008$. Planned t-tests revealed the n-2 repetition cost was not significant for switches from 2 to 0, $t(1973) = 0.07, p = .94.$, but was significant for switches from 0 to 2 (i.e., HEH), $t(1091.4) = 3.69, p = .0002, d = .22$ 95% CI [0.10, 0.34]. Thus, as for the previous two task pairings, the n-2 repetition cost was markedly asymmetric, with significant costs only obtained for HEH switches.

Tasks	Relative Task Difficulty	Switch cost	n-2 repetition cost
0/1	H	10.0	14.5
	E	3.9	1.3
0/2	H	13.7	27.2
	E	5.6	n.s.
1/2	H	13.5	19.6
	E	7.5	n.s.

Table 5.2: Simulation 6b: Switch costs and n-2 repetition costs. Switch costs are calculated as the difference between 1-switch and no-switch mean RTs. N-2 repetition costs are the difference between alternating switch and 2-switch mean RTs. Only statistically significant costs are shown.

Discussion

Significant switch costs occurred for task pairings 0/1, 0/2, and were marginal for 1/2. Given the statistical power resulting from a very large number of trials, however, the switch cost is much less robust than in simulation 4.

Additionally, the asymmetry of the switch cost was reversed in this simulation. For all three task pairings, the switch cost was greater for switches to the harder task. This finding is of particular interest, as it contrasts substantially with the results of simulation 4, which found robust asymmetric switch cost in the typical direction — indeed, switch cost asymmetries were a key assumption in hand-setting parameters for task strengths. Taken together, these findings suggest that switch costs have been substantially reduced by the backward inhibition mechanism, but also that the switch cost asymmetry has been reversed.

Intriguingly, this reversal is consistent with Arbuthnott (2008a), who found the direction of the switch cost asymmetry reversed for one task pairing in experiment 1, and for experiment 2. Of course, the difference between the present simulation and simulation 4 is the inclusion of a backward inhibition mechanism — a manipulation with no obvious empirical analog. Additionally, it is unclear what experimental manipulation present in the experiments of Arbuthnott (2008a) accounts for the reduced asymmetric switch cost. It may be that the inclusion of a third task, alone, is enough to account for this difference. However, it would seem to predict that in situations where backward inhibition is reduced, such as may be observed clinically, the typically observed switch cost asymmetry would be exaggerated (Mayr et al., 2006, however also see Grange et al., 2013). In any event, the asymmetry of the n-1 switch cost clearly continues to be of interest in the three-task switching paradigm, and cannot be assumed in the backward inhibition model.

We turn, next, to the n-2 repetition cost results from this simulation. N-2 repetition costs were significant for all three task pairings. Additionally, the interaction between task and sequence type was significant for all three task pairings, indicating asymmetric n-2 repetition costs. Planned t-test comparisons revealed much larger n-2 repetition costs for the hard task (i.e., HEH alternations), and for two of the three task pairings (0/2, 1/2) n-2 repetition costs failed to reach significance for the easy task (EHE). Comparison with the n-2 repetition costs observed in simulation 6a (e.g., figure 5.4 section 5.4.1) suggests that asymmetric tasks both suppress the n-2 repetition cost for EHE alternations and enhance it for HEH alternations. This finding is in clear opposition to the results obtained by Arbuthnott (2008a).

However, the present simulation has tested results for only one set of parameter settings that affect task difficulty. Furthermore, these parameter settings were hand-set on the assumption that three-task switching should produce n-1 switch cost asymmetries similar to those observed in two-task switching — a finding which the present simulation, and the empirical findings of Arbuthnott (2008a), suggest may be untenable. Therefore, a priority for future simulations is to obtain more general findings regarding both n-1 switch, and n-2 repetition cost asymmetries, independent of specific task strength parameter settings.

5.5 Simulation 6c: Approaches to negative conflict

5.5.1 Rationale

This simulation concerns an issue arising from implementing the conflict units in the backward inhibition mechanism. Specifically, the activation equations used by Gilbert and Shallice (2002), and inherited by the backward inhibition model, scale the minimum/maximum activation of units between -1 and 1, rather than a more typical 0 to 1 (e.g., Cohen & Huston, 1994). Additionally, unit activation does not automatically decay to a 'rest' state. Instead, units are connected to a bias input, which supplies a constant negative input. Thus, in the absence of positive input, net input is negative, leading the activation of units to decay.

This implementational characteristic has a number of implications. Firstly, conflict units take the product of two task demand unit activations as their input. If both task demand units are positive, the conflict unit receives a positive input. If the activation of one task demand unit is negative, and the other positive, the conflict unit receives a negative input, while if both task demand units are negative, the conflict unit once again receives a positive input. Given the activation of units is essentially linear, and arbitrarily scaled between -1 and 1, introducing such a non-linear characteristic in the connections between task-demand and conflict units is theoretically unjustified. In the simulations reported thus far, this problem was solved by gating the signal from each task demand unit to the conflict unit. Thus, the conflict monitoring unit input was taken as the product of two task demand unit activations *only when both task demand units have activation greater than zero*, otherwise the conflict monitoring input was zero. Effectively, this introduces a threshold for triggering conflict monitoring units.

Secondly, and more problematically, activation of units is frequently negative. For example, on a typical trial in the model of Gilbert and Shallice (2002), the activation of output units starts at zero, and decreases over time, with the final response eventually being made by the *least negative* response unit. As a result, the activation of conflict units can become negative, in the absence of positive conflict (i.e., if the two task demand units connected to a conflict unit have near-zero activation for a period of time). While negative activation of conflict units does not pose a conceptual problem, the inhibitory connections between conflict units and task-demand units imply that negative activation of conflict units would lead to a *positive* input from the inhibitory connections. In other words, minimally active conflict units would send excitatory inputs to task-demand units. Conceptually, this mechanism is not consistent with the theory, and thus modification of these inhibitory connections becomes necessary.

Eliminating negative conflict from the model may be resolved through a number of approaches. However, as the issue of negative conflict is purely an implementational issue, and not one which has been proposed in human cognition, any specific mechanism for eliminating negative conflict characteristics from the model should not shape model behaviour in any other way. This raises the question of how one might determine whether such a specific mechanism has shaped model behaviour? Here, we consider two possible treatments of negative conflict, alongside no-conflict and a negative-conflict-allowed conditions. The effects of allowing negative conflict, and the two different treatments for resolving it, may therefore be compared. If the two different methods for resolving negative conflict produce equivalent behaviour, we may consider that the negative conflict resolving mechanism is purely an implementational detail, with no particular effect on the behaviour of the model.

5.5.2 Method

The possible treatments of negative conflict considered here are, firstly, allowing negative conflict (i.e., an identical model and tasks as used in simulation 6a in section 5.4.1), clipping negative conflict (i.e., only the above-zero part of conflict unit activation is allowed to inhibit task demand units), and rescaling negative conflict (i.e., the activation of conflict monitoring units is linearly rescaled from 0 to 1 for the purposes of calculating the input to task-demand units). These three conditions were compared with a no-conflict condition where there is no input from conflict monitoring to task demand units. As in simulation 6a, the simulation consisted of 24,000 triplets (i.e., 1000 of each switch condition and permutation of tasks).

5.5.3 Results

The results are presented in figure 5.8, and the results of t-tests on the n-1 switch cost and n-2 repetition cost for each treatment are presented in table 5.3. As can be assessed graphically, and in terms of the sizes of the effect for switch costs and n-2 repetition costs, the results for the conditions where no biasing occurs (no-conflict) and where negative conflict is allowed, are very similar. Specifically, both produce a robust n-1 switch cost with large effect size (Cohen's *d* of 0.59 – 0.63), and a smaller n-2 repetition facilitation, with negligible effect size (*d* of -0.04 to -0.06, with a confidence interval that approaches zero). Similarly, the results for

Treatment	Switch cost	N-2 repetition cost
Conflict off	28.4	-2.1
Negative conflict allowed	28.2	-3.0
Negative conflict clipped	21.1	7.7
Negative conflict rescaled	22.68	7.8

Table 5.3: Simulation 6c: Approaches to treatment of negative conflict

the two treatments of negative conflict, where it is clipped and rescaled, are also very similar. Both produce an n-1 switch cost, albeit with a reduced effect size (*d* of 0.40 – 0.42). Additionally, both produce a robust n-2 repetition cost, with a smaller effect size (*d* of 0.14, with a confidence interval that does not include 0).

5.5.4 Discussion

These results suggest that occurrence of switch costs is robust to different treatments of conflict, albeit with varying magnitude. Consistent with simulation 6a (section 5.4.1), the n-1 switch cost is suppressed in the backward inhibition version of the model, compared to either no inhibition or the negative-conflict-allowed version. However, the n-2 repetition cost only occurs in the negative-conflict-clipped and negative-conflict-rescaled conditions. While it appears that the theoretically incoherent concept of negative conflict must be excluded from the model, it also seems that there is no difference between the two possible treatments of negative conflict.

5.6 Summary

This chapter has presented a computational model of a backward inhibition mechanism by extending an existing model task switching (Gilbert & Shallice, 2002). Based on the empirical suggestion that backward inhibition is triggered by conflicting task representations, the model implements backward inhibition in terms of conflict units which are sensitive to the co-activation of task-demand units,

Neg Conflict	Switch Costs					n-2 repetition costs				
	t	df	p	d	d 95% CI	t	df	p	d	d 95% CI
Conflict off	34.2	11631.5	$< .0001$	0.63	[0.59, 0.66]	-2.2	11972.1	.03	-0.04	[-0.08, 0.00]
Allowed	32.5	11670.4	$< .0001$	0.59	[0.56, 0.63]	-3.0	11997.5	.0013	-0.06	[-0.09, -0.02]
Clipped	21.8	11896.4	$< .0001$	0.40	[0.36, 0.43]	7.5	11878.9	$< .0001$	0.14	[0.10, 0.17]
Rescaled	22.8	11467.9	$< .0001$	0.42	[0.38, 0.45]	7.6	11884.1	$< .0001$	0.14	[0.10, 0.17]

Table 5.4: Simulation 6c: Statistical tests for the n-1 switch cost and n-2 repetition cost for each treatment of negative conflict. T-test statistics compare the two relevant conditions (i.e., 0SW and 1SW for switch costs, 2SW and ALT for n-2 repetition costs), with effect sizes (Cohen's *d*) and the 95% confidence interval for the effect size.

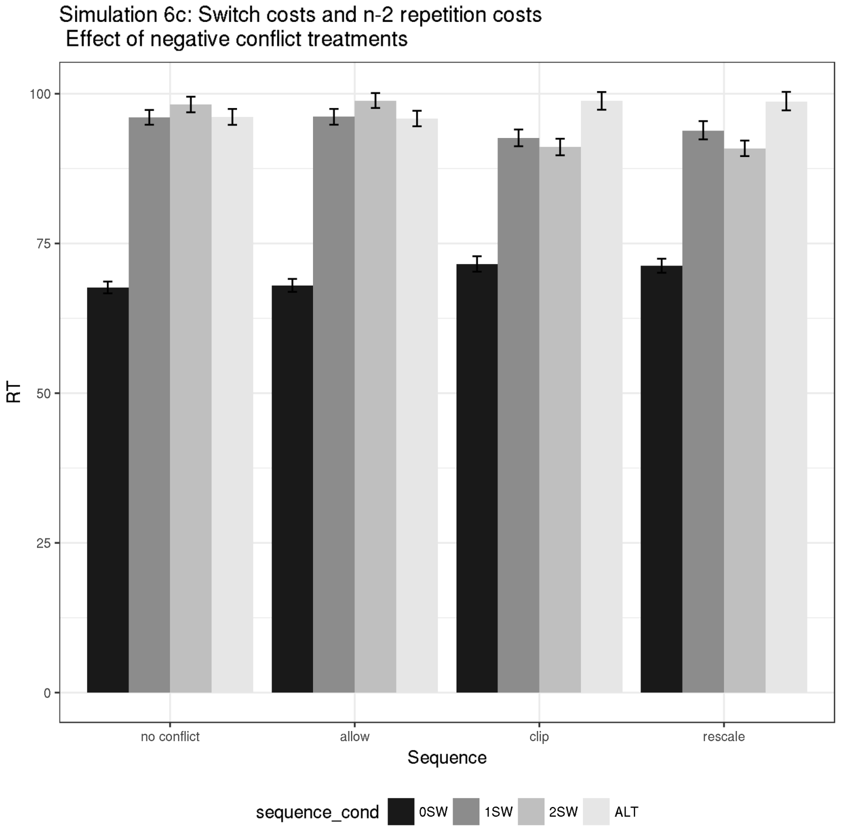


Figure 5.8: Simulation 6c: Switch costs and n-2 repetition costs for various treatments of negative conflict (symmetric tasks). Comparison of no input from conflict monitoring to task demand units (no conflict, i.e., same as simulations 3 and 4), the default where conflict activation ranges from -1 to 1, thus negative conflict excites task demand units (allow), a version where negative conflict is clipped at 0, thus (CM) input to task demand units only occurs when conflict unit activation is above zero, and a version where the activation of conflict units is linearly rescaled between 0 and 1, thus input to task demand units is always inhibitory (with negative weights). Error bars represent 95% confidence intervals.

based on the conflict-monitoring model of Botvinick et al. (2001). This model also highlights the computational similarities between explanations of cognitive phenomena in two relatively distinct empirical literatures. In the model of Gilbert and Shallice (2002), switch cost phenomena are caused by simultaneous activation of the currently relevant task representation and the residually active previous task representation. In the model of Botvinick et al. (2001), conflict is defined as simultaneous non-zero activation of competing representations. In the present model, the conflict units of Botvinick have interactive inhibitory connections with task representations, allowing the model to dynamically inhibit tasks where conflict occurs.

The simulations presented in this chapter have demonstrated that unlike the lateral-inhibition-only model, the backward inhibition model produces n-2 repetition costs, but also a slightly reduced n-1 switch cost (section 5.3). The model also demonstrates that in a three-task paradigm with asymmetric task difficulties, the model predicts asymmetric n-2 repetition costs: specifically, larger costs for switches to the more difficult task (i.e., HEH alternations) and reduced costs for switches to the easier task (i.e., EHE, section 5.4.2). This prediction is in contrast to the limited findings from the existing empirical literature (Arbuthnott, 2008a), and would be falsified if this empirical finding were replicated. Finally, the implementational issue of negative conflict was addressed, with results suggesting that as long as negatively active conflict units are not permitted to inhibit (i.e., excite) task-demand units, the precise implementation does not affect model behaviour.

Thus far, all simulations have been based on hand-set parameters. The model inherits most of its parameters from the previous task switching model of Gilbert and Shallice (2002), where model behaviour was shown to be robust to a wide range of parameter variation. Where possible, the present simulations have used the same parameter settings as the previous model. However, additional parameters are required by the backward inhibition mechanism, and parameters relating to task strengths are required for the simulations of asymmetric tasks. If the model is to demonstrate that the described behaviour is intrinsic to its architecture, and not a specific set of parameters (and that an alternative, opposite pattern of behaviour is not equally consistent with the model), a more systematic approach to parameter setting is required. This approach is the subject of the next chapter.

Chapter 6

Parameter Exploration

This chapter takes the backward inhibition model, which has previously been shown to produce both n-1 switch and n-2 repetition costs as the results of a conflict-driven backward inhibition system, and generalises its performance across a wide range of parameter settings. Specifically, three parameters concerning the conflict system, Bias (β), Gain (γ) and Weight (ω) were varied across a 3-dimensional grid, with qualitative performance patterns (i.e., cost vs. facilitation for n-1 and n-2 switches) mapped across this space. The simulation demonstrates that the empirically observed pattern, the intersection of both n-1 switch and n-2 repetition costs, is widely produced by the model, accepting the parameter constraints that the activation of conflict units must be sensitive to simultaneous activation of competing task-demand units in the model. In the model, error rate costs closely parallel RT costs. However, slower responses are more error-prone than faster responses — an observation which diverges from the empirical data and raises queries about the completeness of the model as an analog of human cognition involved in the task. Additionally, the theoretically meaningless concept of 'negative conflict' is shown to interfere *with model performance, thus versions of the model which prevent any effect of it are considered.*

6.1 Introduction

The previous chapter presented a model of backward inhibition which predicted empirically observed behaviour in two- and three-task switching paradigms. Specifically, the model reproduced the qualitative pattern of n-2 repetition costs and n-1 switch costs. These effects were demonstrated both in highly simplified three-task blocks, and in longer blocks containing multiple switch conditions which are more representative of experiments encountered by human participants (section 5.3).

A major limitation of the simulations presented in the previous chapter is that all were performed by a model using a single set of parameters. Further, the parameters were hand-set, rather than using a more objective or systematic approach. While the architecture of the model was strongly theoretically motivated (drawing from evidence of the role of conflict, and existing conflict-monitoring and task switching models), psychological theory has little to say on the specific parameter settings that form the implementational detail of the model.

Computational modelling in cognitive psychology offers a number of possible approaches to evaluating model performance across a multi-dimensional parameter space. One approach, for example, would be to adopt an algorithmic approach to finding the parameters which produce the best-fit for the model to the empirically observed data. Using this approach, statistics for the fit of the model could be produced, and the ability of different models to fit the data could be compared.

One objection to this approach, in the current context, is the problem of selecting a definitive reference data set. The n-2 repetition cost literature comprises a number of individual studies, which have used different experimental procedures (e.g., 4-choice visual search: Mayr & Keele, 2000; or item classification: Schuch &

Koch, 2003) and different manipulations of the task. Across this range of studies, the magnitude of the n-2 repetition cost varies. Therefore, any reference to empirical behaviour would represent a generalisation across a range of studies, rather than a single quantitative pattern. This makes a quantitative fitting approach problematic. A further objection is that fitting a model to an empirical data pattern does not necessarily imply a good model, if the model could also fit any arbitrary pattern of behaviour with alternative parameter settings (Roberts & Pashler, 2000). Where a model has free parameters, effort should be made to determine how the theoretical content of the model constrains model predictions, and whether there would be any empirical data that would falsify the model (i.e., would be inconsistent with the model's predictions under any parameter settings). The simulations in this chapter therefore pursues a strategy similar to that advocated by Pitt, Kim, Navarro, and Myung (2006), whereby a large portion of the (three-dimensional) parameter space defined by *gain*, *bias*, and *weight* was systematically scanned and partitioned into regions that yielded qualitatively different behaviours.[1] The objective was to understand where in parameter space the behaviour of a model would be qualitatively similar to human participants (i.e., RT costs for n-2 repetitions and n-1 switches) and where it would produce alternative patterns (e.g., n-1 repetition costs and n-2 switch costs, or costs for both n-1 and n-2 switches, or costs for both n-1 and n-2 repetitions).

6.2 Simulation 6d: An exploration of parameter space

The remainder of this chapter presents the analysis of parameter space for the three key parameters that affect the conflict-based backward inhibition system.

---

[1] Note that for simplicity, the *conflict squashing* parameter was not varied in these simulations. However, additional simulations revealed that the effect of varying this parameter was to change the magnitude of the resulting effect sizes but not the qualitative pattern of results.

Additionally, the previous chapter demonstrated that the implementational concept of negative conflict affected model performance. Eliminating this concept from the model appears important for both theoretical reasons and in terms of model behaviour. However the specific means of resolving negative conflict is itself a parameter, and it is important to demonstrate that the specific resolution of negative conflict does not change the overall high-level predictions of the model. The overall performance of the model in terms of RT is considered in section 6.2.2. One advantage of this class of model is that each trial is simulated, producing both a distribution of reaction times, but also occasional errors. Although previous simulations have thus far suggested similar patterns for error rates and reaction times, it remains important to compare both error rates and RTs across parameter space. Error rates are more fully considered in 6.2.3.

Finally, the role of noise within the activation calculations of the model is explored. Noise plays an essential role in the model, not merely in adding variability to the model's behaviour (e.g., producing occasional errors and reaction time distributions rather than simulations proceeding deterministically) but overcoming residual activation in order to produce responses or resolve competition in the model requires a noise component in the activation calculation. It is therefore reasonable to query whether the level of noise affects the behaviour of the model. For example, do higher levels of noise produce larger numbers of errors or slower responses? Do different levels of noise affect higher-level behavioural effects such as switch costs or n-2 repetition costs? For this reason, two different settings of the noise parameter were tested. This section details simulations conducted using a noise parameter of 0.004, while results using the higher noise parameter of 0.006, as used in the model of Gilbert and Shallice (2002), are presented in appendix B.

6.2.1 Method

The current simulation varied three parameters of the conflict system: *gain* (γ, 0 to 100); *bias* (β_c, -40 to 0); and *weight* (ω, -30 to 0). As described above, *gain* and *bias* affect the rate at which conflict unit activation builds up, and decays, respectively. *Weight* affects the amount of biasing that conflict units exert on Task Demand (TD) units. A weight of zero is therefore functionally equivalent to a model with no conflict mechanism (and thus, only lateral inhibition between TD units). The effect that the task inhibition/conflict mechanism has on behaviour is assessed by comparing stronger levels of *weight* with this baseline. For simplicity, simulations were run on blocks of three trials, with the dependent measure being RT in model cycles for the final trial in each triplet. The model was re-initialised after each block. Thus, each block was defined by the sequence of task switches as 0SW, 1SW, 2SW or ALT. Mean RT switch costs and n-2 repetition costs, in model cycles, were calculated from 3000 three-trial blocks of each switch condition, for each point within parameter space.

6.2.2 Results: RT costs

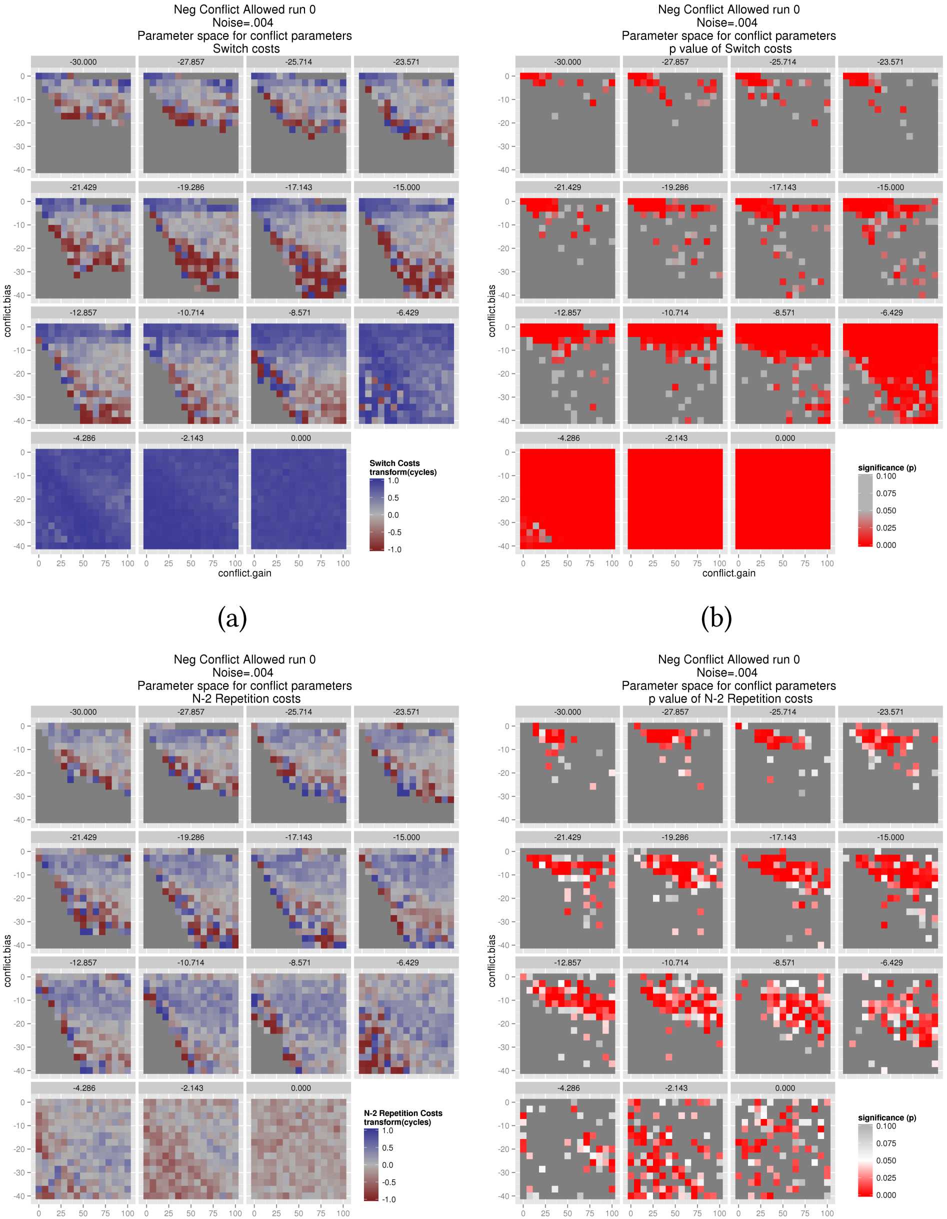
Results for response time (RT) n-1 switch costs and n-2 repetition costs across parameter space are plotted below for three variants of the model. First, where negative conflict is allowed (fig. 6.1), second, where negative conflict is clipped (fig. 6.2), and third, where negative conflict is rescaled (fig. 6.3). For each set of plots, the absolute magnitude of n-1 switch costs and n-2 repetition costs is represented by the left hand plots (upper and lower, respectively). The right-hand shows the statistical significance (p) for a Welch two-samples t-test for each respective voxel (i.e., switch costs as the difference between 0SW and 1SW conditions, n-2 repetition costs as the difference between 2SW and ALT conditions). Dark grey voxels represent no data, indicating that with this combination of parameter settings, the model is unable to correctly complete any trials in at least one condition.

In order to effectively visualise the results using a two-colour palette, the absolute magnitude of the switch cost and n-2 repetition cost was transformed using the logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$.

Switch costs

As demonstrated in previous chapters, switch costs are an intrinsic feature of the model where backward inhibition plays no role. This is also apparent in these simulations: for all three treatments of negative conflict (allowed, fig. 6.1a; clipped, fig. 6.2a; rescaled, fig. 6.3a) where the ω (weight) parameter is zero, positive switch costs are observed, visualised as blue-coloured voxels. For the model allowing negative conflict, switch costs occur for weaker ω (weight) values (i.e., between 0 and -6). For weight values stronger than -6, the model increasingly becomes unable to complete trials in a region with strong β (bias) and weak γ (gain) values, indicated by dark grey voxels. This is because the gain value is insufficient to balance a strong negative bias, leading to a negative saturation of conflict unit activation. This negative conflict activation, which actually excites task demand units (via negatively weighted connections) leads to unstable model behaviour. Aside from increasingly large regions, at stronger ω values, where the model cannot complete trials, parameter space also contains a large region with n-1 switch facilitation, indicated by red voxels, as well as a region where switch costs are close to zero. Clearly, in this treatment of negative conflict, the occurrence of switch costs is highly dependent on specific parameter settings.

For clipped and rescaled treatments, switch costs continue to be robustly predicted for stronger ω values, with two exceptions. For clipped negative conflict, a small region in the upper right of the plot reflects weak β (bias) and high γ (gain) parameter settings. For models with these parameter settings, conflict units tend to accumulate activation leading to a saturation, as a high gain is not influenced by a strong negative bias. For rescaled conflict, switch costs universally occur, except



Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
Switch costs

Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
p value of Switch costs

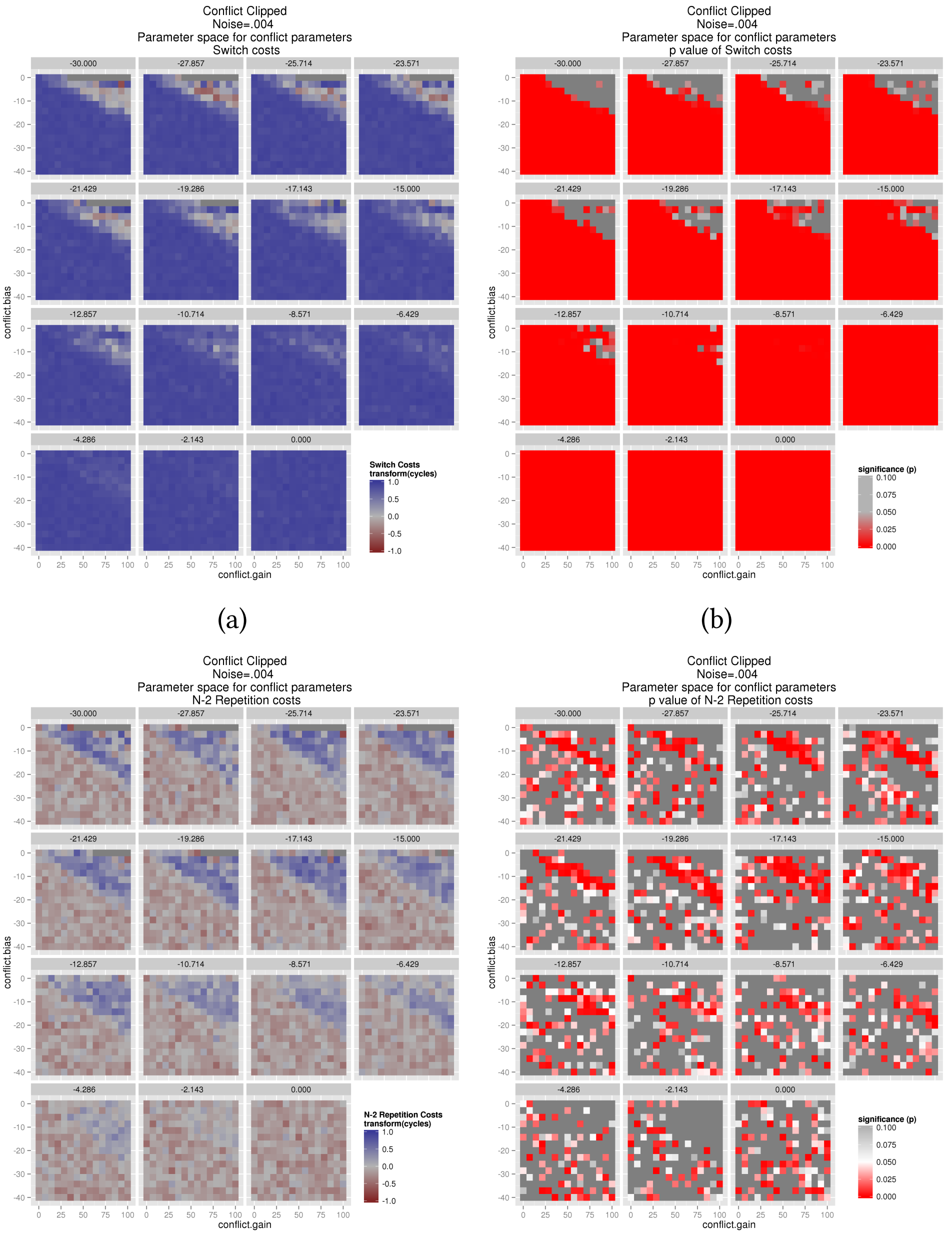
Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
N-2 Repetition costs

Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
p value of N-2 Repetition costs

(a) (b)

(c) (d)

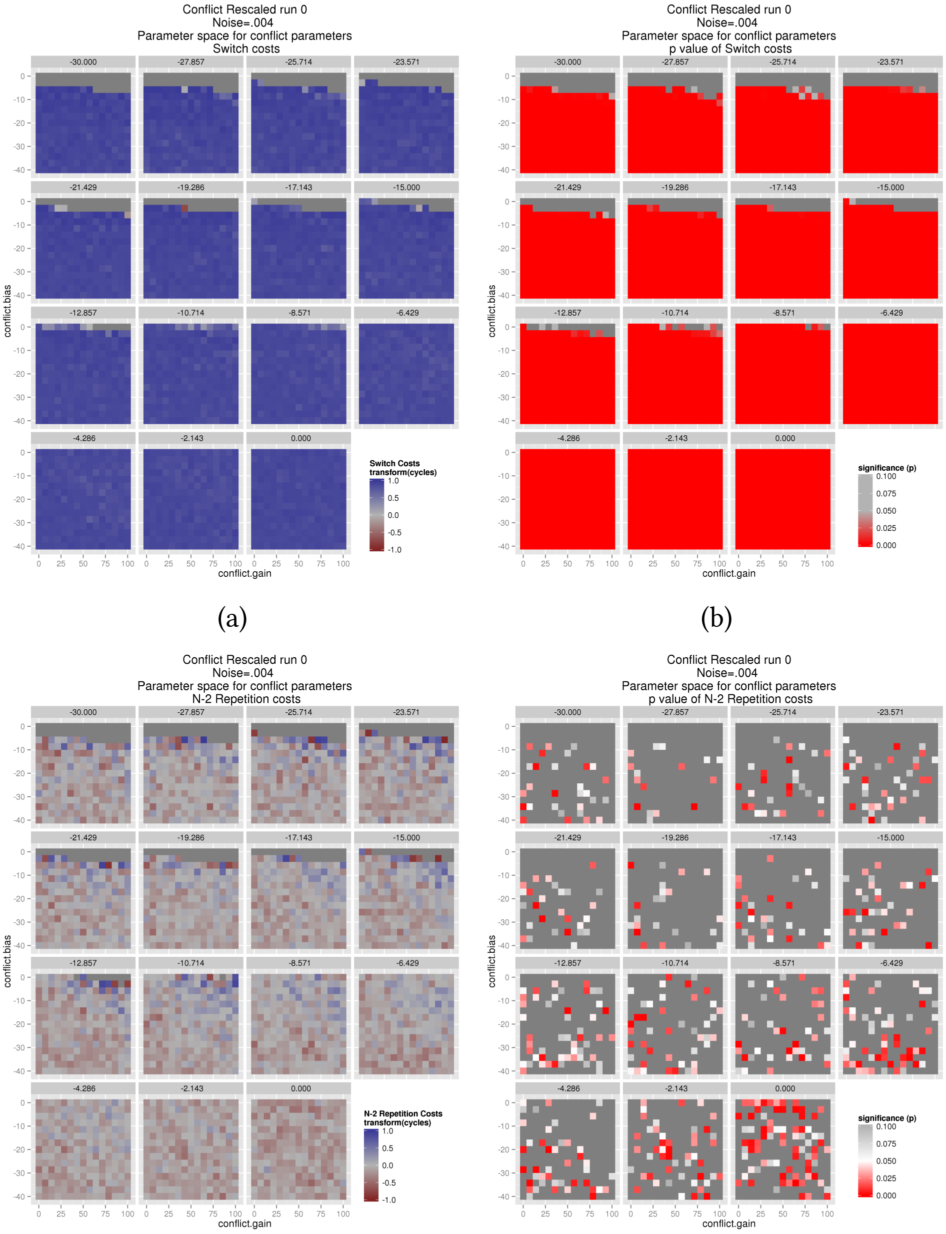
Figure 6.1: Model behaviour (n-1 switch and n-2 repetition costs for RTs) when negative conflict is allowed (noise parameter = .004). Dependent variables (switch costs, 6.1a; n-2 repetition costs, 6.1c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$.



(a) (b)

(c) (d)

Figure 6.2: Model behaviour (n-1 switch and n-2 repetition costs for RTs) when negative conflict is clipped (noise parameter = .004). Dependent variables (switch costs, 6.2a; n-2 repetition costs, 6.2c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$.


Conflict Rescaled run 0
Noise=.004
Parameter space for conflict parameters
Switch costs

Conflict Rescaled run 0
Noise=.004
Parameter space for conflict parameters
p value of Switch costs

Conflict Rescaled run 0
Noise=.004
Parameter space for conflict parameters
N-2 Repetition costs

Conflict Rescaled run 0
Noise=.004
Parameter space for conflict parameters
p value of N-2 Repetition costs

(a) (b) (c) (d)

Figure 6.3: Model behaviour (n-1 switch and n-2 repetition costs for RTs) when negative conflict is rescaled (noise parameter = .004). Dependent variables (switch costs, 6.3a; n-2 repetition costs, 6.3c) are transformed with a logistic function: $f(x) \to \frac{2}{1+e^{(-0.1x)}} - 1$.

for stronger ω and weak β values, where the model cannot successfully complete any trials (dark grey voxels). Thus, for both clipped and rescaled treatments, positive switch costs dominate the parameter space explored by these simulations, except for certain regions where β values are insufficient to balance conflict unit inputs.

N-2 repetition costs

N-2 repetition costs are less robust than switch costs (fig. 6.1c; clipped, fig. 6.2c; rescaled, fig. 6.3c). This is understandable, given that the difference in sequence types occurs on the n-2 trial, and is thus more remote from the trial on which RT is measured, than on an n-1 switch trial.

As previously observed, n-2 repetition costs do not occur when the ω (weight) parameter is zero. As this parameter is increased, all three treatments of negative conflict produce n-2 repetition costs to an extent. When negative conflict is allowed, n-2 repetition costs occur in an initially diffuse, but increasingly coherent region as the weight parameter is increased. This is confined to the top-right region above a diagonal in which performance is unstable (both n-2 repetition costs and repetition facilitation occur in contiguous voxels, producing a “noisy” region of parameter space). Below this diagonal, the model does not complete any trials. As for n-1 switch costs, this is due to the effect of negative conflict on the model: conflict units with a strong β (bias) parameter, and lacking sufficient positive input, become negative saturated, which leads to uniform excitation of the task-demand units. For the clipped treatment of negative conflict, n-2 repetition costs occur in a coherent region only for higher ω values. Again, this region occurs above a diagonal region in the top-right of each subplot, indicating γ (gain) and β (bias) are balanced. As negative activation does not propagate to task-demand units, the region below this diagonal is essentially identical for all values of ω (weight). For the treatment where negative conflict is rescaled, n-2 repetition costs do occasionally occur, but only in sporadic voxels which do not form a coherent, contiguous region.

In summary, although n-2 repetition costs are weaker and occupy a more restricted region of parameter space than switch costs, systematic effects nevertheless occur where negative conflict is either allowed or clipped. For stronger *weight* values, inhibition of TD units by the conflict units produces the empirically observed effect for a contiguous region of parameter space. Note also that, replicating the results of simulation 1, the model does not produce n-2 repetition costs for near-zero *weight* values, demonstrating that lateral inhibition of task representations alone is insufficient to produce n-2 repetition costs.

Intersection of switch costs and n-2 repetition costs in parameter space

From the parameter space maps presented in the preceding sections, both n-1 switch and n-2 repetition costs occur in at least some of the model variants. However, in human participants, the empirical literature suggests that normal behaviour produces both n-1 switch and n-2 repetition costs. The extent to which these two regions overlap is considered in this section. The overlap of n-1 switch and n-2 repetition costs is assessed quantitatively by taking the squared geometric mean of the switch cost and n-2 repetition cost (i.e., the product of the two costs), where these costs are above zero. The resulting measure was then transformed with the same logistic function as in previous plots to a range of between 0 and 1. This is plotted in figure 6.4 for all treatments of negative conflict. It is clear that for all treatments, there exists some region of parameter space conforming to the empirical pattern of data.

However, the parameter space for the model where negative conflict is allowed suggests that where negative conflict saturation occurs, the model simply ceases to produce correct responses. As negative conflict is a property of the specific implementation in the current model, rather than any theoretically grounded concept, future simulations disregard the variant of the model in which it is allowed to occur. Additionally, for the rescaled treatment of negative conflict, although n-2 repetition costs do occur, the previous section suggests that few contiguous voxels represent regions where the cost reaches an $\alpha = .05$ level of statistical significance.

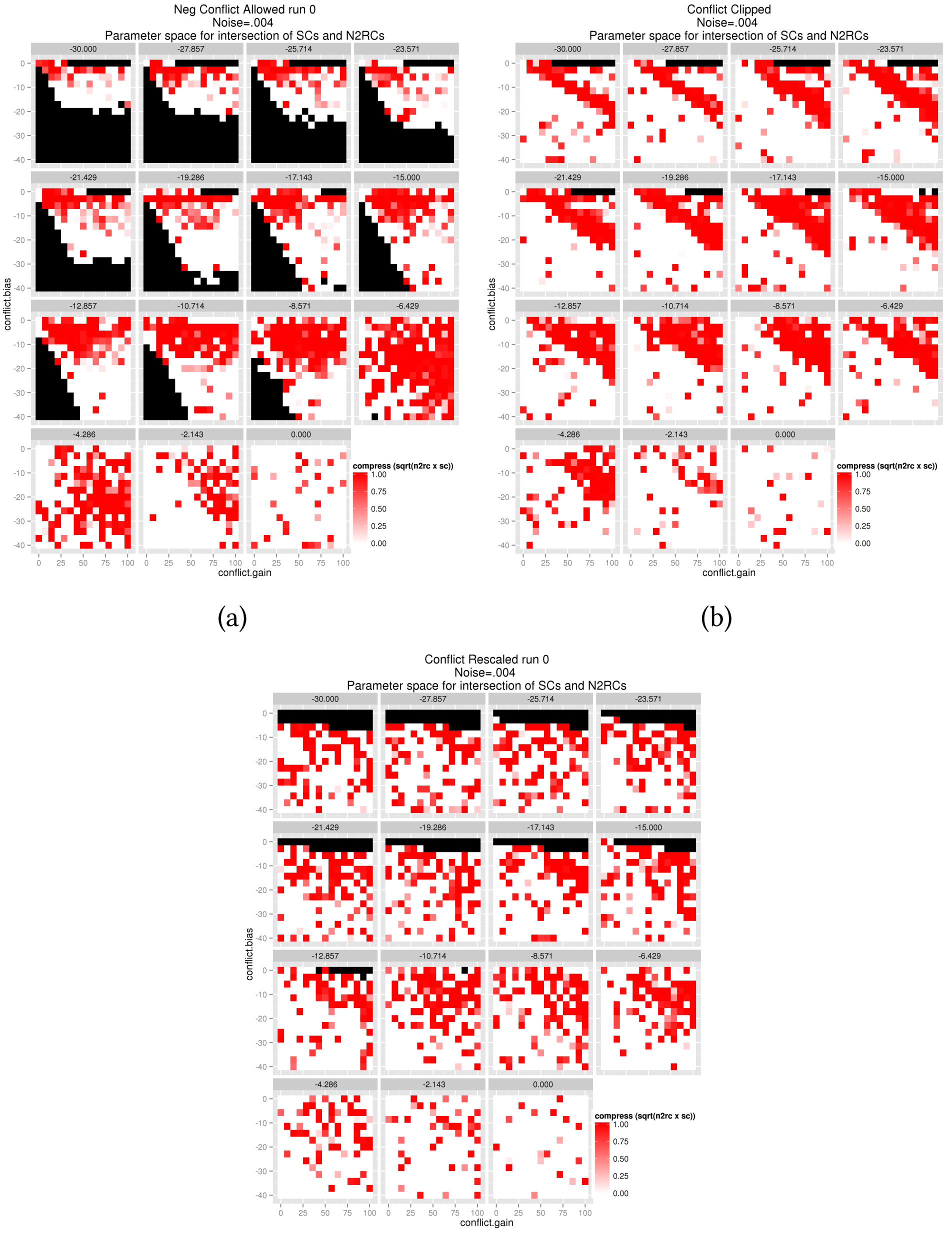
The empirical pattern of data — i.e., an overlap of n-2 repetition costs and n-1 switch costs — can to an extent be considered independent from the exact treatment of conflict. However, this simulation has demonstrated that the version of the model which clips negative conflict best predicts the empirical pattern in a coherent, contiguous region of parameter space. In this treatment of conflict, there also exists a region (below the diagonal) where the model produces positive switch costs but negative n-2 repetition costs — that is, n-2 repetition facilitation. In terms of the criticisms of Roberts and Pashler (2000), the model as it stands can be criticised on the grounds that it does not uniquely predict the empirical pattern of data — with appropriate parameter settings, it could just as easily produce switch costs with n-2 repetition facilitation, although there are some constraints on parameter settings: the empirical region roughly corresponds to a plane following a negative relationship between conflict bias and conflict gain. Informally, gain and bias are required to be balanced in order for conflict monitoring units to be effective. As these two parameters (with this activation function) together control the sensitivity of these units to input and the rate at which their activation decays, it is important that they be appropriately balanced, in order to avoid the units settling into a totally-off (too much negative bias and too little gain) or totally-on (too much gain and too little negative bias) state. The extension of this region across multiple values of weights for conflict to task demand units (except the weakest weights, i.e., weights between zero and -3) suggests that the empirical effect is independent of any specific weight value.

In summary, the two empirically observed phenomena, costs for n-1 switches and n-2 repetitions (fig. 6.4) co-occur in a well-defined region for non-zero *weight* values. Informally, this behavioural pattern is obtained with the constraints that the activation of conflict units must increase given conflict and decrease given lack of conflict. Outside these regions, other behaviour (e.g., switch costs but n-2 repetition facilitation) may be understood either in terms of implementational failure of the model (the parameter settings do not implement the theory of backward inhibition) or in terms consistent with theory (TD unit processing must be biased by inhibitory connections from conflict units). The empirical pattern is produced by any set of parameters in which the model functions according to these theoretical constraints.

Despite this success in matching the empirical pattern of human behaviour, it remains to be demonstrated why this general architecture should be beneficial in task switching. Why should human processing of multiple tasks have a conflict monitoring system which effectively produces slower responses under certain circumstances, unless it provides benefits as well? One possibility is that the human cognitive system adaptively seeks this region in order to minimise the resources, or effort, required to perform the task. Another possibility is that conflict monitoring functions as an error-control system — i.e., it biases the task switching system into a more stable region, more likely to produce correct responses. If either is the case, we might expect this region to also correspond with lower RTs or error rates overall. The next sections address these possibilities by searching the same regions of parameter space, looking at error rates and basic RTs (i.e., RTs on individual trials rather than RT differences on n-1 switch or n-2 repetition trials).

6.2.3 Results: Error rates

The following section addresses error rate performance using the same grid search of parameter space. Given the results of the previous section — that clipping appeared to be the best way of resolving negative conflict in the model — for simplicity, only this treatment of negative conflict is considered in this section. Given the simulation consisted of blocks of 3 trials, error rates, were assessed in two different ways. In calculating error n-1 switch or n-2 repetition cost, only errors which occur on the third trial, where trials 1 and 2 were correct, were considered. This is referred to as $Errors_{(3)}$. Overall error rates (i.e., for assessing global model



(a)

(b)

(c)

Figure 6.4: Graphs showing intersection of regions of parameter space producing switch costs and n-2 repetition costs, for all 3 treatments of negative conflict, based on the above (low noise) simulations. The plotted variable is the product of the positive part of the switch costs and n-2 repetition costs (i.e., negative costs are substituted with zero). This value is then transformed with the logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$.

performance), on the other hand, should take into account errors irrespective of where they occurred. However, as error trials may affect subsequent trials through abnormal residual activation, only the first error in each triplet should be assessed. $Errors_{(1,2)}$ therefore refers to the number of blocks where an error occurred on either the first or second trial. Thus, $Errors_{(1,2)}$ and $Errors_{(3)}$ are cumulative.

Figure 6.5 shows overall error rates in trials 1 and 2 for all conditions. This figure gives an overall impression of error performance across parameter space that is not related to a specific condition. Note, however, that this is not a pure measure — the 0SW (ABB) condition is generally more error prone than the 1SW (AAB) condition. This is because the switch in the 0SW condition occurs on the second trial, whereas the switch occurs on the third trial in the 1SW condition, and is thus not included in this measure. In this conflict condition, errors are restricted to the region at the 'top' of the parameter space (i.e., weak bias), for ω (weight) values stronger than -5. Very low error rates (i.e., 1% or below, visualised in black) occur in the region of parameter space below a weak bias/low gain to intermediate bias/high gain diagonal. Error rates increase around this diagonal, and become very high (i.e., exceeding 10%) at the top of the plot (corresponding with a weak bias).

When compared to the corresponding plot (figure 6.4b), 'intermediate' error rates of around 5% occur in a similar region to the intersection of n-2 repetition and n-1 switch costs. This is understandable in terms of the build up of conflict in the model: The gain and bias parameters balance activation in the conflict units, such that it builds up quickly in response to conflict, and decays in the absence of conflict. Thus, if activation is allowed to accumulate in the conflict units, performance, in terms of global errors, deteriorates. Similarly, activation in the conflict units only affects performance if the weight of the connection from these units to task demand units is sufficiently strong to bias performance, thus for weak conflict-task demand weights, there are no systematic differences in error rates across bias-gain space. Thus, for this treatment of conflict, the differences between different patterns of global error behaviour can be understood in terms of implementational details of the conflict monitoring units. In the space informally described by a 'balance of bias and gain', the model tends to produce uniform error rates. However, it remains to be seen why the empirically observed pattern of behaviour should correspond with an intermediate level of error performance. If the model were optimised to absolutely minimise the number of errors, better performance would be obtained with stronger β or zero ω parameters — i.e., parameters which ensure the conflict units do not bias model performance. As such, global error performance of the model deteriorates with stronger backward inhibition.

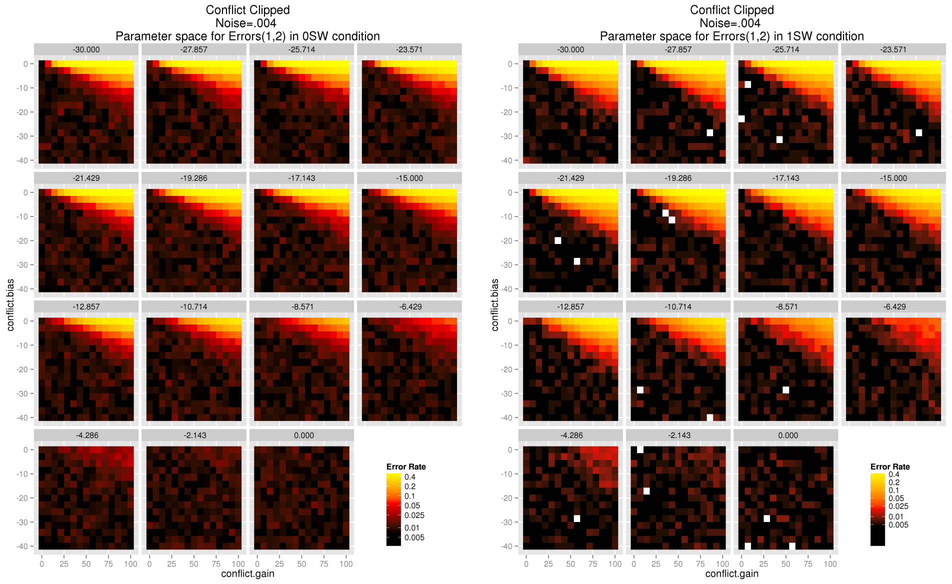
Figure 6.6 shows error rates on trial 3, for triplets where there were no errors on trials 1 or 2, for 2SW and ALT sequences. Graphical inspection suggests that overall performance is broadly similar to errors occurring on trials 1 and 2 (fig. 6.5). Systematic differences due to the task sequence (i.e., costs) were assessed by the difference in error rates between corresponding voxels of parameter space. Error rate switch costs and n-2 repetition costs are plotted in figure 6.7, together with the intersection of these costs, calculated in the same way as for reaction times, above. This figure suggests that for both n-1 switch and n-2 repetition costs, error rates parallel the reaction time data. Positive error rate n-1 switch costs are observed over the majority of parameter space, with n-1 switch facilitation occurring only for very weak β and high γ at strong ω values, similar to reaction time n-1 switch facilitation. Similarly, error rate n-2 repetition facilitation is more prevalent below a diagonal region balancing β and γ parameters, with costs occurring above this region, paralleling reaction time costs. Although error rate data is, by its nature, more stochastic than reaction time data, and therefore these visualisations are somewhat 'noisier' than for reaction times, the intersection of error rate switch costs and n-2 repetition costs (figure 6.7c) is recognisably similar to that for reaction times (figure 6.4c), with the overlaps between the two empirical phenomena occurring in the same region of parameter space.

6.2.4 Results: Basic RT

The previous section, concerning error rate data, suggests that the region of parameter space corresponding with both n-1 switch and n-2 repetition costs for both error rates and RTs is also associated with an intermediate error rate (i.e., between 1% and 10%) rather than a very low or high rate. However, as yet, the model does not suggest why this is the case. One possibility is that of a speed-accuracy trade-off. If faster responses were more error prone, with slower responses more accurate, then selection of a region in parameter space where speed of responding were balanced against accuracy would provide a reasonable explanation.

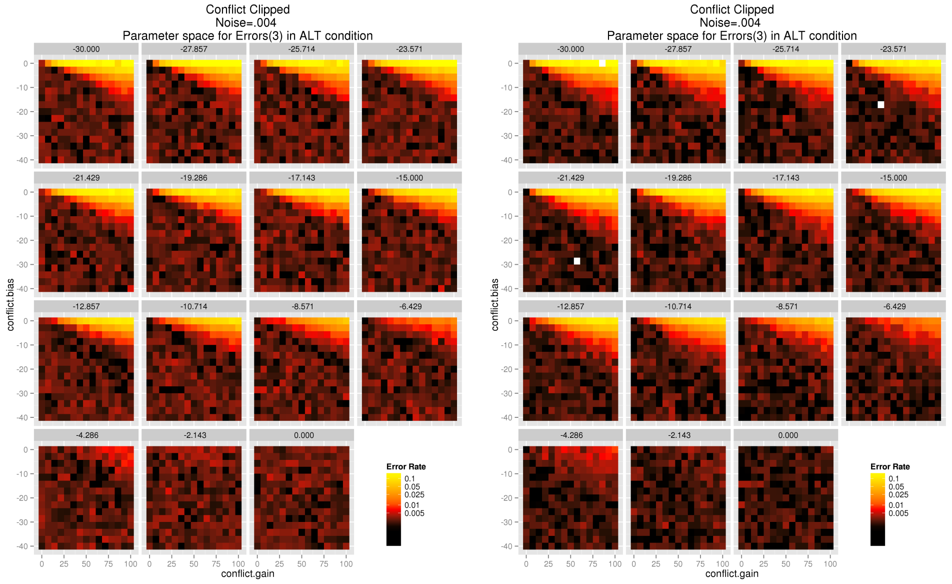
While the preceding sections have dealt with differences in RTs — costs — between conditions, this section considers basic RT (in processing cycles) across parameter space. Figure 6.8 plots RTs for 0SW and 1SW conditions. RTs are visualised across a colour spectrum, with faster RTs plotted in blue, and slower RTs in red. As we are not, here, concerned with RT differences between conditions, and basic RTs are very similar for all four conditions, we shall take these two conditions as representative.

Examination of the global RT data shows slowing in the upper right region of the plot: the region associated with n-2 repetition costs. Thus, fastest responses are obtained where either the ω (weight) parameter is very low, or the β bias parameter is low enough to ensure that backward inhibition does not occur. Slower responses, in the model, also correspond with more error prone regions. Given these results, there is no evidence for a speed-accuracy trade-off across the region of parameter space defined by β, γ and ω. Overall, best performance (in terms of error rates and low RTs) occurs in sections with lower conflict/task-demand weights. The presence of conflict units in these trials therefore leads to poorer performance across all metrics, raising the question of what the functional benefit of conflict monitoring units might be.



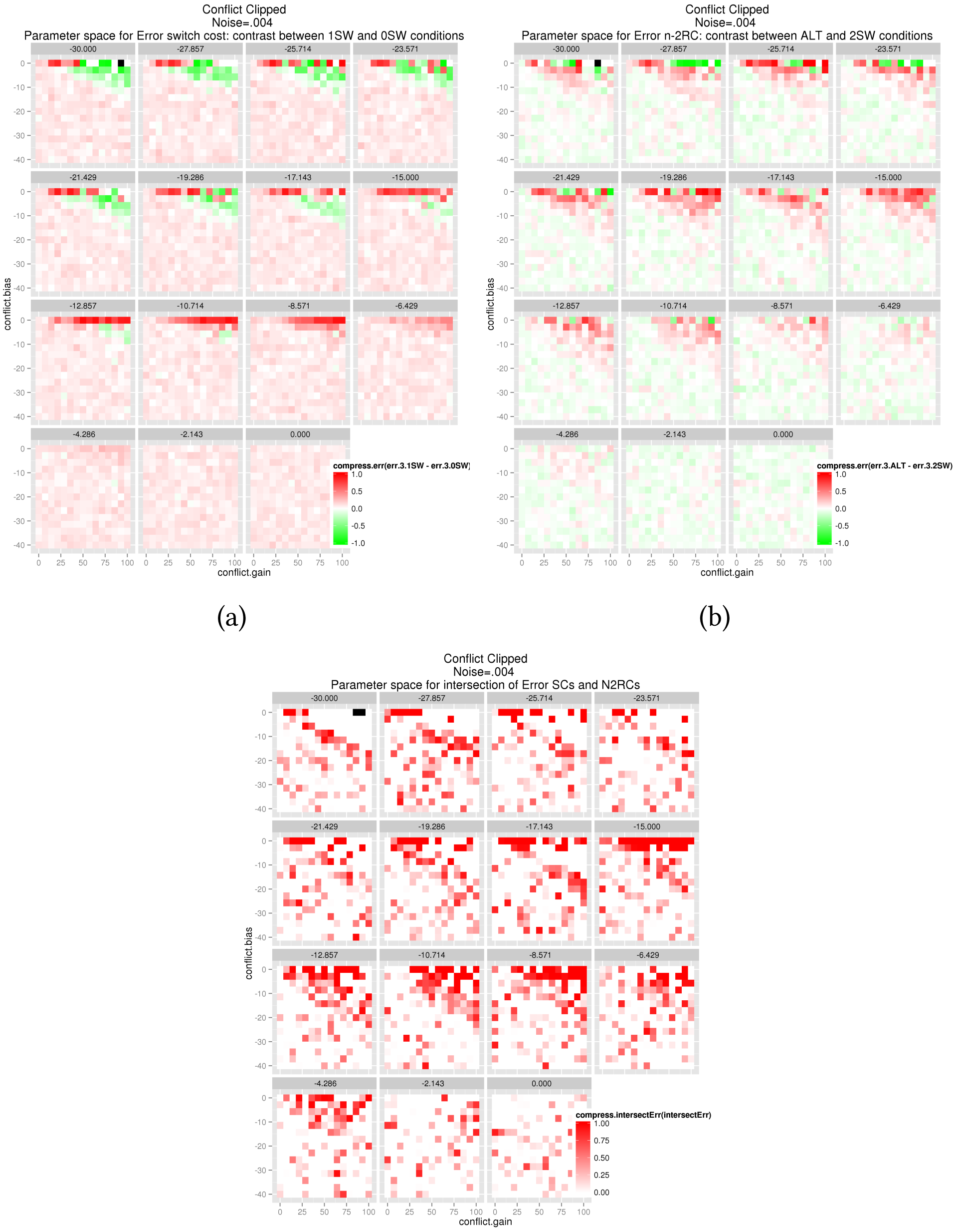
(a) (b)

Figure 6.5: Gridsearch, showing model behaviour, assessed via error rates (occurring on trials 1 or 2), for clipped conflict, low noise (noise parameter = .004), in the 0SW (6.5a) and 1SW (6.5b) conditions.



(a) (b)

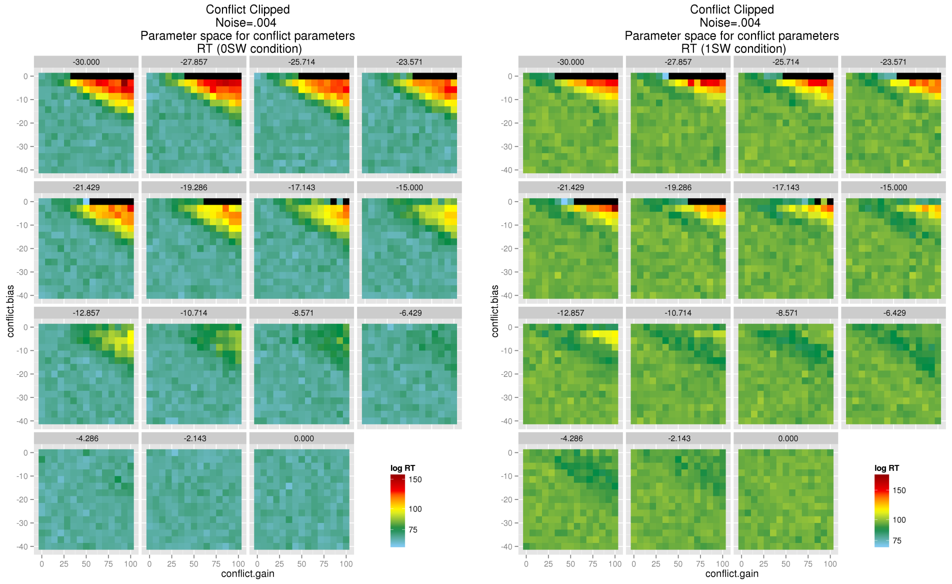
Figure 6.6: Gridsearch, showing model behaviour, assessed via error rates on trial 3, for clipped conflict, low noise (noise parameter = .004), in the ALT and 2SW conditions (6.6a) and 1SW (6.6b) conditions.



(a) (b)

(c)

Figure 6.7: Model behaviour, assessed via error rates on trial 3, for clipped conflict, low noise (noise parameter = .004). Plots show differences between the 1SW and 0SW conditions (i.e., the error switch cost, panel 6.7a) and between the ALT and 2SW conditions (i.e., the error n-2 repetition cost, panel 6.7b). The differences are plotted such that the empirical pattern of behaviour (i.e., 1SW $>$ 0SW, ALT $>$ 2SW) is positive, while the alternative (i.e., 1SW $<$ 0SW, etc.) is in green. The differences in error rates are transformed with a logistic function, $f(x) \rightarrow \frac{2}{1+e^{(-100x)}} - 1$, before plotting, and the actual magnitude of the differences in error rates is very small. The overlap between error rates is assessed as the product of the positive (i.e., red) part of parameter space for SCs and n-2RCs, and is presented in panel 6.7c.



(a)

(b)

Figure 6.8: Gridsearch, showing RT perforamnce for 0SW (6.8a) and 1SW conditions (6.8b). RTs have been plotted on a log scale.

6.3 Simulation 6e: The magnitude of noise

One further query concerns whether the reduced noise (.004 in the simulations reported above rather than the default of .006 used by Gilbert & Shallice, 2002) had any effect on qualitative model behaviour across parameter space — i.e., whether it significantly changed behaviour from a delineated region of space from error switch costs to switch facilitation, or whether the shape of the plotted surface was substantially different to the low noise condition. In order to explore this possibility, the simulations were re-run with noise at the previously used (default) value of .006, and almost identical graphs were produced. Assessed graphically, the main effect of the higher noise parameter was to increase response times, however error plots were strikingly similar. These graphs are contained in appendix B.

6.4 Discussion

In summary, six key findings from the preceding parameter space studies have clarified our understanding of the model. Firstly, the presence of negative conflict in the model interferes with 'proper' operation, both in theoretical terms, and in practical terms of stopping the model from correctly completing trials at all, for certain parameter settings. Secondly, the model nevertheless produces both n-2 repetition costs and n-1 switch costs for all three treatments of conflict. However, clipping negative conflict, rather than rescaling it, produces more coherent behaviour, in terms of grouping similar patterns of performance into contiguous regions of parameter space. Therefore, this version of the model was preferred. Thirdly, for this version of the model, regions of parameter space which did not correspond to the empirical pattern, can be understood in terms of the model's implementation. Specifically, if the activation of conflict units was allowed to saturate, due to very negative bias or very positive gain parameter settings, conflict units became insensitive to the specifics of ongoing trial processing. In these cases, either model performance broke down entirely, or the backward inhibition mechanism was ineffective. However, accepting the constraints that the conflict unit activation must increase following co-activation of task-demand units, and decrease without co-activation, the model produces the empirically observed pattern of behaviour. We may therefore conclude that the empirical pattern is due to the model's architecture and theoretical constraints, rather than specific parameters. Fourthly, error rate n-1 switch and n-2 repetition costs paralleled reaction time costs across parameter space. Fifthly, the 'empirical' region of parameter space was not associated with the most efficient model performance, in terms of the fastest or most accurate responses. Instead, the fastest, most accurate responses were obtained when the backward inhibition mechanism did not affect task processing, such as where the weight parameter was zero. This leaves open the question of what functional advantage is provided by backward inhibition, if it does not appear to improve performance. Sixthly, model performance does not appear to parallel human behaviour in terms of error commission. Specifically, for humans, faster responses tend to be more error prone, while slower responses are more accurate, giving rise to the concept of the speed/acccuracy trade-off.

One striking feature of these results is that RT and error performance correlate across parameter space. That is, there is no evidence that variation of these particular parameters produces a speed/accuracy tradeoff. This is not to suggest the model is incompatible with the idea of a speed/accuracy tradeoff, merely that manipulation of these three parameters does not produce such a tradeoff in itself. If the model is regarded as representative of human backward inhibition, these results also leave open the question of what functional advantage is provided by backward inhibition. The region of parameter space corresponding with the empirical behaviour pattern is both slower, and more error prone, than regions where backward inhibition does not occur.

Overall, however, the model's production of errors diverges from human behaviour: in the model, slower responses are also more error prone. In the empirical literature, slower responses are more accurate (Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988), with slowing occurring on trials that follow errors (Botvinick et al., 2001). One possibility is that the response production mechanism in the present model is too simplified to account for phenomena such as a speed/accuracy tradeoff. In the model, responses are assumed to be produced only when one response node becomes clearly more active than the others. Without a further response execution stage, in which representations of the response are turned into motor representations that produce the physical action of responding, it is difficult to account for other backward inhibition phenomena, such as the influence of a go-no-go signal (Philipp, Jolicoeur, et al., 2007; Schuch & Koch, 2003). An alternative possibility is that the way in which conflict units bias task demand units has been oversimplified — in this model, conflict units inhibit both conflicting task demand units equally, relying on top-down control to provide an increased signal to the selected task. In the model of Brown et al. (2007), intra-trial conflict serves as a signal for slowing the rate of response production. It is possible that conflict, as operationalised in the present model, is also the signal for adjusting some other parameters governing response production, such as the threshold for producing a response, or the monitoring of responses before they are produced. Without theoretical grounds for adding a more elaborate response mechanism to the model, addressing these phenomena remains problematic within the present model.

One scenario in which conflict might be expected to facilitate performance is in over-riding of a prepotent task to perform another, weaker task, such as colour naming of Stroop stimuli. Indeed, Stroop stimuli are routinely cited in the conflict monitoring literature as a source of response conflict (e.g., Botvinick et al., 2001). The triggering of conflict and the subsequent suppression of the irrelevant (prepotent) task should facilitate performance, especially on trials with a high level of conflict (e.g., switch trials). This scenario is addressed in the next chapter.

Chapter 7

Asymmetric Switching in the Three-Task Paradigm: Model Predictions

The parameter-space scanning approach to model performance used in the previous chapter is employed to generate predictions for switching in three-task sequences of asymmetric difficulty. This chapter describes two parallel simulations, in which the two parameters affecting task difficulty are varied together with the parameter affecting the strength of backward inhibition. The two simulations concern an additional modelling assumptions necessitated by the implementation of asymmetric tasks; specifically, whether the strength of backward inhibition is identical for all tasks, or whether it varies depending on the strength of the task. In the first simulation, the model is identical with that used in the previous chapter, while in the second simulation, backward inhibitory weights are scaled in proportion to the top-down control strength of each task. The simulations demonstrated a clear prediction that backward inhibition produces stronger n-2 repeti-

tion costs for hard-easy-hard (HEH) alternations than easy-hard-easy (EHE) alternations. This, and the general effects of backward inhibition, can be understood in terms of the conflict generated on the n-1 trial.

7.1 Introduction

The previous chapters have demonstrated that that a conflict mechanism is sufficient for explaining n-2 repetition costs in human participants, assuming tasks of identical difficulty. However, while they demonstrated that conflict monitoring units produced RT and error n-2 repetition costs in addition to switch costs, and will be shown to predict other empirically observed effects, such as the dependency on specific intertrial intervals (section 9.2), they did not demonstrate a functional benefit that such an architecture brings to task performance. That is, performance typically deteriorated as the weights from conflict to task demand units became stronger, with the best (fastest, least error-prone) performance in the condition where the weights were zero. From this simulation, then, while the conflict units provide a good fit to the empirical data, it is unclear what benefit the conflict units provide to performance. Additionally, although the n-2 repetition literature has predominantly assumed tasks of equal difficulty, in reality human performance may differ between tasks, either as a group, or with individual differences.

Therefore, using the model (of which a close relative has already been successfully applied to the domain of two-task asymmetric switching: Gilbert & Shallice, 2002) to simulate performance in asymmetric switching (i.e., whether n-2 repetition costs vary systematically for the easy or the hard tasks) would be a relatively novel paradigm for validating model predictions against empirically observed human behaviour.

The remainder of this chapter concerns backward inhibition when switching between three tasks of asymmetric difficulty. The model of Gilbert and Shallice (2002) demonstrated that asymmetric n-1 switch costs occur because a greater amount of task demand activation is required to execute a harder task than an easier task, and therefore a greater amount of this activation is carried over to the next trial. This in turn causes greater cross-task interference (i.e., conflict). Consequently switch costs are greater for hard-easy switches than for easy-hard switches. In the case of three-task switching, one might similarly anticipate a high level of task conflict, especially when switching toward the easier task. As backward inhibition is sensitive to task conflict, one might also expect differential levels of backward inhibition to occur. Specifically, when alternating easy-hard-easy, one might anticipate a lower level of conflict would be produced on the middle (n-1) trial, than when alternating hard-easy-hard, and therefore greater backward inhibition would be observed in the latter case.

This chapter extends the asymmetric task paradigm to three task switching, manipulating the between-trial conflict by using two tasks of identical, fixed, intermediate difficulty, while varying the difficulty of the third task.

7.2 Empirical studies of asymmetric switching

To date, only one study has systematically examined n-2 repetition costs for asymmetric tasks. Arbuthnott (2008a) used a digit classification variant, in which the three tasks were judgments about magnitude (low/high), parity (odd/even) and whether the digit was prime (prime/multiple), in increasing order of difficulty as assessed through RT and error rates. In her procedure, the order of tasks was fully randomised for each participant, and each trial was classified by the sequence of switches in relation to previous tasks — possibilities were 0SW (e.g., BAA), 1SW (e.g., BBA), 2SW (e.g., CBA) or ALT (e.g., ABA). Thus, a single procedure was capable of obtaining values for n-1 switch costs (the RT difference between 1SW and 0SW trials) and n-2 repetition costs (the RT difference between ALT and 2SW trials). Arbuthnott reasoned that if the asymmetric n-1 switch cost was due to residual inhibition on the switch trial (that is, greater for hard-easy (HE) switches than easy-hard (EH) switches, because executing the H task requires greater inhibition of the E task than vice versa, and this carries over to subsequent trials), and if this same mechanism was responsible for backward inhibition, then it would lead to greater n-2 repetition costs for EHE than HEH alternations. Alternatively, if the asymmetric switch cost was due to residual priming of the harder task (i.e., greater priming for HE than EH switches), then this would have no differential effect on n-2 repetition costs for EHE compared to HEH alternations[1].

In two experiments, involving either separate or overlapping response sets respectively, Arbuthnott (2008) observed asymmetric n-2 repetition costs, with greater costs for EHE triplets than HEH triplets. That is, the n-2 task received greater backward inhibition when it was easy than when it was hard. This was consistent with Arbuthnott's hypothesis that backward inhibition was caused by the same mechanism as the n-1 switch cost, namely residual inhibition of the preceding task. However, the effect on RTs was not robust, and only reached statistical significance for one pairing of tasks (magnitude-prime switches, not parity-prime switches), and then for non-overlapping response sets only. Additionally, unexpected effects occurred, such as the reversal of direction of the n-1 switch cost asymmetry for magnitude-prime tasks; that is, switching to the prime (harder) task incurred a greater n-1 switch cost than switching to the magnitude (easier) task, a finding inconsistent with the literature on two-task switch costs. It is difficult to see why, if asymmetric backward inhibition is the same mechanism as responsible for asymmetric n-1 switch costs, the direction of the latter effect should be reversed when the former effect is in the hypothesised direction. Moreover,

---

[1] Note that this reasoning assumes that asymmetric n-1 switch costs are exclusively due to either inhibition or activation (priming) of task sets. In models of task switching incorporating inhibitory connections (Brown et al., 2007; Gilbert & Shallice, 2002), activation and inhibition are dynamically interactive, and thus not easily separated.

this pattern of findings is difficult to fully explain using only 'inhibition' as an explanatory mechanism, without recognising the potential dynamic interaction of excitatory and inhibitory processes.

Additionally, the procedure of Arbuthnott (2008a) contained a number of methodological concerns, which are addressed more fully in the following chapter (section 8.1.1). Applying the model to this paradigm would therefore seem potentially informative given the seemingly non-robust nature of these results, potential methodological issues which may affect the results, and the lack of subsequent published replications.

Any model predictions may be compared with the results of Arbuthnott (2008a), either corroborating them (providing a modelled, theoretical explanation alternative to the reasoning in that paper) or producing an alternative prediction. These predictions may then be tested in a conceptual replication of the asymmetric 3-task paradigm that addresses certain methodological concerns to obtain a result which would either validate or falsify the present model.

7.3 Simulations 7 and 8: Methods

As in the model of Gilbert and Shallice (2002), task difficulty in the present model is specified by two parameters — *stimulus input strength (SIS)*, representing the automatic, bottom-up activation of a response by a stimulus (greater for stronger tasks), and *top-down control strength (TDCS)*, specifying the control needed to ensure the task is performed (greater for weaker tasks). The following simulations varied the *SIS* and *TDCS* parameters of task B together with the *weight* parameter of the conflict layer to create a three-dimensional space. The *bias* and *gain* parameters of the conflict layer were fixed at -10.0 and 75.0 respectively, while the task parameters (*TDCS*, *SIS*) of tasks A and C were held at their default values.

For asymmetric tasks, each task sequence (e.g., 0SW) has various permutations

— (e.g., ABB, BAA and BCC). Here, we considered only switches from task B (variable *SIS* and *TDCS*) to task A (fixed). Hence, 0SW sequences are all BAA, 1SW are BBA, 2SW are CBA and ALT are ABA, with only the n-1 task being of variable difficulty (except for the 0SW condition). By varying the parameters of task B, the following simulations therefore test the effect on behaviour for both hard-easy-hard (HEH) and easy-hard-easy (EHE) switches.

Similarly to the simulations reported in the previous chapter, parameter space was mapped in 3 dimensions. Top-down control strength, stimulus input strength and conflict inhibitory weight were mapped on the X, Y and Z axes respectively. By systematically varying task parameters for one task, the 'task strength' (a composite of stimulus input and top down control that gives an optimal level of control) was manipulated such that tasks A and B/C could be varied between easy-hard and hard-easy. An intermediate level of input strength was therefore selected for tasks B and C.

By varying *SIS* and *TDCS* factorially, such that for either variable, task A can be greater or less than B, four quadrants are effectively created. $SIS(A > B)$, $TDC(A > B)$ (i.e., the upper-right quadrant of the plots) is the region in which the both the input and control strength are greater for task A. Effectively, the task is over-controlled, such as might occur in a particularly high-pressure situation or one in which mistakes are very costly. $SIS(A > B)$, $TDC(A < B)$, the upper-left of the plot, represents the region in which task A is stronger, but less controlled, than task B, as is typical for a stronger task (e.g., word reading compared to colour naming). $SIS(A < B)$, $TDC(A > B)$ is the region in which A is weaker, but more controlled, than task B, as is typical for a weaker task (colour naming compared to word reading). Finally, $SIS(A < B)$, $TDC(A < B)$ represents the under-controlled region, where a weak task is coupled with weak control, such as might correspond to an absent-minded or under-controlled task performance. One might expect performance to deteriorate markedly in this quadrant. Given that error costs have been shown to parallel RT costs previously, for simplicity RTs are the only dependent variable considered in these simulations.

The implementation of asymmetric tasks requires an additional modelling assumption. Specifically, whether the weight of the conflict unit-task demand unit connection is independent of, or related to, the level of top-down control used to perform a task. Two parallel simulations were therefore conducted. In simulation 7, a simplifying assumption is that the inhibitory *weight* parameter is identical for all three tasks. This may be a problematic implementation, however. If conflict units are to facilitate performance by biasing task demand activation in response to between-task conflict, then it should benefit performance most in situations where more than one task demand unit is highly active. In the task switching model, this occurs on switch trials, due to residual activation of the previously relevant task, and especially on easy switch trials, compared to hard switch trials, due the the longer time it takes the weaker *TDCS* for the easy task (e.g., word reading) to overcome the residual activation, compared to the stronger *TDCS* for the harder task (e.g., colour naming) (see Gilbert & Shallice, 2002, figure 5). This feature of the model's activation dynamics is responsible for the characteristic asymmetry in switch costs, with the cost for switches to word reading being greater than that for switches to colour naming. Therefore, if conflict units suppress task demand activation when both tasks are active, one might expect switch costs to be smaller for stronger conflict-task demand strengths.

In the present implementation of the model, used in simulation 7, conflict units send an equal inhibitory signal to both conflicting task demand units, given that the inhibitory signal sent to the currently relevant TD unit will be counteracted by a top-down control input for the currently active task. In cases where tasks are of asymmetric difficulty, the stronger task (i.e., weaker *TDCS*) will in effect be inhibited more by a given conflict signal than the stronger task. Thus, switching to the weaker task may not benefit as much from a conflict signal (especially a strong conflict signal) as would be anticipated if the inhibitory effect of conflict on task demand units was more selective.

To address this possible confounding effect, in simulation 8 the parameter controlling the weight of the connection from conflict to task demand units, rather than being used directly as the weight for all connections, is instead multiplied by the *TDCS* parameter to generate unique weights for each TD unit. Thus, an overall conflict weight of 1.5 to TD unit A (with a TD control strength of 12) would have a weight of 18, while the same weight to TD unit B (with TD control strength of 7) would be 10.5. Effectively, this implementation normalises the strength of the inhibitory conflict signal against TD control strength. In other respects, simulation 8 was identical to simulation 7. The results of both simulations are discussed in parallel, below.

7.4 Simulations 7 and 8: Results and discussion

7.4.1 Basic RT

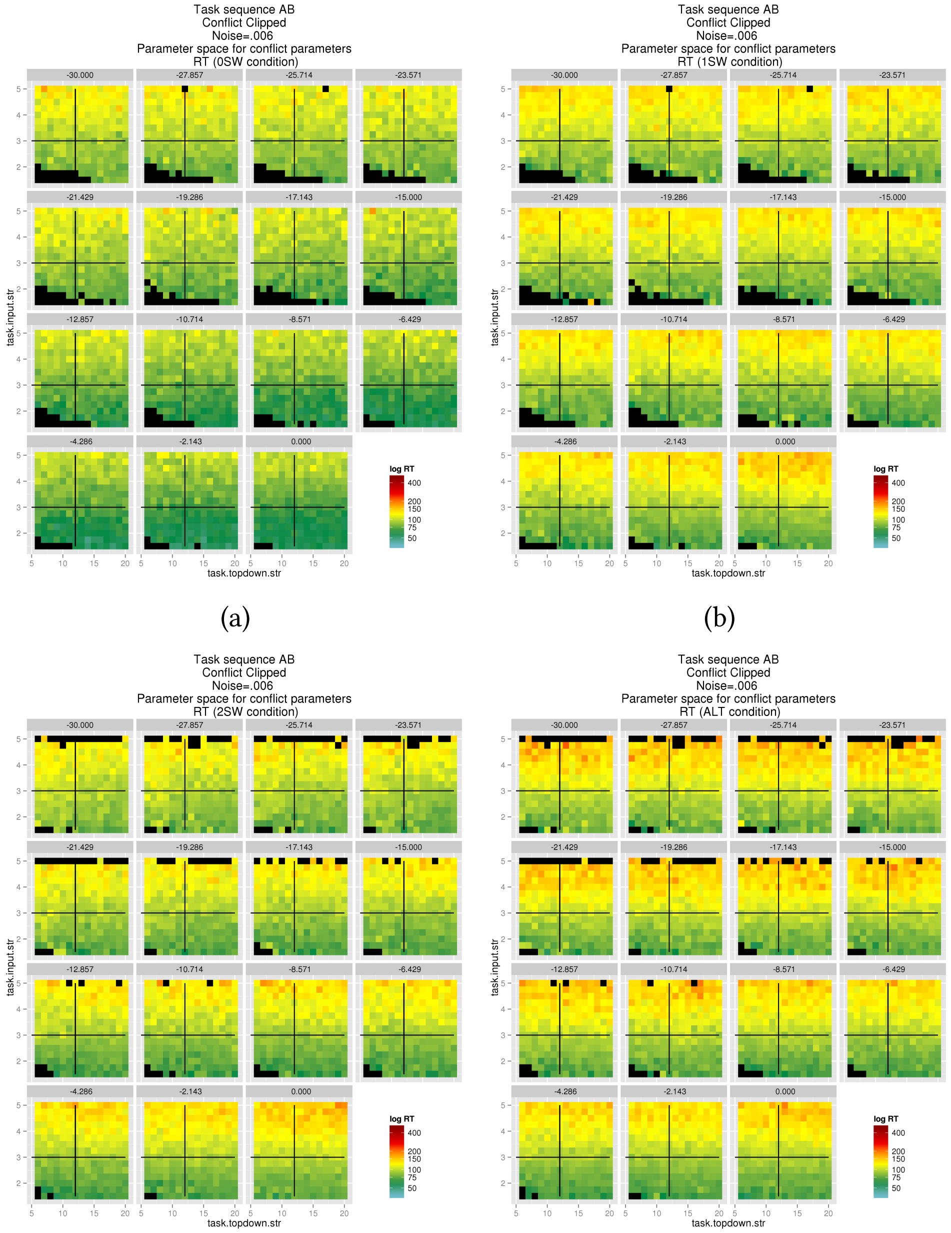
Firstly, let us consider the effect of the conflict system, in interaction with varied task asymmetry, on the absolute RT. In all four switch conditions (simulation 7, figure 7.1; simulation 8, figure 7.7), TDCS has an effect on trial 3 RTs, producing slower responses in general, particularly for 0SW trials. While the slowing has a less pronounced effect on 2SW trials, it has a selective effect on the upper quadrants of 1SW and ALT conditions (i.e., easy-easy-hard or hard-easy-hard sequences) — that is, on an easy-hard switch, a situation in which less conflict is observed on the middle trial (i.e., producing smaller n-1 switch costs) than hard-easy switches.

Thus far, results have been comparable for both simulations 7 (figure 7.1) and 8 (figure 7.7). In contrast, the effect of the TDCS parameter has a differing effect in the two simulations. As TDCS affects the level of activation of task-demand units, and thus the ability of the model to overcome residually active tasks, one might expect a stronger TDCS on the n-1 trial to produce a slowing on trial n, by producing a higher level of activation (and thus, residual activation) in the task demand unit for task B, which becomes the irrelevant task on trial n. In simulation 7, TDCS has a limited effect on RTs. (Recall, that this is the primary cause of the asymmetric switch cost in the model of Gilbert & Shallice, 2002, .) Right-hand quadrants, representing sequences where task B has greater TDCS than task A (i.e., hard-easy switches) are not slower than left-hand quadrants. Simulation 8, in contrast, shows faster RTs in right-hand quadrants. That is, hard-easy sequences, that produce more conflict, are faster than easy-hard sequences. This is because the TDCS parameter also affects the strength of backward inhibition in this simulation. On the final trial of a sequence ending in a BA switch, activation of the AB conflict unit provides greater inhibition to tasks with a higher TDCS parameter. Thus, if task B has greater TDCS than task A (i.e., right-hand quadrants) the residually active B task-demand unit will receive greater backward inhibition from the conflict unit than the currently active A task demand unit. Conversely, if task B has lower TDCS than task A, the residually active conflict unit will receive less backward inhibition than the currently active task demand unit.

It should be emphasised that these findings relate to absolute RTs, and not to the relative slowing caused by the backward inhibition system, or costs associated with contrasting particular sequences. We consider these effects next.

7.4.2 N-1 switch and n-2 repetition costs

To illustrate the effect of the conflict/task-inhibition mechanism on performance, figures 7.2 (simulation 7) and 7.8 (simulation 8) (also see figure 7.5, panels 3 to 6), plot the modulation of the trial 3 RTs, relative to a baseline where the *weight* of conflict-task demand connections is zero. That is, RTs equal to those obtained when the *weight* parameter is zero are plotted in grey, with slower RTs plotted in red, and faster RTs plotted in blue. The absolute magnitude of n-1 switch and

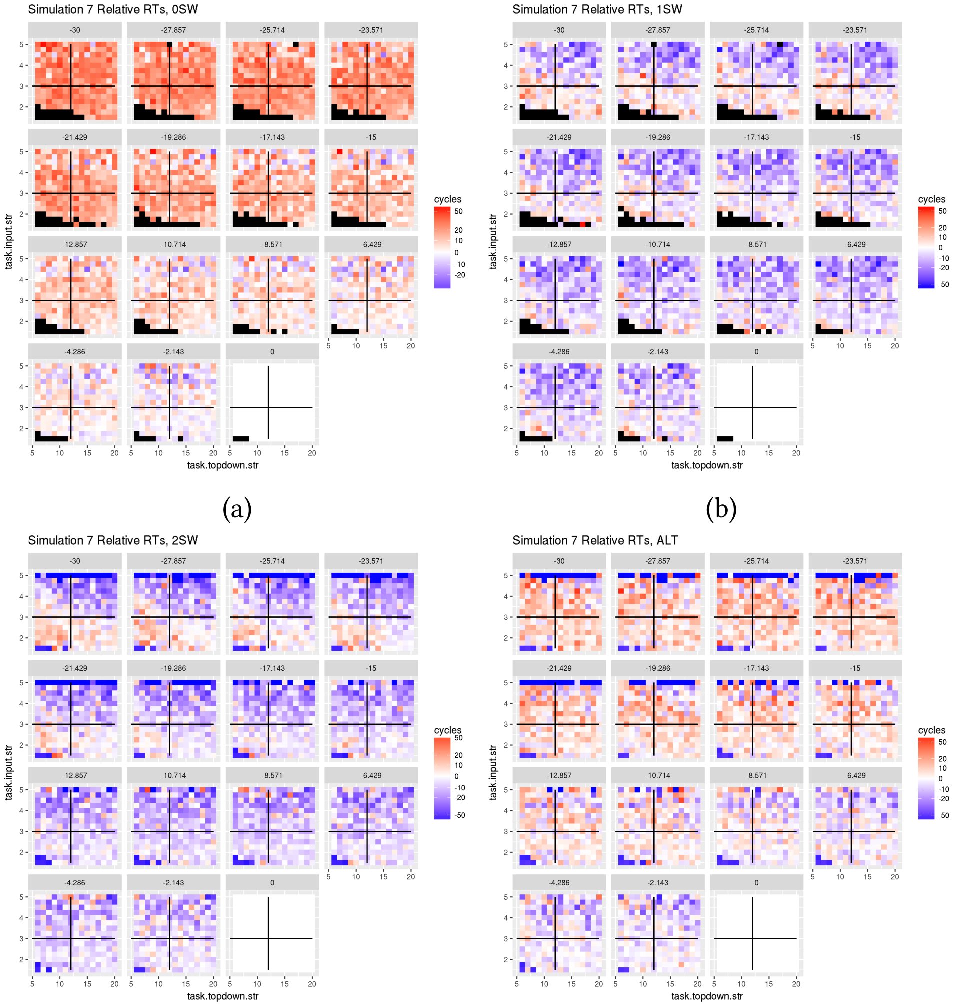


(a) (b)

(c) (d)

Figure 7.1: Simulation 7: Parameter space for asymmetric tasks, plotting basic RTs for varied task input strength and topdown control strength, for different levels of conflict weight. The black lines on the plot correspond with the parameters for task A, i.e., the centre of the cross represents symmetrical switching, with upper left and lower right quadrants representing stronger, more controlled, and weaker, less controlled, tasks, respectively.

n-2 repetition costs are presented in figures 7.3 and 7.9, for simulations 7 and 8, respectively, with positive costs represented in blue, and negative costs (i.e., facilitation) in red. The modulation of n-1 switch and n-2 repetition costs is plotted in



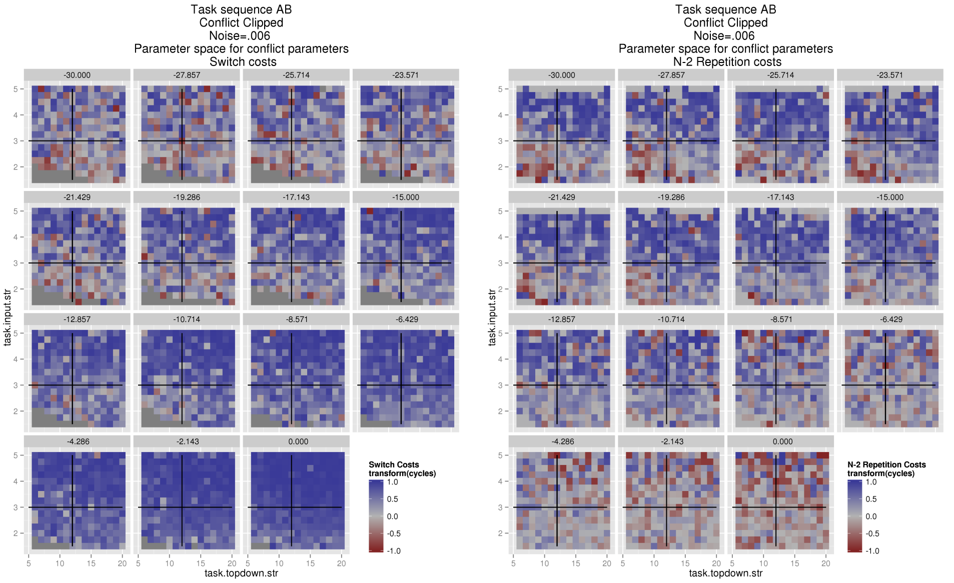
(a) (b)

(c) (d)

Figure 7.2: Simulation 7: Relative effects of conflict on basic RTs: difference against baseline of conflict-TD weight = 0.

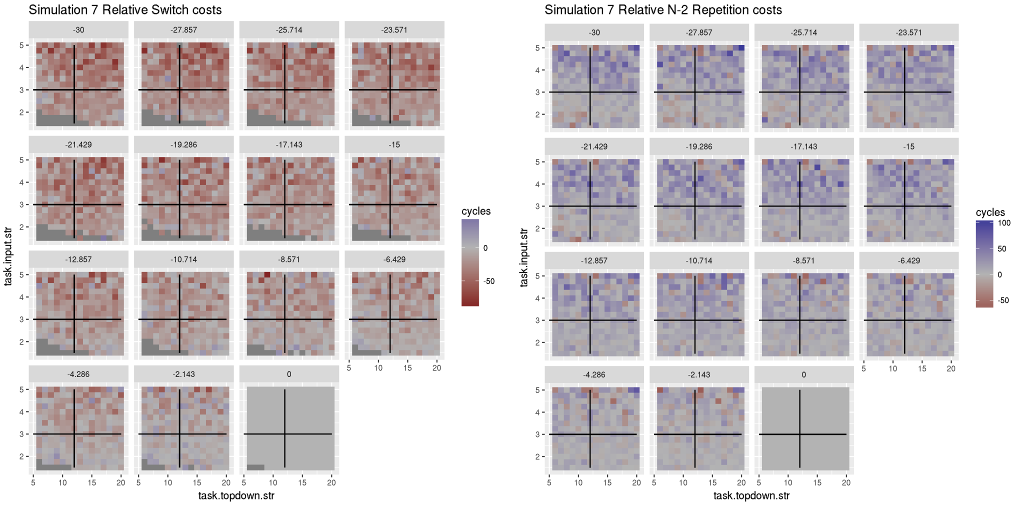
figures 7.4 and 7.10 (see also figure 7.6).

The plots of relative switch costs (figures 7.4 & 7.10, upper panels) show that for both simulations 7 and 8, stronger *weight* values produce smaller switch costs, especially for HEH switches (upper quadrants). The effect is qualitatively modulated by the strength of inhibitory biasing: for the weakest *weight*, the effect is minimal, while for stronger *weight* values, the reduction in switch costs is greatest where the n-1 task receives a high level of activation (top right quadrant) due to both a high level of control and high input strength. This is because greater task-demand activation on the n-1 task leads to more residual conflict on the n (switch)



(a) (b)

Figure 7.3: Simulation 7: Parameter space for asymmetric tasks, plotting RT costs for n-1 switches (7.3a) and n-2 repetitions (7.3b). The black lines on the plot correspond with the parameter settings for task A, i.e., the centre of the cross represents symmetrical switching, with upper left and lower right quadrants representing stronger, more controlled, and weaker, less controlled, tasks, respectively.



(a) (b)

Figure 7.4: Simulation 7: Relative effects of conflict on n-1 switch and n-2 repetition costs: difference against baseline of conflict-TD weight = 0.

trial, and the resulting task inhibition reduces interference, leading to a reduction in relative switch cost. Interestingly, this selective reduction in switch costs exaggerates the switch cost asymmetry (i.e., it reduces costs more for EH than HE

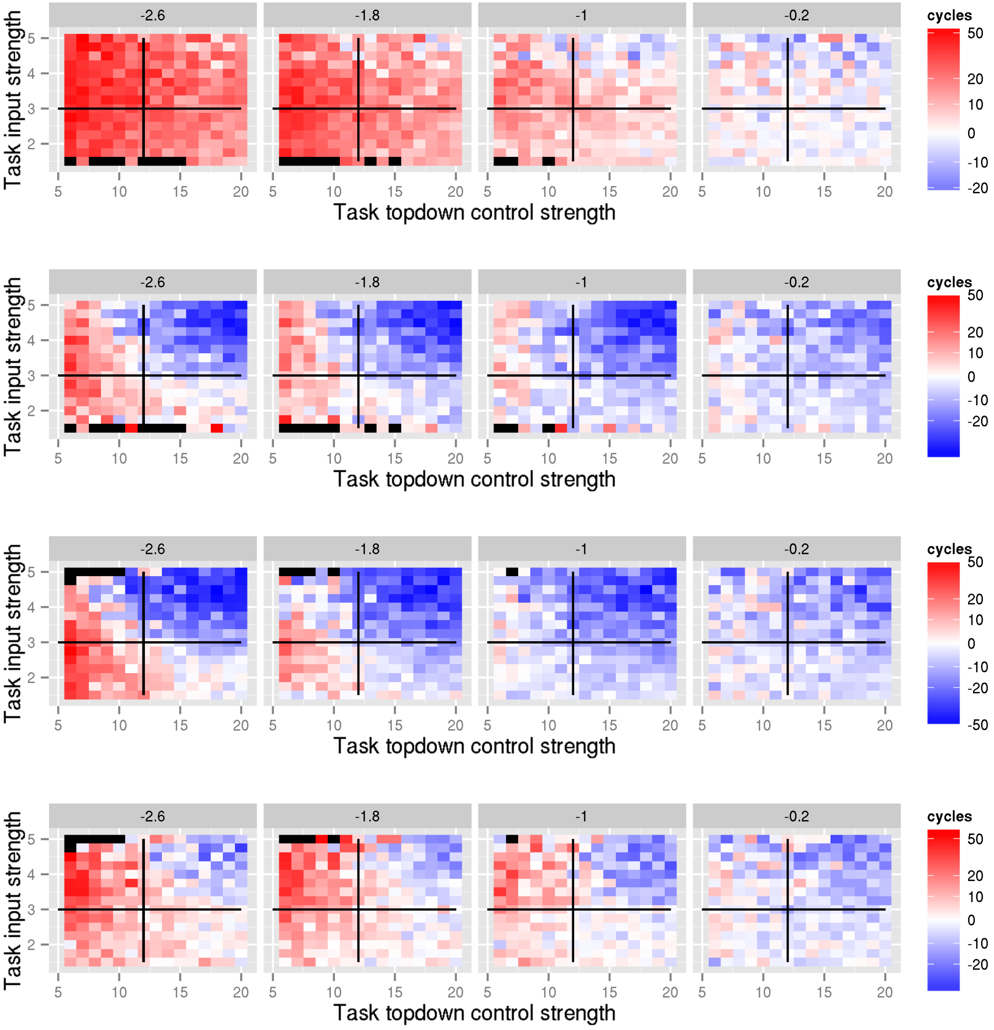


Figure 7.5: Simulation 8: Modulation of RT for individual 0SW, 1SW, 2SW and ALT conditions (from top to bottom), for successive values of *weight*, from strong (left) to weak (right). All values are relative to a zero weight baseline. Black squares represent no data, and the model's failure to produce correct responses with these parameter settings.

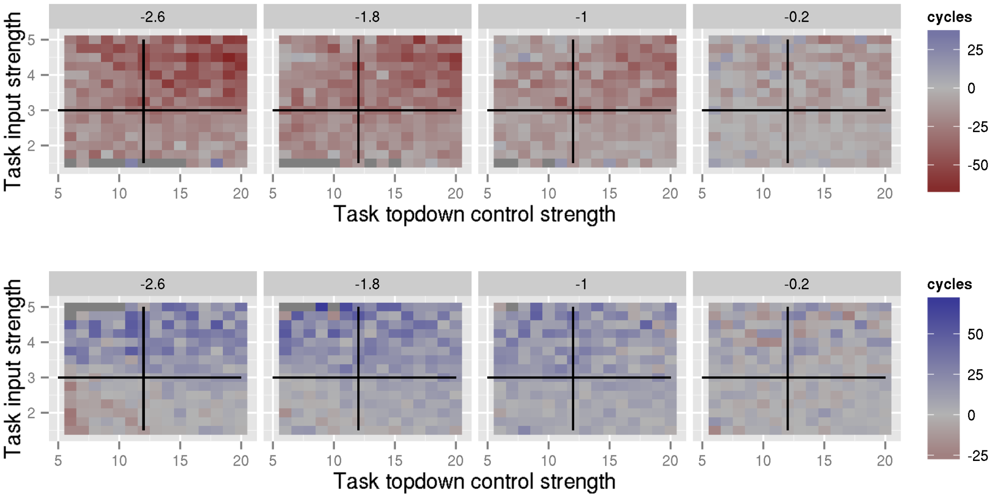


Figure 7.6: Simulation 8: Modulation of n-1 switch costs (upper panel) and n-2 repetition costs (lower panel) for successive values of *weight*, from strong (left) to weak (right). All values are relative to a zero weight baseline.

switches), suggesting that in a task-switching system with such a mechanism, a component of the switch cost asymmetry may be attributable to task inhibition.

To understand the basis for this asymmetry, we next consider the effect of the backward inhibition mechanism on 1SW and 0SW sequences individually. The simplest case is the 1SW (e.g., BBA) condition (figures 7.2b & 7.8b), where the switch occurs on trial n. Here, lower *weight* levels produce RT facilitation, particularly for switches from highly controlled tasks (e.g., HE switches, right hand quadrants). On the switch trial, more conflict is generated in activating the A task demand unit when B is harder than when it is easier, due to greater residual task-demand activation when performing the harder task on trial n-1. This is the reason for the asymmetric n-1 switch cost (i.e., larger costs for HE than EH switches) in the model of Gilbert and Shallice (2002). Due to this asymmetry in conflict, processing on trial n is facilitated when task B is harder than task A. When B is easier than A, more activation of task demand unit A is required to produce a response. However, because task demand unit B is still residually active, task inhibition counteracts the activation by top-down control of task demand unit A, leading to interference on trial n, and slower responses.

In the task-repeat 0SW condition (e.g., BAA, figures 7.2a & figures 7.8a), stronger *weight* values produce longer RTs on the final trial, especially following switches from easier tasks (left quadrants). As for the switch on trial n in the 1SW sequence, conflict occurs between the A and B task demand units on the switch trial, excepting that for 0SW sequences this occurs on the n-1 trial. Due to persistent conflict, the A/B conflict unit is still residually active on trial n, inhibiting the A task. Thus, slowing in 0SW trials is due to residual conflict from previously occurring task switches. At stronger *weight* values, there is greater slowing for EH (top left) than for HE (bottom right) switches. Because 0SW is the baseline condition for establishing the n-1 switch cost, slower responses lead to a reduction in the EH switch cost.

For both 0SW and 1SW switches, these effects of backward inhibition are qualitatively similar in both simulations 7 and 8. Taken together, the reduction in switch cost, which is greater for EH switches, occurs for two reasons: firstly, 1SW trial facilitation, particularly for switches from more controlled tasks (including HE switches); and secondly, 0SW trial interference, particularly for switches from less controlled tasks (including EH switches). Overall, the switch cost is reduced for both HE and EH switches, but the effect is greater for EH switches, exaggerating the switch cost asymmetry.

In contrast to the effects on n-1 switch costs, stronger *weight* values produce larger n-2 repetition costs in both simulation 7 (figure 7.4b) and simulation 8 (figure 7.10b). This effect is modulated by input control strength, and is larger for HEH alternations (i.e., upper quadrants) than EHE alternations. In the ALT condition (figures 7.2d & figures 7.8d), the effect is modulated by n-1 task difficulty in simulation 8, but not simulation 7.

While in both simulations 7 and 8, conflict units generally facilitate performance for low *weight* values, in simulation 8 they selectively impair HEH alternations (top left quadrant), while for the strongest values they also begin to slow

EHE alternations. The reason for this difference is twofold. Firstly, there is an effect of *TDCS* that occurs in simulation 8 but not in simulation 7. The reason for this difference is that in simulation 8, the strength of backward inhibition is scaled by *TDCS*, hence when task B is easier than task A (i.e., HEH) the A task receives more inhibition than task B, and thus the RT is slower. Conversely, when task B is harder than task A (i.e., EHE) the A task receives less inhibition than task B, and thus RTs are faster. Secondly, there is an effect of *SIS* that occurs for both simulations. Essentially, a greater SIS for the irrelevant task leads to more conflict being generated on the n-1 trial for HEH sequences than EHE sequences (for the same reasons that HE switches generate more conflict than EH switches in 1SW trials, above). This results in slower responses for the lower quadrants, reflecting harder tasks, and can be observed in both simulations.

In the 2SW condition (figures 7.2c & 7.8c), mild and intermediate *weight* values produce generalised facilitation. Once again, for stronger *weight* values, the effect on RTs is highly modulated by top-down control in simulation 8 but not simulation 7, with interference and facilitation caused by low-control and high-control n-1 trials, respectively, and with most pronounced effects for under- and over-activated trials (i.e., bottom left and top right quadrants). The fact that, in both simulations, facilitation dominates irrespective of *SIS* or *TDCS* for all but the strongest *weight* values, suggests that the mechanism beneficially reduces the amount of control required to achieve good performance when frequently switching between multiple tasks. In contrast to the effect on ALT trials, on a 2SW trial the difficulty of the n-1 task has a much less pervasive effect on trial n performance, because the task demand unit for trial n has not recently been inhibited by conflict units.

Together, the 2SW and ALT results in both simulations 7 and 8 explain the larger n-2 repetition costs observed in HEH alternations than in EHE alternations — it is a composite of stronger facilitation for switches from easier/more controlled tasks in the 2SW condition, and greater interference when switching from those same tasks in the ALT condition.

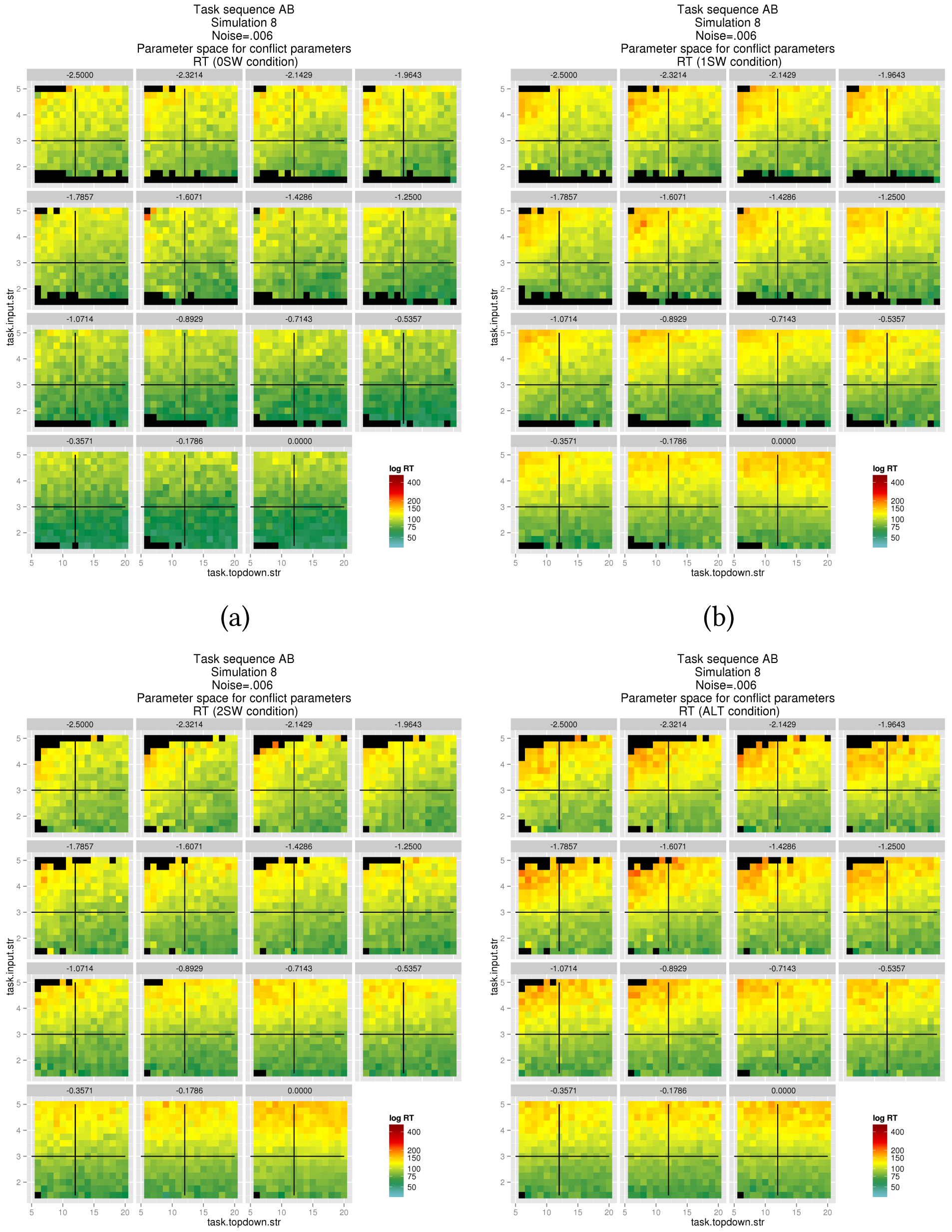
Interestingly, although the modification to the model in simulation 8 (namely normalising the weight of backward inhibitory connections by the *TDCS* parameter) has an effect on RTs, modulated by the *TDCS*, this does not have an effect on the qualitative pattern of n-2 repetition costs. In other words, simulations 7 and 8 both exhibit clearly smaller n-2 repetition costs for EHE than HEH switches. This is because the effect of *TDCS* appears to largely cancel out for ALT and 2SW sequences. Thus, we may conclude that the exact weighting of the backward inhibitory connection is largely independent of the overall effect on n-2 repetition costs.

7.5 Comparison of models in the present and previous chapters

Given the modification to the conflict-task demand connection weights in the simulation 8 (i.e., multiplying the *weight* parameter by the *TDCS* for each task demand unit, to obtain individual weights for conflict-task demand unit connections), it is important to verify that the change to the architecture has not had a serious effect on parameter space. Simulation 6e, from the previous chapter, was therefore replicated using the modified model from the present chapter. These results are presented in appendix C. Comparison of the parameter space partitioning graphs with the equivalent simulation (appendix B) suggests no differences in the shape of parameter space. The parameter settings used in the present chapter[2] were still in a region reasonably central for the intersection between RT switch costs and n-2 repetition costs (see figure C.2). In conclusion, the change to the model architecture in simulation 8 does not make a qualitative change to the shape of parameter

---

[2] Conflict gain γ = 75.0, conflict bias β = -10.0, conflict/task-demand connection weight ω = [-2.5 0]


Task sequence AB
Simulation 8
Noise=.006
Parameter space for conflict parameters
RT (0SW condition)

Task sequence AB
Simulation 8
Noise=.006
Parameter space for conflict parameters
RT (1SW condition)

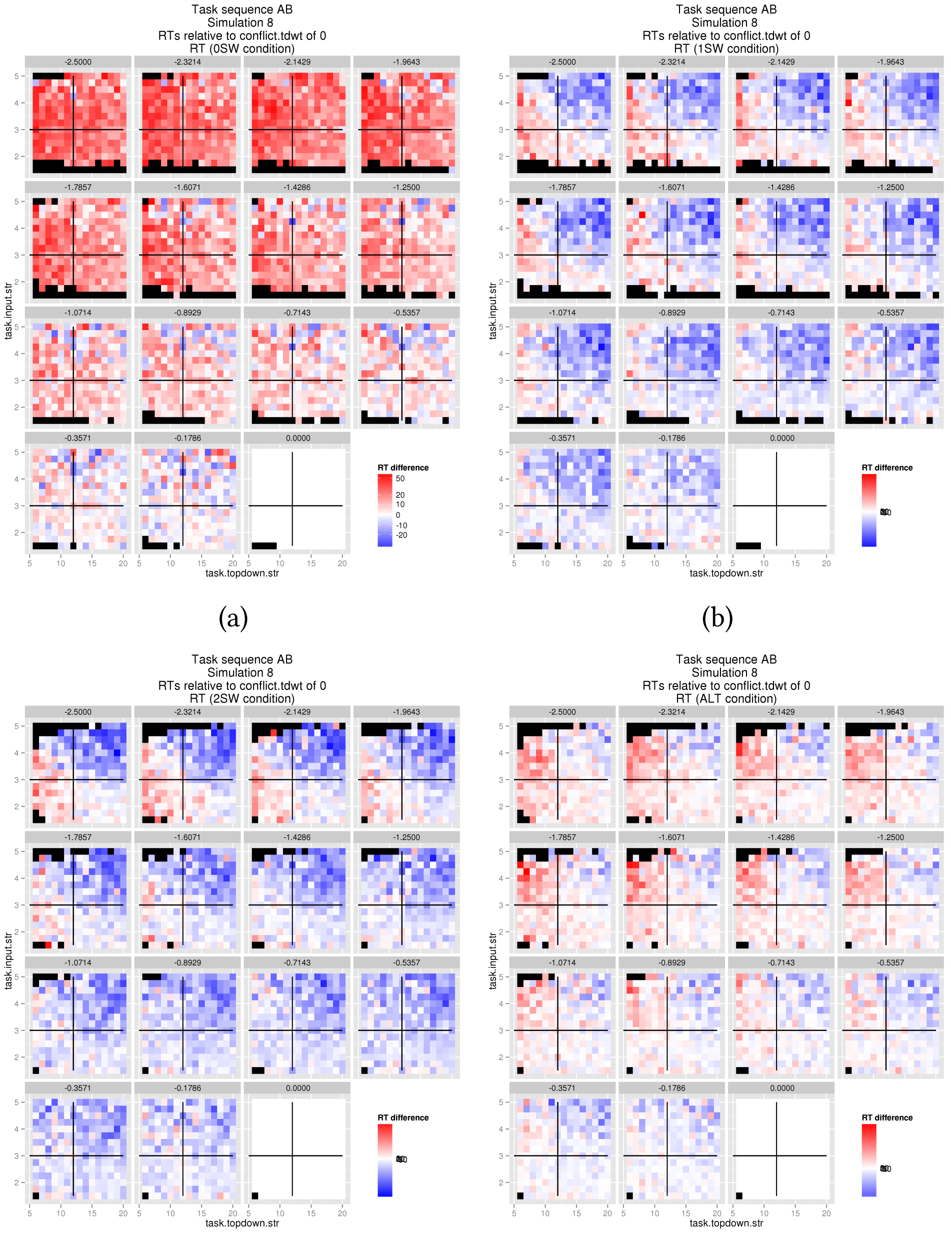
Task sequence AB
Simulation 8
Noise=.006
Parameter space for conflict parameters
RT (2SW condition)

Task sequence AB
Simulation 8
Noise=.006
Parameter space for conflict parameters
RT (ALT condition)

(a) (b) (c) (d)

Figure 7.7: Simulation 8: Parameter space for asymmetric tasks, plotting basic RTs for varied task input strength and topdown control strength, for different levels of conflict weight. In simulation 8, the conflict weight parameter is multiplied by the TD control strength for each task to obtain the weight from the conflict to task demand unit.

space.



(a)

(b)

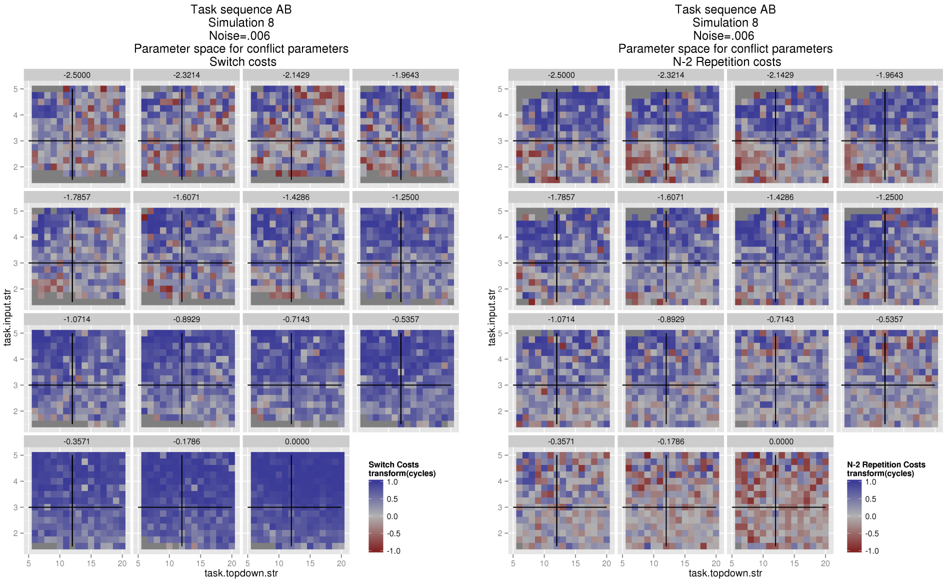
(c)

(d)

Figure 7.8: The effect of increasing conflict on basic RTs: difference against baseline of conflict-TD weight=0.

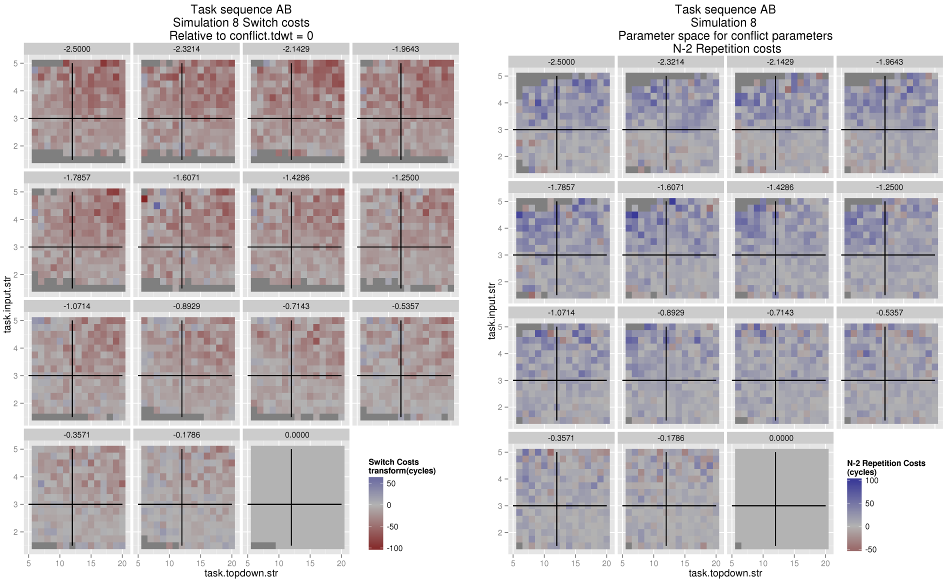
7.6 Summary

In general, the effects of backward inhibition on all four types of trial can be understood in terms of the conflict generated on switch trials by simultaneously active task demand units, and the asymmetry in activation dynamics between easy-hard and hard-easy switches. Because the model regulates task inhibition on the basis



(a) (b)

Figure 7.9: Simulation 8: Parameter space for asymmetric tasks, plotting RT and error switch costs for varied task input strength and topdown control strength, for different levels of conflict weight. Contrasted are task sequence AB (ie., AAB vs. ABB) and BA (BBA vs. BAA). The black lines on the plot correspond with the parameters for task B, ie. the centre of the cross represents symmetrical switching, with upper left and lower right quadrants representing stronger, more controlled; and weaker, less controlled, tasks, respectively.



(a) (b)

Figure 7.10: Difference in costs when conflict-TD weights are zero, and greater values.

of detected conflict, the model predicts an asymmetric effect on the size of the n-2 repetition cost, specifically, that it is larger for HEH than EHE alternations, and that this is a residual effect from the additional conflict generated by HE switches than EH switches, occurring on the n-1 trial. This prediction is directly tested in experiments reported in the next chapter.

Chapter 8

Empirical Studies of Three-Task Asymmetric Switching

Three experiments are reported which test the predictions of simulations in the preceding chapter — that n-2 repetition costs are modulated by asymmetric task difficulties. Specifically, the experiments test that the RT n-2 repetition cost is larger in the n-2 repetition condition (i.e., ABA sequences) for hard-easy-hard (HEH) triplets than for easy-hard-easy (EHE) triplets. This prediction runs counter to the extant empirical literature, although previous studies may be compromised by methodological issues. Experiment 1 details a new n-2 repetition paradigm in which participants classify letters according to three dimensions: colour, font and alphabetical position. The results are inconclusive, with participants tending to maintain a consistent RT across the three tasks, with varying error rate. Experiments 2 and 3 addressed this issue by encouraging participants to maintain a consistent error rate through the use of performance feedback immediately following error trials, and at the end of each block. Experiment 2 tested the hypothesis using transparent task cues, while experiment 3

used non-transparent task cues. Consistent with the predictions of the model, the results of experiments 2 and 3 both showed significantly larger n-2 repetition cost for HEH than EHE triplets.

8.1 Methdological preliminaries

As discussed in the previous chapter, the only previous study that has systematically examined n-2 repetition costs for asymmetric tasks is that of Arbuthnott (2008a). Given the difficulties previously identified in interpreting her results, this section discusses some issues in the design of an appropriate experimental paradigm that is likely to yield clear empirical effects.

8.1.1 Methodological issues with the study of Arbuthnott (2008a)

There are three methodological concerns regarding the procedure used by Arbuthnott (2008a). Firstly, and most simply, at least one study suggests that procedures which include immediate task repeats may reduce or eliminate n-2 repetition costs (Philipp & Koch, 2006). In investigating modulations of the n-2 repetition cost, therefore, it seems prudent to design task sequences to avoid immediate task repetitions, at least in one condition.

The second issue relates to the possibility of graded response congruence effects in the digit classification task, in which single digits are classified according to three different criteria: In this case, parity (odd or even), magnitude (greater or less than 5), or prime (prime or non-prime), with two possible responses signalled by pressing left or right response keys. The mapping of stimuli to responses was as follows: for magnitude, *2,3,4 / 6,7,9*; for parity, *3,7,9 / 2,4,6*; for prime, *2,3,7 / 4,6,9*. Given six stimuli and three possible dimensions, the stimuli have different levels of congruence between the correct response, and the responses to the alternative tasks afforded by the stimulus. For example, for magnitude judgments, the stimulus *6* has a correct response (right) congruent with both the odd-even task's and the prime-nonprime task's response (i.e, fully congruent), *7* is congruent on only one of the irrelevant tasks (semi-congruent), while *4* is incongruent with both irrelevant tasks (fully incongruent). Thus, specific stimulus items are conflated with the degree of response conflict generated on each task. In fact, there is recent empirical evidence that this type of graded response congruence indeed affects reaction times (Longman, Lavric, Munteanu, & Monsell, 2014; Schneider, 2014) in three-task versions of the digit classification task. While differing levels of stimulus response congruence might be assumed to counterbalance for each task for symmetric difficulties, when difficulty varies by task, stimulus response congruence also varies per task, thus confounding stimulus congruence with task difficulty. This issue could be overcome using a design that offers three possible responses to each task, rather than the typical two. In this case, the confound may be removed if the responses to all three stimulus dimensions are mutually incongruent. However, to date, no backward inhibition study using classification tasks has taken this approach.

The final issue is the assumption that the task sets involved do not overlap. It is questionable, however, whether prime number judgment is really independent from the other two tasks. For example, if a digit is even it is guaranteed to be nonprime unless it is 2, providing a heuristic method for determining prime/nonprime status using an easier odd/even judgment. If responding to a prime trial involved even partial activation of the parity task-set, then any task-set inhibition occurring on the next trial may be applied to both prime and parity task-sets. In sum, in three-task switching paradigms that incorporate only two possible responses to each task, pairings of task difficulty may be confounded with response congruence and response set overlap from the switched-away-from task. Avoiding these problems within a three-task switching paradigm requires tasks with (at least) three possible responses.

8.1.2 On cue transparency

It is known that cue characteristics, such as spatial cue location, can contribute to task conflict (Arbuthnott, 2008b). Arbuthnott argued that spatial cues formed location-task representations, which increased cue conflict when locations overlapped. Other cue characteristics also appear to contribute to task conflict, including cue type (i.e., spatial vs. verbal, Arbuthnott, 2005) or degree of transparency (i.e., verbally related vs. arbitrary, Arbuthnott & Woodward, 2002).[1] Houghton et al. (2009) compared easy vs. hard cue-task translation conditions (e.g., iconic cues vs. vebal cues), finding that n-2 repetition costs were only significant for verbal cues. In their experiment 3, non-transparent cues (arbitrary geometric shapes) identified the target dimension. The n-2 repetition cost was greater for non-transparent (arbitrary) cues than for verbal cues. In their experiment 4, verbal non-transparent cues were contrasted against verbal transparent cues, again producing a greater backward inhibition effect (with no n-2 repetition cost when the cue was identical with the target). Overall, these studies suggest that iconic cues do not produce n-2 repetition costs: some degree of cue-target translation is necessary. However, the task used by Houghton was the location of a deviant stimulus — in essence, the task was identical on all trials, but the target dimension changed. This contrasts with the present case of digit classification, in which the task changes as well as the target dimension. Given that the task used by Houghton was visual search, it is possible there may be less 'task' related processing to occur after the target dimension is identified than when a classification task is identified by a verbal cue (e.g., 'CENTRAL/PERIPHERAL')

To conclude, in order to obtain reliable n-2 repetition costs, iconic cues should be avoided entirely. Some transparency (i.e., verbal cueing of the task) is regarded as reasonable, where it refers to the name of the task rather than the response options (i.e., 'PARITY' rather than 'ODD/EVEN', although, note that the latter type

---

[1] See the introduction of Houghton et al. (2009) for a good review.

of cue has been used extensively, see Mayr, 2002; Mayr, 2006; Mayr & Keele, 2000; Houghton et al., 2009).

8.2 Experiment 1

8.2.1 Introduction

Experiment 1 tested the hypotheses regarding the modulation of the n-2 repetition cost by asymmetric task difficulties, formulated in the previous chapter — n-2 repetition costs are larger for easy-hard-easy (EHE) triplets than for hard-easy-hard (HEH) triplets. Bearing in mind the previously discussed shortcomings of the extant empirical studies, experiment 1 used a new procedure based on the switching variant of the Stroop task. In a switching version of the task, participants switch between reading the word (the word task) and naming the colour the word is printed in (the colour task), with the colour task typically more difficult (slower and more error prone). This is the task used to obtain classic effects of switch cost asymmetry (Allport et al., 1994; Allport & Wylie, 2000), hence it seems a suitable basis for testing for task asymmetry effects in the 3-task paradigm. Thus a third task was added to the procedure: participants named the colour of a frame which appeared around the word.

8.2.2 Method

Procedure

On each trial, one of three possible cues (‘WORD’, ‘COLOUR’ or ‘FRAME’) appeared 1000ms in advance of the target, transparently indicating which task to perform. The target stimulus appeared in the centre of the screen, and consisted of a coloured word (‘RED’, ‘GREEN’, or ‘BLUE’) enclosed by a coloured rectangular frame.

Participants responded to one of three tasks by pressing a key corresponding to a colour. The correct response for each task was as follows: For the colour task, the key corresponding to the colour the word appears in; for the word task, the response matching the name of the word, and for the frame task, the colour of the frame. For all three tasks, there were three possible responses (red, green or blue), each of which was mapped to a single key for all tasks (left-arrow, down-arrow, or right-arrow, respectively). Thus, this procedure involves an overlap of response sets at the conceptual level, as well as overlapping response mappings, as is typical in the digit classification procedure.

The order of tasks was randomised, with the constraint that immediate task repeats were not allowed (hence, on each trial both switch tasks appear with 50% likelihood). The mapping of stimuli was also randomised, with two constraints. Firstly, each possible response appeared exactly once in the stimulus (e.g., the word 'GREEN' appearing in red ink, inside a blue frame), controlling for irrelevant dimension congruency (each target dimension is accompanied by two incongruent irrelevant dimensions). Secondly, response repetitions were disallowed. Hence, on each trial, the correct response had a 50% likelihood of being either response that was not the previous correct response.

The procedure was administered using PsychoPy version 1.82.01 (Peirce, 2009) on a GNU/Linux PC. Stimuli were presented on 380mm × 300mm LED monitor, placed approximately 750mm from the participant. With the exception of the target word, all text was displayed in white, against a grey background. Responses were made using the bottom row of arrow keys on a standard PC keyboard, as described above.

Participants were instructed to perform the task 'as quickly and accurately as possible'. To practice the tasks, participants performed 10 trials of each single task with no switches, and feedback provided on incorrect responses (e.g., 'Oops! the correct response was: left. Press space to continue'). Participants then performed

20 practice trials where the task switched, as in the main procedure. Participants then performed four experimental blocks of 102 trials, each block followed by a break, with the participant free to rest as long as desired. Feedback was not provided on incorrect responses in the experimental blocks.

Participants

35 participants were recruited through the Birkbeck psychology participant database, of whom 19 were female. Informed consent was obtained from each participant. The median participant age was 33 years, with a range of 39 years. Participants took part in exchange for their choice of course credit or a cash sum of £6.

Data Analysis

Each trial was classified as either 2SW (e.g., CBA) or ALT (e.g., ABA) depending on the sequence of preceding trials. Additionally, each trial (n) was identified by its task (word, colour, or frame) and the previous (n-1) task. Given that task repeats were excluded, the n $\times$ n-1 task factors give 6 possibilities. For example, where task n is colour, the n-1 task may be either word or frame. Taking the combination of n and n-1 tasks as a single factor therefore yields a 2 $\times$ 6 within-subjects design, with the other factor being the direction of the alternation (hard-easy vs. easy-hard).

8.2.3 Results

Of the 35 participants, 15 had error rates exceeding an a-priori threshold of 10% on at least one of the three tasks, and were excluded from further analysis. The RTs and error rates for each task for the remaining 20 participants are presented in table 8.1.

The asymmetry in task difficulty between the three tasks was assessed using a one-way, within-subjects ANOVA, with the three tasks as the only factor. The effect of task was significant, $F(2, 34) = 13.37, p < .0001$. Pairwise t-tests (Bon-

Task	RT	Error rate
Frame	1254 *(663)*	2.77% *(2.70%)*
Colour Naming	1912 *(1191)*	5.70% *(2.80%)*
Word Reading	1910 *(1320)*	4.48% *(2.70%)*

Table 8.1: Mean (SD) of RTs (milliseconds) and error rates for each task

ferroni corrected) revealed the difference between colour and word was not significant ($p = 1.00$), but there were significant differences between word and frame, ($p = .0052$) and colour and frame ($p = .0016$). The effect of the task on error rates was also significant, $F(2, 34) = 9.74, p < .0001$. Pairwise comparisons revealed significant differences in error rate between colour-frame ($p = .001$) and word-frame ($p = .026$) but not colour-word ($p = .372$).

The tests for task asymmetry indicated a statistically significant difference between the colour and frame tasks in both reaction times and error rates, and for word and frame tasks, in reaction times only. Therefore, n-2 repetition costs oberved in colour/frame and word/frame alternations were compared for different switch directions (i.e., CFC vs. FCF; WFW vs. FWF). Given that no significant difference was found between colour-word tasks in reaction time, colour-word pairings were not tested for n-2 repetition cost asymmetry.

The model predicted a larger n-2 repetition cost for hard-easy-hard (HEH) triplets than for easy-hard-easy (EHE) triplets. This was tested using a 2 × 2 ANOVA, with within-subjects factors of task pairing (colour/frame vs. word/frame) and switch direction (easy-hard-easy vs. hard-easy-hard). Neither the main effect of task, $F(1, 32) = 0.001, p = .958$, nor the main effect of switch direction, $F(1, 32) = 2.00, p = .167$, reached significance. The task × switch direction interaction was also not sigificant, $F(1, 32) = .06, p = .804$. The nonsignificant difference in means between EHE and HEH n-2 repetition costs was in the opposite direction to that hypothesized, with EHE n-2 repetition costs larger (mean = 138ms, sd = 744ms) than HEH n-2 repetition costs (mean = -27ms, sd = 497ms).

Mean N-2 repetition costs for the colour/frame task were calculated as sub-

Task Pairing	Easy-Hard-Easy	Hard-Easy-Hard
Colour-Frame	134 *(354)*	-44 *(481)*
Word-Frame	346 *(1300)*	-82 *(730)*

Table 8.2: Mean (SD) n-2 repetition costs for different switch directions of colour-frame and word-frame pairings

tracting mean RTs on trial 3 for WFC trials from CFC trials; and WCF trials from FCF trials, with the costs for word/frame tasks similarly obtained. These data are presented in table 8.2. These were compared using Welch two-sample t-tests. The difference in means was not statistically significant for either the colour/frame tasks, $t(31.2) = 1.27, p = .214$, or the word/frame tasks, $t(26.8) = 1.22, p = .233$. Note that both differences in means, although nonsignificant, were counter to the experimental hypothesis.

8.2.4 Discussion

The results did not support the hypothesis that n-2 repetition costs for HEH triplets would be larger than for EHE triplets. The three tasks were selected to be relatively easy, intermediate, or hard in terms of difficulty, and also to induce inter-task conflict. The three tasks conflicted in that the set of possible responses for each task were the same. However, while the colour naming and word reading tasks clearly differed in terms of the task being performed, the colour naming and frame colour identification tasks, it could be argued, were insufficiently differentiated — both tasks involved naming a coloured stimulus, with the only switch being the target stimulus. There are a number of potential issues with the selection of these three tasks. Firstly, if the colour and frame task are sufficiently similar in terms of their task-set (taken to be a cognitive representation of S-R bindings), then it could be that the task-set is not as forcefully inhibited when abandoning the colour task-set in a CF switch, than in a CW switch, which involves a more complete switch of task-set. Secondly, Arbuthnott (2008b) found that target stimuli needed to spatially overlap in order to generate n-2 repetition costs; here, the frame and target word stimuli always appeared in the same, non-overlapping locations.

Thirdly, there were not reliable differences in RTs between the three tasks. While RTs differed between the frame-word and frame-colour, there was not a reliable RT difference between colour-word. However, there was a significant pairwise difference in error rates between colour/frame and word/frame. This presents a problem, as the present analysis takes RT as a proxy for task difficulty. It was assumed that participants would regulate the amount of attention paid to each task, appropriate to the difficulty of the task. However, these results suggest that participants regulated their performance such that RT was relatively uniform and error rates differed, rather than vice versa. As no feedback was given for error trials during the experimental blocks, there was effectively nothing to stall performance. In addition, it should be noted that this experimenter's subjective experience of performing the procedure was that the presentation of the cue and stimulus, with uniform response-cue and cue-stimulus intervals, established a particular rhythm of responding. This rhythm may have encouraged forced-time responding, resulting in too-quick responses to the harder tasks, leading to an elevated error rate but a less elevated RT.

These problems were addressed in experiment 2. Specifically, the problem of participants possibly regulating their speed/accuracy tradeoff to prioritise speed was addressed in experiment 2 by providing error-feedback during the experimental block, on a screen which introduced a pause in the procedure while the feedback was displayed. The problems with the tasks (insufficient differentiation between colour and frame tasks) were addressed by using three different tasks.

8.3 Experiment 2

8.3.1 Introduction

Participants classified letters according to the dimensions of colour (red, green or blue), alphabetical position (beginning, middle, or end) and font (bold, regular, italic). Importantly, each task has three possible responses to each stimulus, rather than two. This has the advantages that, firstly, response repetition effects can be eliminated (even if response repetitions never occur, there are still two possible responses to each task), and secondly, as the three stimulus dimensions (colour, alphabet, font) can be manipulated independently, irrelevant-dimension congruency effects can be controlled such that all stimulus dimensions are always mutually incongruent.

8.3.2 Method

Participants

39 participants were recruited through the Birkbeck psychology participant database, of whom 22 were female. Informed consent was obtained from each participant. The median participant age was 28 years, with a range of 46 years. All participants spoke English fluently, although nine reported a first language using a non-Roman alphabet. Participants received their choice of either £7.50 cash or course credit.

Design

The experiment compared n-2 repetition costs for each of three possible pairings of tasks. The relative ordering of task difficulty was determined by comparing reaction times and error rates. For each task pairing, n-2 repetition costs could then be determined independently for easy-hard-easy (EHE) or hard-easy-hard (HEH) alternations. For example, n-2 repetition costs for the colour-alphabet pairing were calculated as the difference between colour-alphabet-colour and font-alphabet-colour sequences (EHE), and alphabet-colour-alphabet and font-colour-alphabet sequences (HEH). Thus, the experimental design had two factors: switch direction (EHE vs. HEH) and task pairing (colour-font vs. colour-alphabet vs. alphabet-font).

Tasks

Participants performed one of three tasks in response to a target letter stimulus. The target letter, rendered in the Latin Modern Roman font, appeared in the centre of the screen, approximately 20mm × 30mm in size. Simultaneously with the target stimulus, a cue presented above the target, 'colour', 'alphabet' or 'font' in lower case and approximately 7mm in height, indicated which task to perform. The response to all three tasks was made by pressing one of the left, down, or right cursor keys on a standard PC keyboard, thus the set of possible responses was fully overlapping. The colour task was to identify the colour of the target, indicating red, green, or blue, by pressing left, down or right respectively. The font task was to indicate whether it was rendered in a bold (left), regular (down), or italic (right) font. The alphabet task was to indicate whether the letter occurred toward the beginning, (**d, e, f, g, h**; left), middle, (**k, l, m, n, o**; down) or end (**s, t, u, v, w**; right) of the alphabet.

Procedure

The procedure was administered using PsychoPy version 1.82.01 (Peirce, 2009) on a GNU/Linux PC. Stimuli were presented on a 380mm × 300mm LED monitor, placed approximately 750mm from the participant. With the exception of the target letter, unless otherwise stated, all text was displayed in white against a black background.

Participants were instructed to perform the tasks 'as quickly and accurately as possible'. Instructions for each task were worded as follows: 'In the colour task, press the key corresponding to whether the letter appears in red, green, or blue';

‘In the alphabet task, press the key corresponding to where the letter appears in the alphabet: toward the beginning (d e f g h), middle (k l m n o), or end (s t u v w)’; ‘In the font task, press the key corresponding to whether the font of the letter is bold, regular, or italic’. Simultaneously with the instruction, the three response mappings (red/green/blue, beginning/middle/end, bold/regular/italic) appeared in the bottom left, bottom centre, and bottom right of the screen, to correspond with the response key mappings.

After receiving instructions for each task, participants performed a practice block of 10 trials. Participants then performed 20 practice trials where the task switched, as in the main procedure. For all practice blocks, specific feedback was provided on incorrect responses (e.g., ‘Oops! the correct response was: left. Press space to continue’).

Following the practice blocks and a break screen, in which participants could rest as long as desired, participants performed six blocks of 103 trials, with a break screen between each block. The order of tasks was randomised, with the constraint that immediate task repeats were not allowed (hence, given a specific preceding task, the two other switch tasks may each occur with 50% likelihood). Stimulus dimensions were randomised, subject to the constraint that all three possible response keys were represented in each stimulus (e.g., a bold, green letter *t*) to avoid stimulus congruency effects.

Given the concerns raised in interpreting the results of experiment 1, error feedback throughout the experimental blocks was considered an important element of the procedure, and was provided in the following ways. Firstly, immediate feedback was given following incorrect responses. This was in the format of an additional message which appeared on screen (‘Oops!’) and a reminder of the response mappings for all three tasks, appearing in the bottom left, bottom centre, and bottom right, of the screen. This message remained on screen for 4,000ms, before trials resumed. This additional pause following an incorrect response was

Task	RT	Error rate
Colour	1544 *(391)*	3.74% *(2.83%)*
Font	1774 *(411)*	5.86% *(3.61%)*
Alphabet	1849 *(500)*	7.98% *(5.99%)*

Table 8.3: Mean (SD) of RTs (milliseconds) and error rates for each task

intended to break any rhythm in responding, as well as providing a mild penalty for an incorrect response. Secondly, overall feedback was given following each block on the accuracy rates for each task. This was in the format of e.g., ‘colour accuracy: 95%’. This message appeared in white if accuracy was 90% or higher, but in red if accuracy dropped below 90%. If the accuracy for any task was below 90%, the message ‘Please take more care!’ also appeared on screen, while if all were above 90%, the message read ‘Well done!’.

8.3.3 Results

A single participant with a mean RT more than 2.5 SDs above the group mean was excluded from the analysis, for generally very slow responses (overall mean RT 4131ms, compared to a group mean of 1782, SD 553ms). In addition, error trials and the subsequent two trials, and trials with very slow RTs (over 20s) were all excluded from the RT analysis. After trimming, the mean proportion of trials retained for analysis was 83.2%. This proportion did not vary substantially across cells of the experimental design, with the least trials retained for colour-alphabet switches (81.4%) and the most for font-alphabet switches (84.9%). The means and SDs for the remaining trials for each task are presented in table 8.3.

The asymmetry of the three tasks was established using a 1 $\times$ 3 within-subjects ANOVA. For RTs, the effect of task was significant, $F(1.66, 61.57) = 23.1, p < .001$ (Geenhouse-Geisser corrected for non-spherical data). Pairwise t-tests (Holm-Bonferroni corrected for multiple comparisons) indicated significant RT differences between colour and font, $t(37) = 5.39, p < .001$, colour and alphabet, $t(37) = 5.43, p < .001$, and a marginally significant difference for font and alphabet, $t(37) = 1.89, p = .067$. Parallel tests for error rates revealed the same pattern: the effect of task was significant, $F(1.73, 64.16) = 13.95, p < .001$ (Greenhouse-Geisser corrected), with significant pairwise differences between colour and font, $t(37) = 3.38, p = .003$, colour and alphabet, $t(37) = 4.87, p < .001$, and font and alphabet, $t(37) = 2.40, p = .022$.

However, individual differences did exist in the ordinal difficulty of the three tasks. In increasing difficulty, assessed by RTs/error rates, were *colour-font-alphabet* (21 participants for RTs/20 for error rates) *colour-alphabet-font* (8/10), *font-alphabet-colour* (2/2), *font-colour-alphabet* (2/2), *alphabet-colour-font* (4/4) and *alphabet-font-colour* (2/2). Nevertheless, given the statistically significant differences in task difficulty, it was concluded that the three tasks were indeed of asymmetric difficulty, with the colour task easier than the font task, which in turn was easier than the alphabet task. The analysis then proceeded to test the main hypothesis.

Mean and standard deviations of n-2 repetition costs (for RTs and error rates), for each task pairing and direction of alternation, are given in table 8.4. The analysis of n-2 repetition costs consisted of a 3 × 2 within-subjects ANOVA, with factors of task pairing (colour-font vs. colour-alphabet vs. alphabet-font) and switch direction (EHE vs. HEH), to test the hypothesis that n-2 repetition costs for hard-easy-hard triplets would be greater than for easy-hard-easy triplets. For RT, there was a significant main effect of switch direction, $F(1, 37) = 4.441, p = .042$, partial $\eta^2 = 0.022$, consistent with the main hypothesis. The main effect of task pairing was not significant, $F(2, 74) = 1.06, p = .353$, and nor was the task × switch direction interaction ($F(2, 74) = 0.218, n.s.$). For error rates, the main effect of switch direction failed to reach significance ($F(1, 37) = 2.815, p = .102$). The main effect of task pairing was also not significant ($F(2, 74) = 0.064, n.s.$). However, the task × switch direction interaction was significant, $F(1.98, 73.19) = 3.400, p = .039$, partial $\eta^2 = 0.034$ (Greenhouse-Geisser corrected).

Given the statistically marginal nature of the results, the individual differences

Task Pairing	Easy-Hard-Easy		Hard-Easy-Hard	
	RT (ms)	Error rate (%)	RT (ms)	Error rate (%)
Colour-Font	7 *(202)*	-0.95 *(4.01)*	98 *(292)*	1.82 *(4.96)*
Colour-Alphabet	16 *(249)*	-0.45 *(4.67)*	100 *(266)*	0.82 *(4.57)*
Font-Alphabet	-10 *(211)*	-0.96 *(4.23)*	31 *(263)*	-0.47 *(5.51)*

Table 8.4: Experiment 1: Mean (SD) n-2 repetition costs for RT and error rates by task pairings and switch direction

in task difficulty present one area for potential concern. For example, the larger n-2 repetition costs observed for HEH alternations may have been partially driven by the 18 participants for whom the average ordinal difficulty of tasks was reversed. The results were therefore analysed again according to relative task difficulty, with the three tasks for each participant ranked as primary, secondary, and tertiary tasks on the basis of mean RTs, in ascending order of difficulty. For RT, there was no statistically significant main effect of task, $F(2, 76) = 1.02, p = .36$. However, the main effect of switch direction was significant, $F(1, 38) = 4.58, p = .039, \eta_p^2 = 0.030$. The task × direction interaction was not significant, $F(2, 76) = 1.09, p = .34$. A parallel analysis of error rates revealed no statistically significant effects (main effect of task: $F(1, 38) = 1.88, p = .16$; main effect of direction: $F(1, 38) = 0.00, p = .95$; interaction: $F(2, 76) = 1.10, p = .34$).

8.3.4 Discussion

The results, a significant main effect of switch direction for RTs, and a nonsignificant trend for error rates, with greater n-2 repetition costs for HEH triplets than EHE triplets, contrast with those of Arbuthnott (2008a), who reported greater inhibition of the easier task (larger n-2 repetition costs for EHE switches). They are, however, consistent with the main hypothesis, and the prediction of the model.

How might one account for the differences in direction of effect between this experiment, and the findings of Arbuthnott (2008a)? As discussed above, there are a number of differences in task design. Specifically, the current procedure used tasks with three possible responses, allowing the procedure to remove stimulus congruency and varying response set overlap as possible confounding variables. Additionally, the three tasks (font, alphabet, and colour) were chosen to avoid potential overlap in their task-sets.

One feature of the present results is that the magnitude of the n-2 repetition costs observed are rather small. However, the tested hypothesis concerned the modulation of the n-2 repetition cost, with no prediction as to the magnitude of the n-2 repetition cost overall. One potential concern is that the absence of an n-2 repetition cost might suggest the absence of task inhibition, precluding its modulation in the present experiment. While there is reason to believe n-2 repetition cost of zero does not represent zero task inhibition, merely reduced task inhibition (cf. Grange et al., 2013, and as also predicted by the present model, chapter 4), small task inhibition effects may be one reason for the marginally significant results, and small effect size. It is therefore reassuring that the additional analysis conducted for ordinal task difficulty, in which individual differences in task difficulty are accounted for, produced a slightly larger effect size for reaction times than the first analysis. If the procedure were modified to produce larger n-2 repetition costs, it is reasonable to expect a larger effect size for the modulation of this cost by task difficulty. The next experiment therefore aimed to obtain stronger support for the model by modifying the procedure to produce larger n-2 repetition costs overall.

8.4 Experiment 3

8.4.1 Introduction

One experimental factor known to produce larger n-2 repetition costs is the use of non-transparent or abstract task cues, where the relationship between the cue and the task is arbitrary rather than bearing a resemblance (e.g., using geometric shapes to represent which task to perform next, as in Houghton et al., 2009). Experiment 3 featured a modified procedure with non-transparent cues with the aim of obtaining larger overall n-2 repetition costs.

8.4.2 Method

Participants

As in the previous experiment, 40 participants were recruited through the Birkbeck psychology participant database, of whom 28 were female. Participants received £7.50 in cash.

Design

The design was identical to experiment 2.

Tasks

The tasks were identical to experiment 2, with the exception that tasks were cued by a frame appearing simultaneously to, and enclosing the target letter. The frame could take the shape of a circle, square or hexagon. The association between frame shape and target task was randomised for each participant.

Procedure

As participants were required to learn cue-task associations, they received additional practice trials (compared to experiment 2), performing 15 trials for each task, before completing a practice switching block of 20 trials. The main experiment consisted of 8 blocks of 60 trials.

8.4.3 Results

Of the 40 participants, data was not analysed from two participants who failed to complete the task, and five who exceeded an error rate of 20%, leaving data from

Task	RT	Error rate
Colour	1866 *(498)*	5.49% *(3.63%)*
Font	2079 *(446)*	7.45% *(4.43%)*
Alphabet	2095 *(490)*	6.66% *(5.31%)*

Table 8.5: Mean (SD) of RTs (milliseconds) and error rates for each task

33 participants. The data were analysed as for experiment 2. For consistency with the previous experiment, trials exceeding 20s were excluded from the reaction time analysis. However, the pattern of data was robust to stricter RT thresholds of 10s and 5s. Additionally, error trials and the two subsequent trials were also excluded from the RT analysis. Mean RTs and error rates for the three tasks are presented in table 8.5.

Unlike experiment 2, the difference in difficulty of the three tasks was equivocal. For RT, ANOVA suggested significant differences in task difficulty, $F(2, 64) = 7.70, p = 0.010$, however pairwise t-tests suggested significant differences between colour and alphabet ($t(32) = 3.27, p = 0.007$) and colour and font ($t(32) = 3.29, p = 0.007$) pairings, but not alphabet and font ($t(32) = 0.26, p = 0.796$). For error rates, ANOVA revealed overall significant differences between tasks, $F(2, 64) = 4.44, p = 0.016$, but pairwise t-tests suggested only significant differences between colour and alphabet ($t(32) = 3.13, p = 0.011$) but not colour and font ($t(32) = 1.52, p = 0.277$) or alphabet and font ($t(32) = 1.37, p = 0.277$, all Holm-Bonferroni corrected for multiple comparisons).

Further inspection of the data revealed substantial individual differences in task difficulty. The ordinal ranking of task difficulty, in increasing difficulty, assessed by RTs/error rates, were *colour-alphabet-font* (11 participants for RTs/6 for error rates) and *colour-font-alphabet* (10/12), followed by *font-alphabet-colour* (5/6), *font-colour-alphabet* (4/3), *alphabet-colour-font* (2/5) and *alphabet-font-colour* (1/1).

The remaining analysis therefore proceeded according to ordinal task difficulty, as was also conducted for the previous experiment. On this basis, n-2 repetition costs (RT and error rates) for the critical conditions are shown in table 8.6. The

Task Pairing	Easy-Hard-Easy		Hard-Easy-Hard	
	RT (ms)	Error rate (%)	RT (ms)	Error rate (%)
Primary-Secondary	112 *(250)*	0.50 (5.85)	184 *(303)*	0.40 (6.42)
Primary-Tertiary	79 *(247)*	1.30 (6.03)	255 *(315)*	2.31 (7.72)
Secondary-Tertiary	77 *(387)*	0.05 (5.31)	153 *(308)*	1.22 (6.81)

Table 8.6: Experiment 3: Mean (SD) n-2 repetition costs for RT and error rates by task pairings and switch direction

mean proportion of trials retained was 81.3%. Like the previous experiment, this proportion was consistent across individual cells, with the fewest trials retained for tertiary-primary switches (79.9%) and the most for tertiary-secondary switches (82.8%).

Once again, the analysis of n-2 repetition costs consisted of a 3 × 2 within-subjects ANOVA, with factors of task pairing (primary-secondary vs. primary-tertiary vs. tertiary-secondary) and switch direction (EHE vs. HEH), testing the hypothesis that n-2 repetition costs for hard-easy-hard triplets would be greater than for easy-hard-easy triplets. For RTs, the results revealed that the main effect of task was not significant ($F(2, 64) = 0.59, p = 0.557$), but the main effect of direction was ($F(1, 32) = 5.92, p = 0.021$, $\eta_p^2 = 0.031$). The interaction was not significant ($F(2, 64) = 0.64, p = 0.53$). Pairwise t-tests revealed these differences reached significance for the primary-tertiary task pairing, $t(60.49) = 2.52, p = 0.014$ but not for primary-secondary, $t(61.75) = 1.05, p = 0.298$, or secondary-tertiary, $t(60.93) = 0.89, p = 0.379$.

No significant effects were found in the analysis of error rates (main effect of task: $F(2, 64) = 0.79, p = .46$; main effect of direction: $F(1, 32) = 0.81, p = .37$; interaction: $F(2, 64) = 0.17, p = .85$).

8.4.4 Discussion

Once again, as predicted, a significant effect of switch direction, with a greater n-2 repetition cost for HEH than EHE triplets, was found. Although the finding was only significant for the task pairing with the greatest differences in difficulty for each participant (the primary-tertiary pairing), the trend was observed for each of the task pairings.

As anticipated, mean reaction times and error rates were greater for experiment 3 than experiment 2, reflecting the increased difficulty introduced by non-transparent task cueing. Unexpectedly, non-transparent task cueing appeared to modify the relative difficulty of each of the tasks, introducing substantial individual differences. However, as our hypothesis concerns the relative difficulty of tasks, rather than differences between specific tasks, the inferences from these results remain valid.

8.5 General Discussion

Both experiments 2 and 3 were consistent with the prediction of the model, that HEH alternations produce larger n-2 repetition costs than EHE alternations, although this was inconsistent with the results of Arbuthnott (2008a). Experiment 2 did not produce large n-2 repetition costs overall, particularly for EHE alternations for which they were negligible. However, the prediction of the model was that the n-2 repetition cost would be modulated by task difficulty. The model made no prediction about the magnitude of the cost overall. Consistent with this, using non-transparent cues, experiment 3 replicated these results with larger n-2 repetition costs. These two results empirically confirm the predictions of the model.

The model predicts that HEH switches should produce greater n-2 repetition costs than EHE, for the same reason that the model of Gilbert and Shallice (2002) predicts switch cost asymmetries when switching between two tasks. Firstly, executing a harder task against interference from easier, but irrelevant, stimulus dimensions requires a greater degree of biasing by task-demand units than executing an easier task against interference from a harder task. Secondly, if the next trial is a switch, HE switches have a greater switch cost than EH switches because more task demand activation is carried over (Gilbert & Shallice, 2002). Finally, in a system with a backward inhibition mechanism, the simultaneous activation of task demand units on the n-1 trial means more conflict is detected on the n-1 trial for HEH switches than EHE switches. As a result the n-2 task receives more inhibition. In contrast, on a two-switch trial, the difficulty of the n-1 task has a much less pervasive effect on performance on trial n. This overall effect, greater n-2 repetition costs for HEH trials, is precisely what was observed.

These empirical results extend the growing literature on n-2 repetition costs by showing that the effect is modulated by the relative difficulty of the tasks involved. The typical size of the basic n-2 repetition cost (e.g., 31 – 35 msec, as reported by Mayr & Keele, 2000) is smaller than the typical size of the n-1 switch cost (e.g., 200 msec, as reported by Rogers & Monsell, 1995), and consequently it is harder to obtain positive evidence for the cost.[2] The relative difference in size is not surprising given that n-2 repetition costs relate to effects that persist across two switches (and a longer temporal window), as opposed to one, but our finding that the n-2 repetition cost is also sensitive to task difficulty (albeit in the opposite direction to the only previous study of asymmetric repetition costs; though see the earlier discussion for potential reasons why) further suggests that considerable methodological care is required to isolate such costs.

---

[2] There is evidence of substantial individual differences in n-2 repetition costs, with Grange and Juvina (2015) recently reporting costs of up to 200ms for certain individuals.

Chapter 9

Further Simulations of Backward Inhibition

This chapter extends the range of behavioural predictions made by the model with two sets of additional simulations. The first concerns the effect of manipulating the response congruency of the target stimuli. For tri-dimensional stimuli used in the three-task switching paradigm, for any target dimension relevant to the current task, two irrelevant dimensions afford responses that may be fully congruent, fully incongruent, or one-congruent and one-incongruent (semi-congruent). These graded levels of congruency are assessed for their effects on response times and n-2 repetition costs, when they occur on the first, second and third trial of a triplet. In addition, the data is also analysed to determine whether the model predicts the alternative empirically observed inhibitory effect of dimension inhibition/competitor rule suppression. The model predicts results consistent with empirically observed data for trial n congruency, and offers a novel prediction of an effect of trial n-1 congruency on n-2 repetition costs, albeit with very small effect size. The second simulations concern the modulation of the *n-2 repetition cost when the two intertrial intervals (RCI_{n-1} between trials n-2 and n-1, and RCI_n between n-1 and n) are varied systematically. Empirically, n-2 repetition costs are affected by RCI_{n-1} but not RCI_n, a finding that has widely been interpreted as supporting the theory that the n-2 repetition cost is related to conflict occurring on the n-1 trial. The simulations reveal that the backward inhibition model does indeed predict this empirical pattern.*

9.1 Simulation 9: Response congruency effects

9.1.1 Trial n graded response congruency effects

Response congruency refers to the congruence or incongruence between the correct responses to each task afforded by a multidimensional stimulus. For example, when categorising the digit ‘3’ according to the two categories of *odd/even* and *low/high*, the responses are said to be congruent if the odd and low responses are mapped to the same physical response, such as the same response key.

While a number of studies have examined the effect of varied levels of response congruency on task switching, and some have even offered computational accounts (e.g., Brown et al., 2007; Schneider & Logan, 2009), only a small number have extended this line of research to the three-task paradigms. While the correct response to a bi-dimensional stimulus, with respect to a specific task, may be congruent or incongruent, tri-dimensional stimuli have three possible congruency states: congruent, semi-congruent, and incongruent. For example, given the three possible digit classification tasks *odd/even*, *low/high* and *prime/nonprime* mapped to left and right response keys respectively, the stimulus *6* has a correct response (right) congruent with both the odd-even task’s and the prime-nonprime task’s response (i.e., fully congruent), *7* is congruent on only one of the irrelevant tasks (semi-congruent), and *4* is incongruent with both irrelevant tasks (fully incongruent).

Behavioural studies have found that graded response congruence indeed affects reaction times in three-task versions of the digit classification task. Arbuthnott (2005), using a three digit classification tasks, analysed only fully- and semi-congruent responses, finding an effect of response congruence (fully congruent faster than semi-congruent) that did not interact with n-2 repetition cost. Longman et al. (2014) reported two experiments which used a three-task digit classification paradigm, in which three possible stimuli were presented simultaneously in different locations, with a cue indicating both the relevant stimulus and the task to perform (tasks were invariantly mapped to each location). Both experiments revealed an effect of graded response congruence on RT and errors, with no reliable interaction with switch costs. Significant n-2 repetition costs were restricted to errors only for experiment 1,[1] although the unreliability of this effect may have been due to parameters of the experiment known to reduce n-2 repetition costs, such as the presence of immediate task repetitions (Philipp & Koch, 2006) and long RCIs (Gade & Koch, 2005). Finally, Schneider (2014) reported similar findings, finding significant linear increases in RT from fully-incongruent, semi-congruent, to congruent stimuli, and a corresponding linear decrease in error rate. While n-1 switches were significantly slower and more error prone than n-1 repeats, neither n-1 switch cost was modulated by response congruency. In his experiment, Schneider did not find a significant n-2 repetition cost for RT or error rate.

In sum, the available behavioural evidence consistently suggests graded effects of response congruence: fully incongruent stimuli evoke slower and more error prone responses than semi-congruent stimuli, which in turn are slower and more error prone than fully congruent stimuli.

One important implication of these findings concerns the design of three-task

---

[1] This finding is in contrast to that of Arbuthnott (2009), who found n-2 repetition costs under similar conditions, where the cue location was constant and overlapping and the location of the target varied.

switching studies. A large number of the studies in the empirical literature have used three classification tasks with two overlapping responses. Given the empirical and simulation results above, it is clear that in this case, specific stimulus items are conflated with different levels of response conflict. Eliminating these effects requires a modified procedure, such as tasks affording three possible responses, where responses may be mapped such that response congruence effects are independent of specific stimuli. For example, through the use of composite stimuli such as digit-letter combinations displayed in various colours.

9.1.2 Trial n-1 response congruency effects including Dimension Inhibition/Competitor Rule Suppression

The previously discussed response congruency effect is a intra-trial, or non-sequential effect: that is, the effect on RT is due to the congruence status of the trial in question. However, inter-trial sequential congruence effects may also affect reaction time. The competitor rule suppression (CRS, Meiran, Hsieh, & Dimov, 2010) or dimension inhibition (Goschke, 2000) effect is where slower RTs occur for a task in which the response afforded by that task was previously incongruent with the correct response, compared to when the task's response was congruent with the correct response. This may be represented schematically as follows, given a multi-dimensional stimulus affording tasks A and B, each with possible responses 1 and 2 (or, in general, $*$), with the cued task in bold .

The following two-trial sequence:

$$\mathbf{A}_1 B_2 ... A_* \mathbf{B}_*$$

Produces slower RTs for all responses (indicated by B_*), than the following sequence:

$$\mathbf{A}_1 B_1 ... A_* \mathbf{B}_*$$

While this effect is well known in the two-task switching literature (e.g., Goschke, 2000), a few recent studies have examined the effect in three-task switching paradigms, and thus related the inhibitory effect to backward inhibition Thus, Meiran et al. (2010) suggested that this effect reflected an inhibitory process occurring as a result of incongruence on trial *n-1*, acting on the task-set of the irrelevant task (in this case, task B) in order to suppress interference. This inhibition persists into trial *n*, leading to slower responses for the now-relevant task B. In order to demonstrate that this effect reflects inhibition directed at a specific task-set, it is necessary to rule out generalised effects of slowing following incongruent, rather than congruent trials, such as have been attributed to conflict (e.g., Botvinick et al., 2001; Brown et al., 2007). In support of this view, Meiran et al. (2010) reported an experiment utilising four tasks, finding a significant graded intra-task congruency effect (with significant pairwise differences between each congruency level, of which there were four due to four tasks) for both RT and errors, a significant CRS (inter-task congruency) effect for both RT and errors, and a significant n-2 repetition cost for RT and errors. However, as there was no indication of any interaction between the n-2 repetition cost and the CRS effect, the authors invoked the principle of additivity (Sternberg, 1969) to infer that backward inhibition and CRS affected different processing stages, and thus were independent inhibitory effects.

More persuasively, a number of studies have demonstrated that n-2 repetition costs and CRS effects are variably affected by different manipulations. Regev and Meiran (2016) manipulated cue-type, which affected the difficulty of retrieving task-response mappings. Dimension cues referred only to the specific task (i.e., ‘colour’) while mapping cues referred to specific keys (‘red/green’ indicated left or right keys respectively). Regev and colleagues found that while the n-2 repetition cost was not affected by cue type, CRS was larger when dimension cues were used, compared to mapping cues, both when cue-type varied by block, and by trial. This finding was interpreted to suggest that CRS and BI are two different inhibitory phenomena, of which CRS (but not BI) is modulated by making the retrieval of task-cue mappings from WM more difficult. Given that earlier studies suggest backward inhibition is modulated by making the cue-task retrieval more difficult (i.e., by manipulating transparent vs. non-transparent cues (e.g., Houghton et al., 2009), this pattern of results together suggests that it is the cue-response key-mapping aspect of the task-set that is crucial for CRS, while cue-task retrieval affects BI. One test of this possibility might be to see whether a cue-transparency manipulation such as that used by Houghton et al. (2009) affects CRS, or merely BI. At present, no study using a cue-transparency manipulation appears to have systematically examined both effects in the same experiment. If cue-mapping retrieval were entirely separate from cue-task retrieval, one might expect it to affect CRS but not BI; if, however, task-set retrieval were hierarchically structured, involving cue-task retrieval prior to cue-mapping retrieval, the manipulation could conceivably influence both CRS and BI.

Similar behavioural dissociation of the two effects comes from Astle, Jackson, and Swainson (2012), who conducted three experiments with varying task parameters which dissociated putative dimension inhibition from backward inhibition effects. In a first experiment involving letter classification, CTIs were predominantly long (1200ms), with some short (200ms) to encourage early engagement (analysis was only conducted on all-long sequences), with the cue disappearing with target onset. This experiment found marginally significant effects of dimension inhibition for error rates only, but no effect of backward inhibition. A second experiment, involving classification of a single item within a three-item (letter/digit/symbol) compound stimulus, predominantly using short-CTI (100ms) trials, found the reverse effect: no evidence of dimension inhibition, but a significant effect of backward inhibition for error data only. A third experiment, which combined the task from the first experiment with the design of the second, found similar results to the first, namely a main effect of dimension inhibition for RT but not errors (although a significant task × dimension inhibition interaction was observed for both RT and errors, suggesting the effect was observed for some tasks but not others), but no evidence for backward inhibition. While the observed effects were somewhat marginal, with inconsistencies as to whether the effects were observed in the RT or error rate data, the authors concluded that backward inhibition and dimension inhibition were dependent on the different tasks used, and not the other design parameters. Backward inhibition required the task involving classification of unitary stimuli, and did not occur for compound stimuli. Dimension inhibition, on the other hand, only occurred for compound stimuli. Thus, Astle et al. (2012) provided more direct evidence than Meiran et al. (2010) that backward inhibition and dimension inhibition/CRS are separate inhibitory processes, however their empirical results are relatively weak and would be strengthened by replication.

9.1.3 Simulation rationale

Tentative empirical evidence suggests that backward inhibition and dimension inhibition are two independent inhibitory effects observable from sequential analysis of three-task switching paradigms. Nevertheless, response congruency presents an issue for the computational model developed in this thesis. When conducting simulations, it is necessary to specify the congruency between all stimulus dimensions for each trial. Furthermore, as response congruency has a large effect on model reaction times, it is necessary to carefully control congruency for all simulations. Except where specified, simulations have been conducted by including equal numbers of all permutations of response congruency (fully congruent, semi-congruent and fully incongruent) for each position in a block of three trials. However, further systematic analysis is necessary to fully understand model behaviour — averaging across different congruency conditions may mask effects if they interact with response congruency on a particular trial. Due to bottom-up activation in the model, irrelevant stimulus dimensions that are incongruent or congruent with the currently relevant stimulus, activate output units, which in turn activate task-demand units irrelevant task-demand units. Therefore, stimulus congruency does influence the degree of conflict at the task-demand level, and thus would contribute to backward inhibition. The extent to which this takes place, and its effect on model behaviour, requires further systematic simulation.

Furthermore, the model is ideally placed to generate predictions about the empirical effects of sequential response congruency within a task switching model with a backward inhibition mechanism of the type previously described. Of specific concern is whether model behaviour matches that described in the empirical literature for *n* graded response congruency, *n-1* response congruency (dimension inhibition/CRS).

9.1.4 Method

The model was tested on 38,400 blocks of three trials, consisting of either ABA or CBA sequences. For each trial in the block, stimulus congruency was either semi-congruent (IC) or fully incongruent (II). For simplicity, fully congruent trials were not considered, although it may be assumed that without any incongruent stimulus dimensions, model processing will proceed more quickly than the semi-congruent sequences. Thus, the simulation contained four possible factors: sequence type (ABA vs. CBA), and a congruency factor for each of trials 1, 2 and 3 within the block.

9.1.5 Results

Consistent with previous simulations, when considering only sequence type and collapsing across the three congruency factors, the model produces clear n-2 repetition costs, as illustrated in figure 9.1. This n-2 repetition cost was tested using a Welch two sample t-test. ABA sequences were significantly slower than CBA sequences (77.10 cycles vs. 71.28 cycles), $t(37102.02) = 19.59, p < .0001$.

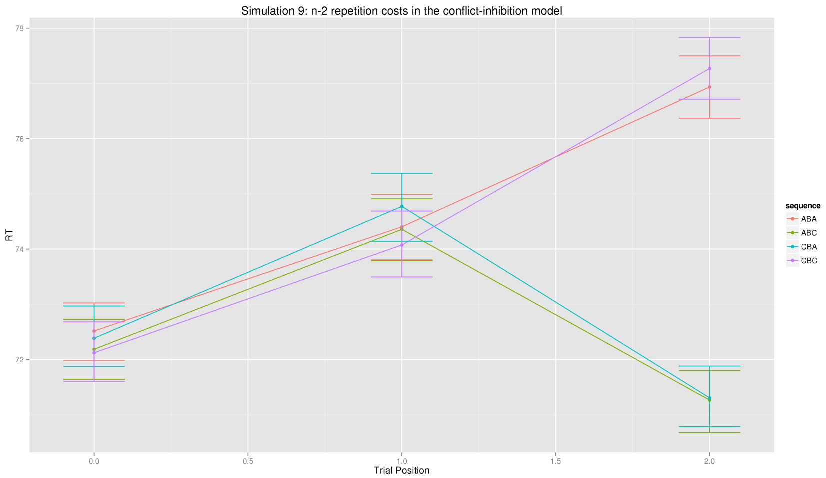


Figure 9.1: Simulation 9: n-2 repetition effects when collapsed across congruency condition.

Next, we consider the effect of the three response congruency factors alongside sequence type. Figure 9.2 plots the three congruency factors. The two levels of trial 3 congruency, the factor with the most direct influence on performance, may be compared via the two vertical panels. Trial 2 congruency may be compared via the first and second, and via the third and fourth horizontal panels, while trial 1 congruency may be compared via the first and third, and via the second and fourth panels. Although this figure graphically represents the complete data for this simulation, due to its complexity, the effect of various sub-combinations of factors is next considered by collapsing the other factors.

Trial *n* graded response congruency effects

The direct intra-trial effect of response congruency on model performance is reflected by the trial 3 response congruency factor, by comparing the top and bottom panels of figure 9.2. As error bars represent 95% confidence intervals, it is clear graphically that trial 3 response congruency has a large effect on model reaction time, with semi-congruent trials faster than fully-incongruent trials. This was assessed via a 2×2 ANOVA (sequence type × trial 3 congruency) which found highly significant main effects of sequence, $F(1, 37396) = 251.99, p < .0001, \eta_p^2 = .007)$, 3rd trial congruency, $F(1,37396) = 17236.19, p < .0001, \eta_p^2 = .315$), and also a highly significant $sequence \times 3rd\ trial\ congruency$ interaction, $F(1,37396) = 23.64, p < .0001, \eta_p^2 < .001$. Thus, both n-2 repetition costs and trial 3 congruency effects were found to be significant, although respective effect sizes suggest the congruency effect is by far the larger effect. While trial 3 congruency did significantly modulate the n-2 repetition cost in the simulation, the effect size was relatively negligible.

Although only semi-congruent and fully-incongruent trials were represented in the model (unlike Schneider, 2014, who modelled, and measured empirically, all three possible levels of response congruency), the present model behaviour shows significant differences between the congruency conditions. Given that model RT is dependent on competition between competing response units (and task-demand units, via bottom-up activation), and fully congruent stimuli would evoke no bottom-up competing activation whatsoever, we may reason that fully congruent stimuli would be faster than semi-congruent stimuli in all cases.

Trial *n-1* sequential congruency effects

Next, we turn to the sequential effects of congruency on trial n-1. Firstly, we consider the effects of graded response congruency on trial n-1 (i.e., the overall effect of fully vs. semi-congruent stimuli). As a test of all possible effects, a 4-way ANOVA was conducted: sequence (S, ABA vs. CBA) × trial 1 congruency (C_{n-2}, IC vs. II), × trial 2 congruency (C_{n-1}, IC vs. II) × trial 3 congruency (C_n, IC vs. II). The results were as follows:

Main effects: Sequence, $F(1,37119) = 575.72, p < .0001, \eta_p^2 = .0153$, 1st trial congruency, $F(1,37119) = 4.70, p < .030, \eta_p^2 = .0001$, 2nd trial congruency, $F(1,37119) = 199.45, p < .0001, \eta_p^2 = .0053$, 3rd trial congruency, $F(1,37119) = 18055.73, p < .0001, \eta_p^2 = .3272$.

Two way interactions: The interactions between sequence and irrelevant

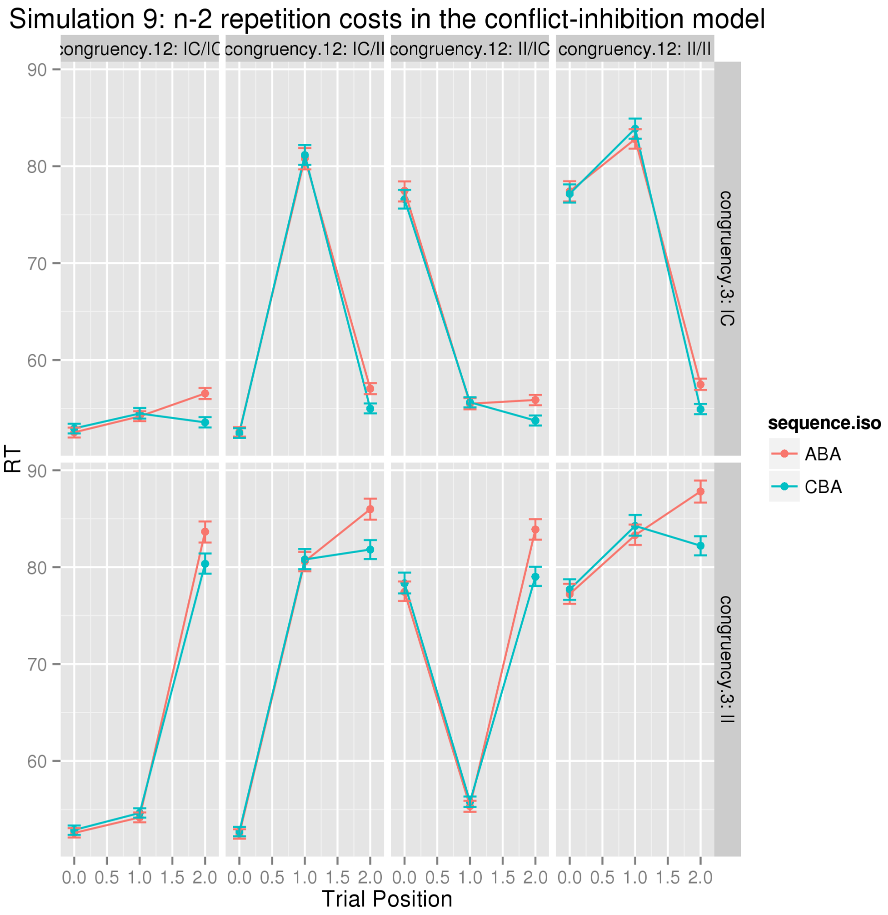


Figure 9.2: Simulation 9: n-2 repetition effects by congruency condition.

stimulus congruency suggest a modulation of the n-2 repetition cost: $S \times C_{n-2}$, $F(1, 37119) = 65.12, p < .0001, \eta_p^2 = .0017$; $S \times C_{n-1}$, $F(1, 37119) = 132.45, p < .0001, \eta_p^2 = .0035$; $S \times C_n$, $F(1, 37119) = 40.68, p < .0001, \eta_p^2 = .0011$.

Interactions between trial congruency were not generally significant, apart from that between 2nd and 3rd trial congruency: $C_{n-2} \times C_{n-1}$, $F(1, 37119) = 1.78, p = 0.18$; $C_{n-1} \times C_n$, $F(1, 37119) = 0.45, p = 0.50$; $C_{n-1} \times C_n$, $F(1, 37119) = 24.22, p < 0.0001, \eta_p^2 = .0007$.

The **higher-order interactions** between sequence and trial congruency were not, generally, significant: $S \times C_{n-2} \times C_{n-1}$, $F(1, 37119) = 0.08, p = 0.78$; $S \times C_{n-2} \times C_n$, $F(1, 37119) = 0.78, p = 0.38$; $S \times C_{n-1} \times C_n$, $F(1, 37119) = 5.60, p = 0.018, \eta_p^2 = .0001$; $C_{n-2} \times C_{n-1} \times C_n$, $F(1, 37119) = 2.33, p = 0.13$. Finally, the

4 way interaction, $S \times C_{n-2} \times C_{n-1} \times C_n$ was not significant, $F(1, 37119) = 0.46, p = 0.49$.

The trial *n* congruency effect, is therefore significant, as were a basic n-1 and n-2 congruency effect. However, the size of the trial *n* congruency effect was several orders of magnitude larger ($\eta_p^2 = 0.327$) than more remote effects of congruency ($\eta_p^2 < .01$), and detection of these latter effects empirically seems unlikely. Additionally, all three basic congruency effects interacted with the n-2 repetition cost, albeit with empirically negligible effect sizes ($\eta_p^2 < 0.01$). It is of interest that the biggest effect on the n-2 repetition cost was n-1 congruency, which supports the suggestion that the conflict generated on the n-1 trial is critical to triggering backward inhibition. The higher-order interactions, where significant, were also of negligible effect size so they will not be considered further.

Dimension Inhibition/Competitor Rule Suppression

We now turn to the so-called dimension inhibition effect. As described previously, dimension inhibition is putatively represented by the contrast between the response time on trial n, following two types of semi-congruent stimulus on trial n-1, for example in a sequence involving a switch from task B to task A. In the inhibited condition, the n-1 stimulus relevant to task A is incongruent with the relevant stimulus (to task B). Thereafter, when switching to task A, responses are slower. In the control condition, the n-1 stimulus relevant to task A is congruent with the task B stimulus, leading to faster responses. Thus, the dimension inhibition effect may be observed on all sequences where trial n-1 is a semi-congruent stimulus.

The trial 3 RT data was analysed with a 3-way ANOVA, with factors of sequence (ABA vs. CBA), Competitor rule suppression (CRS- vs. CRS+) and trial *n* congruency (II vs. IC). As the CRS condition only occurs when trial 2 is semi-congruent, all sequences with a fully-incongruent trial 2 were excluded from the analysis.

The analysis revealed a significant main effect of sequence, $F(1, 18738) = 122.37$, $p < .0001$, $\eta_p^2 = 0.007$, and a significant main effect of trial *n* congruency, $F(1, 18738) = 8205.7$, $p < .0001$, $\eta_p^2 = 0.305$. The main effect of CRS was not significant, however, $F(1, 18738) = 0.99$, $p = .321$. Neither did CRS interact with trial *n* congruency, $F(1, 18738) = 1.30$, $p = .254$, or sequence, $F(1, 18738) = 2.28$, $p = .131$. There was, however, a sequence $\times$ trial *n* congruency interaction, $F(1, 18738) = 6.70$, $p = .009$, $\eta_p^2 < .0001$, which was discussed previously. The three-way interaction was not significant, $F(1, 18738) = 0.13$, $p = .717$. These results are presented graphically in figure 9.3.

9.1.6 Interim discussion

Overall, then, the simulation showed no evidence for a dimension inhibition effect, despite other effects of trial *n-1* congruency. This result would seem to support the tentative empirical evidence from Meiran et al. (2010) and Astle et al. (2012) that the source of the dimension inhibition effect is not backward inhibition, at least not from the type implemented in the present model, nor can it be attributed merely to a lateral inhibition effect.

However, the model predicted a somewhat different trial *n-1* congruency effect. Instead of depending on whether the currently relevant stimulus (trial n) is congruent with a previously irrelevant dimension, as in the CRS effect, it depends on whether the currently congruent irrelevant dimension (on semi-congruent trials) was performed on the previous task.

Consider the effect of whether the congruent irrelevant dimension on an IC trial corresponds with the previously performed, or nonperformed, task. It is possible that this may have a direct effect on performance on that trial, and/or an indirect (sequential) effect on subsequent trial. For example, if the congruent dimension matches the previously performed task, one might expect that if the task

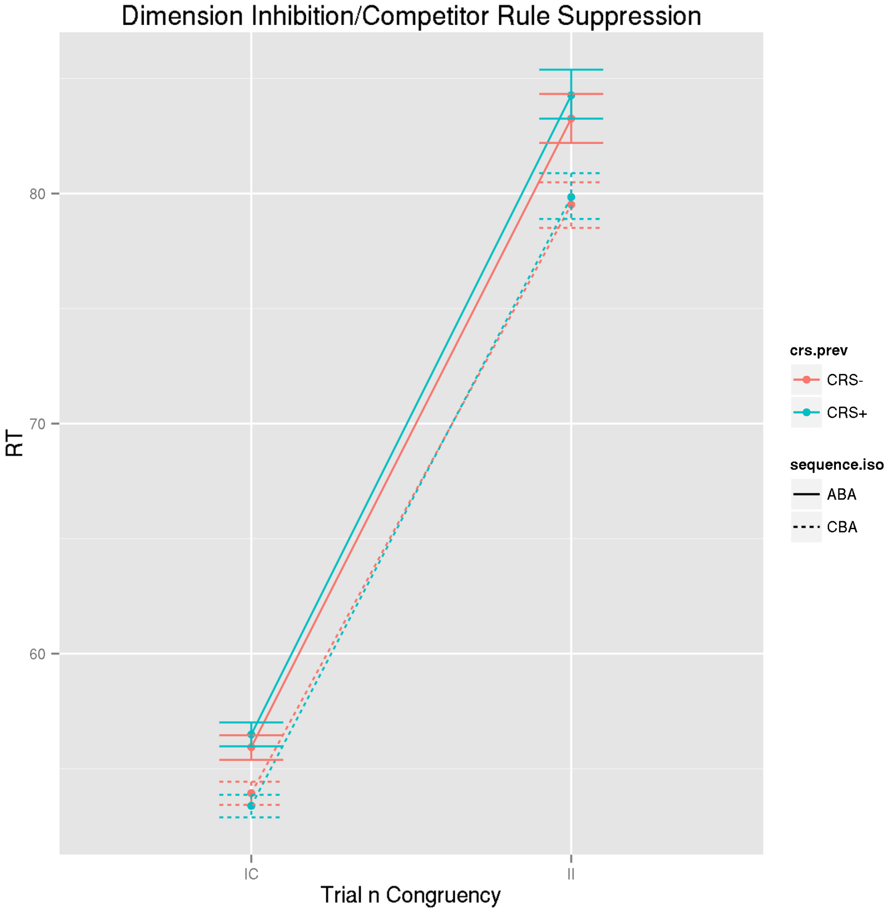


Figure 9.3: Analysis of competitor rule suppression effects in the model.

is inhibited when switching to a new task, the congruent dimension would evoke less cross-task interference than if it corresponded with the previously nonperformed task. On the other hand, residual activation of the previously performed task would also produce activation of that task's response nodes, hence it might amplify the facilitatory effect of the congruent irrelevant dimension. Given these interactive components of the model, it is difficult to arrive at a clear and unambiguous behavioural prediction, thus it is useful to examine the performance of the model in these situations.

9.1.7 Further analysis of sequential congruency effects

Given that this effect only applies to IC trials, and never (in the simulation) to the first trial in a triplet, the design of the simulation affords four possible comparisons. If both trials 2 and 3 were IC, there would be four possible combinations of whether the congruent trial corresponds with the previously performed task (TRUE) or nonperformed task (FALSE): TRUE/TRUE, TRUE/FALSE, FALSE/TRUE and FALSE/FALSE. However, if either or both of trials 2/3 is II, there are fewer conditions. Here, we consider the effect of matching trial 2/3 congruent dimensions to an immediately preceding task, on trial 3 RTs.

II/IC/IC, IC/IC/IC (i.e., trials 2 and 3 are both semi-congruent; upper far left, upper mid right): A 3-way anova was conducted with factors sequence (ABA vs. CBA), trial 2 previous task irrelevant-stimulus congruency, and trial 3 previous task irrelevant-stimulus congruency. There were highly significant main effects of sequence, $F(1.9592) = 92.787, p < .0001, \eta_p^2 = .0096$, trial 3 previous task irrelevant-stimulus congruency, $F(1.9592) = 387.20, p < .0001, \eta_p^2 = .0388$, and trial 2 previous task irrelevant-stimulus congruency, $F(1.9592) = 4.33p = .038, \eta_p^2 = .0005$. Additionally, there was a significant sequence $\times$ trial 2 previous task irrelevant-stimulus congruency interaction, $F(1, 9592) = 4.52, p = .033, \eta_p^2 = .0005$. None of the other interactions approached significance. These results suggest that the n-2 repetition cost is modulated by the previous task irrelevant-stimulus congruency on the final trial only, and is thus a direct but not sequential effect, but with a very small effect size.

IC/II/IC, II,II,IC (i.e., trial 3 only is semi-congruent; upper mid left, upper far right): A 2-way ANOVA was conducted with factors sequence (ABA vs. CBA) and trial 3 previous task irrelevant-stimulus congruency. There were highly significant main effects of sequence, $F(1, 9592) = 74.29, p < .0001, \eta_p^2 =$

$.0077$ and trial 3 previous-task irrelevant stimulus congruency, $F(1, 9592) = 937.4, p < .0001, \eta_p^2 = .0890$. The sequence × trial 3 previous-task irrelevant stimulus congruency interaction approached significance, $F(1, 9592) = 2.87, p = .090$.

IC/IC/II, II,IC,II (i.e., trial 2 only is semi-congruent; lower far left, lower mid right): A 2-way ANOVA was conducted with factors sequence (ABA vs. CBA) and trial 2 previous task irrelevant-stimulus congruency. The main effect of sequence was highly significant, $F(1, 9142) = 58.45, p < .0001, \eta_p^2 = .0064$, while the main effect of trial 2 previous task irrelevant-stimulus congruency and the interaction did not approach significance.

Thus, in the model, for semi-congruent stimuli it matters which irrelevant stimulus dimension is congruent and which incongruent, in that trials were performed quicker when the dimension affording the previous task was congruent with the currently task dimension, than trials where the congruent dimension was not performed previously. However, this effect was largely independent of the n-2 repetition cost.

The only detected effects in the simulation with $\eta_p^2 > .001$ were the main effect of sequence (i.e., n-2 repetition costs) and trial 3 previous task irrelevant stimulus congruency (i.e., non-sequential effects). While trial 2 previous task irrelevant stimulus congruency was statistically significant for II/IC/IC sequences, indicating some sequential effect of congruency from the n-1 task, and some interactions with n-2 repetition costs. These effects were of very small effect size and even if the model predictions were correct, are unlikely to be detectable empirically.

9.1.8 Discussion

The model produces trial n response congruency effects consistent with the empirical literature (e.g., Schneider, 2014; Longman et al., 2014). However, neither

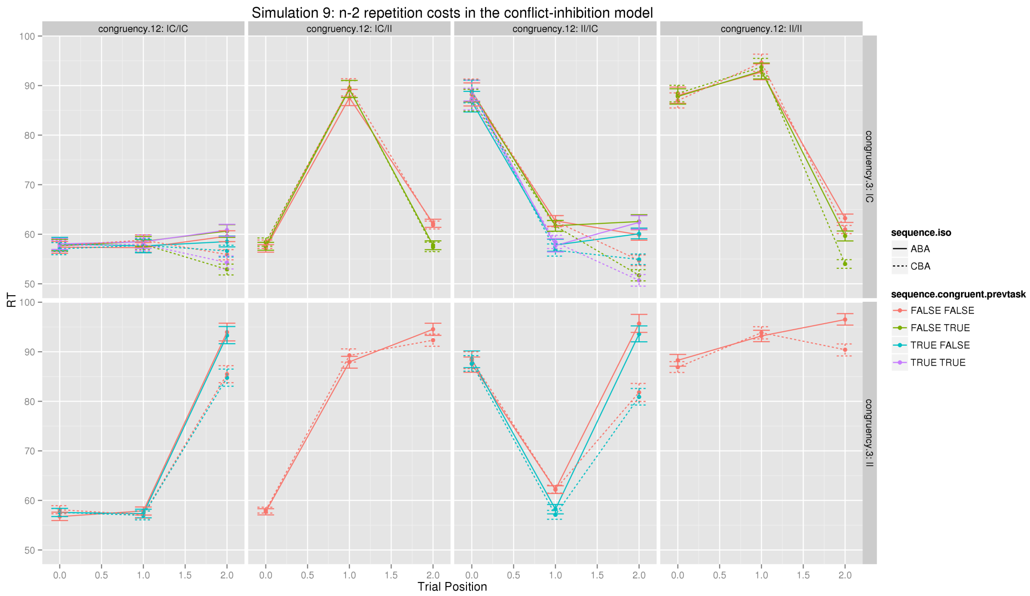


Figure 9.4: Simulation 9: n-2 repetition effects by congruency condition, split by whether congruent dimension matches (TRUE) or does not match (FALSE) the task dimension of the previous trial.

of those studies found a robust n-2 repetition cost, and nor did they explore the potential interaction between the n-2 repetition cost and the response congruency effect. While the model does predict such an interaction, the negligible effect size makes detecting it empirically unlikely.

While congruency on any of the three trials affects the n-2 repetition cost, n-1 congruency has the biggest effect on the n-2 repetition cost. This is consistent with our intuitive understanding of the operation of the model: incongruency represents a source of conflict, through a process of bottom-up activation, and greater conflict produces a larger backward inhibition effect. Importantly, in the model, response congruency is mediated through the activation of output units and then task demand units. At both output and task-demand unit level, incongruent possibilities are partly resolved through mutual lateral inhibition. However, conflict at the task-demand activation also triggers backward inhibition of task-sets.

There was no evidence for a dimension inhibition/CRS effect in the model. Meiran et al. (2010) argued, on the basis of the non-interaction between their observed n-2 repetition cost and the CRS effect that the two reflected different inhibitory processes, and that conclusion is supported by these results. Meiran et al. (2010) proposed that the target of the CRS effect is a specific task rule, while the backward inhibition effect is the representation of the task more generally. In the present model, task demand units represent sources of activation to all response units for a particular task, and inhibition to all response units for competing tasks. Thus, activating a single task demand unit both activates and inhibits mutually congruent response units. Within the model, task demand units, conceptually, seem closer to being an abstract representation of the task, rather than a specific task implementation, which presumably would affect a specific response. While the source of the CRS effect remains outside the model (i.e., it is unlikely to be due to residual conflict between task demand units, otherwise we might expect to see evidence of the effect in the model) it is also unclear what type of representation is the target of the CRS effect: whether it acts as another source of inhibition on the task demand units (i.e., acting on task representations more generally) or affecting specific task-demand response connections, such as by temporarily attenuating connection weights. This remains an open theoretical question.

9.2 Simulation 10: Variable intertrial intervals

One of the key empirical findings suggesting that n-2 repetition costs are an inhibitory phenomenon, rather than merely reflecting interference from a residually active irrelevant task, comes from studies which manipulated the intertrial interval (RCI, response-cue interval). When RCIs are manipulated blockwise, n-2 repetition costs tend to decrease with longer intervals (Koch et al., 2004; Meiran et al., 2000). This finding would be consistent with n-2 repetition costs being caused either by residual activation (i.e., of the B task in an ABA sequence), or residual inhibition (of the A task in the same sequence), which dissipates over time and is thus greater for shorter RCIs. Gade and Koch (2005), however, hypothesized that if task inhibition is primarily affected by decay of the n-1 task, n-2 repetition costs should mainly be affected by manipulation of the RCI preceding trial n (τ_2)— that is, task inhibition should decrease with longer intervals before the return to the inhibited task. Alternatively, if task inhibition were mainly affected by conflict between co-active tasks, n-2 repetition costs should mainly be affected by the RCI preceding trial n-1 τ_1— that is, task inhibition should decrease with longer intervals immediately before the trial that causes the 'A' task to be inhibited. In their experiments, when τ_1 and τ_2 could each be either short (i.e., 0.1s) or long (1.4s or 1.9s in experiments 1 & 2 respectively), n-2 repetition costs were affected by τ_1 but not τ_2. This finding has been taken to support the view that n-2 repetition costs are an instance of cognitive inhibition, and linked to intertrial conflict generated on the n-1 trial (Koch et al., 2010). As the present model represents an implementation of the inter-task conflict hypothesis, it should readily produce this same empirical pattern. This was tested in simulations 10a and 10b.

9.2.1 General methods

In the model, the concepts of residual activation and inhibition are modelled by two features. As in Gilbert and Shallice (2002), residual task activation is simulated by a proportion of task demand unit activation (determined by the task demand squashing parameter, here designated as ϕ), which is carred forward to subsequent trials. Inhibition is caused by the residual activation of conflict units, modelled by a proportion of conflict unit activation (conflict squashing parameter, χ) carried forward to subsequent trials. Thus, residual activation in the model, act_r, resulting from the previous trial, is calculated as:

$$act_r = \begin{cases} act_{task} \times \phi & \text{for task demand units} \\ act_{conflict} \times \chi & \text{for conflict units} \end{cases}$$

where act_{task} is the task demand unit activation, and $act_{conflict}$ is the final conflict unit activation, at the end of the preceding trial.

These calculations assume that RCI is constant and do not consider how the decay of activation over time will be affected by RCI. In this simulation, we assume that the equation calculates decay after unit time ($\tau = 1$). More generally, decay after time τ may be calculated as:

$$act_r = \begin{cases} act_{task} \times \phi^{\tau} & \text{for task demand units} \\ act_{conflict} \times \chi^{\tau} & \text{for conflict units} \end{cases}$$

Thus, the model parameters which affect residual activation and its decay are the task demand squashing parameter ϕ, the conflict squashing parameter χ, and the two values of τ used as short and long RCI.

9.2.2 Simulation 10a: Method

This simulation varied τ while maintaining ϕ and χ at their default values (0.2 and 0.5 respectively). Values of τ of 0.9 and 1.5 were used for the short and long RCIs, respectively. The simulation was run on the four different combinations of τ_1 and τ_2 for 2,000 3-trial blocks (1,000 of each ABA and CBA sequences).

9.2.3 Simulation 10a: Results

The RTs for trial 3 of 2SW and ALT sequences are plotted in figure 9.5. As can be observed from the figure, for a given set of parameters, ALT trials have longer RTs than 2SW trials, hence all produce an n-2 repetition cost. The τ_1 parameter is represented by line colour: short/long values represented by black/grey lines respectively. It is evident that longer values of τ_1 produce a greater n-2 repetition cost, in addition to producing longer RTs overall. The τ_2 parameter is represented by line style, with solid/dashed lines representing short/long values. Different val-

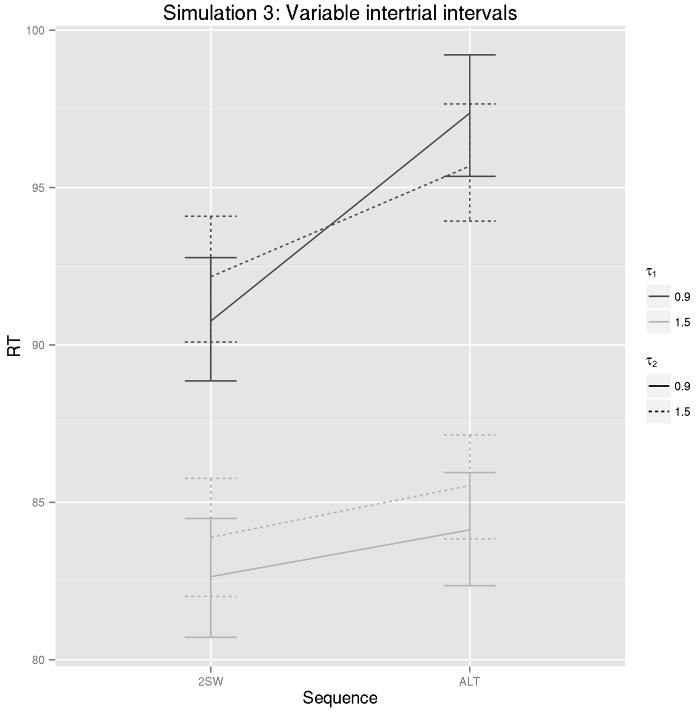


Figure 9.5: Simulation 10a: RTs for 2SW and ALT conditions. Black/grey lines represent short/long values of τ_1, while solid/dashed lines represent short/long values of τ_2. Error bars represent bootstrapped 95% confidence intervals.

ues of this parameter have a relatively insignificant effect on either overall RTs or n-2 repetition costs.

This graphical analysis was assessed quantitatively using a 3-way ANOVA with factors of $sequence \times RCI_{n-1} \times RCI_n$. The main effect of sequence (n-2 repetition costs) was significant, $F(1, 7301) = 23.09, p < .0001$, as was the main effect of τ_1, $F(1, 7301) = 213.88, p < .0001$. However, the main effect of τ_2 was not significant, $F(1, 7301) = 0.83, p = .36$. The effect of interest, that of RCI on n-2 repetition costs, was assessed by the two $\tau \times$ sequence interactions. First the $\tau_2 \times sequence$ interaction was not significant, $F(1, 7301) = 1.13, p = .29$. However, the $\tau_1 \times sequence$ interaction was highly significant, $F(1, 7301) = 6.71, p = .009$. Finally $\tau_1 \times \tau_2$ interaction, $F(1, 7301) = 1.14, p = .29$, was not significant, niether was the three-way $\tau_1 \times \tau_2 \times sequence$ interaction, $F(1, 7301) = 1.43, p = .23$. Thus, for these two values of τ, the model confirms the hypothesis and reproduces

Sequence		Intercept		$\beta(\tau_1)$		$\beta(\tau_2)$
2SW	105.32	[104.59 – 106.04]	-12.45	[-12.81 – -12.10]	0.14	[-0.20 – +0.49]
ALT	112.56	[111.81 – 113.31]	-14.57	[-14.93 – -14.20]	0.14	[-0.22 – +0.49]

Table 9.1: Regression coefficients [and 95% confidence intervals] for RT for 2SW and ALT blocks

the empirical finding, that n-2 repetition costs are affected by the RCI preceding the n-1 (τ_2), but not the n (τ_1), trial.

9.2.4 Simulation 10b: Method

To determine the respective effects of τ_1 and τ_2 on n-2 repetition coosts more generally, a further simulation was run in which the two were varied factorially, from 0.3 to 2.4 in increments of 0.3. For each of these 64 combinations, 1000 blocks of three trials were run for both ALT and 2SW sequences.

9.2.5 Simulation 10b: Results

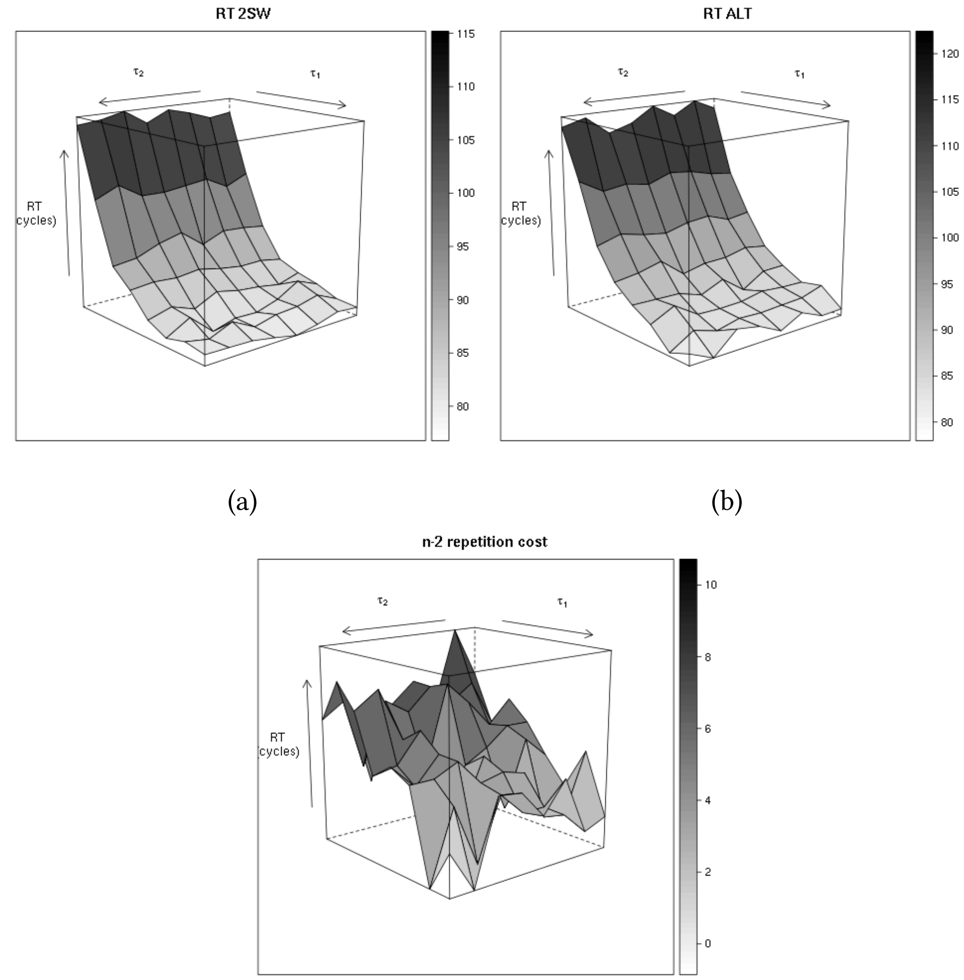
The mean RTs are represented as 3D plots in figure 9.6a and 9.6b, and the resulting n-2 repetition costs are plotted in figure 9.6c. Linear regression was used to estimate the effect of τ_1 and τ_2 on trial 3 RT, for both 2SW and ALT sequences. Estimates for β coefficients for τ_1 and τ_2 (and their bootstrapped 95% confidence intervals) are given in table 9.1. Note that the confidence intervals for $\beta(\tau_1)$ do not overlap, while both confidence intervals for $\beta(\tau_2)$ include the other estimated τ_2 coefficient, and indeed are nearly identical. The estimated coefficients yield the following regression equations, which may be combined to calculate n-2 repetition costs:

$$2SWRT = 105.22 - 12.45\tau_1 + 0.14\tau_2$$

$$ALTRT = 112.56 - 14.57\tau_1 + 0.14\tau_2$$

$$\ldots$$

$$\text{n-2 repetition cost} = 7.34 - 3.12\tau_1$$



(a) (b)

(c)

Figure 9.6: Simulation 10b: Effects on RT of varying RSI preceding n-1 and n trials (τ_1, τ_2) factorially. 9.6a: 2SW RTs 9.6b: ALT RTs. 9.6c: n-2 repetition costs.

Importantly, the n-2 repetition cost is dependent on τ_1, while the τ_2 term is cancelled out.

9.2.6 Discussion

The model's predictions clearly parallel the behavioural pattern observed by Gade and Koch (2005). Most critically, longer values of τ_1 reduce RTs of both 2SW and ALT trials. However, the former are reduced by less than the latter. This has the effect of decreasing n-2 repetition costs. τ_2 has no reliable effect on RTs. The general increase in RTs for both 2SW and ALT trials for shorter values of τ_1, but not τ_1 is also observed by Gade and Koch (2005), and consistent with blockwise RCI manipulations (e.g., Grange & Houghton, 2009; Koch et al., 2004).

This simulation demonstrates that a characteristic empirical pattern is predicted by the model with no changes to its architecture, beyond the assumption that residual activation decays over time (and hence decays more with longer intertrial intervals). It also suggests that the claim that backward inhibition is a residual effect caused by conflict occurring when switching away from a task is a viable theoretical proposal, and that this does indeed predict the observed empirical pattern. In short, it validates the verbally specified model of Koch et al. (2010), demonstrating that the theoretical proposals are indeed consistent with the data.

9.3 General Discussion

This chapter presents two simulations that extend the predictions of the backward inhibition model, in situations that affect task conflict in the model. Both manipulations of response congruency and intertrial intervals affect the n-2 repetition cost, when the manipulation increases conflict on the n-1 trial. In the case of simulation 10, this prediction matches the empirical results observed in the study of Gade and Koch (2005), and thus offers empirical validation of the model.

Chapter 10

Strategic Adaptation of Backward Inhibition

Thus far, it has been assumed that one general task switching model can be applied to both two and three task switching paradigms. However, Grange et al. (2013) argued that classic empirically observed two- and three-task switching effects (i.e., n-1 switch and n-2 repetition costs) rarely occur within the same experimental block. One possibility is that humans strategically adapt task switching to optimise for the task context: specifically, whether immediate task repeats occur. This chapter examines the current empirical evidence for such strategic adaptation, and moves toward a model of strategic adaptation via a parameter tuning process of the strategic adaptation system. It is demonstrated that tuning any of the three conflict-system parameters would produce such an adaptation effect, and that cumulative conflict may be an appropriate environmental cue to trigger such tuning. However, a learning algorithm which produces stable behaviour over time in contexts where tasks may either repeat or not, has not been established.

10.1 Introduction

One question so far unaddressed is the extent to which backward inhibition may be regulated in different situations. There is some evidence that backward inhibition may be enhanced in situations where it is beneficial, such as in switching paradigms where tasks never repeat, or reduced in situations where it is counterproductive, such as where tasks repeat frequently (Philipp & Koch, 2006).

Theoretically, this question also has a bearing on the locus of task inhibition. While there is robust empirical evidence for both n-1 switch costs and n-2 repetition costs in a range of behavioural paradigms, Grange and colleagues have favoured a model which does not produce these two phenomena simultaneously. Instead, in certain advantageous situations, such as where tasks continually switch and do not repeat, task-sets may be self-inhibitory, producing n-2 repetition costs instead of switch costs (Grange et al., 2013; Grange & Juvina, 2015).

In the development of the present model, we have taken as a basic empirical constraint that the model must produce both n-1 switch costs and n-2 repetition costs. It has been previously demonstrated that within certain parameter constraints, the model produces both these behavioural effects. However, the model is also capable of producing behaviour that does not meet these constraints. For example, if the weight of the conflict-task demand unit inhibitory connections is attenuated, or set to zero, the size of the n-2 repetition cost decreases, albeit with an increase to the magnitude of the n-1 switch costs (cf. figure 5.3). To the extent that we have accurately modelled the cognitive mechanisms of interest, it has been an open question how a naturalistic cognitive system, such as the human mind, sets its control parameters in such as way as to produce both empirical costs. One possibility which has been explored by the empirical literature is whether there is an adaptive element to backward inhibition – that is, whether the degree of backward inhibition applied in a given context is adapted to provide an advantage, such as more efficient cognitive processing, or leaving the system in a 'ready state' such that it produces quicker and more accurate responses in the current context. In this chapter, we will refer to such adaptation of cognitive mechanisms to a given context as ‘strategic adaptation’.

The notion that some strategic adaptation takes place has a fair degree of empirical support. Some studies have shown that n-2 repetition costs occur when n-1 repeats are possible (Arbuthnott, 2005), while others have shown that the n-2 repetition cost is significantly reduced in such circumstances (e.g., Philipp & Koch, 2006), with yet others suggesting that consistent n-1 switch and n-2 repetition costs tend not to occur in the same block of trials (see Philipp & Koch, 2006, for a review). Whether the degree of strategic adaptation observed empirically is sufficient to support the modelling assumption of Grange and Juvina (2015) — that self-inhibition of task sets is the mechanism for producing n-2 repetition costs, but itself precludes n-1 switch costs — is unclear, and the exact mechanisms and conditions for producing such adaptation, to date, remain unspecified.

In summary, The possibility of strategic adaptation of a backward inhibition mechanism to the structure of the task environment is an intriguing suggestion, but the cognitive system responsible for such strategic (re)configuration remains to be specified in theoretical terms, and empirical support remains somewhat inconclusive. The occurrence of both n-2 repetition costs and n-1 switch costs within the same cognitive mechanism is a central criterion for assessing computational models of task switching.

10.2 Empirical studies of strategic adaptation

The occurrence of n-2 repetition costs in behavioural paradigms that also allowed n-1 task repeats, was reviewed by Philipp and Koch (2006). There is evidence that the backward inhibition process is largely automatic, and occurs even when participants should know it will be disadvantageous. Thus, Mayr and Keele (2000) found task inhibition occurred even for pre-planned sequences, such as when the participant knew they would return to a task after a single intervening trial.

A number of three-task switching studies have been conducted where the presence of immediate task repeats was included (Arbuthnott & Frank, 2000; Arbuthnott & Woodward, 2002; Mayr & Keele, 2000; Mayr, 2001, 2002; Dreher & Berman, 2002). Of these studies, three found both n-1 switch costs and n-2 repetition costs (Mayr, 2001, 2002; Mayr & Keele, 2000). Mayr and Keele (2000) reported experiments both with (exp. 4) and without (exp. 1) task repeats, finding n-2 repetition costs for both. However, no between-experiment comparison on the size of the n-2 repetition cost was reported. One study found n-2 repetition costs but no significant n-1 switch costs (Arbuthnott & Frank, 2000), while another found n-2 repetition costs but did not report potential n-1 switch costs (Dreher & Berman, 2002). A further study found reliable n-1 switch costs but unreliable n-2 repetition costs (Arbuthnott & Woodward, 2002).

More systematically, Philipp and Koch (2006) conducted two experiments which compared n-2 repetition costs for blocks with repeats vs. no repeats. In a first experiment, using a between-subjects design in which the presence of immediate repetitions in the sequence of tasks was either allowed or disallowed, ANOVA results indicated a significant main effect of sequence, indicating an overall n-2 repetition cost, and importantly, a significant $sequence \times group$ interaction, indicating a significant difference in n-2 repetition costs between the groups. Post-hoc t-tests indicated n-2 repetition costs were significant for the no-repetition group but not the repetition group for RTs, while error rates showed the same trend but did not reach significance. No effect size statistics were reported.

A second experiment used a within-subjects design with alternating repetition vs. no-repetition blocks. The block number (1/2 vs. 3/4 vs. 5/6 vs. 7/8) was used as a factor, to account for gradual adaptation over the course of the experiment, to determine whether adaptation occurs rapidly for each block (i.e., switching be-

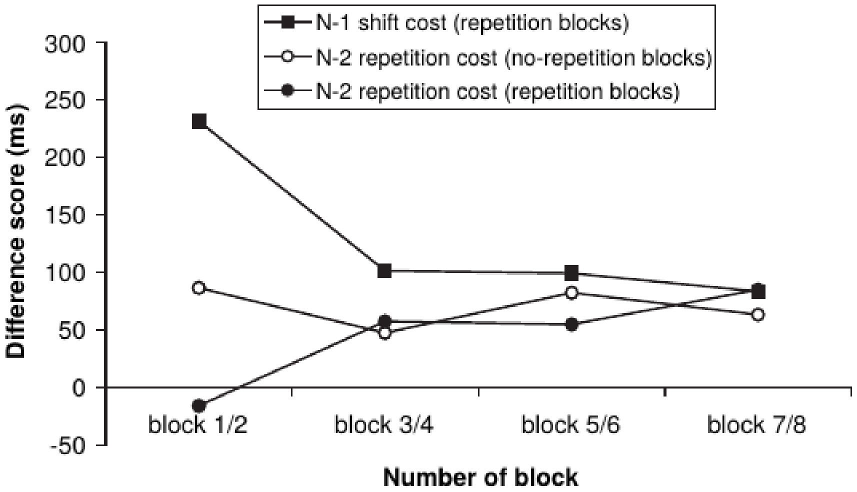


Figure 10.1: Experiment 2 from Philipp and Koch (2006): Change in n-1 switch costs and n-2 repetition costs over the course of 8 blocks, as a result of alternating between repetition vs. no-repetition blocks. Reprinted from Philipp and Koch (2006).

tween different adaptations for repeating vs. non-repeating blocks), or settles to a compromise setting for both repetition and non-repetition blocks. The empirical data from Philipp and Koch (2006) experiment 2 (figure 10.1) shows that after adaptation occurred, n-1 switch costs were significant in all blocks, although they were significantly larger in blocks 1/2 (as indicated by the significant $block \times sequence$ interaction). While n-2 repetition costs were significantly larger for blocks with no repetitions than repetitions (the $sequence \times block\ type$ (repetitions vs. no repetitions) interaction), and n-2 repetition costs increased over a number of blocks ($sequence \times block\ type \times block\ number$), separate analyses by block number showed that n-2 repetition costs were only different for repetitions/no repetitions in blocks 1/2, with no significant $block\ type \times sequence$ interaction in blocks 3/4, 5/6, or 7/8, suggesting n-2 repetition costs did occur even after adaptation had taken place, and that after adaptation, the presence of repetitions did not affect n-2 repetition costs.

Making sense of their results, Philipp and Koch (2006) suggested dual mechanisms for resolving interference in task switching: Task activation and task inhibition. Adaptation to the presence of task repeats consists of either increasing task activation (where repeats are possible) or inhibition (where repeats do not occur). This explanation is a reasonable high-level description of the proposal by Grange and Juvina (2015): task representations become more highly active after their execution where repeats are possible, while they self-inhibit in contexts where repeats do not take place. However, as argued above, this conceptual explanation does not necessarily imply that a model operating in this manner would fit the empirical data. In particular, the data from Philipp and Koch (2006, experiment 2) suggests that both n-1 switch costs and n-2 repetition costs are robust to the presence of task repetitions once adaptation has occurred, a finding which is problematic for a model which produces exclusively either n-1 switch costs or n-2 repetition costs.

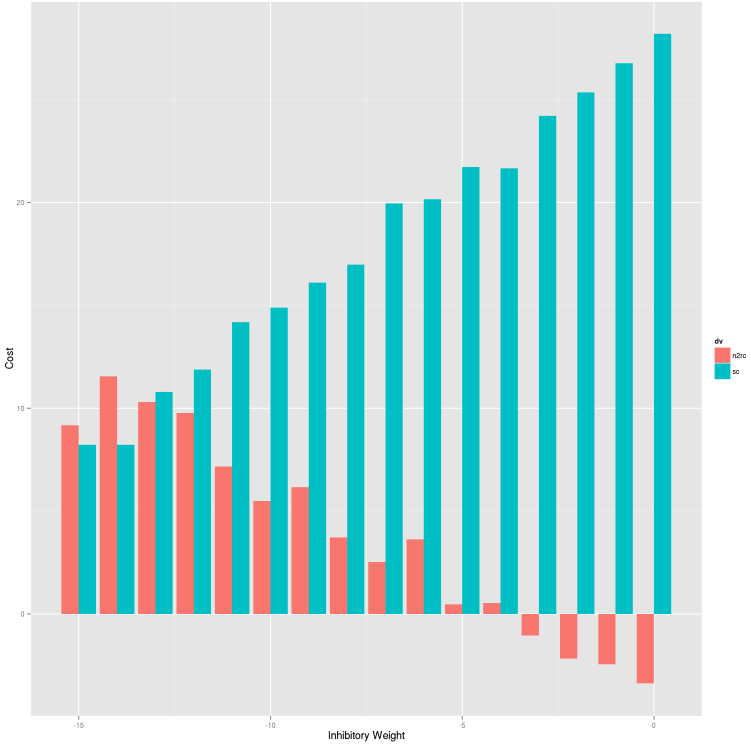
In addition, rather than proposing two mechanisms for adaptation, it would be more parsimonious to account for this effect with a single mechanism. For example, adapting the weight of conflict-task demand connections and thus the strength of inhibitory biasing, might plausibly produce both effects, given their appropriate conditions. It is notable that over the course of 8 blocks (figure 10.1), in *repetition* blocks the n-1 switch cost and n-2 repetition cost behave as, in simulation 6d, as if adaptation were occurring by increasing the weight of inhibitory conflict-task demand unit connections. In this case, n-1 switch costs become smaller while n-2 repetition costs become larger, as a result of alternating with no-repeat blocks. The current empirical evidence for strategic adaptation presents a clear picture, although it relies on results from only a single study. Replication of these empirical results is, therefore, desirable. One outstanding empirical question concerns whether the adaptation taking place is truly strategic (i.e., applies equally to all tasks) or whether it is task specific. Answering this question could be addressed with an experiment where different tasks repeat with varying rates. In such a situation, does the presence of repetitions of task A only lead to reduced n-2 repetition costs for task A only, or for all tasks?

10.3 Simulation 11: Strategic adaptation as parameter tuning

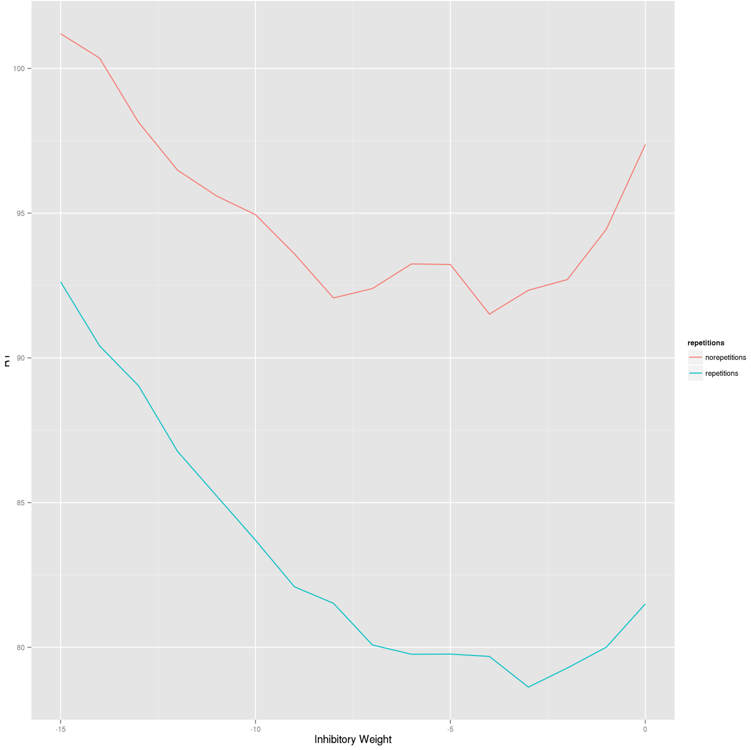
To begin the process of modelling strategic adaptation, let us first assume that such an adaptation to the presence, or absence, of repeat trials, such as described previously, may be accomplished via a single mechanism. We have also previously noted that by modulating the *weight* parameter in simulation 6d (chapter 6), which controls the level of backward inhibition applied to task demand units, the model may produce smaller (or negative) n-2 repetition costs and greater switch costs, or vice-versa. Indeed, as noted above this pattern is similar to the adaptation seen in experiment 2 of Philipp and Koch (2006).

The purpose of this set of simulations is to explore which, if any, single model parameters might produce the empirically observed pattern, if they were the subject of a strategic adaptation process. To achieve this, we re-ran the analysis from simulation 6d (chapter 6), which used symmetric switching, and varied three parameters affecting the backward inhibition system, *bias*, *gain*, and *weight*. Parameter settings were varied between the following limits: *gain*: 0 to 98; *bias*: -42 to 0; *weight*: -21 to 0.

First, in simulation 11a *bias* and *gain* parameters were kept at default values (*gain* = 77, *bias* = -9) to examine effect that varying task demand weight has on switch costs and n-2 repetition costs (figure 10.2a). When *weight* is varied over a small range (-15 to 0) the model reproduces the empirical pattern. That is, at small weights (*weight* close to 0) the model produces large switch costs but negligible n-2 repetition costs. When the weight becomes stronger (*weight* approaches -15) switch costs are reduced and n-2 repetition costs increase. Thus, if the weights of inhibitory connections between conflict and task demand units were attenuated in response to the presence of repeat trials, it is plausible that the model would produce results similar to those reproduced in figure 10.1.



(a) Switch cost and n-2 repetition cost (model cycles) when inhibitory *weight* is varied.



(b) Aggregate RTs for repetition (1SW, 0SW) and no-repetition (2SW, ALT) conditions when inhibitory *weight* is varied.

Figure 10.2: Simulation 11a, varying *weight* on horizontal axis. (*gain* = 77; *bias* = -9; blocksize = 1500.)

What would prompt such an adaptation? One possibility is that the human cognitive system tries to optimise performance, by decreasing RT or error rate. This may occur if a different weight were optimal for repeat and no-repeat trials. In our simulations, repeat trials are considered to be those with an immediate repetition in the trial and its two preceding trials — the 0SW and 1SW conditions — while no-repeat trials include 2SW and ALT sequences. RTs are plotted aggregated for repetition and no-repetition trials in figure 10.2b, and individually per condition in figure 10.3. Note that the optimal (i.e., lowest RT) *weight* setting for repetition and non-repetition conditions is different, with lowest RT obtained for repetitions around -3.5, and the lowest RT for non-repetitions at -9.5. The reason for this is illustrated by figure 10.3, which shows that 0SW trials are optimised with a *weight* of zero, while 1SW and 2SW trials both show optimal performance between -9 and -15. ALT trials are optimised around -9. Thus, including 0SW sequences in a switching block would have the effect of altering the *weight* setting required to give minimised RT performance, with the degree of shift in ideal *weight* dependent on the proportion of 0SW trials. Note that in human participants, it is also plausible that 0SW trials may have a disproportionate effect on adaptation, as they may be more salient than the other conditions (which all involve a switch of task).

The previous analysis demonstrates that adjustment of the *weight* parameter represents a viable adaptation to deliver optimised (i.e., minimised RT) performance. Next, a similar analysis was carried out for systematically varying the *bias* (simulation 11b) and *gain* (simulation 11c) parameters, to establish what effect, if any, these have on performance.

The graphs shown in figure 10.4 (simulation 11b) illustrate that adjusting the *gain* parameter over a portion of its range has a similar trade-off of switch costs and n-2 repetition costs. In figure 10.4a, a *gain* of 25 produces large switch costs and negligible (slightly negative) n-2 repetition costs, while a *gain* of 75 (the model

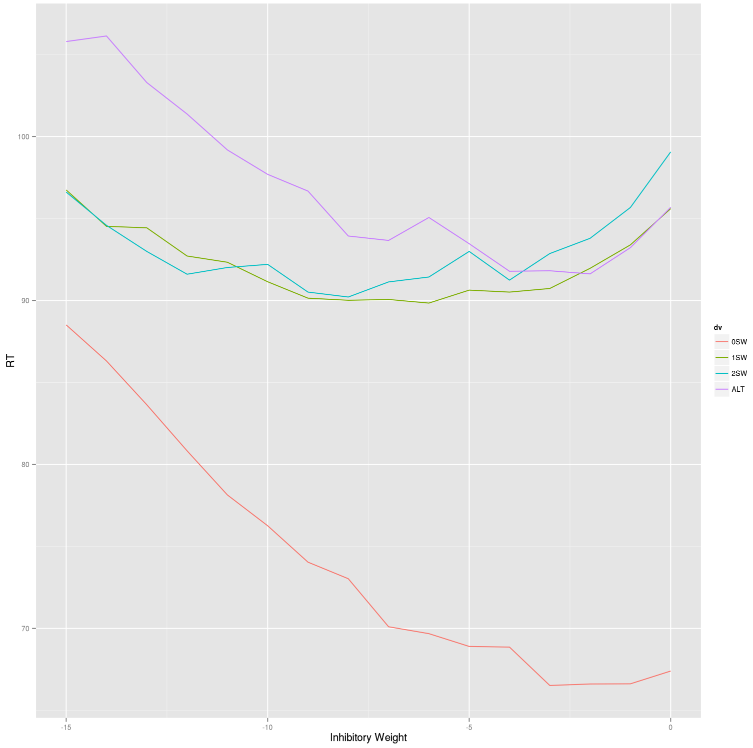
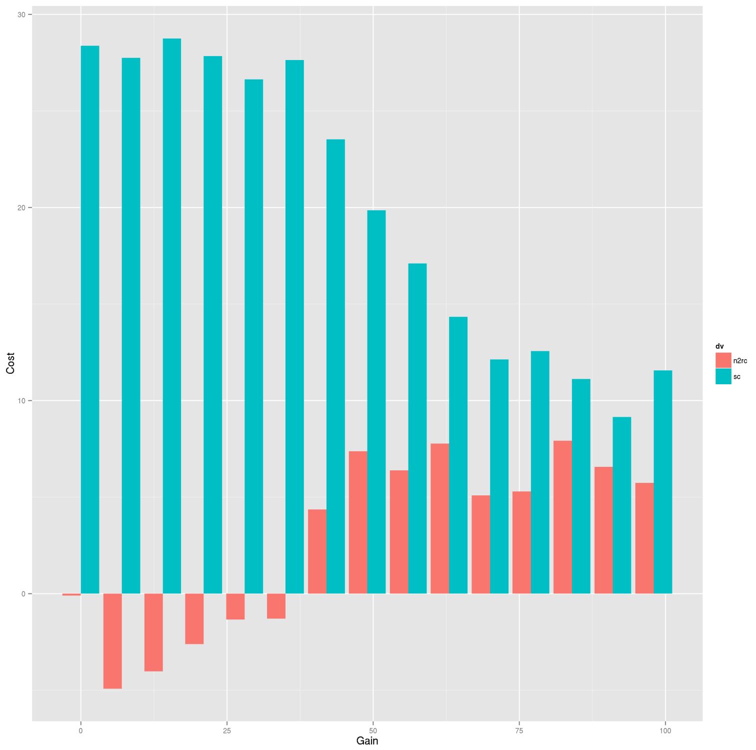


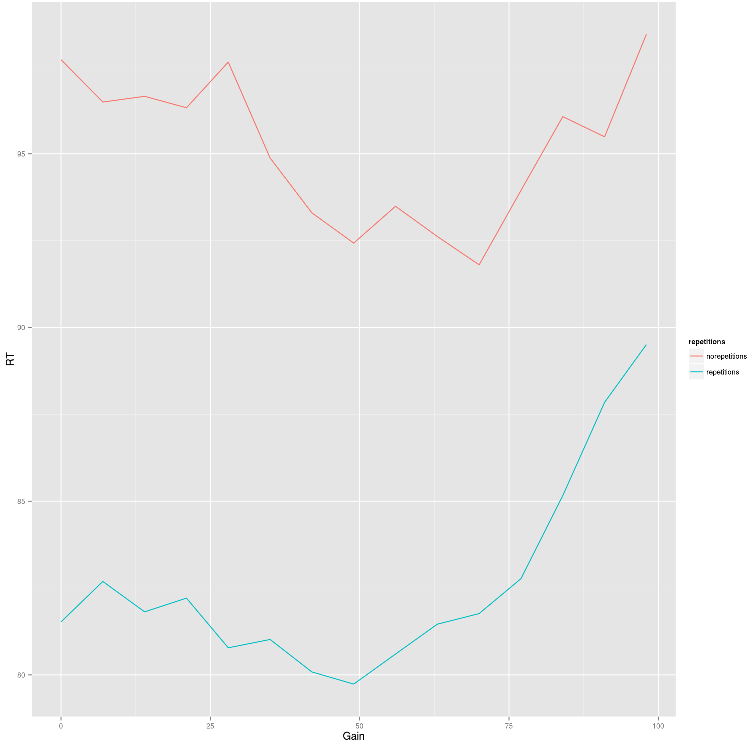
Figure 10.3: Simulation 11a, varying *weight*. RTs individually per condition. Parameters as in figure 10.2.

default) produces smaller switch costs, and substantial n-2 repetition costs (but still smaller than the switch costs). Examining figure 10.4b suggests that a low *gain* value of approximately 45 is optimal for repetition sequences, while a *gain* of between 50 - 60 is optimal for no-repetition sequences. While, the irregular non-monotonicity of the no-repetition RT would not necessarily provide an ideal signal for adapting parameter settings, the range of optimal settings (i.e., 45 to 60) still matches the empirical data, when assessed graphically. Thus, manipulation of the *gain* parameter is also a viable means of adapting task inhibition to the sequential structure of the task block.

Finally, figures 10.5a and 10.5b illustrate the varying of the *bias* parameter on the horizontal axis (simulation 11c). Similarly to the manipulation of *gain*, adjusting *bias* over a portion of its range produces the pattern of costs observed in the empirical data, that of large switch costs and negligible or negative n-2 repetition costs (e.g., *bias* = -40 to -20) to smaller switch costs and significant n-2 repetition



(a) Switch cost and n-2 repetition cost (model cycles) when *gain* is varied.



(b) Aggregate RTs for repetition (1SW, 0SW) and no-repetition (2SW, ALT) conditions when *gain* is varied.

Figure 10.4: Simulation 11b, varying *gain* on horizontal axis. (*bias* = -9; *weight* = -10.5; blocksize = 500.)

costs (*bias* = -5). Examining figure 10.5b suggests that very large negative *bias* of around -40 to -28 (functionally, disabling the task inhibition system) produces optimal repetition RTs, while a *bias* of -11 produces optimal non-repetition RTs. The respective costs produced for these parameter settings are also a qualitative match to the empirical pattern in figure 10.1.

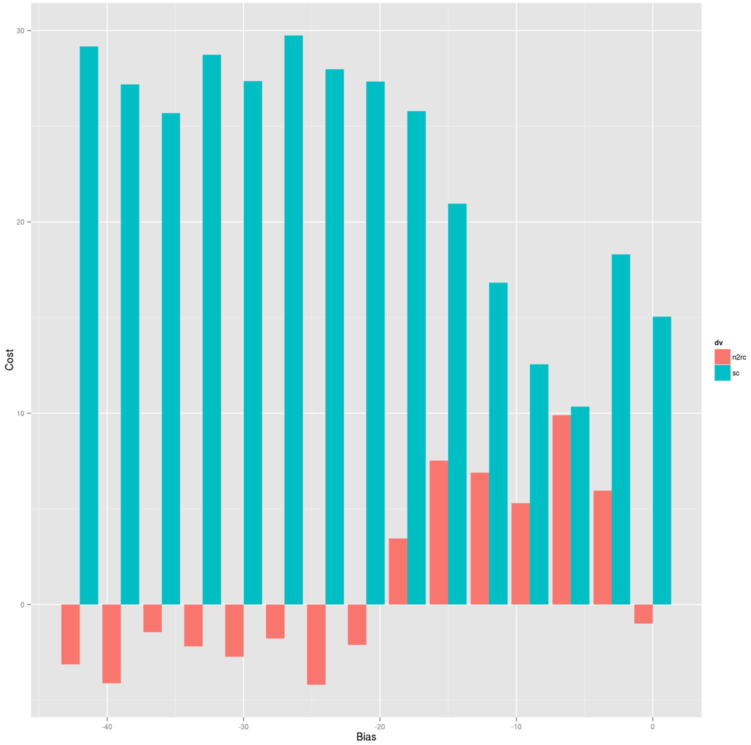
In sum, manipulating *bias*, *gain* and *weight* all offer a viable means of adapting or configuring the task inhibition system appropriate to sequences containing either repeating, or non-repeating trials. The RTs obtained for each type of sequence suggest that if the cognitive system adapted performance by adjusting any of these three parameters to minimise RT, the effects on n-1 switch and n-2 repetition costs would be similar to the adaptation observed in the empirical data.

For modelling strategic adaptation of task inhibition, these results demonstrate that a mechanism for implementing adaptation, of the type observed in the empirical data, is not specific to a particular arbitrarily selected parameter within the model, but relates to the degree of biasing exerted by the backward inhibition system in general. Manipulating any of the three parameters, broadly speaking, has a functionally equivalent effect of increasing or decreasing the modulatory influence of conflict units on task demand units, presenting a variety of ways of achieving an appropriate setting for achieving optimal (i.e., fastest) performance.

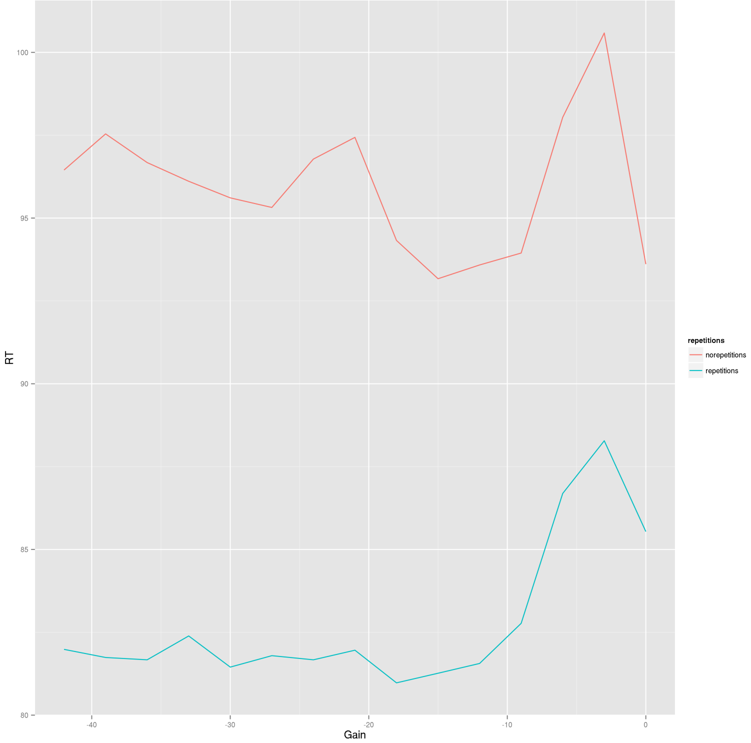
In order to be a plausible mechanism for such strategic adaptation, however, it remains to be demonstrated that this adaptation can be learned by the model as the result of exposure to different block types (i.e., the presence or absence of repeats in the task sequence.)

10.4 Exploring potential cues for adaptation

Simulation 11 has demonstrated that by optimising performance to RT, the model can produce the empirically observed adaptations. However, rather than an ex-



(a) Switch cost and n-2 repetition cost (model cycles) when *bias* is varied.



(b) Aggregate RTs for repetition (1SW, 0SW) and no-repetition (2SW, ALT) conditions when inhibitory *biast* is varied.

Figure 10.5: Simulation 11c, varying *bias* on horizontal axis. (*gain* = 77; *weight* = -10.5; blocksize = 500.)

ternally observed metric of performance, such as reaction time or error rate, such a mechanism would be cognitively more plausible if it used a signal internal to the model on which to base adaptation. Some salient possibilities relate to conflict — for example, the final level of conflict at the end of each trial, or the cumulative degree of conflict experienced over the duration of the trial. The next step in developing a strategic adaptation mechanism, therefore, is to systematically examine the relative levels of conflict typical for different types of sequence, within the model, and implement a model which adjusts weights (or some other parameter) on the basis of this conflict. Assuming that conflict is monotonic with RT, such an adaptation would involve making weights larger in response to more conflict, and attenuating weights in response to less conflict.

10.4.1 Simulation 12a: Final trial conflict as a cue for strategic adaptation

The aim of simulation 12a is to investigate a link between the within-model variable of conflict (operationalised as final trial conflict, as a proxy for cumulative conflict which is harder to measure) and reaction time. If the values correlate, as seems plausible, then conflict might also plausibly serve as a a signal for training a mechanism for strategic adaptation. Such a model would adjust any one of three parameters — *bias*, *gain*, or the *weight* of conflict-task demand connections — such that following high conflict trials (i.e., task switches) the biasing influence of task inhibition is increased, and following low-conflict trials, the biasing influence is decreased.

Are there systematic differences between sequences in final trial 3 conflict, as there are in RT? Figure 10.6 suggests that the type of sequence is indeed the source of significant differences in total final trial conflict. However, the magnitude of the differences is small, and it is doubtful whether such a small difference could produce significant adaptation in the model. Additionally, the ordering of sequences

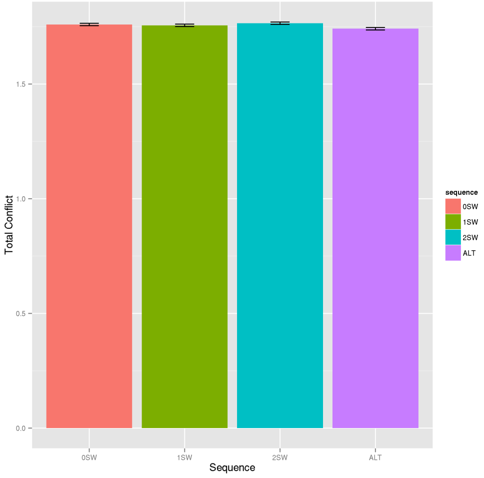


Figure 10.6: Simulation 12a: Total conflict by sequence. Error bars represent 95% confidence intervals

according to final conflict does not correspond with ordering according to RT. In particular, 0SW sequences generate the greatest amount of final trial conflict, even though the final trial is a repeat trial. This may be due to conflict generated between irrelevant response units: e.g., for the 0SW sequence BAA, conflict generated between BC tasks is irrelevant to the performance of task A, and in fact a large amount of irrelevant task conflict would facilitate task A performance.

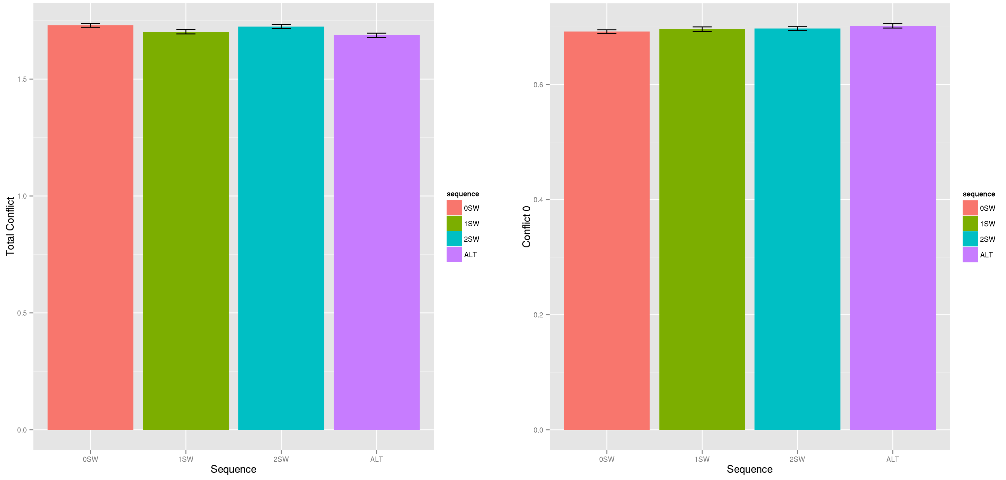
10.4.2 Simulation 12b: Task specific conflict as a cue for strategic adaptation

Given the conclusions from simulation 12a, a more accurate measure of conflict may be relevant task conflict, i.e., on the AB conflict detecting unit only. Simulation 12b looks at conflict affecting specific conflict units, rather than total conflict. Here, therefore, conflict is considered as specific to a particular task, rather than a global variable.

In this simulation, all switches are symmetric and unidirectional — i.e., all 0SW sequences are BAA, all 1SW are BBA, all 2SW are CBA and all ALT are ABA. Therefore, conflict between tasks A and B (i.e., the AB conflict unit, unit 0 in the model) is the crucial unit for managing relevant-task conflict. The final trial conflict for trial 3, summed for all units, is plotted per sequence in figure 10.7a. Additionally, conflict for specific conflict units is plotted in figures 10.7b (unit AB), 10.7c (unit BC), and 10.7d (unit CA). Although differences in total conflict are the reverse of reaction times, with greater conflict for repeating sequences, AB conflict follows the reverse trend, with higher conflict on non-repeating sequences. Why, then, does total conflict not follow the same trend? The answer follows from the final activation of units BC and CA, where differences in activation per sequence are greater than the differences observed in the activation of the AB unit (which should be sensitive to the increased conflict associated with the ABA alternation in the ALT condition). This finding suggests that final trial conflict, calculated on the 3rd trial of each triplet, is not a good analog for a measure of efficiency such as reaction time.

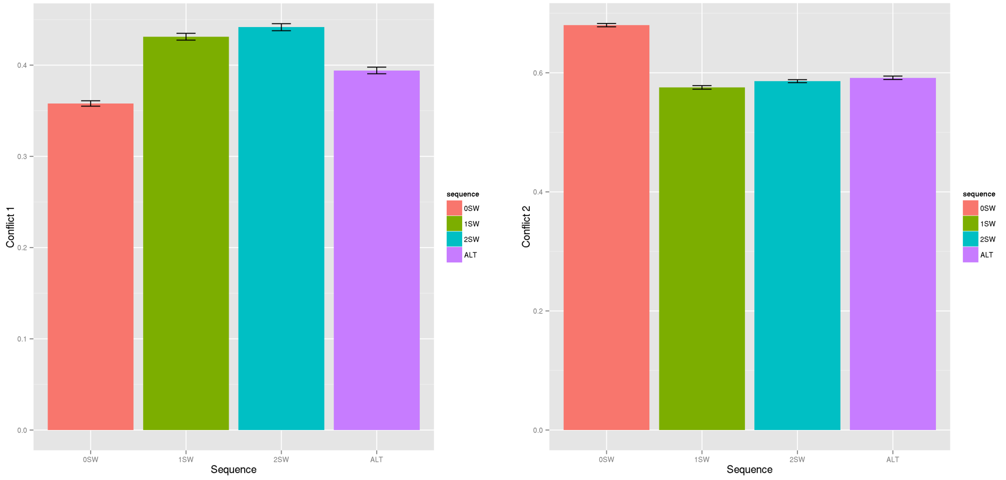
One further possibility is worthy of consideration. Rather than restricting the adaptation to final trial conflict on the third trial of each block, which would be excessively artificial, let us consider conflict calculated on all trials. Final trial conflict for all trials is therefore plotted in figure 10.8. Although this means of calculating conflict produces differences per sequence, for specific conflict units, overall conflict summed for all three units produces similar results when summed over 3 trials, compared to the final trial: the average conflict for repeating sequences is slightly higher than for non-repeating sequences. However, this represents an extremely small difference.

In conclusion, final-trial conflict does not provide a compelling analog to the behavioural measure of reaction time. Thus, final-trial conflict is unlikely to be a source for an adaptation signal.



(a) Total conflict

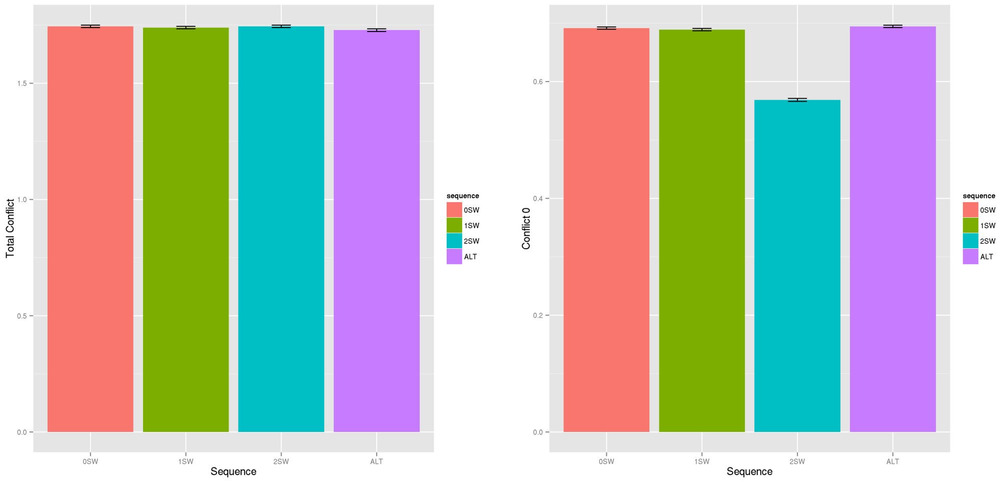
(b) AB conflict



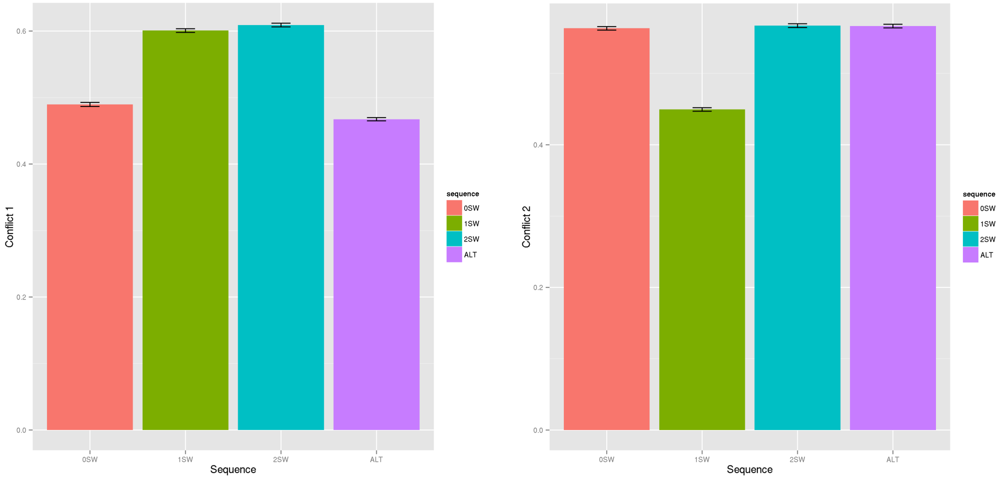
(c) BC conflict

(d) CA conflict

Figure 10.7: Simulation 12b: Conflict by sequence type. Conflict is generally higher for non-repeating sequences than repeating sequences, except 0SW conflict is high in 10.7d.

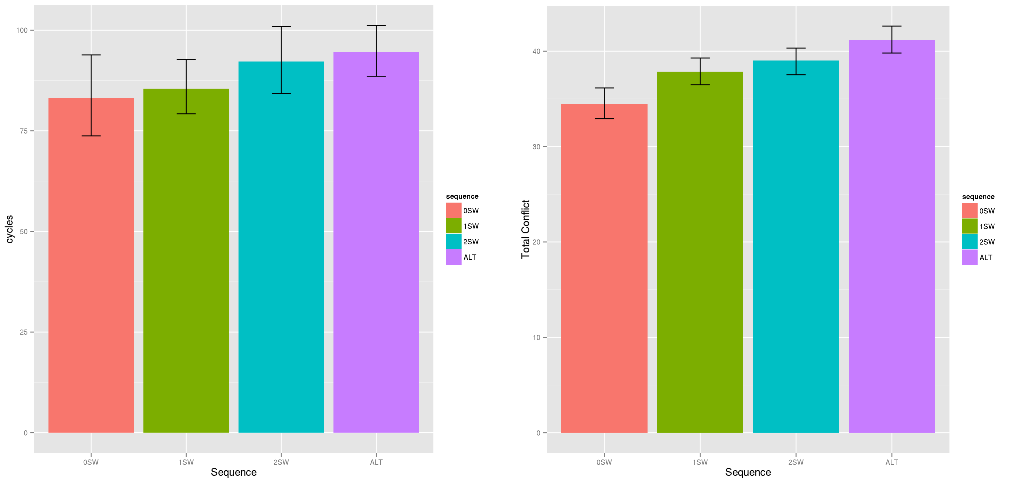


(a) Total conflict (b) AB conflict



(c) BC conflict (d) CA conflict

Figure 10.8: Simulation 12b: Conflict by sequence. Repeats figure 10.7 but calculates total conflict from all trials, rather than just at end of 3rd trial.



(a) (b)

Figure 10.9: Simulation 12c: Total conflict and RT by sequence

10.4.3 Simulation 12c: Cumulative conflict as a cue for strategic adaptation

Simulation 12c, like simulation 12a, aimed to find a link between a within-model variable and reaction time. Simulation 12a did not find a persuasive link between final-trial conflict and reaction time. One explanation for the absence of such a link is provided by looking at the difference between trials at different ends of the RT distribution: long RT trials compared with short RT trials. On long RT trials, final-trial conflict was low because the model had longer to resolve the conflict. On short RT trials, final-trial conflict was low because a response could be generated without a large degree of between-task interference.

Thus, instead of looking at final trial conflict, simulation 12c instead examined cumulative conflict. This was calculated as the cumulative sum of conflict on all model cycles. Thus, if activation of conflict units were static, more cumulative conflict would be observed on long RTs compared with short RTs. Total cumulative conflict (summed for all conflict units) per sequence type is plotted in figure 10.9b. This may be compared with reaction times (figure 10.9a). Visual comparison of the two suggests that total conflict may indeed be a suitable analog for reaction time.

It should be emphasised at this point that there is no empirical evidence that suggests cumulative conflict (or final-trial conflict, for that matter) is tracked or serves as the basis for any adaptation by the human cognitive system. In that respect, these simulations have been somewhat exploratory. The aim of modelling strategic adaptation, however, is not to provide a definitive cognitive theory of the behavioural phenomenon, but to demonstrate that such an adaptation could plausibly take place on the basis of information contained within the model. Although, at this stage, we have not adapted the model to include such features, the necessary computational units for tracking cumulative conflict, such as accumulator units, are not new to this type of model, and have often been used for thresholds (i.e., a response is deemed to have been made when a certain amount of activation, or evidence, has been accumulated; see, e.g., Ratcliff & Rouder, 1998).

10.5 Towards a model of strategic adaptation

In the following simulations, the model was run on large blocks of trials, where the sequence of tasks was generated randomly but with the constraint that immediate repeats were either allowed, or not allowed, depending on condition. If the assumptions that strategic adaptation occurs on the basis of a conflict signal available to the model, and that this adaption acts by adjusting the *weight* (or another parameter of the conflict monitoring system) are correct, exposing the model to large blocks of trials either containing, or not containing, immediate task repetitions, should cause different adaptations. Specifically, following repeat blocks, we should expect the *weight* parameter to settle at approximately -3. Following no-repeat blocks, it should settle between -4 and -7. This is because these values represent local RT minima when the *weight* parameter is varied (cf. figure 10.2a).

The following simulations detail different versions of the learning algorithm and different parameter settings. All parameters were hand-set. While this approach to parameter setting does not rigorously demonstrate that the behaviour of the model is universal, the aim at this stage was to demonstrate that with certain parameters, it was at least possible for the model to achieve different adaptations in the *weight* parameter based on cumulative conflict.

10.5.1 Simulation 13a

This simulation introduced a strategic adaptation algorithm into the model. After each trial, this algorithm adjusted the *weight* parameter of the model in response to changes in the cumulative conflict, compared to the trial before. On the first trial, the model made an adjustment to the *weight* parameter in a random direction. On subsequent trials, if the level of conflict was greater than on the previous trial, the model adjusted the *weight* parameter in the opposite direction to the previous adjustment. If the conflict was less than on the previous trial, it adjusted the *weight* parameter in the same direction as the previous adjustment.

Additionally, the algorithm contained a momentum term aiming to minimise the adaptation to short term trial-to-trial differences in conflict, and instead keep the model adapting to longer term trends.

$$\Delta weight_n = \begin{cases} \rho \Delta weight_{n-1} + \lambda (C_{n-1} - C_n) & \text{if } \Delta weight_{n-1} \geq 0 \\ \rho \Delta weight_{n-1} + \lambda (C_n - C_{n-1}) & \text{if } \Delta weight_{n-1} < 0 \end{cases} \tag{10.1}$$

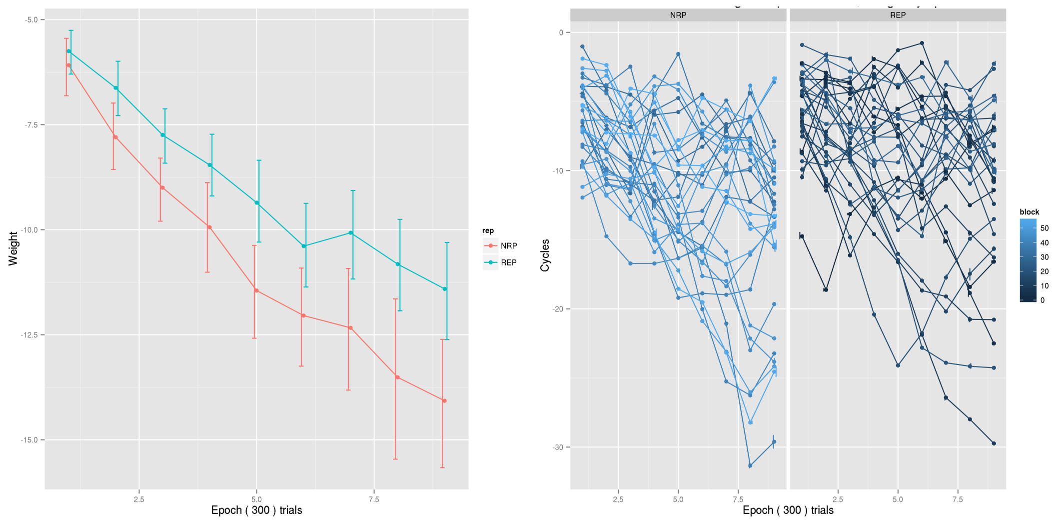
C_n is the cumulative conflict on this trial, C_{n-1} is the cumulative conflict on the previous trial, and λ is a learning rate parameter (set to 0.01). The starting weight (i.e., $weight_0$) was -3.0.

Two versions of the simulation were run for 300 trials (figure 10.10) and 800 trials (figure 10.11). Of these graphs, 10.10a plots the *weight* parameter per epoch (in this case, averaged over blocks of 300 trials). Error bars represent 95% confidence intervals. In most cases, the model initially sharply adjusts weights downwards from their starting value of -3.0. This is, as previously, because the algorithm aims to minimise differences in conflict between successive trials, and weights closer to zero produce bigger between-trial differences in conflict (although less absolute levels of conflict overall). In this instance, the model adapted differently to repeating blocks than non-repeating blocks, although the final weights were significantly different to those expected based on the minima in the RT curve (figure 10.2b), with final weights approximately -11 and -13 for repeating and non-repeating blocks respectively. However, in a second run with 800 trials per block (figure 10.10b), the model did not adapt differently to the two types of block, suggesting that this result is less than robust.

Figure 10.10b plots the changing weight parameter for individual models, allowing a visual representation of whether the previously described averages are being driven by outlier models. In this case, it is visually evident that despite differences between average weights over time, there is significant overlap for individual models between the two sequences. Figure 10.10c plots RTs and costs per epoch. If the model were adapting to repeating and non-repeating blocks as expected, we would expect switch costs to increase and n-2 repetition costs to decrease for repeating blocks, and vice-versa for non-repeating blocks. While there is some evidence of this pattern, it is not robust over time, and is not evident when the simulation is run with more models (figure 10.11). Finally, figure 10.10d plots conflict over time, illustrating that the model is indeed reducing conflict for each sequence. However, the reduction in conflict does not appear to be producing the expected adaptation.

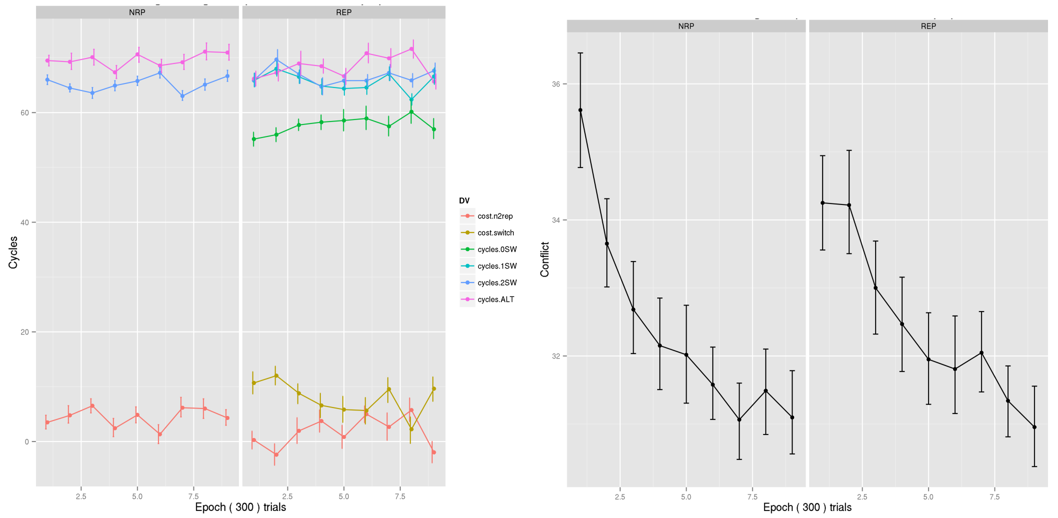
10.5.2 Simulation 13b

In this simulation, the learning algorithm was further adapted, in an aim to improve the model's adaptation to longer-term trends in conflict, rather than specific trial-to-trial changes, which tend to produce more negative weights over time. Rather than comparing current-trial conflict with previous trial conflict, therefore,



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.



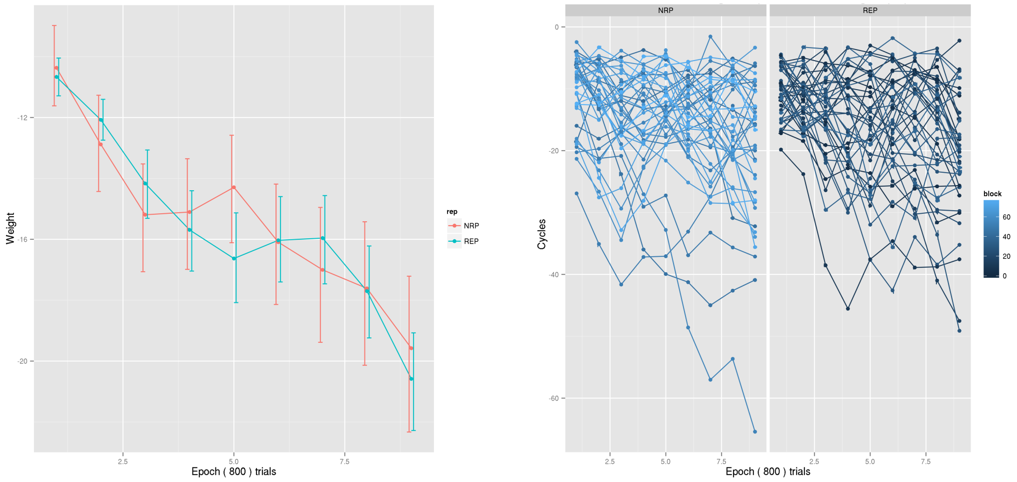
(c) RTs and costs per epoch.

(d) Conflict over time.

Figure 10.10: Simulation 13a: Adaptation by reducing cumulative conflict. ($\lambda = 0.005$, $\rho = 0.8$, 10 × 300 trial blocks, starting Conflict-TD weight = -3.0.)

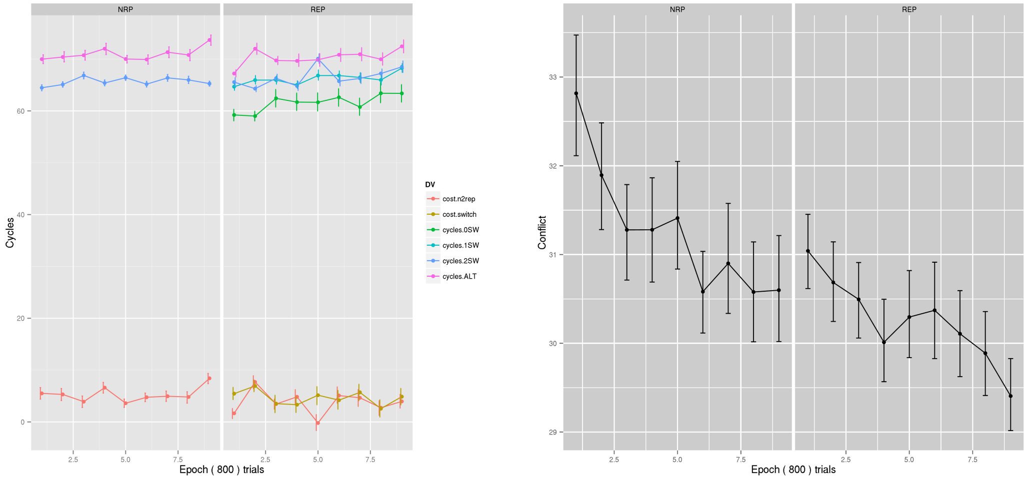
the algorithm compares with a weighted average of a number of previous trials conflict:

$$\Delta weight_n = \begin{cases} \rho\Delta weight_{n-1} + \lambda(Ch_{n-1} - C_n) & \text{if } \Delta weight_{n-1} \geq 0 \\ \rho\Delta weight_{n-1} + \lambda(C_n - Ch_{n-1}) & \text{if } \Delta weight_{n-1} < 0 \end{cases} \tag{10.2}$$



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.



(c) RTs and costs per epoch.

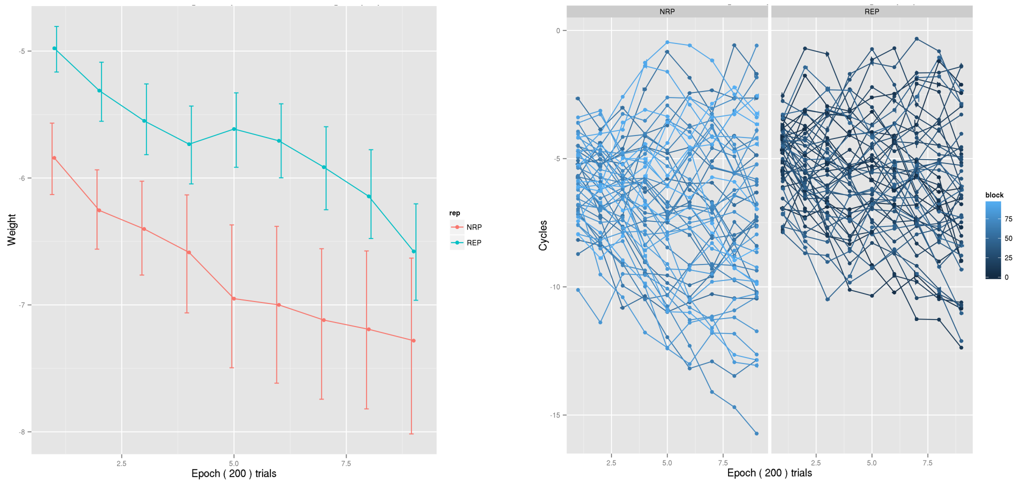
(d) Conflict over time.

Figure 10.11: Simulation 13a: Adaptation by reducing cumulative conflict. ($\lambda = 0.0025$, $\rho = 0.95$, 10×800 trial blocks.)

C_n is the cumulative conflict on this trial, Ch_{n-1} is the historical conflict, a value with is iteratively updated slowly for each trial, according to the equation:

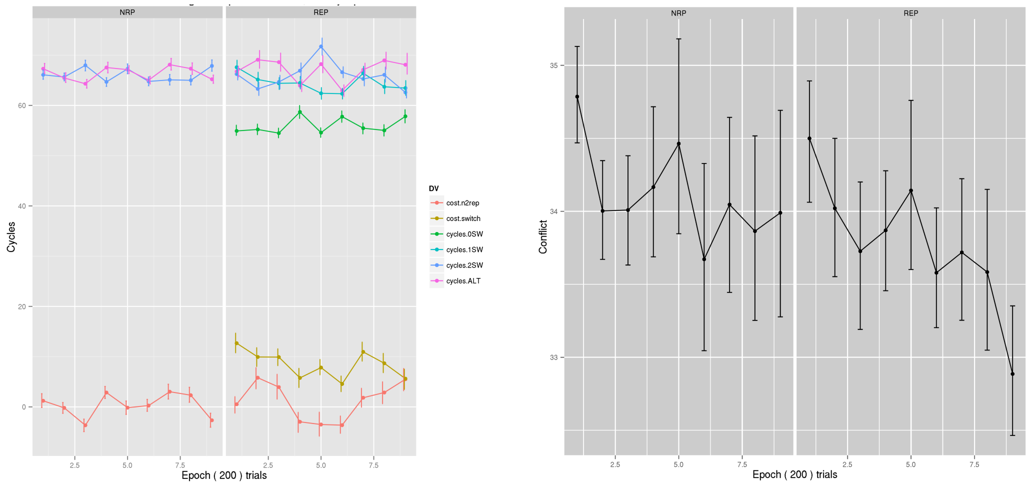
$$Ch_n = \frac{\omega(Ch_{n-1} + C_n)}{\omega + 1} \tag{10.3}$$

These simulations did suggest separate adaptation to repeating and non-repeating blocks (see figs. 10.12a, 10.13a, 10.14a), which did produce differences in switch and n-2 repetition costs in some simulations (figs. 10.12c, 10.13c, 10.14c), although in others the differences in n-2 repetition costs, the main empirically observed effect,



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.



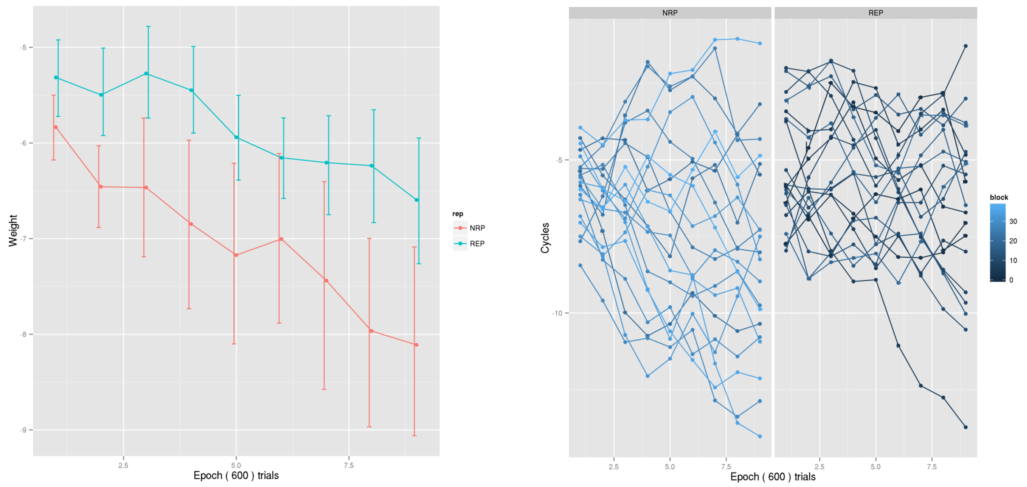
(c) RTs and costs per epoch.

(d) Conflict over time.

Figure 10.12: Simulation 13b: Adaptation by reducing conflict ($\lambda = 0.0020$, $\rho = 0.80$, $\omega = 0.80$, 50 models, 2,000 trials each, starting Conflict-TD weight = -5.0.)

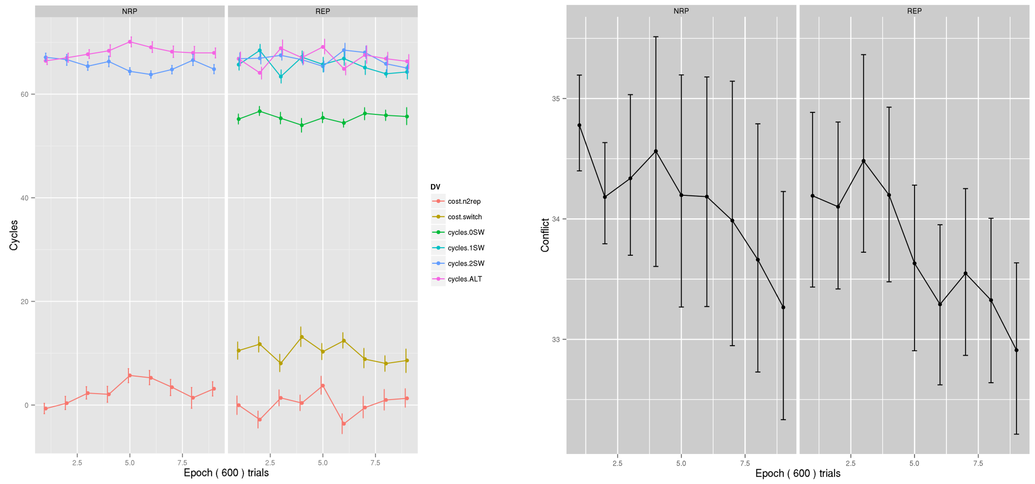
was less clear (e.g., fig. 10.12c), and again the weight parameter did not reach final stable values.

Additionally, when the simulation was run with more models (40) for significantly longer blocks of trials (10,000 trials, figure 10.15), the results suggested that over this duration, weights did not stabilise, and instead results from different models adapting to either repeating or non-repeating sequences diffused sufficiently to eliminate average differences between the two groups of models (figs. 10.15a, 10.15b).



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.

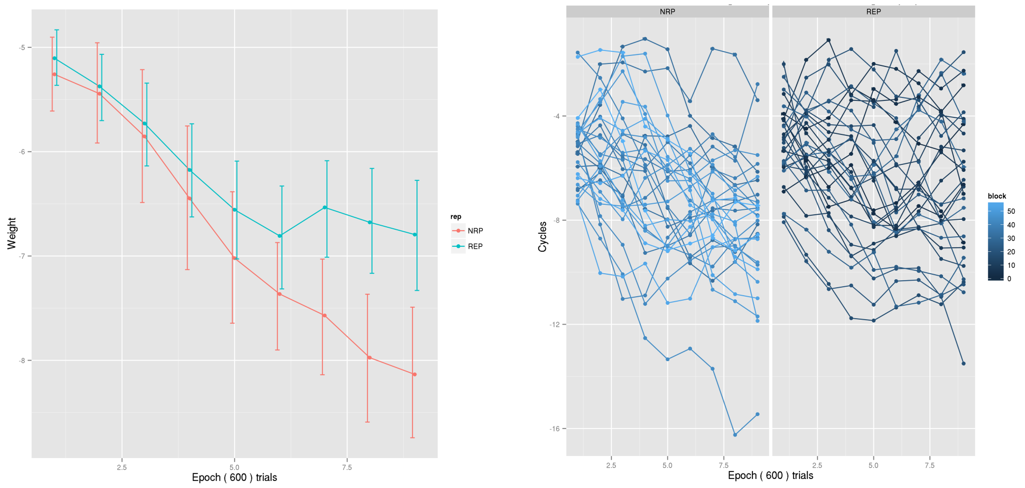


(c) RTs and costs per epoch.

(d) Conflict over time.

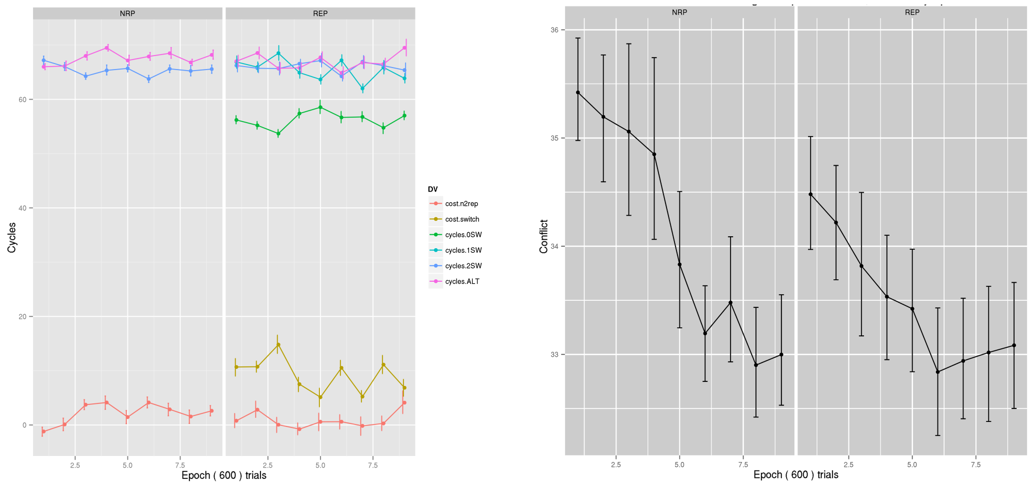
Figure 10.13: Simulation 13b: Adaptation by reducing conflict ($\lambda = 0.0005$, $\rho = 0.95$, $\omega = 0.80$, 20 models, 6,000 trials each, starting Conflict-TD weight = -5.0.)

In conclusion, simulation 13b suggests that the model with the strategic adaptation mechanism does adapt differently to repeating, than non-repeating sequences. While these differences could reach statistical significance (as evidenced by the non-overlapping 95% confidence intervals) this adaptation was not robust, and for a population of models, there was substantial overlap between the adaptations of the repeating and non-repeating groups.



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.



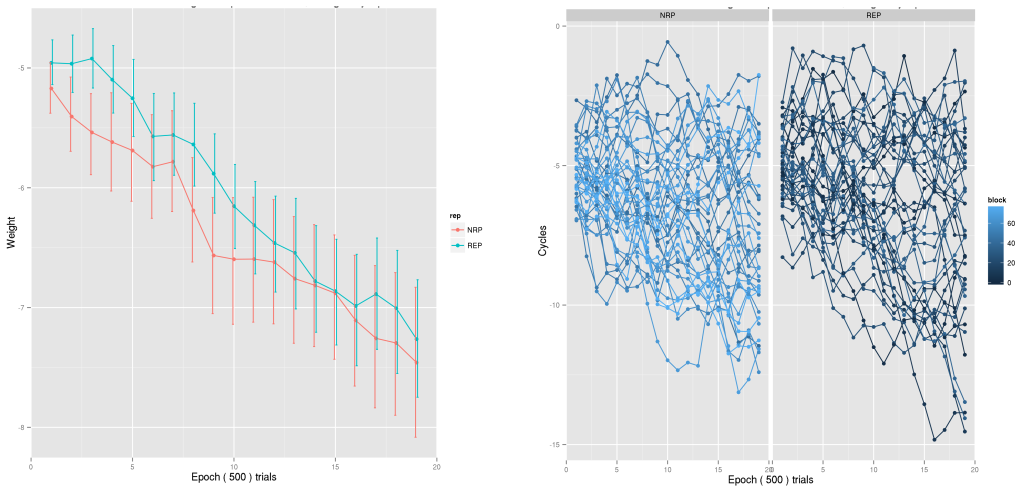
(c) RTs and costs per epoch.

(d) Conflict over time.

Figure 10.14: Simulation 13b: Adaptation by reducing conflict ($\lambda = 0.0005$, $\rho = 0.95$, $\omega = 0.80$, 30 models, 6,000 trials each, starting Conflict-TD weight = -5.0.)

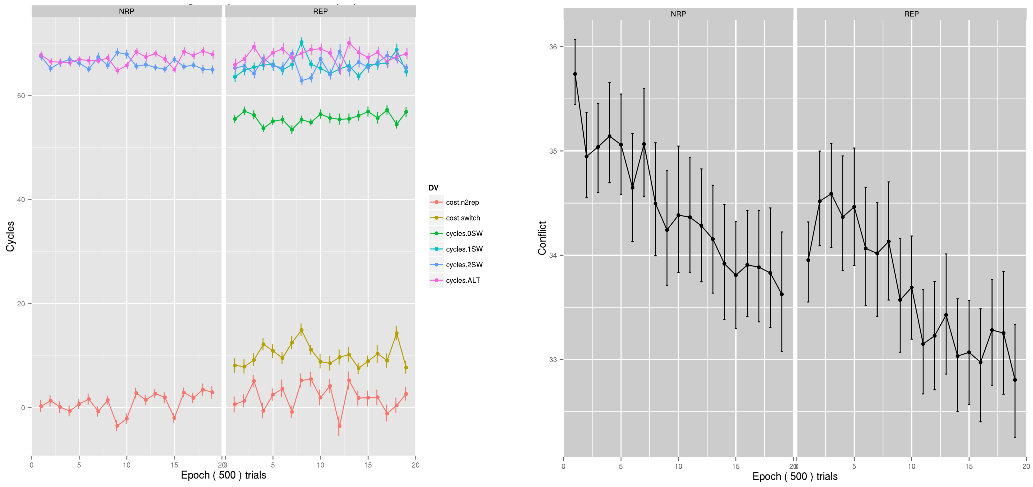
10.6 Discussion

The strategic adaptation simulations presented in this chapter produced mixed results. While certain simulations did produce separate adaptations for repeating and non-repeating sequences, by making small adjustments to the *weight* parameter, this adaptation did not stabilise over time, but instead continued to grow in magnitude. In a human cognitive system, it is highly unlikely that such an unstable adaptation would occur, and indeed in the only systematic empirical study, the effects did appear to stabilise over the course of multiple blocks (figure 10.1). While



(a) *Weight* as a function of epoch (averaged over models).

(b) *Weight* as a function of epoch for individual models.



(c) RTs and costs per epoch.

(d) Conflict over time.

Figure 10.15: Simulation 13b: Adaptation by reducing conflict ($\lambda = 0.00035$, $\rho = 0.97$, $\omega = 0.90$, 40 models, 10,000 trials each, starting Conflict-TD weight = -5.0.) Note that with very long blocks, models have diffused too much and results start to overlap, e.g., after 5,000 trials.

the model operated by minimising differences in cumulative conflict, and it was anticipated that this would lead to the *weight* parameter for each model settling on a value producing the lowest conflict for the type of sequence each model was performing, reflected by local minima in RTs, in practice this did not occur. Additionally, strategic adaption did not prove to be robust. When the model was run for large numbers of models or over long blocks of trials, the adaptations of individual models diffused, such that the adaptations of the group trained on repeating sequences overlapped with those of the non-repeating group. This diffusion increased over time, leading to a reduction in differences in behavioural measures between the two groups.

One possible cause of this failure to capture strategic adaptation is that the differences in cumulative conflict between repeating and non-repeating sequences did not provide a strong enough signal for the models to make stable adaptations and settle on optimal settings. Instead, adaptations drifted akin to a random-walk. Another possibility is limitations in the specific strategic adaptation algorithm used. This algorithm attempted to minimise differences in conflict between trials. One way of achieving this was to make the *weight* parameter very negative, which does minimise differences in conflict between different types of trial, even though the absolute level of conflict is higher. Despite a number of features of the algorithm (i.e., addition of a momentum and an averaged historical conflict term) intended to allow the mechanism to adapt to longer-term trends in the changes of conflict, it may still have been too sensitive to trial-by-trial changes in conflict such as those produced from moving between a 0SW trial and a 2SW trial. This may have been responsible for the model's lack of stability. One approach to a more successful strategic adaptation mechanism would be to devise some algorithm capable of minimising absolute conflict over a block of trials, rather than changes in conflict.

Nevertheless, in some simulations the model did produce adaptations sufficient to produce differences in n-2 repetition cost, the key empirical effect, between repeating and non-repeating sequences. While these effects were not fully robust, either over time or across different models, it should be noted that the empirically observed strategic adaptation effect may be similarly non-robust. Of multiple studies, only a certain proportion have reported differences in n-2 repetition cost between repeating and non-repeating sequences. Therefore, alongside future modelling work, additional empirical studies of strategic adaptation should be conducted.

Chapter 11

General Summary and Conclusions

11.1 Contributions

The research reported as the subject of this thesis has contributed to the task switching literature in a number of respects. These are specifically as follows:

1. The task switching model of Gilbert and Shallice (2002) was successfully replicated, by reconstructing the model from the published details and without re-use of their original code. This type of replication confirms that the originally published data from the model is indeed due to the details of the model, and not to any errors or bugs in the model. While an attempted replication of the model of Yeung and Monsell (2003) was unsuccessful, this thesis details limitations in the published detail of the model which prevented successful replication.

2. The two-task switching model of Gilbert and Shallice (2002) was generalised to three tasks, without any additional cognitive mechanisms. This simple three-task switching model established that while n-1 switch costs remained a behavioural property of the model, n-2 repetition costs were not. This demonstrated that the existing model mechanisms — specifically between-task interference implemented as the residual activation of task de-

mand units, and lateral inhibition between cognitive representations at the task-demand and response levels — did not account for n-2 repetition costs. This finding supported the simpler simulations of Grange et al. (2013) that lateral inhibition alone was insufficient to account for the phenomenon.

3. A working computational model of backward inhibition has been presented, in the form of an additional cognitive mechanism to the previous simple three-task switching model. This consists of a layer of conflict-monitoring units, similar to those used within the conflict monitoring literature (Botvinick et al., 2001), with the exception that in the present model they bias model processing interactively, in keeping with the existing implementation of the model. These conflict units detect simultaneous activation of units at the task demand level, and have reciprocal inhibitory connections. The model implements residual activation of conflict units, consistent with the residual activation of task demand units used in the model of Gilbert and Shallice (2002) that accounts for the n-1 switch cost. These additional mechanisms account for both n-1 switch and n-2 repetition costs in the same model.

4. The scope of the empirical effect and its independence from specific model parameter settings has been explored. Specifically, the three parameters governing the conflict system (gain (γ), bias (β) and weight (ω)) were systematically searched in a three-dimensional grid, of which a large central region corresponded to the empirically observed behavioural pattern. Outside of this region, the model's failure to produce the empirical pattern can be understood in terms of theoretical or implementational failure of the model (e.g., where the gain parameter is so small that the activation of conflict units does not increase even when a large degree of simultaneous task-demand activation is occurring). Additionally, multiple simulations demonstrated that the reproduction of the empirically observed pattern is independent of spe-

cific implementational details of the model, such as the treatment of negative conflict. While negative conflict may occur within the model due to a particular activation function, the concept is not theoretically meaningful. While simulation of the empirically observed behavioural pattern requires preventing negative conflict from influencing model performance, two possible treatments of negative conflict suggest that the specific way that this is implemented in the model does not affect the model's behavioural properties at a gross, qualitative level.

5. Rather than an ad-hoc computational model, the model presented in the present thesis has been implemented as an extension of a literature of existing IAC models. In an attempt to additionally constrain modelling, principles of GRAIN models were adhered to (McClelland, 1993). Further, the non-theoretically relevant aspects of the model were left unchanged from earlier models where possible. While more sophisticated artificial neural network models are presently being developed (e.g., O'Reilly, Hazy, & Herd, 2013), the simplicity of the present model makes clear the elements of the model which are both necessary and sufficient for backward inhibition, in its current formulation.

6. The model has been extended to additional simulations of behavioural phenomena already reported in the backward inhibition literature, or novel predictions:

   Intertrial Intervals It has been reported that the n-2 repetition cost is modulated by the size of the interval preceding the n-1 task, but not the current task (n) itself (Gade & Koch, 2005). This evidence has been interpreted as suggesting that backward inhibition is related to conflict occurring on the n-1 trial, rather than the current trial. In the present model, variable intertrial intervals were simulated by modulating the level of residual task-demand and conflict unit activation. The simulated results clearly predict the empirical pattern: the n-2 repetition cost is related to the interval preceding the n-1 task, but not the n task.

Graded Response Congruency Effects While the effects of graded response congruency (i.e., fully congruent, semi-congruent, fully incongruent) on n-1 switch costs has been studied, the effects on n-2 repetition costs have not. The model was used to simulate graded response congruency conditions on the n-2, n-1, and n trials. While the results are consistent with the empirical literature (i.e., effects of trial n) they also predict an effect of n-1 congruency on the n-2 repetition cost, albeit with very small effect size. Given the predicted size of the effect, an empirical test of this prediction was not attempted as part of this thesis. The simulated results also validated concerns that response congruency is a significant confound for the behavioural three-task classification paradigm. The empirical studies conducted in the present thesis used three possible responses, allowing all trials to be fully incongruent and removing congruency as a confounding factor. However, these simulated results clearly indicate that empirical studies using classification tasks should include at least three possible responses, allowing stimuli to be fully incongruent.

Dimension Inhibition/Competitor Rule Suppression As part of the simulation of the effects of response congruency, it was additionally shown that the backward inhibition mechanisms in the model do not account for the inhibitory effect of dimension inhibition (competitor rule suppression). This suggests the involvement of an additional inhibitory mechanism. Alternatively, it leaves scope for a model of backward inhibition which parsimoniously accounts for these effects with a single mechanism.

Asymmetric task difficulties The model was used to simulate tasks of variable difficulty. This was simulated systematically using a gridsearch approach, using the two task-relevant parameters of stimulus input strength (SIS) and top-down control strength (TDCS) as well as the weight (ω) parameter affecting the strength of inhibitory biasing by conflict units. The results clearly predicted that for the n-2 repetition condition (i.e., ABA sequences) hard-easy-hard alternations had larger n-2 repetition costs than easy-hard-easy alternations. These predictions diverge from the extant theoretical literature (Arbuthnott, 2008a).

7. The predictions of model behaviour for switches between three asymmetric tasks was empirically tested. A new experimental version of the digit classification paradigm was developed, in which participants classified letters according to dimensions of colour, font, and alphabet. The results of two experiments were inconsistent with the null hypothesis, providing validation of the model.

8. Finally, the thesis addressed the question of strategic adaptation, or whether the model can adapt to respond more advantageously to different task contexts. It has been suggested that the human cognitive system may adapt situationally — whether tasks immediately repeat, or do not — applying backward inhibition in the latter case. If this were the case, the constraint that the same model must predict the empirical behavioural pattern in two- and three-task contexts, may not apply. Indeed, there is some empirical support for this type of strategic adaptation. A version of the model was developed which modulated the effectiveness of task inhibition in response to ongoing task performance. Specifically, whether measures of conflict could provide the information a model would need to drive such adaptation. Systematic differences were found in the levels of conflict between repeating and

non-repeating sequences, suggesting that conflict could, in principle, provide such a learning signal. However, the present simulations did not find a learning algorithm which could produce stable states corresponding to repeating and non-repeating contexts. This remains a project for the future.

11.2 Outstanding theoretical issues and areas for future research

Although this thesis has contributed to an understanding of backward inhibition by giving one account of the phenomenon in computational terms, a number of theoretical issues remain outstanding.

It remains unclear where backward inhibition fits in an overall architecture of human cognitive control processes. Within the executive function literature, task or set shifting remains an important high-level control function which accounts for a range of human abilities on a variety of tasks (Miyake & Friedman, 2012). However, the mechanistic processes remain yet to be specified. The present thesis has clarified the function of backward inhibition as a means of managing interference between competing task representations. Moreover, it has been demonstrated that backward inhibition is computationally ‘cheap’, in that it can be accomplished by a single layer of units reciprocally connected to task representations. This type of relatively automatic inhibition does not seem to fit the criteria for an executive function. It might, however, be regarded as one constituent process of task/set switching, were that process to be further fractionated.

A related unanswered question is whether backward inhibition represents a form of proactive (cf. Kuhns et al., 2007) or reactive (Schuch & Koch, 2004) control, given these present differing computational implications. In the present model, conflict acts as a signal for the requirement of cognitive control, and backward inhibition provides an automatic inhibitory response. As the effects of conflict persist and only decay gradually over time, backward inhibition is exerted and released accordingly, explaining the n-2 repetition cost. This explanation is not consistent with a more proactive view of backward inhibition. Additionally, it remains unclear whether backward inhibition can be exerted strategically in response to changing environmental requirements. For example, some empirical evidence suggests that backward inhibition may be somewhat adaptive, with n-2 repetition costs only occurring in task contexts where immediate repeats do not occur. However, the empirical support for this type of adaptation is, at present, mixed. In the present model, backward inhibition facilitates switches but presents a disadvantage for immediate task repeats. It remains a possibility that the level of backward inhibition could be modulated, either deliberatively or reactively/automatically by the human cognitive system. Simulations have demonstrated that in the model this could be accomplished by modulation of any one of the three key parameters controlling the conflict units (bias (β), gain (γ) or weight (ω)). Whether such a process of adaptation occurs, what level it occurs at (i.e., whether it represents a top-down 'executive' control process) and how it is triggered, remain theoretical and empirical questions of significant interest.

A further question concerns the stage of cognitive processing at which backward inhibition is triggered. Various empirical studies have provided evidence for cue-encoding, response-selection, or response execution processes. In their review, Koch et al. (2010) propose that backward inhibition is a flexible mechanism, which may be deployed at whatever level of the cognitive system conflict takes place. How this proposal might be implemented computationally, however, is unclear. One possibility is that multiple conflict-inhibition loops exist, operating in a similar manner to the conflict system operating in the present model. As has been demonstrated, such a system is computationally 'cheap', requiring only a single layer of units to monitor conflict between task-relevant representations, with reciprocal inhibitory connections. One possibility is that this is a relatively common arrangement deployed at various points in the cognitive system. An alternative possibility would consist of a single backward inhibition system with multiple conflict detectors, and multiple inhibitory connections. How such a system would be arranged and would operate has yet to be specified.

The present model does not emphasise either cue encoding processes (task-demand units merely receive ‘top down control’ activation which is assumed to be all-or-none) or response execution processes. Implementation of either of these mechanisms would require a wide range of additional modelling assumptions. Thus, simulation of the modulation of the n-2 repetition cost by either cue-encoding or response execution processes is beyond the scope of the present model. Instead, detailed simulation of these phenomena would be better suited to a backward inhibition model based within an architectural framework such as the models of Grange and colleagues (Grange et al., 2013; Grange & Juvina, 2015). Indeed, as one of the strengths of the ACT-R architecture is memory retrieval, the focus of those models on cue-based memory encoding processes makes them well suited for simulating cue-based phenomena.

Further research on this question may require further empirical and computational approaches, firstly, to empirically determine the limiting conditions for the modulation of the n-2 repetition cost due to cue-based and response-based processes, and secondly, to empirically determine whether backward inhibition triggered by cue-based conflict and response-based conflict reflects the operation of a single system or multiple systems. One simplistic approach might be to devise a paradigm in which cue-based and response-based manipulations modulate the n-2 repetition cost. If backward inhibition were a single cognitive mechanism, one might expect these manipulations to interact, while if backward inhibition consisted of multiple systems, one might expect them to be independent. A more sophisticated approach might use computational modelling to make distinct quantitative predictions based on different models representing alternative hypotheses (e.g., single conflict system, multiple conflict systems).

One area of difference between the model and the present empirical data is the simulation of error trials. In human participants, a trade-off is observed between speed and accuracy: it appears that participants regulate their level of performance to give either fast, but more error-prone responses, or slower, more accurate responses. Indeed, balancing this speed/accuracy tradeoff by encouraging participants to regulate their performance appropriately for each task was an important factor in the design of experiments 2 and 3, in order to prioritise good RT data. In the model, however, error trials tend to occur with very long RTs, while the fastest responses tend to be the most accurate. This is because in the model responses are not produced until one response unit's activation is sufficiently more than that of the next largest, incompatible unit. Ordinarily, correct responses may occur quickly. However, due to a combination of residual task-demand activation (irrelevant task interference) and noise, errors may occur if an incorrect response unit is able to exceed that of the correct unit. In this case, intense competition typically occurs, with the activation of both units at a similar level, before one unit 'wins' and its activation exceeds the response threshold. This mechanism for producing responses, and of error commission, does not seem well suited to capturing the speed/accuracy tradeoff that characterises human behaviour.

One alternative possibility is provided by substantially more complex model of Brown et al. (2007). In that model, response production is governed by tonic excitation of the response layer, generating regular responses. This tonic signal is disrupted by the detection of various conflicts in the model (e.g., provided by incongruency or across-trial change detectors) producing a slowing effect and reinforcing the activation of the current task set. Thus, slower and more accurate responses are produced following changes or conflicts presented in the task, making the model substantially better suited to modelling speed/accuracy tradeoffs characteristic of human behaviour. How can the present model be reconciled with the model of Brown et al. (2007)? The simplicity of the computational mechanisms accounting for backward inhibition in the present model would make incorporating these elements into a significantly more complex model challenging. The model of Brown et al. (2007) already incorporates mechanisms for conflict (incongruency) detection and response slowing at the task level. However, the complexity of the model's architecture and parameter set would require a substantial survey of the model's behaviour in order to determine how the model as presented by the authors would perform in the three-task paradigm, and how it might be modified to incorporate backward inhibition.[1] In this regard, incremental development of a backward inhibition model from earlier task switching models, within the significantly simpler GRAIN framework, presents advantages in terms of making clear which mechanisms exactly are responsible for the model's success in simulating empirically observed behaviour.

11.3 Final Words

This thesis has presented one possible model of backward inhibition, which offers one explanation for the n-2 repetition cost phenomenon. While it is related to existing verbal backward inhibition theories, it also draws on constraints and computational principles inherited from the IAC modelling framework, as well as previous models of task switching. It therefore represents a cumulative approach to theory building, by extending existing theories of task switching. While computational cognitive models may have many roles within cognitive science, this thesis has aimed to develop a model which will prove useful for the understanding of human cognition. To this end, it aims to present as simplified a model of its subject as possible, while still containing the necessary and sufficient elements to explain the phenomenon in cognitive terms. It has not been the intention of this

---

[1] Although, it remains a possibility that backward inhibition is already a property of the model out-of-the-box.

thesis to present a complete model of human cognition, although it is hoped that the knowledge gained as a part of this thesis will contribute to the development of such models in the future.

Appendix A

Model Parameter Settings

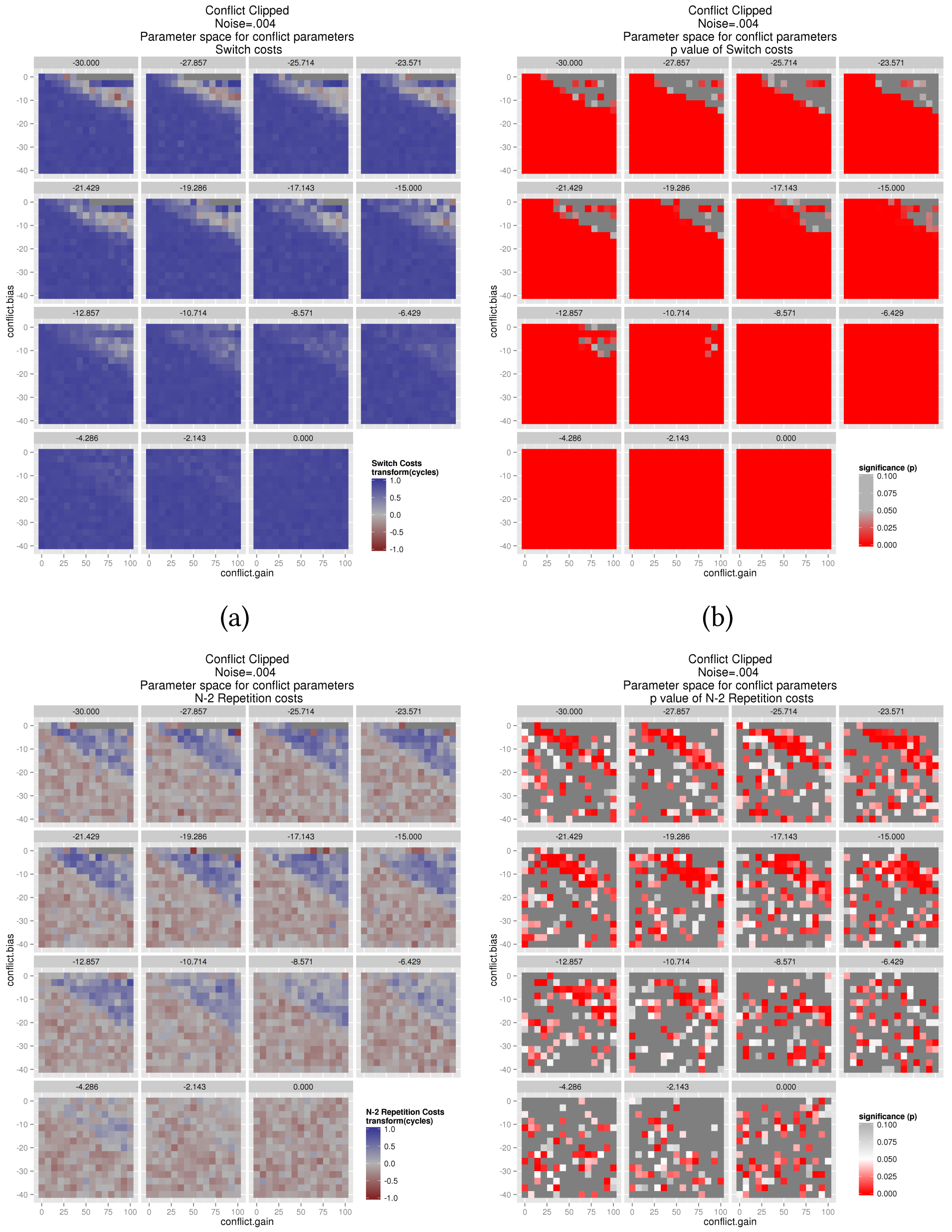
Parameter	Simulation 5	Simulation 6d	Simulation 10	Simulation 8
Unit activation max	1.0	1.0	1.0	1.0
Unit activation min	-1.0	-1.0	-1.0	-1.0
Response Threshold	0.15	0.15	0.15	0.15
Step Size	0.0015	0.0015	0.0015	0.0015
Squashing Parameter (task demand units)	0.8	0.8	0.8	0.8
Noise	0.006	0.006	0.006	0.006
Bias (output units)	-6.0	-6.0	-6.0	-6.0
Bias (task demand units)	-4.0	-4.0	-4.0	-4.0
Stimulus Input Strength (A)	3.0	3.0	3.0	3.0
Stimulus Input Strength (B)	3.0	3.0	3.0	1.5 — 5.0
Stimulus Input Strength (C)	3.0	3.0	3.0	3.0
Topdown Control Strength (A)	12.0	12.0	12.0	12.0
Topdown Control Strength (B)	12.0	12.0	6.0	6.0 — 20.0
Topdown Control Strength (C)	12.0	12.0	12.0	12.0
Output unit lateral weights (incongruent)	-2.0	-2.0	-2.0	-2.0
Output unit lateral weights (congruent)	2.0	2.0	-2.0	2.0
Output unit - task demand weights (congruent)	1.0	1.0	1.0	1.0
Output unit - task demand weights (incongruent)	-1.0	-1.0	-1.0	-1.0
Task demand - output inhibitory weights	-2.5	-2.5	-2.5	-2.5
Task demand - output excitatory weights	2.5	2.5	2.5	2.5
Squashing Parameter (conflict units)	0.5	0.5	0.5	0.5
Bias (conflict units)	-10.0	-40 — 0	-7.5	-10
Gain (conflict units)	75.0	0 — 100	39.0	75.0
Conflict - task demand unit weights	-10.0	-30 — 0	-14.0 / 0	-2.6 — 0

Table A.1: Model parameter settings

Appendix B

Simulation 6e: Parameter search with default noise parameter

This appendix presents parallel results to those from simulation 6d (chapter 6). While those simulations were conducted with a reduced noise parameter (0.004) compared with that used in the model of Gilbert and Shallice (2002), the same simulations were also conducted with the original parameter setting of 0.006. Comparison of results indicates that the noise parameter has no qualitative effect on the shape of parameter space.



(a) (b)

(c) (d)

Figure B.1: Model behaviour when negative conflict is clipped, noise = 0.006 (cf. figure 6.2). Dependent variables (switch costs, B.1a); n-2 repetition costs, B.1c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$

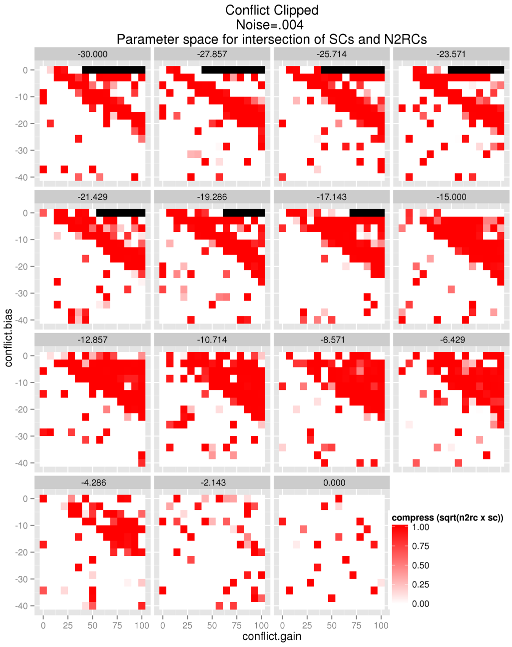
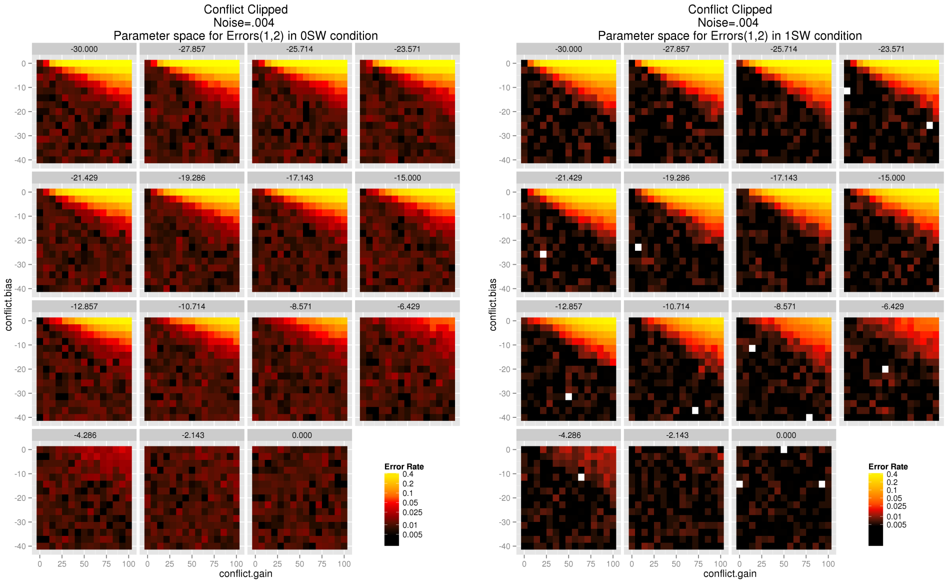


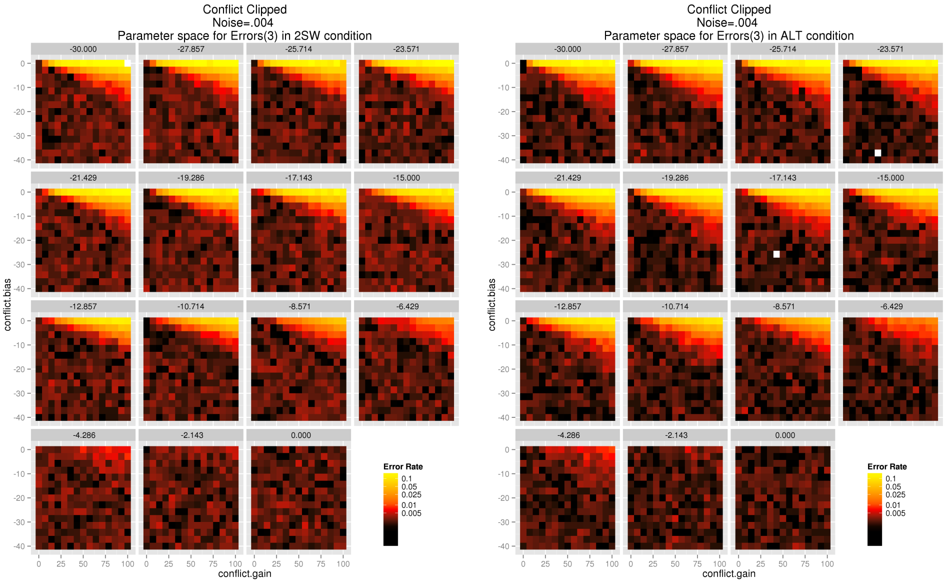
Figure B.2: Intersection of RT switch cost and n-2 repetition cost, negative conflict clipped, noise = 0.006 (cf. figure 6.4)



(a)

(b)

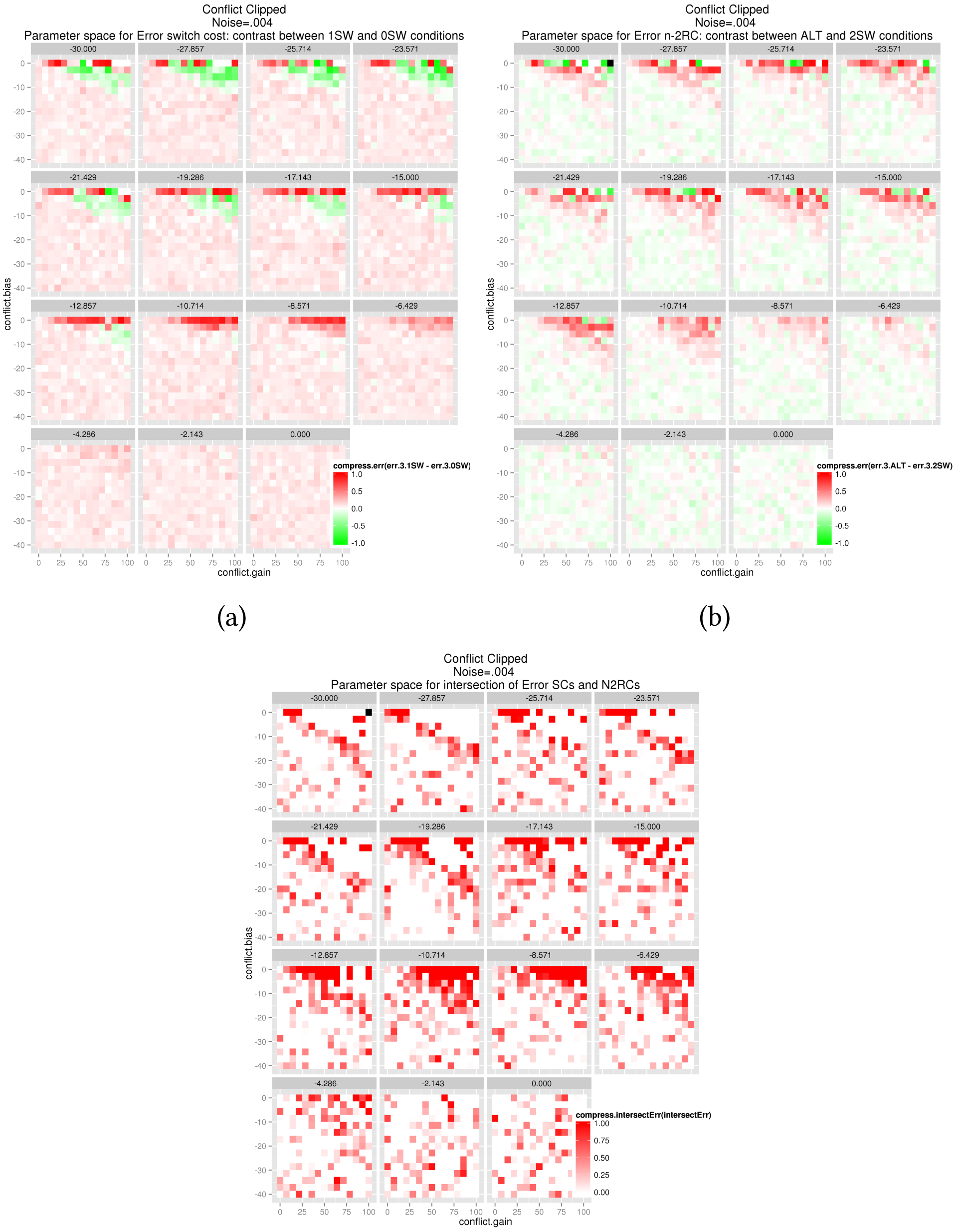
Figure B.3: Model behaviour, assessed via error rates (trials 1&2), negative conflict clipped, noise = 0.006, in the 0SW (B.3a) and 1SW (B.3b) conditions (cf. figure 6.5).



(a)

(b)

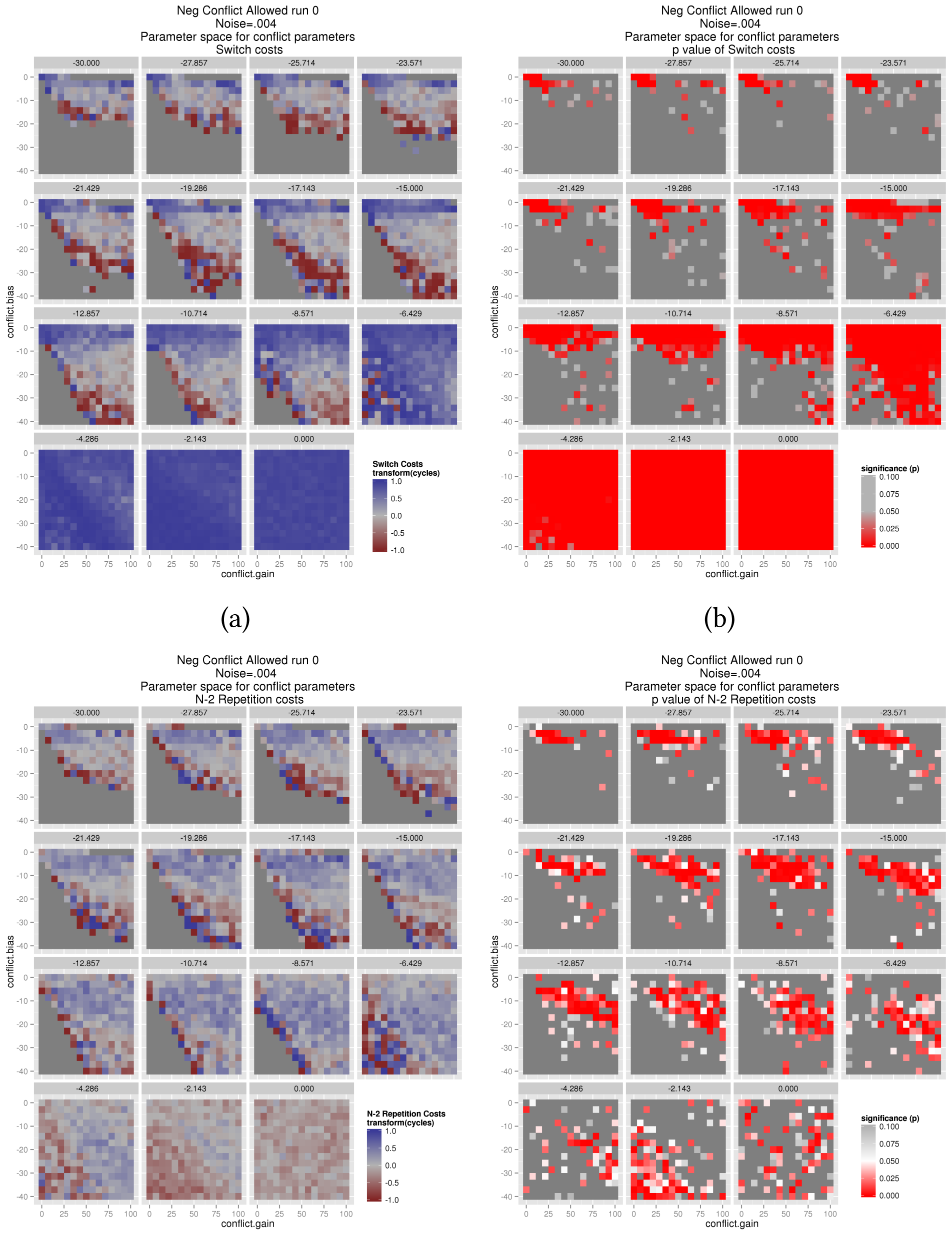
Figure B.4: Model behaviour, assessed via error rates on trial 3, negative conflict clipped, noise = 0.006, in the ALT (B.4b) and 2SW conditions (B.4a).



(a) (b)

(c)

Figure B.5: Model behaviour, assessed via error rates on trial 3, negative conflict clipped, noise=.006. Plots show differences between the 1SW and 0SW conditions (i.e., the error switch cost, panel B.5a) and between the ALT and 2SW conditions (i.e., the error n-2 repetition cost, panel B.5b). The differences are plotted such that the empirical pattern of behaviour (i.e., 1SW $>$ 0SW, ALT $>$ 2SW) is positive, while the alternative (i.e., 1SW $<$ 0SW, etc.) is in green. Note that the differences in error rates are transformed with a logistic function, $f(x) \rightarrow \frac{2}{1+e^{(-100x)}} - 1$, before plotting, and the actual magnitude of the differences in error rates is very small. The overlap between error rates is assessed as the geometric mean of the positive (i.e., red) part of parameter space for SCs and n-2RCs, and is presented in panel B.5b.


Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
Switch costs

Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
p value of Switch costs

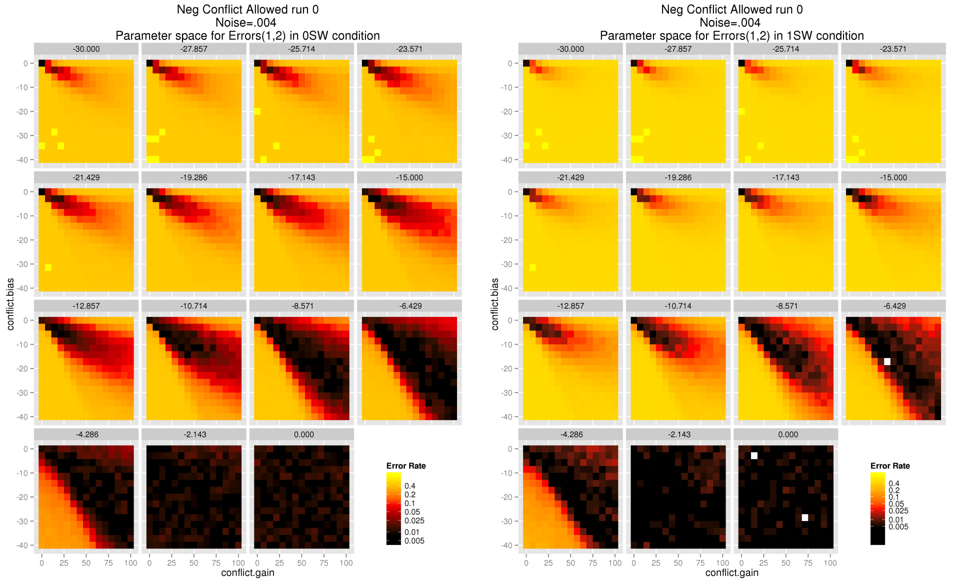
Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
N-2 Repetition costs

Neg Conflict Allowed run 0
Noise=.004
Parameter space for conflict parameters
p value of N-2 Repetition costs

(a) (b)

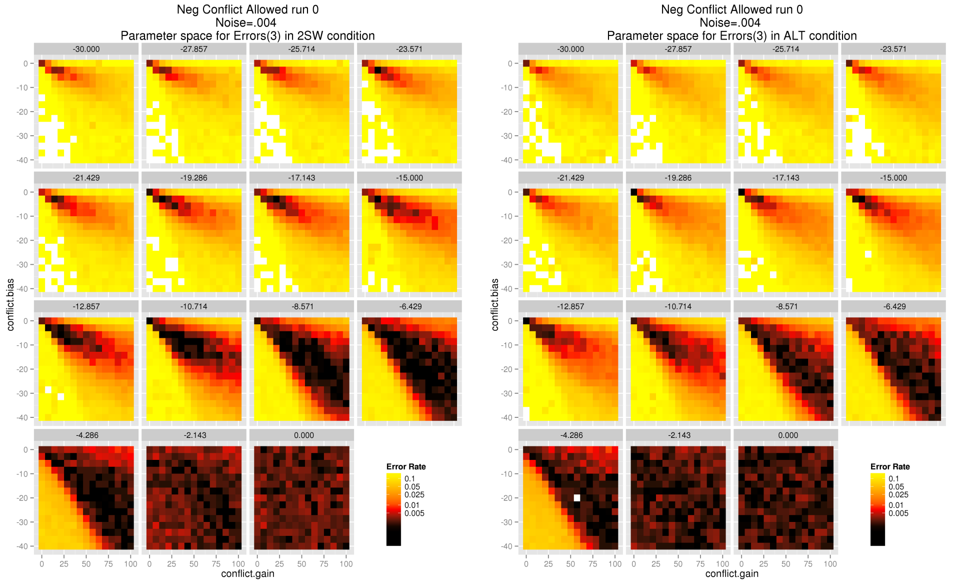
(c) (d)

Figure B.6: Model behaviour when negative conflict is allowed, noise = 0.006 (cf. figure 6.1). Dependent variables (switch costs, (B.6a); n-2 repetition costs, (B.6c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$



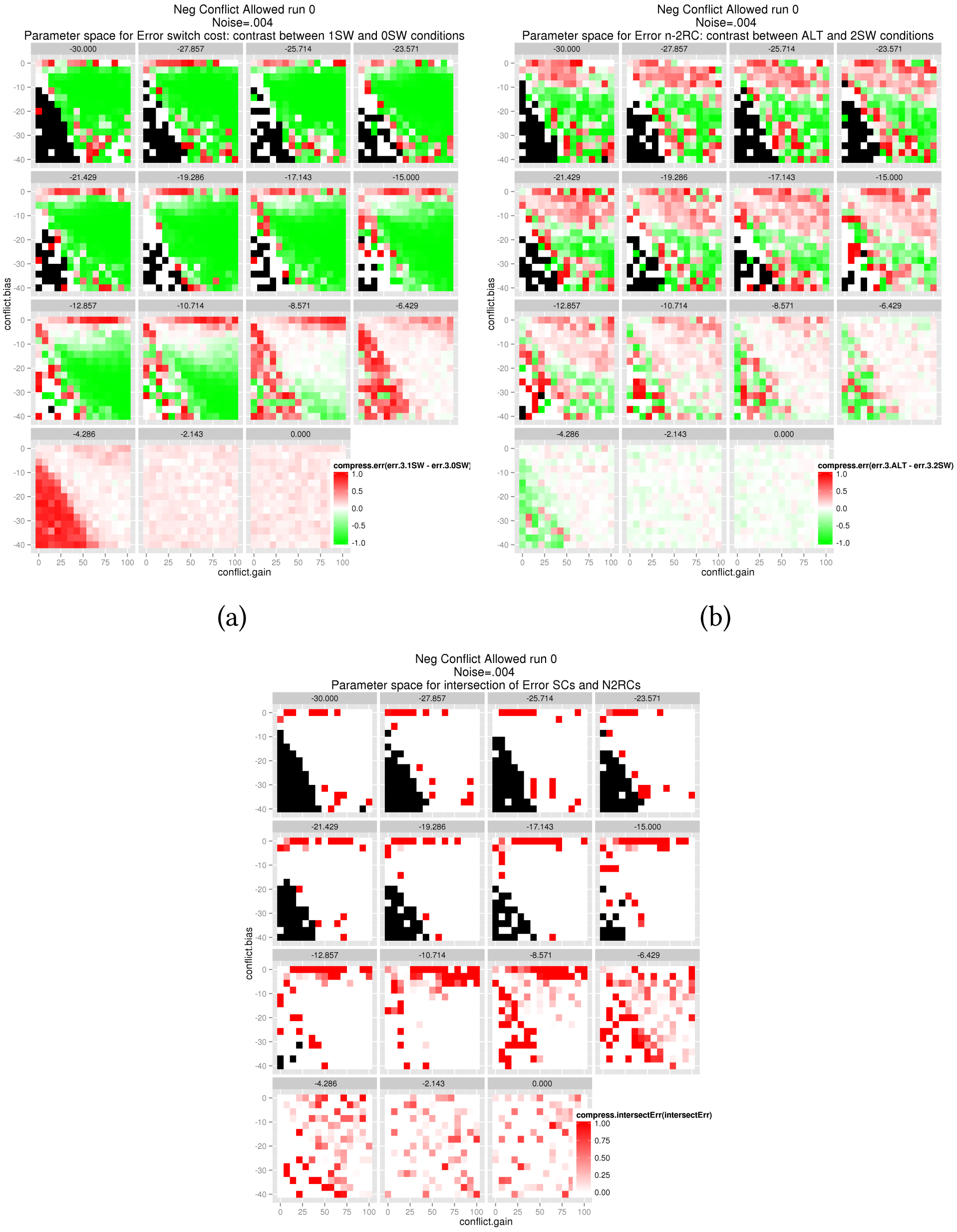
(a) (b)

Figure B.7: Model behaviour, assessed via error rates (trials 1&2), negative conflict allowed, noise = 0.006, in the 0SW (B.7a) and 1SW (B.7b) conditions.



(a) (b)

Figure B.8: Model behaviour, assessed via error rates on trial 3, negative conflict allowed, noise = 0.006, in the in 2SW (B.8a) and ALT (B.8b) conditions.

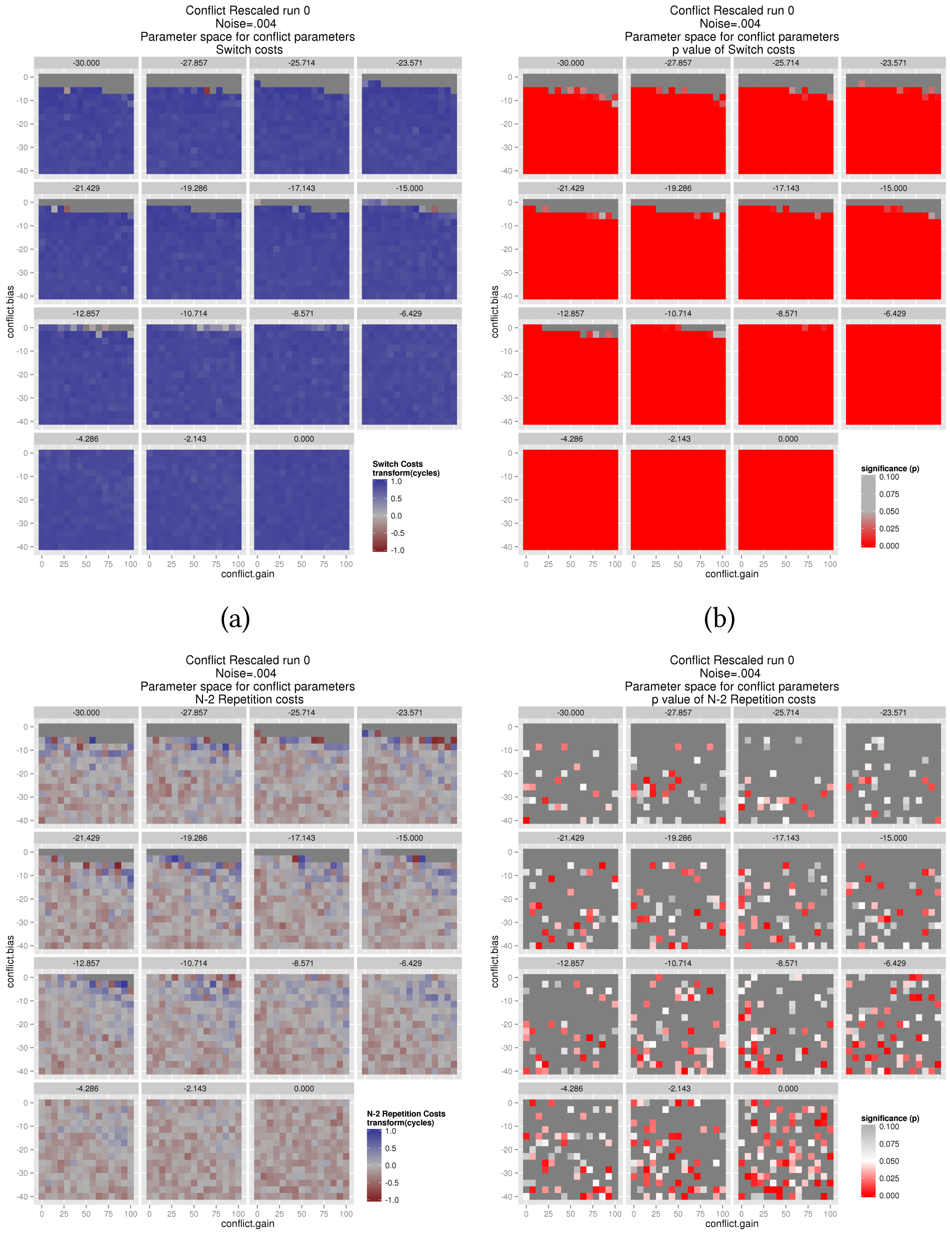


(a)

(b)

(c)

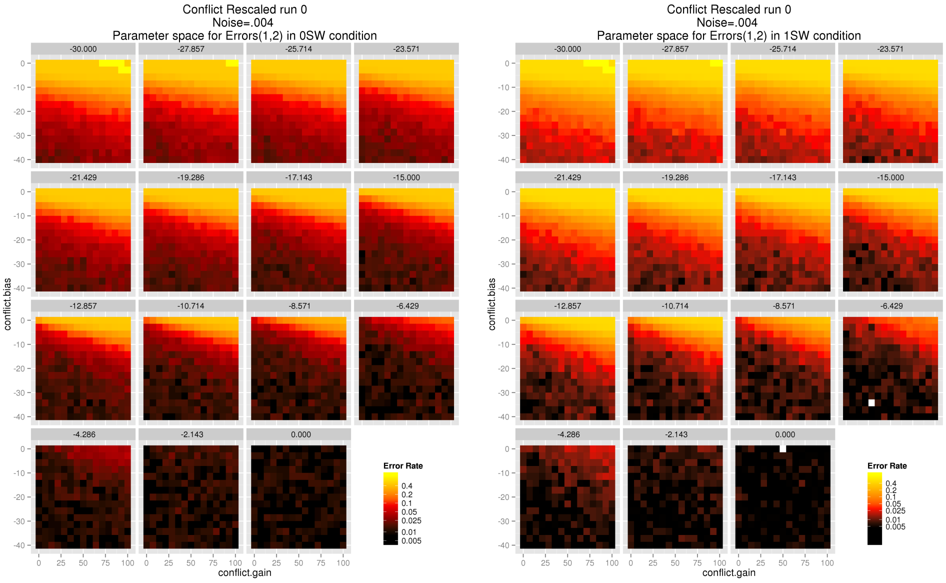
Figure B.9: Model behaviour, assessed via error rates on trial 3, negative conflict allowed, noise=.006. Plots show differences between the 1SW and 0SW conditions (i.e., the error switch cost, panel B.9a) and between the ALT and 2SW conditions (i.e., the error n-2 repetition cost, panel B.9b). The differences are plotted such that the empirical pattern of behaviour (i.e., 1SW $>$ 0SW, ALT $>$ 2SW) is positive, while the alternative (i.e., 1SW $<$ 0SW, etc.) is in green. Note that the differences in error rates are transformed with a logistic function, $f(x) \rightarrow \frac{2}{1+e^{(-100x)}} - 1$, before plotting, and the actual magnitude of the differences in error rates is very small. The overlap between error rates is assessed as the product of the positive (i.e., ed) part of parameter space for SCs and n-2RCs, and is presented in panel B.9c.



(a) (b)

(c) (d)

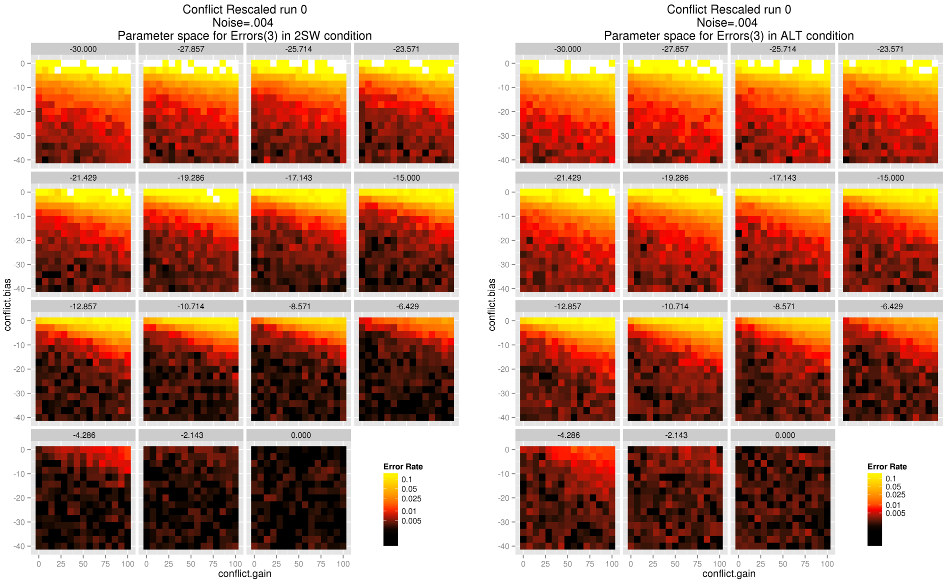
Figure B.10: Model behaviour when negative conflict is rescaled, noise = 0.006 (cf. figure 6.3). Dependent variables (switch costs, (B.10a); n-2 repetition costs, (B.10c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$


Conflict Rescaled run 0
Noise=.004
Parameter space for Errors(1,2) in 0SW condition

Conflict Rescaled run 0
Noise=.004
Parameter space for Errors(1,2) in 1SW condition

(a) (b)

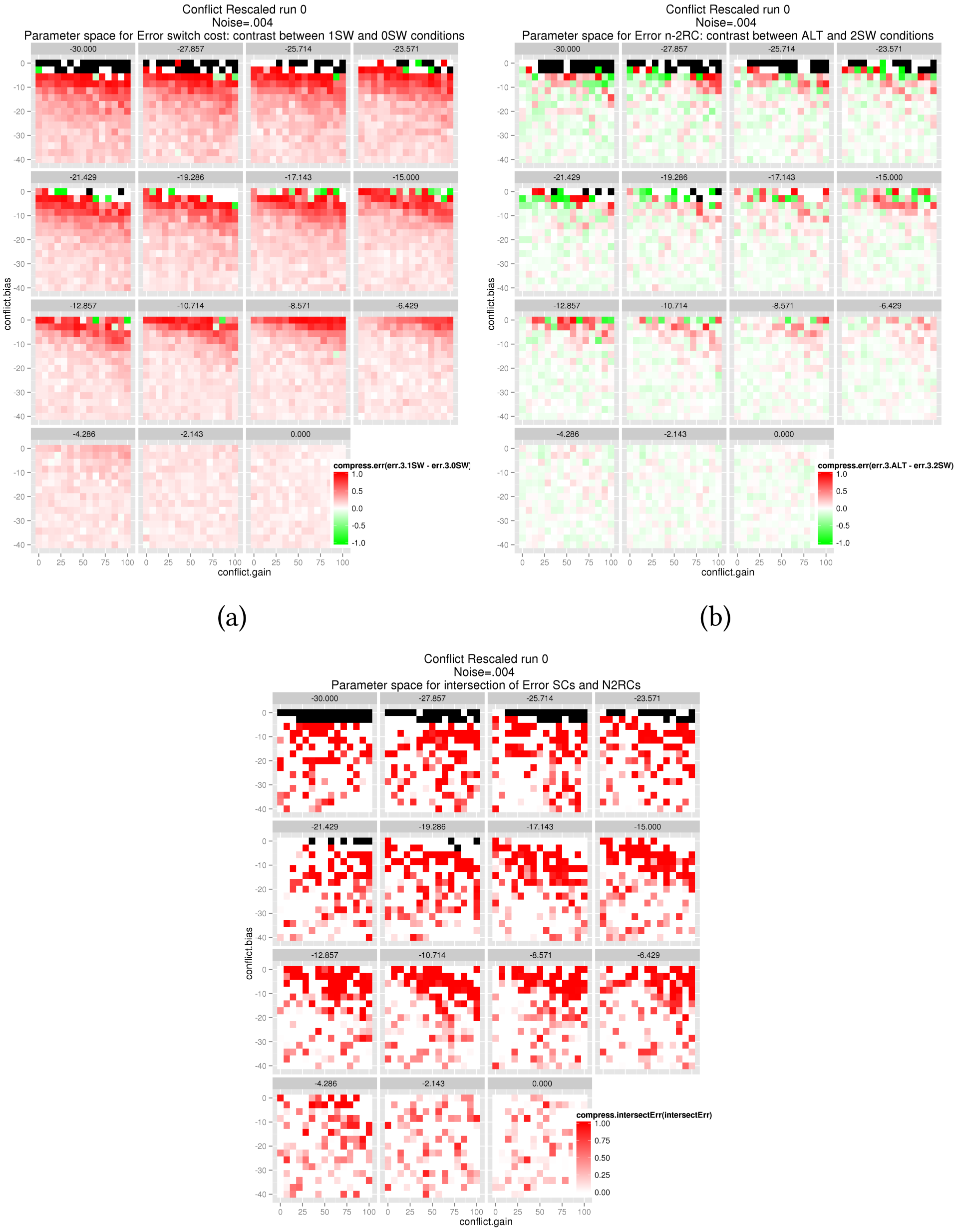
Figure B.11: Model behaviour, assessed via error rates (trials 1&2), negative conflict rescaled, noise = 0.006, in the 0SW (B.11a) and 1SW (B.11b) conditions.


Conflict Rescaled run 0
Noise=.004
Parameter space for Errors(3) in 2SW condition

Conflict Rescaled run 0
Noise=.004
Parameter space for Errors(3) in ALT condition

(a) (b)

Figure B.12: Model behaviour, assessed via error rates on trial 3, negative conflict rescaled, noise = 0.006, in the in 2SW (B.12a) and ALT conditions (B.12b).


Conflict Rescaled run 0
Noise=.004
Parameter space for Error switch cost: contrast between 1SW and 0SW conditions

Conflict Rescaled run 0
Noise=.004
Parameter space for Error n-2RC: contrast between ALT and 2SW conditions

Conflict Rescaled run 0
Noise=.004
Parameter space for intersection of Error SCs and N2RCs

(a)

(b)

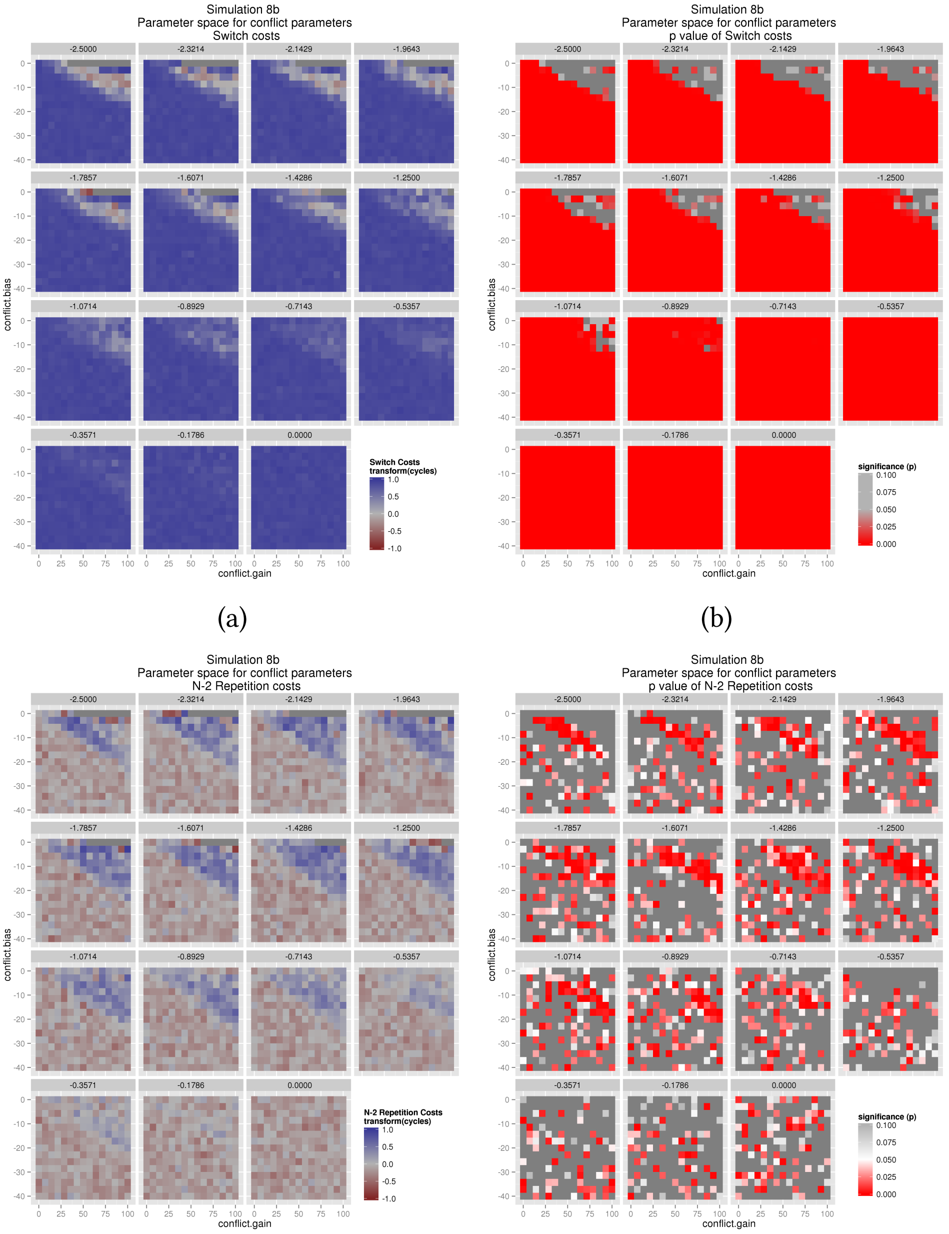
(c)

Figure B.13: Model behaviour, assessed via error rates on trial 3, negative conflict allowed, noise=.006. Plots show differences between the 1SW and 0SW conditions (i.e., the error switch cost, panel B.13a) and between the ALT and 2SW conditions (i.e., the error n-2 repetition cost, panel B.13b). The differences are plotted such that the empirical pattern of behaviour (i.e., 1SW $>$ 0SW, ALT $>$ 2SW) is positive, while the alternative (i.e., 1SW $<$ 0SW, etc.) is in green. Note that the differences in error rates are transformed with a logistic function, $f(x) \to \frac{2}{1+e^{(-100x)}} - 1$, before plotting, and the actual magnitude of the differences in error rates is very small. The overlap between error rates is assessed as the product of the positive (i.e., red) part of parameter space for SCs and n-2RCs, and is presented in panel B.13c.

Appendix C

Simulation 8b: Parameter search with a modified model

Chapter 7 presented a modified version of the backward inhibition model adapted to tasks of asymmetric difficulty. Specifically, the top-down control strength parameter was scaled by the weight parameter for each task. This appendix presents results when the parameter studies conducted in chapter 6 are re-run using the modified model. The figures presented here demonstrate that the qualitative shape of parameter space is not changed for this version of the model.



(a) (b)

(c) (d)

Figure C.1: Model behaviour (n-1 Switch & n-2 repetition costs for RTs) when negative conflict is clipped, noise =.006 (cf. figure 6.2) Dependent variables (switch costs, (C.1a); n-2 repetition costs, (C.1c) are transformed with a logistic function: $f(x) \rightarrow \frac{2}{1+e^{(-0.1x)}} - 1$

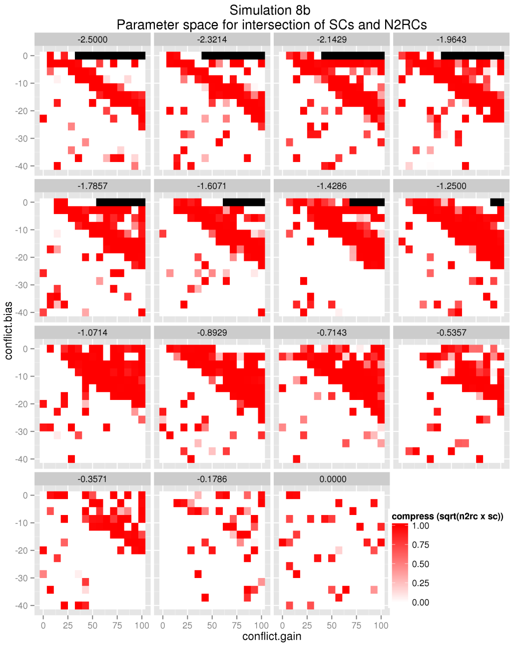
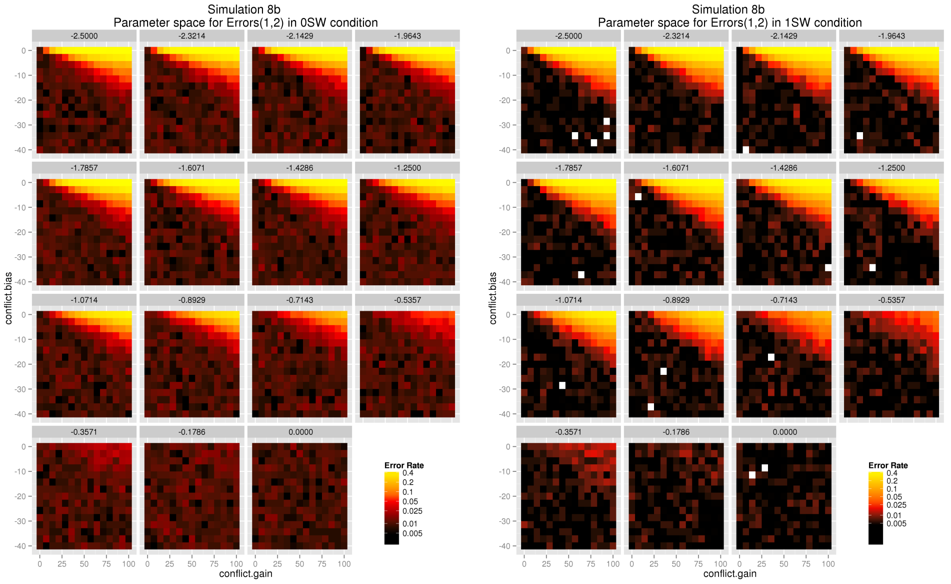
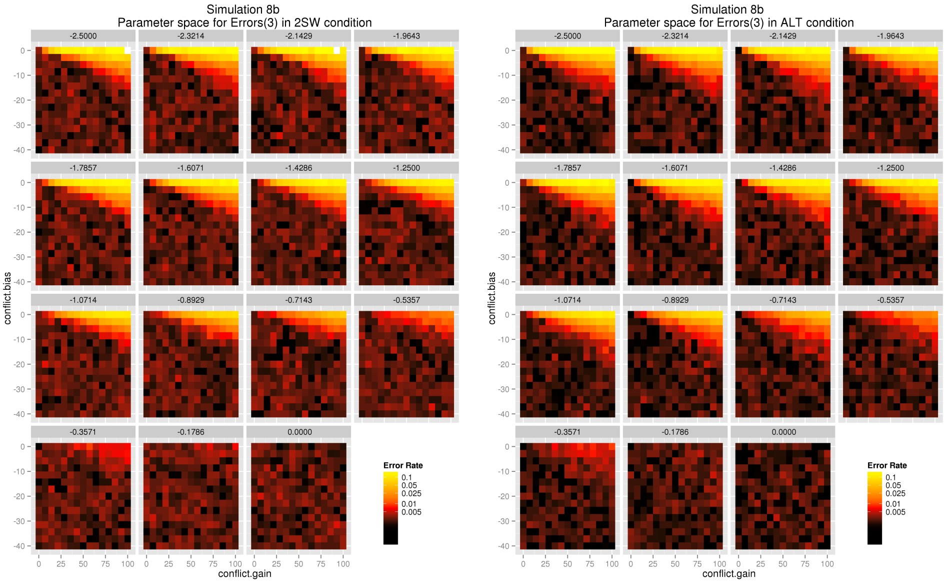


Figure C.2: Model behaviour: Intersection of RT switch cost and n-2 repetition cost (cf. figure 6.4)



(a) (b)

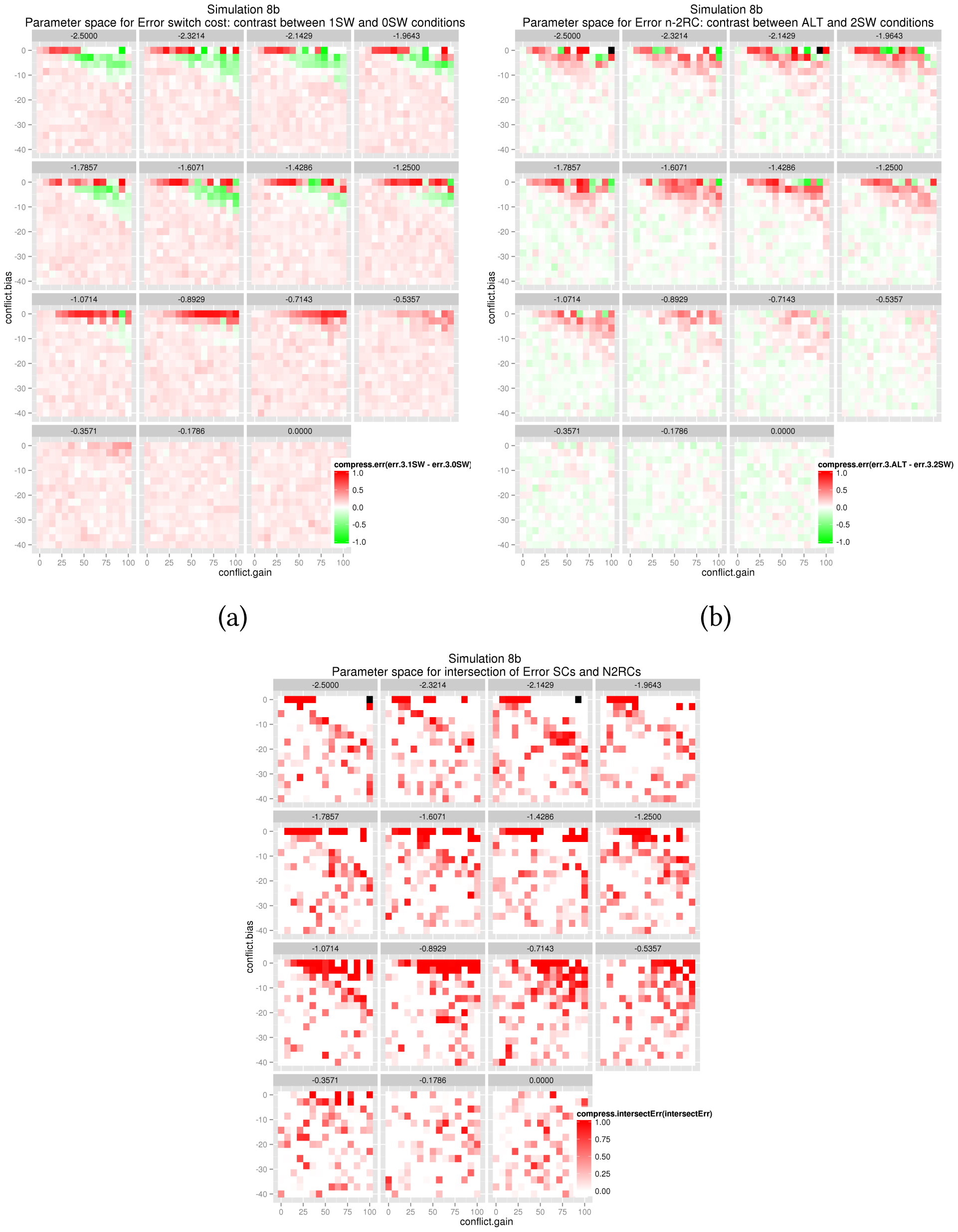
Figure C.3: Model behaviour (error rates), for clipped conflict, in the 0SW (C.3a) and 1SW (C.3b) conditions (cf. figure 6.5)



(a) (b)

Figure C.4: Model behaviour: Trial 3 error rates in the ALT (C.4b)and 2SW (C.4a) conditions (cf. figure 6.6).

between ALT and 2SW conditions)


Simulation 8b
Parameter space for Error switch cost: contrast between 1SW and 0SW conditions

Simulation 8b
Parameter space for Error n-2RC: contrast between ALT and 2SW conditions

Simulation 8b
Parameter space for intersection of Error SCs and N2RCs

(a)

(b)

(c)

Figure C.5: Error rate differences between the 1SW and 0SW conditions (i.e., the error switch cost, panel C.5a) and between the ALT and 2SW conditions (i.e., the error n-2 repetition cost, panel C.5b). cf. figure 6.7. The differences are plotted such that the empirical pattern of behaviour (i.e., 1SW $>$ 0SW, ALT $>$ 2SW) is positive, while the alternative (i.e., 1SW $<$ 0SW, etc.) is in green. Note that the differences in error rates are transformed with a logistic function, $f(x) \rightarrow \frac{2}{1+e^{(-100x)}} - 1$, before plotting, and the actual magnitude of the differences in error rates is very small. The overlap between error rates is assessed as the product of the positive (i.e., red) part of parameter space for SCs and n-2RCs, and is presented in panel C.5c.

References

Allport, A., Styles, E., & Hsieh, S. (1994). Shifting intentional set - exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance xv: Conscious and nonconscious information processing* (Vol. 15, pp. 421–452). Cambridge, MA: MIT Press.

Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford University Press.

Allport, A., & Wylie, G. (2000). Task switching, stimulus-response bindings, and negative priming. In *Control of cognitive processes attention and performance XVIII* (pp. 35 – 70). Cambridge, MA: MIT Press.

Altmann, E. M. (2004). Advance preparation in task switching. *Psychological Science (Wiley-Blackwell)*, *15*(9), 616–622.

Altmann, E. M. (2006). Task switching is not cue switching. *Psychonomic Bulletin & Review*, *13*(6), 1016–1022.

Altmann, E. M. (2007a). Comparing switch costs: Alternating runs and explicit cuing. *Journal of Experimental Psychology-Learning Memory and Cognition*, *33*(3), 475–483.

Altmann, E. M. (2007b). Cue-independent task-specific representations in task switching: Evidence from backward inhibition. *Journal of Experimental Psychology-Learning Memory and Cognition*, *33*(5), 892–899.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, *115*(3), 602–639.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Arbuthnott, K. D. (2005). The influence of cue type on backward inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1030–1042.

Arbuthnott, K. D. (2008a). Asymmetric switch cost and backward inhibition: Carryover activation and inhibition in switching between tasks of unequal difficulty. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(2), 91–100.

Arbuthnott, K. D. (2008b). The effect of task location and task type on backward inhibition. *Memory & Cognition (pre-2011)*, *36*(3), 534–43.

Arbuthnott, K. D. (2009). The representational locus of spatial influence on backward inhibition. *Memory & Cognition*, *37*(4), 522–528.

Arbuthnott, K. D., & Frank, J. (2000). Executive control in set switching: Residual switch cost and task-set inhibition. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*(1), 33–41.

Arbuthnott, K. D., & Woodward, T. S. (2002). The influence of cue-task association and location on switch cost and alternating-switch cost. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *56*(1), 18–29.

Ashby, F. G. (1982). Deriving exact predictions from the cascade model. *Psychological Review*, 599-607.

Astle, D. E., Jackson, G. M., & Swainson, R. (2012). Two measures of task-specific inhibition. *The Quarterly Journal of Experimental Psychology*, *65*(2), 233–251.

Bao, M., Li, Z.-H., Chen, X.-C., & Zhang, D.-R. (2006). Backward inhibition in a task of switching attention within verbal working memory. *Brain Research*

Bulletin, *69*(2), 214–221.

Blais, C., Robidoux, S., Risko, E. F., & Besner, D. (2007). Item-specific adaptation and the conflict-monitoring hypothesis: A computational model. *Psychological Review*, *114*(4), 1076–1086.

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive Affective & Behavioral Neuroscience*, *7*(4), 356–366.

Botvinick, M. M., Braver, T., Barch, D., Carter, C., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*(12), 539–546.

Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology*, *55*(1), 37–85.

Bryck, R. L., & Mayr, U. (2008). Task selection cost asymmetry without task switching. *Psychonomic Bulletin & Review*, *15*(1), 128–134.

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. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance xv: Conscious and nonconscious information processing* (pp. 453 – 476). MIT Press.

Cooper, R. P. (2010). Cognitive control: Componential or emergent? *Topics in Cognitive Science*, *2*(4), 598–613.

Cooper, R. P., & Davelaar, E. J. (2010). Modelling the correlation between two putative inhibition tasks: A simulation approach. In *Proceedings of the 10th* *international conference on cognitive modeling.* Philadelphia, PA: Drexel University.

Cooper, R. P., Fox, J., Farringdon, J., & Shallice, T. (1996). A systematic methodology for cognitive modelling. *Artificial Intelligence*, *85*(1-2), 3–44.

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297–338.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and l2 learners. *Journal of Memory and Language*, *50*(4), 491–511.

De Jong, R. (2000). An intentional-activation account of residual switch costs. In *Control of cognitive processes attention and performance XVIII.* Cambridge, MA: MIT Press.

Dick, P. K. (1969, October). The electric ant. *The Magazine of Fantasy and Science Fiction.*

Dreher, J.-C., & Berman, K. F. (2002). Fractionating the neural substrate of cognitive control processes. *Proceedings of the National Academy of Sciences*, *99*(22), 14595–14600.

Dreisbach, G., Haider, H., & Kluwe, R. H. (2002). Preparatory processes in the task-switching paradigm: Evidence from the use of probability cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 468–483.

Druey, M. D., & Hübner, R. (2007). The role of temporal cue-target overlap in backward inhibition under task switching. *Psychonomic Bulletin & Review*, *14*(4), 749–754.

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.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology. General*, *137*(2), 201–225.

Gade, M., & Koch, I. (2005). Linking inhibition to activation in the control of task sequences. *Psychonomic Bulletin & Review*, *12*(3), 530–534.

Gade, M., & Koch, I. (2007). The influence of overlapping response sets on task inhibition. *Memory & Cognition*, *35*(4), 603–609.

Gade, M., & Koch, I. (2008). Dissociating cue-related and task-related processes in task inhibition: Evidence from using a 2:1 cue-to-task mapping. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(1), 51–55.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*(3), 297–337.

Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes. attention and performance XVIII* (p. 331-355). Cambridge, MA: MIT Press.

Grange, J. A., & Houghton, G. (2009). Temporal cue-target overlap is not essential for backward inhibition in task switching. *Quarterly Journal of Experimental Psychology*, *62*(10), 2068–2079.

Grange, J. A., & Houghton, G. (2010a). Cue-switch costs in task-switching: cue priming or control processes? *Psychological Research-Psychologische Forschung*, *74*(5), 481–490.

Grange, J. A., & Houghton, G. (2010b). Heightened conflict in cue-target translation increases backward inhibition in set switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 1003–1009.

Grange, J. A., & Houghton, G. (2011). Task preparation and task inhibition: a comment on koch, gade, schuch, & philipp (2010). *Psychonomic Bulletin & Review*, *18*(1), 211–216.

Grange, J. A., & Houghton, G. (2014). Models of cognitive control in task switching. In *Task switching and cognitive control* (pp. 160–199). Oxford University Press.

Grange, J. A., & Juvina, I. (2015). The effect of practice on n–2 repetition costs in set switching. *Acta Psychologica*, *154*, 14–25.

Grange, J. A., Juvina, I., & Houghton, G. (2013). On costs and benefits of n-2 repetitions in task switching: towards a behavioural marker of cognitive inhibition. *Psychological Research-Psychologische Forschung*, *77*(2), 211–222.

Grange, J. A., Kowalczyk, A. W., & O'Loughlin, R. (2017). The effect of episodic retrieval on inhibition in task switching. *Journal of Experimental Psychology: Human Perception and Performance*.

Gratton, G., Coles, M. G., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre-and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology: Human perception and performance*, *14*(3), 331.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709.

Houghton, G., Pritchard, R., & Grange, J. A. (2009). The role of cue–target translation in backward inhibition of attentional set. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 466–476.

Hübner, M., Dreisbach, G., Haider, H., & Kluwe, R. H. (2003). Backward inhibition as a means of sequential task-set control: Evidence for reduction of task competition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 289–297.

Jersild, A. T. (1927). Mental set and shift. *Archives of psychology*.

Jost, K., De Baene, W., Koch, I., & Brass, M. (2013). A review of the role of cue processing in task switching. *Zeitschrift für Psychologie*, *221*(1), 5–14.

Jost, K., Mayr, U., & Roesler, F. (2008). Is task switching nothing but cue priming?

evidence from ERPs. *Cognitive Affective & Behavioral Neuroscience*, *8*(1), 74–84.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874.

Kimberg, D. Y., Aguirre, G. K., & D'Esposito, M. (2000). Modulation of task-related neural activity in task-switching: an fMRI study. *Cognitive Brain Research*, *10*(1), 189–196.

Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1474–1486.

Koch, I. (2005). Sequential task predictability in task switching. *Psychonomic Bulletin & Review*, *12*(1), 107–112.

Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition*, *34*(2), 433–444.

Koch, I., & Brass, M. (2013). Task switching—maturation of a paradigm. *Zeitschrift für Psychologie*, *221*(1), 1–4.

Koch, I., Gade, M., & Philipp, A. M. (2004). Inhibition of response mode in task switching. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *51*(1), 52–58.

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review*, *17*(1), 1–14.

Koch, I., & Philipp, A. M. (2005). Effects of response selection on the task repetition benefit in task switching. *Memory & Cognition (pre-2011)*, *33*(4), 624–34.

Koch, I., Philipp, A. M., & Gade, M. (2006). Chunking in task sequences modulates task inhibition. *Psychological Science (Wiley-Blackwell)*, *17*(4), 346–350.

Kuhns, D., Lien, M.-C., & Ruthruff, E. (2007). Proactive versus reactive task-set inhibition: Evidence from flanker compatibility effects. *Psychonomic Bulletin & Review*, *14*(5), 977–983.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, *29*(3), 575–599.

Logan, G. D., & Bundesen, C. (2004). Very clever homunculus: Compound stimulus strategies for the explicit task-cuing procedure. *Psychonomic Bulletin & Review*, *11*(5), 832–840.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393–434.

Logan, G. D., & Schneider, D. W. (2006). Priming or executive control? associative priming of cue encoding increases "switch costs" in the explicit task-cuing procedure. *Memory & Cognition*, *34*(6), 1250–1259.

Logan, G. D., & Schneider, D. W. (2010). Distinguishing reconfiguration and compound-cue retrieval in task switching. *Psychologica Belgica*, *50*(3), 413-433.

Logan, G. D., Schneider, D. W., & Bundesen, C. (2007). Still clever after all these years: Searching for the homunculus in explicitly cued task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 978–994.

Longman, C. S., Lavric, A., Munteanu, C., & Monsell, S. (2014). Attentional inertia and delayed orienting of spatial attention in task-switching. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1580.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. *Psychology of Learning and Motivation: Advances in Research and Theory*, *43*, 163–214.

Mayr, U. (2001). Age differences in the selection of mental sets: the role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and aging*, *16*(1), 96-109.

Mayr, U. (2002). Inhibition of action rules. *Psychonomic Bulletin & Review*, *9*(1), 93–99.

Mayr, U. (2006). What matters in the cued task-switching paradigm: Tasks or cues? *Psychonomic Bulletin & Review*, *13*(5), 794–799.

Mayr, U. (2009). Sticky plans: Inhibition and binding during serial-task control. *Cognitive Psychology*, *59*(2), 123–153.

Mayr, U., Diedrichsen, J., Ivry, R., & Keele, S. W. (2006). Dissociating task-set selection from task-set inhibition in the prefrontal cortex. *Journal of Cognitive Neuroscience*, *18*(1), 14–21.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, *129*(1), 4–26.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 362–372.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: The MIT Press.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological Review*, *88*(5), 375–407.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*(3), 211–253.

Meiran, N., Hsieh, S., & Dimov, E. (2010). Resolving task rule incongruence during task switching by competitor rule suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 992–1002.

Meiran, N., & Kessler, Y. (2008). The task rule congruency effect in task switching reflects activated long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 137–157.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40.

Milán, E. G., Sanabria, D., Tornay, F., & González, A. (2005). Exploring task-set reconfiguration with random task sequences. *Acta Psychologica*, *118*(3), 319–331.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science (Sage Publications Inc.)*, *21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 493–516.

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342.

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: Is it easier to switch to the weaker task? *Psychological Research*, *63*(3), 250.

Newell, A. (1990). *Unified theories of cognition.* Harvard University Press.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, R. Schwartz, & R. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory IV.* Plenum Press.

O'Reilly, R. C., Hazy, T. E., & Herd, S. A. (2013). The leabra cognitive architecture: How to play 20 principles with nature and win! In *Oxford handbook of cognitive science.* Oxford University Press.

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*, 10.

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, *19*(3), 395–416.

Philipp, A. M., Jolicoeur, P., Falkenstein, M., & Koch, I. (2007). Response selection and response execution in task switching: Evidence from a go-signal paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1062–1075.

Philipp, A. M., & Koch, I. (2005). Switching of response modalities. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1325–1338.

Philipp, A. M., & Koch, I. (2006). Task inhibition and task repetition in task switching. *European Journal of Cognitive Psychology*, *18*(4), 624–639.

Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1187–1195.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*(1), 57.

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

Regev, S., & Meiran, N. (2016). Cue-type manipulation dissociates two types of task set inhibition: backward inhibition and competitor rule suppression.

Psychological Research, *80*(4), 625–639.

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

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 763–797.

Scheil, J. (2016). Effects of absolute and relative practice on n-2 repetition costs. *Acta Psychologica*, *164*, 65–69.

Scheil, J., & Kleinsorge, T. (2014). N - 2 repetition costs depend on preparation in trials n - 1 and n - 2. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 865–872.

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 222–250.

Schneider, D. W. (2007). Task-set inhibition in chunked task sequences. *Psychonomic Bulletin & Review*, *14*(5), 970–976.

Schneider, D. W. (2014). Modeling graded response congruency effects in task switching. *Acta Psychologica*, *153*, 160–168.

Schneider, D. W., & Anderson, J. R. (2010). Asymmetric switch costs as sequential difficulty effects. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1873–1894.

Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: a short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General*, *134*(3), 343.

Schneider, D. W., & Logan, G. D. (2009). Selecting a response in task switching:

Testing a model of compound cue retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 122–136.

Schneider, D. W., & Verbruggen, F. (2008). Inhibition of irrelevant category–response mappings. *The Quarterly Journal of Experimental Psychology*, *61*(11), 1629–1640.

Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 92–105.

Schuch, S., & Koch, I. (2004). The costs of changing the representation of action: Response repetition and response-response compatibility in dual tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 566–582.

Sdoia, S., & Ferlazzo, F. (2008). Stimulus-related inhibition of task set during task switching. *Experimental Psychology*, *55*(5), 322–327.

Sohn, M.-H., & Anderson, J. R. (2001). Task preparation and task repetition: two-component model of task switching. *Journal of Experimental Psychology: General*, *130*(4), 764.

Stafford, T., & Gurney, K. N. (2011). Additive factors do not imply discrete processing stages: a worked example using models of the stroop task. *Frontiers in Psychology*, *2*, 287.

Steinhauser, M., & Hübner, R. (2006). Response-based strengthening in task shifting: Evidence from shift effects produced by errors. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 517–534.

Steinhauser, M., & Hübner, R. (2008). How task errors affect subsequent behavior: Evidence from distributional analyses of task-switching effects. *Memory & Cognition (pre-2011)*, *36*(5), 979–90.

Steinhauser, M., Maier, M., & Hübner, R. (2007). Cognitive control under stress how stress affects strategies of task-set reconfiguration. *Psychological Science*, *18*(6), 540–545.

Sternberg, S. (1969). The discovery of processing stages: Extensions of donders' method. *Acta psychologica*, *30*, 276–315.

Sternberg, S. (1998). *Discovering mental processing stages: The method of additive factors.* The MIT Press.

Sternberg, S. (2013). The meaning of additive reaction-time effects: some misconceptions. *Frontiers in Psychology*, *4*, 744.

Tornay, F. J., & Milán, E. G. (2001). A more complete task-set reconfiguration in random than in predictable task switch. *Quarterly Journal of Experimental Psychology: Section A*, *54*(3), 785–803.

Vandierendonck, A. (2013). On the role of dissipating inhibition in task switching. *Zeitschrift für Psychologie*, *221*(1), 61–62.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626.

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience & Biobehavioral Reviews*, *33*(5), 647–661.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus–task bindings in task-shift costs. *Cognitive Psychology*, *46*(4), 361–413.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 455–469.