# Task-set control and procedural working memory

Submitted by Félice Maria van ‘t Wout to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Psychology
in May 2012

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature:…………………………………………………….

# Abstract

Flexible and goal-driven behaviour requires a process by which the appropriate task-set is selected and maintained in a privileged state of activation. This process can be conceptualised as loading a task-set into a *procedural working memory (PWM)* buffer. Task switching experiments, which exercise this process, reveal “switch costs”: increased reaction times and error rates when the task changes, compared to when it repeats. The process of loading a task-set into PWM may be one source of these costs. The switch cost is reduced with preparation, suggesting that at least some of the processes involved in a successful change of task can be achieved in advance of the stimulus.

The aim of this thesis was to investigate the properties of PWM, and its contribution to task-set control. One account of PWM distinguishes between the level at which recently exercised (but currently irrelevant) task-sets are represented, and the level at which only the currently relevant task-set is maintained in a most active state. To distinguish between these levels of representation, and to assess the extent to which the process of getting a task-set into a most-active state (loading it into the PWM buffer) is subject to a capacity limit at each level, the experiments varied the number of tasks participants switched among (Experiments 1 and 2), and the complexity of individual task-sets (Experiments 3-6) in a task-cueing paradigm.

In Experiments 1 and 2, participants switched among three or five tasks, in separate sessions. There was no effect of the number of tasks on the switch cost, or its reduction with preparation, provided that recency and frequency of task usage were matched. When recency and frequency were not matched, there appeared to be a larger switch cost with five tasks at a short preparation interval, suggesting that the time consumed by getting a task-set into a most active state is influenced by its recency and frequency of usage, not the number of alternatives per se.

However, Experiment 3 showed that the time required to select an S-R mapping within a task-set does increase as a function of the number of alternatives (even when stimulus frequency and recency are matched), suggesting that representation of the most active task-set in a PWM buffer is subject to a strict capacity limit. Experiments 4-6 further investigated the capacity limit of this PWM buffer, and found that task-set preparation was more effective for task-sets that are less complex (i.e. specified by fewer S-R rules). These findings suggest that only very few S-R rules can be maintained in a most active state in the PWM buffer.

Finally, Experiments 7-9 investigated whether S-R rules are represented phonologically for task-set maintenance and preparation, by manipulating the phonological properties of the stimulus terms. But task-cueing performance was not affected by the name length (Experiment 7) or phonological similarity (Experiments 8 and 9) of the stimulus terms. These results suggest that phonological representations of S-R rules do not make a functional contribution to task-set control, possibly because the rules are compiled into a non-linguistic PWM.

The results of these experiments are discussed in terms of a procedural working memory which is separate from declarative working memory, and distinguishes between two levels of task-set control: the level of task-sets, which are maintained in a capacity unlimited state of representation, and the level at which the currently relevant task-set is maintained in a most-active but highly capacity limited state of representation.

# Acknowledgements

Firstly, thanks go to Stephen Monsell, my first supervisor. I have happy memories of our animated discussions about procedural working memory. I would also like to thank my second supervisor, Aureliu Lavric, for his encouragement and expert advice.

Thanks also to my good friends and fellow (former) PhD students – particularly Heike Elchlepp, Charlotte Forrest and Tobias Stevens. There is nothing like a swim in the sea or a trip to the pub with good friends to make you forget that you just spent a week collecting data in a windowless cubicle.

During the third year of my PhD I also spent two very fun and productive weeks in Klaus Oberauer's lab in Zurich: I would like to thank Klaus and his group for welcoming me into their lab, and Miriam Gade and her lovely family for welcoming me into their home.

Lastly, I would like to thank my parents, who always encouraged me to do whatever made me happy; and Peter Lown, my almost-husband, for everything.

The research was supported by a studentship from the Economic and Social Research Council.

# Index

# List of Figures

# List of Tables

# Declaration

The research reported in this thesis was carried out at the University of Exeter between October 2008 and May 2012, and was supervised by Prof. Stephen Monsell and Dr. Aureliu Lavric.

This dissertation has not been submitted, in whole or in part, for any other degree, diploma or qualification at any university. Chapters 2 and 5 are articles that will be submitted to scientific journals. Chapter 2 will be submitted to Journal of Experimental Psychology: Human Perception and Performance by van ‘t Wout, F., Monsell, S., and Lavric, A. I conducted the experiments, wrote the first draft and prepared the figures and tables. My co-authors have edited the manuscript. Chapter 5 will be submitted to Journal of Experimental Psychology: Learning, Memory and Cognition. I conducted the experiments, wrote the first draft and prepared the figures and tables. My co-authors have edited the manuscript.

Two second year interns, Beth Whiteman and Michael Clayton, kindly helped to collect the data for Experiment 6.

Félice van ‘t Wout
Exeter, May 2012.

# Chapter 1: General Introduction

Human behaviour is characterised by our ability to respond flexibly to a changing environment. Our actions are driven both by internal factors (our goals and intentions) and external influences. Among the many cognitive skills that enable flexible goal-driven behaviour, one appears particularly important: any one situation or object typically affords a range of possible actions, but how do we select the appropriate action from among its competitors? Imagine a person with perfect knowledge of the rules of a language, but who lacks the ability to apply each rule as it is needed. Clearly, knowing *how* to do something is of no use when you are unable to apply a particular rule or behaviour when it is required.

To act in accordance with our goals and environment we must apply cognitive control; we must instantiate an appropriate organisation of mental resources, a "task-set", which specifies what to attend to and when, and what the appropriate course of action is given particular input. When required, this task-set must be maintained in a privileged state of activation, until a change in our external environment or our intentions calls for a different course of action. Indeed, daily life frequently requires us to shift between different tasks. This process has been studied extensively in the laboratory using task switching paradigms, which require participants to classify stimuli according to one of several sets of rules. Several theories have been proposed to account for the different phenomena generated by these paradigms. There are of course important differences between these theories, but most theories require a process by which one task-set temporarily dominates all other task-sets. One way of conceptualising this process is to consider the most activate task-set as being represented in a *procedural working memory* buffer. In contrast to the study of task switching, which has become increasingly popular over the last two decades or so, the properties of procedural working memory remain poorly specified. One of the major aims of this thesis was to develop a bridge between research on task-set control and that on procedural working memory. In this chapter I will firstly review the research on task-set, followed by a review of the procedural working memory literature. This chapter will end with a description of the questions addressed by the experiments reported in this thesis.

## 1.1 Task-set control

This section will first review paradigms that have been used to study task-set control. Then I will describe two prominent theories that have been used to explain the cost of switching between tasks, followed by a summary of the key phenomena in the literature, and the extent to which each theory is able to account for the empirical phenomena. The final section of this first part will consider the proposed sub-processes of task-set reconfiguration in more detail, and introduce the concept of ‘procedural working memory’. This will then be the focus of the second part of this chapter.

### 1.1.1 Task switching paradigms

Task-set control can be measured using a variety of procedures. The first of these procedures was Jersild’s (1927) list paradigm, which required participants to cycle through a list of numbers, either alternating between two tasks (mixed list condition, e.g. +6 and -3), or performing only one of those tasks. Jersild found what he called a “shift loss”: longer completion times for mixed lists than for pure lists. But it was not clear to what extent increased list completion times for mixed lists reflects the cost of switching between tasks, and to what extent it reflects the requirement to keep two task-sets active (Rogers & Monsell, 1995), or increased interference between those task-sets (Rubin & Meiran, 2005). Consequently, the cost of switching among tasks relative to performing just one task, as found in these list paradigms, is more commonly referred to as the mixing cost (e.g. Rubin & Meiran, 2005), or “global switch cost” (Kray & Lindenberger, 2000).

In order to circumvent the limitations associated with the list procedure, Rogers and Monsell (1995) developed the alternating runs paradigm. This paradigm requires participants to switch or not switch predictably between two (or more) tasks within blocks of trials. For example, Rogers and Monsell (1995) presented participants with a letter-digit pair in one of four positions of a quadrant. The stimulus location moved clockwise on successive trials, and depending where they were in a cycle of four trials, participants determined either whether the letter was a vowel or a consonant (letter task) or whether it was an odd or even number (number task). For each participant, the task sequence was AABBAA etc., and the switch cost was calculated by subtracting the mean RT (or error rate) on repeat trials (i.e. AA or BB) from that on task-switch trials

(AB or BA). The key advantage of the alternating runs paradigm (in comparison to the list paradigm) is that it allows for a calculation of the switch cost within a block. Furthermore, the successive presentation of individual stimuli on a computer screen also allows for the manipulation of the time available between the last response and the upcoming stimulus (response-stimulus interval; RSI). Altmann (2007) has criticised the alternating runs paradigm, arguing that the switch cost it obtains is overinflated. The existence of restart costs (i.e. slower RTs on the first trial of a run, regardless of transition) supports his argument, although substantial switch costs remain in corrected calculations (e.g. comparing only trials that are the first in a run). A more important limitation is it cannot be determined whether the effect of increasing RSI is the result of active preparation or of the passive decay of the previous task-set state.

Probably for this reason, the most popular paradigm now is the task-cueing procedure in which each stimulus is preceded by a task cue instructing participants which task to perform; until the cue is presented, whether the task will change or repeat is unpredictable. In this procedure, the time available for preparation can be varied by manipulating the cue-stimulus interval (CSI). This can be done independently of the RSI; hence the time available for preparation can be varied independently of the time since the last response. This particular feature of the cueing paradigm has played an important role in distinguishing between different theories of task-set control, as will become apparent.

Other (less frequently) used paradigms include the intermittent instruction paradigm, in which participants continue to perform a task until a cue instructs a change of task (e.g. Gopher, 1996); and the voluntary switching paradigm, in which participants choose which task to perform on a trial to trial basis (e.g. Arrington & Logan, 2004; 2005).

The different paradigms reported above all have their advantages and disadvantages. Jersild’s list paradigm is still used on occasion, particularly in experiments investigating the contribution of linguistic representations to task switching performance (section 1.2.3). Here it performs a valuable role, as it can help determine the contribution of speech to ‘keeping track of the task sequence’, which is not required in the other paradigms. Because the task-cueing paradigm allows opportunity for preparation (CSI) to be varied independently of passive carry over from the previous trials (RSI) this is the paradigm best suited to the study of preparatory processes, and it was hence the paradigm of choice in this thesis.

### 1.1.2 Theories of the switch cost

The key observation in task switching experiments is that RTs are longer and error rates are greater on a trial requiring a change of task, compared to when the task repeats. This RT (and error) difference between switch and repeat trials is known as the "switch cost". Two distinct but not mutually exclusive theories were proposed in the mid 1990s to account for the switch cost: task-set inertia (TSI, Allport, Styles & Hsieh, 1994) and task-set reconfiguration (TSR, Rogers & Monsell, 1995). According to the TSR theory, the switch cost reflects the time it takes to reconfigure task-set on a switch trial. This is thought to involve processes such as retrieving the stimulus-response (S-R) rules (Mayr & Kliegl, 2000), goal shifting (e.g. Rubinstein, Meyer & Evans, 2001) and reallocating attention (Meiran, 2000). The processes involved in TSR are discussed in more detail in section 1.1.4.

Task-set inertia theory, on the other hand, assumes that on a task switch trial, activation from the previous (now irrelevant task) carries over and conflicts in some way with the operation of the current task-set. According to TSI then, the switch cost reflects the amount of time consumed by overcoming the persisting activation of the irrelevant task-set. An important difference between these theories is that whereas TSR theory assumes that active control processes are the root of the switch cost, TSI theory attributes the switch cost to passive interference effects instead. Each theory has been supported (and compromised) by particular phenomena often observed in task switching experiments. Some of these phenomena are described below (in a review that does not aim to be exhaustive but focuses on the findings most relevant to this thesis).

### 1.1.3 Task switching phenomena

*Reduction in switch cost with preparation* If the switch cost indeed reflects (at least in part) the time consumed by an activate reconfiguration process, then the switch cost should be reduced when participants are given the opportunity to prepare for an upcoming task switch. Consistent with this notion, Rogers and Monsell (1995) showed that increasing the response-stimulus interval (RSI) in an alternating runs paradigm substantially reduces the switch cost. Because the task sequence was predictable, preparation could be initiated immediately after the previous response. Monsell, Sumner and Waters (2003) have since shown that the reduction in switch cost (RISC) with preparation is in size similar for predictable and unpredictable switching, other things being equal (though whereas performance recovered to baseline on the first repeat trial for predictable switches, it took two or three trials to do so with unpredictable switches). Although finding a RISC effect with predictable switching is consistent with the notion of advance reconfiguration, increasing the RSI also allows for passive decay of the irrelevant task-set. For this reason, preparation effects are most commonly studied using the cueing paradigm, which allows for the manipulation of the preparation interval (CSI), whilst keeping time available for passive decay (RSI) constant (Meiran, 1996). Indeed, numerous studies have subsequently demonstrated that robust preparation effects are still found when RSI is held constant, and some (e.g. Meiran, 2000; Monsell & Mizon, 2006) have used several CSIs to trace out the RISC function, which is typically asymptotic by a CSI of around three quarters of a second. Though these findings emphasise the contribution of control processes to preparation effects, the cue also requires interpretation. Logan and Bundesen (2003; also see Mayr & Kliegl, 2003) have pointed out that many early task-cueing experiments confounded task and cue switches: whenever the task switched, the cue switched, too, suggesting that task switch costs may simply be cue switch costs. Indeed, this is what Logan and Bundesen (2003) have argued: they have attributed the switch cost in cueing paradigms to cue encoding processes. However, numerous subsequent studies that have used two cues per task have shown that task-switch costs remain regardless of cue-switch costs (Mayr & Kliegl, 2003; Schneider & Logan, 2007), again providing clear evidence for TSR. In order to ensure that a “pure” estimate of task switch costs was obtained, the experiments reported in this thesis always used two cues per task and the cue never repeated from one trial to the next.

Finally, it is worth briefly considering one result which questions the generalisability of the RISC. Altmann (2004) failed to find a RISC using a between-subjects design, leading him to argue that the RISC observed in within-subject designs is the result of strategic modulation. But findings from our lab (Elchlepp, Lavric, Mizon & Monsell, 2012; Monsell & Mizon, under revision) demonstrate that the RISC can be found in between-subject designs, provided that participants are sufficiently motivated. It may be that a participant is more likely to appreciate the benefit of advance preparation when they have experienced both short and long CSIs.

*Residual switch cost* Though switch costs are largely reduced with the opportunity for preparation, they are (almost) never eliminated: a so-called residual switch cost remains.

The existence of the residual switch cost suggests that there is an element of the switch cost which cannot be reduced by preparation. Naturally, any theory of the switch cost must also be able to explain the residual switch cost.

Residual switch costs are easily explained by TSI theory, according to which they reflect interference from the irrelevant task. It has perhaps been more difficult for TSR theories to explain the residual switch cost. Theories that attribute the residual switch cost to reconfiguration processes can be broadly classified into two classes: those that assume a structural limit (two-stage models of switching) and those that do not (e.g. de Jong, 2000). The first variant of the former type of account was proposed by Rogers and Monsell, who viewed the residual switch cost as "... the completion of a stagelike process of reconfiguration (...) that can be triggered only exogenously by the arrival of a stimulus suitably associated with the task" (p. 229). This distinction between exogenous and endogenous components of the switch cost is also central to Rubinstein et al.'s (2001) account of the switch cost, which differentiates between goal shifting (under endogenous control, can be performed in advance of stimulus) and rule activation (exogenously triggered, must await arrival of the stimulus). The authors provided support for this theory by manipulating aspects of the task switching paradigm that should selectively affect only one of those stages: specifically, task switch costs decreased with cueing (presumably this facilitates the goal shifting stage) and increased with rule complexity (which should increase the duration of the rule activation stage). These latter experiments will be discussed in more detail in the Introduction of Chapter 4. Although Rubinstein et al.'s (2001) assertion that rule activation must await stimulus presentation accounts for the residual switch cost, it is not obvious why this should be the case. Indeed, experiments reported in this thesis (Chapter 4) show that participants *do* activate S-R rules prior to stimulus onset.

A different two-stage model proposed by Meiran (2000) argued that task-set reconfiguration happens in two stages: shifting the attentional weights applied to task-relevant dimensions (e.g. colour versus form, or parity versus size or value), and reallocating the meaning of the response mappings (e.g. left key now means "small" rather than "odd"); the former could be done is advance of the stimulus, while the latter requires one performance of the task before it is established, and this is responsible for the switch cost.

De Jong's (2000) "failure-to-engage" theory provides one alternative to the two-stage model. According to the Jong (2000), the residual switch cost arises because preparation is successful on some trials, but completely fails on others. Two studies using RT distributions support this theory (de Jong, 2000; Nieuwenhuis & Monsell 2002): prepared switch trials were very similar to prepared repeat trials for fast RTs (indicative of successful preparation), but they are more similar to unprepared switch trials for slow RTs (indicative of completely unsuccessful preparation), and the RT distribution for prepared switch trials was well fit by a probabilistic mixture from unprepared switches and prepared repeats. But if this theory is correct and task-set preparation is not hindered by a structural limit, one might suppose that participants should be able to prepare successfully on (almost) every trial, provided that they are sufficiently motivated. Nieuwenhuis and Monsell (2002) repeated one of Rogers and Monsell's (1995) experiments with several changes designed to maximise incentive, but although their RT distributions were consistent with the de Jong model, they were able to increase the estimated preparation probability from 0.5 only to about .67. And Karayanidis, Provost, Brown, Paton and Heathcote (2011) found that a cue-locked switch positivity (an ERP component thought to reflect TSR, e.g. Lavric, Mizon & Monsell, 2008) varied gradually with an increasing switch cost, which is not consistent with the “all-or-none” approach taken by de Jong (2000). Could there be some other "trick" to achieving complete preparation? Verbruggen, Lieoofghe, Vandierendonck and Demanet (2007) claimed use of a very brief cue encouraged complete preparation (because participants had to "use it or lose it"), on the basis of the absence of a significant residual switch cost, but a small numerical switch cost did remain with preparation. And experiments in our own lab have not found that brief cues are especially effective in minimising the residual cost.

Whereas de Jong (2000) proposed that preparation succeeds or fails completely, Lien, Ruthruff, Remington and Johnston (2005) suggested preparation may be partially successful, but all-or-none for individual S-R rules. They argue that because of capacity limitations, only a subset of the complete set of S-R mappings can be prepared. Consistent with this proposal, Lien et al. (2005) had participants switch between tasks with three response categories mapped to a row of keys, and found that residual switch costs were essentially zero for the leftmost S-R mappings, and larger for the others. Although this result has not been replicated in other experiments with a similar S-R structure, (Monsell & Mizon, 2006; Lindsen & de Jong, 2010), the idea that the *completeness* of task-set preparation is subject to capacity constraints is also central in this thesis (Chapter 4), and will be returned to later.

*Congruency effect* Another index of interference between tasks can readily be observed in experiments on switching between two tasks which use *bivalent* stimuli, i.e. stimuli which afford a response in both tasks (such as coloured shapes in the case of a colour and shape task). Such experiments show response congruency effects: RTs and error rates are increased for incongruent stimuli (require a different response in both tasks) and congruent stimuli (require the same response in both tasks). The congruency effect reflects interference from the stimulus-response rules of the currently irrelevant task-set. Moreover, it is typically larger on switch trials, on which the current task-set is less well-established. Though congruency effects are normally attributed to competition at the level of stimulus-response rules, the finding that RTs are faster to univalent stimuli (which require a response in only one task) than congruent stimuli (which require the same response in both tasks) suggests there must also be some competition at the task-set level (e.g. Aron et al, 2004; Rogers & Monsell, 1995; Steinhauser & Hübner, 2007).

Can the task-set competition reflected by the congruency effect be reduced by advance preparation? Considering the robustness of the switch cost and its reduction with preparation the answer to this question is perhaps surprisingly unclear. Several studies have failed to find a reduction in congruency effect (RICE) with preparation (Allport et al., 1994; Meiran, 1996; Rogers & Monsell, 1995; see Kiesel et al., 2010, p. 864), yet others have found a RICE (e.g. Meiran, 2000; Monsell & Mizon, 2006, in some experiments). This discrepancy in findings could relate to the probability of a task switch (RICE was found when p-switch = 0.25 but not when p-switch = 0.5, Monsell & Mizon, 2006). Another possibility is that the presence/absence of a RICE is influenced by the nature of the tasks: one might expect interference from the irrelevant task-set to be smaller when the stimulus dimensions are spatially separable compared to when they are not – when stimulus dimensions are spatially separable, spatial attention can be used to filter out the irrelevant value.

*Asymmetric switch costs* When switching between colour naming and word reading, switch costs are larger when switching towards the more familiar word reading task (Allport et al., 1994; Yeung & Monsell, 2003). Similarly, in bilinguals switch costs are larger when switching from the second to the first language in a naming experiment (Meuter & Allport, 1999). This effect can be explained in terms of inhibition: when switching away from the more familiar task it requires more inhibition, which then later interferes when switching back to that task. However, other findings suggest that there is more to asymmetric switch costs than inhibition: for example, Yeung and Monsell

(2003) found that the switch cost asymmetry reversed when interference between task-sets was reduced (also see Bryck & Mayr, 2008).

*N-2 repetition effect* One other finding that has been taken as evidence for task-set inhibition is the n-2 repetition effect: RTs are slower for task A when it was recently switched away from (i.e. ABA) than when it was not (CBA) (Mayr & Keele, 2000). But this effect is mostly found when all (or most) trials are switch trials, and in the absence of an appropriate baseline (repeat trials) it is hard to determine to what extent inhibition contributes to the switch cost. Moreover, when repeat trials are included the n-2 repetition effect is often not found (e.g. Philipp & Koch, 2006, also see Chapter 2). Hence the contribution of inhibition to the switch cost remains far from unambiguous. Indeed, when examining the effects of task recency beyond a lag of 2 quite a different picture emerges: Ruthruff, Remington and Johnston (2001) found RTs to increase with increased lag, whereas Monsell et al. (2003) did not. According to Sumner and Ahmed (2006), the cause of this discrepancy was a difference in the requirement for control: lag effects should only be found when the need for control is minimal, as is the case when univalent stimuli are used (as in Ruthruff et al., 2001).

*Effects of position in run* In addition to examining effects of lag, it is also informative to consider the effect of position in run: is there any further acceleration (or, as it may turn out, slowing), beyond the second trial of a run (with the first trial being a task switch)? Monsell, Sumner and Waters (2003, p. 328) review a number of studies which have analysed their data as a function of position in run, with mixed results. Whereas some found RTs to decrease with position in run beyond the first repeat trial (Salthouse, Fristoe, McGuthry & Hambrick, 1998; Mayr, 2001; Meiran, Chorev & Sapir, 2000), others found no improvement (e.g. Rogers & Monsell, 1995; Keele & Rafal, 2000). Moreover, Altmann (2002; Altmann & Gray, 2002; 2008) have actually observed *within-run slowing,* though it is peculiar that such slowing has not been found in other experiments (e.g. Monsell et al., 2003; also see Lewandowsky, Oberauer & Brown, 2009) .

Monsell et al. (2003) argue that these discrepancies might reflect a difference between paradigms: particularly, the effect of position in run may depend on whether switching is predictable or unpredictable. Their results confirm this: performance only improves beyond the first repeat for random switching, not predictable switching. Monsell et al. (2003) attribute this to strategic modulations, arguing that in unpredictable switching, "… participants to some degree voluntarily attenuate or restrain the increment in readiness that would otherwise result from one performance of the task; they adopt a less extreme task-set bias" (p. 340). Although strategic modulation may explain position in run effects, other factors such as a capacity limit may also contribute to run length effects as will become apparent in this thesis (Chapter 4).

*Summary* When the considering the findings described above in light of theories of the switch cost, it is clear that there is evidence for task-set reconfiguration and task-set interference. The RISC observed with preparation in the cueing paradigm is the strongest evidence for TSR theory, and, provided that RSI is held constant, the RISC cannot be explained by TSI theory. The residual switch cost, on the other hand, is more easily explained by task-set interference. Such an integrative view is accepted by most (see Monsell, 2003; Vandierendonck, Liefooghe & Verbruggen, 2010). Even though most theories of task switching now acknowledge the contribution of reconfiguration processes, very few studies have explicitly addressed the processes that constitute TSR. As a result, very little is known about this process. Below I will review some of the processes which are typically thought to contribute to TSR.

### 1.1.4 What does task-set reconfiguration exist of?

As noted above, most theories of task-set control now assume that TSR contributes to the switch cost. But what does TSR consist of? Monsell (2003) proposed that TSR includes shifting attention, uploading/removing sets of S-R rules, retrieving goal states and adjusting response criteria. Others have mentioned adjusting attentional settings (Meiran, 2000), goal shifting (Rubinstein et al., 2001) and task-set retrieval (Mayr & Kliegl, 2000). Yet very few studies have directly investigated the processes that constitute TSR. There is, however, some evidence for a contribution of (1) task-set retrieval, (2) goal activation, (3) linguistic representations and (4) attentional selection to TSR, as reviewed below.

The contribution of task-set retrieval to task-set preparation is best illustrated by Mayr and Kliegl's (2000) LTM retrieval hypothesis, according to which the switch cost reflects the time it takes to upload the task rules from long term memory into working memory. They claimed their hypothesis was supported by a series of experiments which manipulated retrieval demand. In Experiment 1, Mayr and Kliegl (2000) employed an alternating runs paradigm which required participants to switch between a task requiring episodic retrieval, and a task requiring semantic retrieval. In the semantic, low-retrieval demand condition, participants switched between size and animosity decisions. In the "episodic", high-retrieval demand condition, participants classified stimuli based on pre-experimentally learned qualities (i.e. colour and location during the practice session). As the authors expected the retrieval demand to be higher in the episodic condition, switch costs were hypothesised to be increased for this task. Retrieval difficulty was manipulated within a block, as to enable a comparison between engaging and disengaging from task-sets with varying retrieval demands. The results confirmed their prediction: switch costs were higher for the episodic task. Experiment 3 provided further support to the LTM retrieval hypothesis: when the preparation interval was sufficiently long (allowing task-set retrieval to take place prior to stimulus onset) or when the task-cue explicitly displayed the S-R mappings (thus eliminating the need for LTM retrieval), the effect of retrieval demand was eliminated. These results show that task-set retrieval is an important part of TSR. However, although Mayr and Kliegl (2000) claimed to have manipulated retrieval demand, they actually compared *qualitatively* different kinds of retrieval ("episodic" versus semantic). A *quantitative* manipulation of task-set retrieval demands might shed further light on the process of retrieving or activating a task-set, and it was for this reason that we chose to manipulate the number of tasks in play (Chapter 2) or the number of S-R rules that specify a task-set (Chapter 4).

Prior to the activation or retrieval of task rules, some theorists (e.g. Rubinstein et al., 2001) distinguish a goal activation stage (“classify the colour of the stimulus”) which is separate from retrieving a task rule representation (“if the stimulus is red, press left, if it is green, press right”).[1] In connectionist models, goals are represented by task nodes (Gilbert & Shallice, 2000). Very little research has directly addressed the contribution of goal representation to task-set control, perhaps because it is assumed that without goal activation, task-set control would not be possible, at least with bivalent stimuli. For example, without activating the task goal, how would one know whether to respond to the colour of a stimulus, or to its shape? Experiments with bivalent stimuli have also shown that competition between tasks exists on a task-set level as well as on an S-R level: RTs are longer on congruent trials compared to neutral trials, presumably because there is competition at the task-set level for the former but not the latter (Rogers

---

[1] In the social/motivational psychology literature, a similar distinction is made between goal intentions (which describe the desired goal) and implementation intentions (which describe the behaviour required to achieve that goal, e.g. Sheeran, Webb & Gollwitzer, 2005).

& Monsell, 1995; Aaron, Monsell, Sahakian & Robbins, 2004; Steinhauser & Hübner, 2007). Furthermore, Rubinstein et al. (2001) found that explicit task cues reduced the switch cost (reflecting reduced goal activation time), whereas the switch cost increased with rule complexity (reflecting increased rule activation time). Effects of cue and rule complexity were additive, suggesting that the rule activation and goal shifting stage are separate. Finally, Dreisbach and colleagues have also shown that goal representations play an important role in task-set control. For example, Dreisbach and Haider (2008) required participants to classify eight words (items of clothes) superimposed on line drawings of related (clothes) or unrelated (animals facing left or right) items. Participants were either informed of a task rule (covers top versus bottom half of body) or learned the individual S-R mappings. Semantically related distractors resulted in compatibility effects in both groups (faster RTs for pictures of clothes compatible with the response afforded by the word), but compatibility effects for the semantically unrelated distractors were only found in the S-R group (faster RTs when the animal's spatial orientation was compatible with the response assignment). Dreisbach and Haider concluded that task goals serve the purpose of shielding performance from irrelevant information. Dreisbach and Wenke (2011) have subsequently shown that task-set shielding is relaxed on task switch trials (presumably this is necessary to allow for the required change of task). Chapter 4 of this thesis will show that this proposed vulnerability to interference on a switch trial is further modulated by the complexity of the task-set.

The role of goal representation in task-set control is arguably facilitated by linguistic self-instruction. A number of studies have found that switch costs increase when participants engage in irrelevant speech. One interpretation of this finding is that inner speech is used to retrieve the relevant task goal (Miyake, Emerson, Padilla & Ahn, 2004), though others have proposed alternative roles for linguistic representations to task-set control, such as keeping track of the task-sequence (e.g. Bryck & Mayr, 2005), or rehearsing the S-R mappings (Liefooghe, Vandierendonck, Muyllaert, Verbruggen & Vanneste, 2005). Chapter 5 further investigates the contribution of linguistic representations to task-set control.

Finally, shifting attention is often highlighted as a potential sub process of TSR (Monsell, 2003; Meiran, 2000). Very few studies have directly addressed the contribution of attentional selection to task-set preparation. Recent research from our lab (Longman, Lavric & Monsell, submitted) required participants to identify either a face or a letter displayed on its forehead, and measured eye fixations to investigate the contribution of attentional selection to the switch cost. The onset of fixations towards the relevant regions was delayed on switch trials and the number of fixations on the task-irrelevant region decreased as a function of preparation, suggesting that reorientation of attention plays an important role in TSR (also see Elchlepp, 2011).

Hence the evidence described above suggests that task-set retrieval, goal activation, linguistic self-instruction and attentional selection all contribute to task-set reconfiguration. The key aim of this thesis was to further our understanding of TSR, focusing in particular on the process by which the S-R rules constituting the relevant task-set are selected, or promoted into a limited-capacity *procedural working memory (PWM)* buffer. Compared to the task switching literature, much less is known about PWM. Furthermore, research on task-set control and research on PWM have often been conducted in isolation from one another, despite the fact that they are closely related. Indeed, experiments investigating the properties of PWM have important implications for theories of task-set control, as will soon become apparent.

## 1.2 Procedural working memory

Despite continuing disagreement about the origin of the switch cost, most theories of task switching require a process by which the new task-set becomes the most active.[2] Symbolic models have conceptualised this process of task-set promotion in terms of *retrieval*: the task-set is “loaded” into a PWM store. For example, according to Mayr and Kliegl’s (2000) LTM-retrieval hypothesis, the currently relevant task-set is “loaded” from LTM into working memory (the switch cost then reflects the time consumed by this process). Rubinstein et al.’s (2001) two-stage model is another example: their rule activation stage consists of “loading the next task’s rule into procedural working memory” (p. 771). Within such symbolic conceptions of task-set activation one can further distinguish between accounts that assume the whole task-set is retrieved as a “package” (e.g. de Jong, 2000), and accounts that assume task-sets can be partially retrieved (Lien et al., 2005).

Connectionist accounts of task switching have conceptualised this process of task-set promotion in terms of *activation*: For example, in the model developed by Gilbert and Shallice (2002), which simulates task switching in a Stroop environment,

---

[2] Except perhaps accounts of the switch cost such as Logan and Bundesen’s (2005) according to which the task switch cost as measured in the task-cueing paradigm is not the result of task-set switching.

input units (representing the colours in the colour pathway and the colour names in the word pathway) can activate output units, which in turn can be modulated by task-demand units receiving a top down control input. The most active task-set then, is determined by the task-demand unit with the highest activation.

In this way, both symbolic and connectionist models incorporate a representation of the most active task-set. Although throughout this thesis I will speak of "retrieving (or uploading) S-R rules into a PWM buffer" (hence borrowing terminology from symbolic models), this thesis does not aim to (nor is it able to) decide between the appropriateness of connectionist and symbolic models. Rather, the aim of the research reported here is to further characterise the process of task-set promotion which is inherent to both connectionist and symbolic models.

Even though the concept of a PWM is (implicitly) incorporated into most theories of task switching, its properties remain poorly specified. Oberauer's (2009) recent model of working memory is an exception. Based on empirical evidence, Oberauer has distinguished between procedural and declarative working memory and, further, between three levels of representation within each of these. The levels differ from each other both in terms of accessibility and in terms of capacity constraints (see Figure 1.1). As a result, this model yields some testable hypotheses. Some of these hypotheses are directly addressed in this thesis, hence Oberauer's (2009) model is described in some detail here.

First of all, Oberauer distinguishes between declarative and procedural working memory (DWM and PWM). Whereas the former holds information available for processing – operating upon – and the products of those operations, the latter enables processing through the implementation of sets of procedures (task-sets). Although the distinction between declarative and procedural WM is certainly not new (e.g. James, 1890; Milner, 1962; Anderson, 1983), it is not one that features in most models of working memory (Miyake & Shah, 1999).

According to Oberauer's model, both PWM and DWM each consist of three analogous subcomponents or "levels of representation". The first level of representation is the activated part of LTM, which holds a subset of representations which are activated above baseline/more than others. Oberauer (2009) does not explicitly assume a distinction between DWM and PWM at this level, but the two kinds of content are clearly quite different, and neuropsychological dissociations between procedural and declarative memory (e.g. Milner, 1962) would appear to support a distinction that is not just functional but anatomical. In the context of a task switching experiment, the activated part of LTM holds a representation of all task-sets currently in play. Important features of LTM are that it is thought to be capacity unlimited, and that representations held within it cannot directly control action. However, recently activated representations are held in semi-active state which allows easy retrieval when required. Moreover, ease of retrieval is determined by the level of activation: retrieval will take less time for highly activated (sets of) procedures (similar assumptions underlie ACT/ACT-R, Anderson, 1983, Anderson & Lebiere, 1998). In the context of task switching, it is reasonable to assume that this baseline activation reflects in part recency and frequency of usage. Because task-sets represented in the activated part of LTM are not subject to capacity constraints, Oberauer predicts that increasing the number of tasks among which a person is required to switch beyond two should not, per se, increase the switch cost.

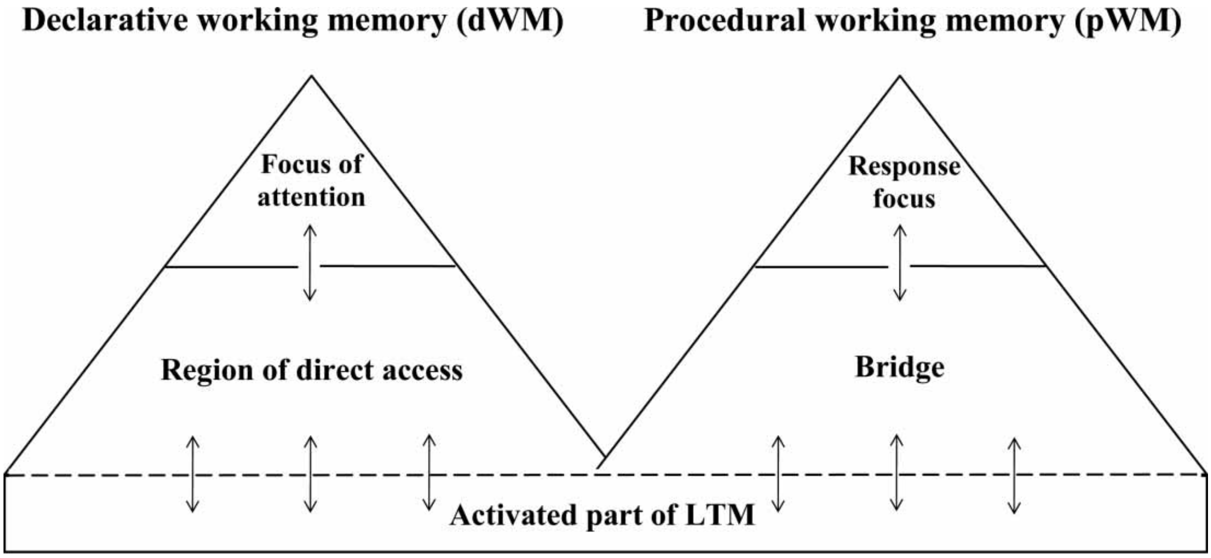


**Figure 1.1** Oberauer’s model of working memory (from da Silva Souza, Oberauer, Gade & Druey, in press). This model distinguishes between declarative working memory (left) and procedural working memory (right). Both DWM and PWM consist of three levels of representation. Their position within this figure indicates their accessibility and capacity limit: representations in the top level (focus of attention/response focus) are most accessible, but also subject to the most severe capacity constraint.

All procedures held in LTM compete for the control of action, but only the most active task-set controls action. In Oberauer’s model, the most active set of procedures constitutes the content of the “bridge” (i.e. bridge between stimulus and response). The task-set held in bridge controls behaviour, acting like a "prepared reflex" (Hommel, 1998). The bridge is also where new procedures are established and novel task-sets are formed by compilation from instructions (held in declarative working memory). Importantly, the bridge specifies not only which task-set controls action, but also which task-sets *do not* control action (by virtue of their absence from the bridge). This is particularly relevant when objects afford multiple (conflicting) actions, as is the case in task switching experiments which use bivalent stimuli. Although only the task-set held in the bridge can control action, response congruency effects suggests that recently activated competing task-sets (in the activated part of LTM) are able to activate response representations in the response selection process.

In contrast to LTM, the bridge is thought to have a limited capacity, normally restricted to a single task-set (especially when bivalent stimuli are used, as explained above). However, very little experimental evidence has directly addressed the capacity limit of the bridge: Oberauer (2010) assumes the capacity is a single-task set, but simultaneously acknowledges that the complexity of this task-set must surely also have an effect, as is apparent from Hick’s (1952) law (increased RTs with log of number of S-R mappings) – as will be discussed in more detail in Chapter 3. According to Oberauer, task switch costs reflect the time required to load the new task-set into (and remove the old task-set from) the bridge. Because the bridge has a limited capacity, it is possible to derive some predictions from Oberauer’s model about how this process is influenced by the complexity of the task-sets; this will be discussed in more detail in Chapter 4.

The third and most “narrow” component of Oberauer’s PWM is the response focus: task-sets held in the bridge typically consist of several S-R mappings, and Oberauer’s response focus represents a mechanism by which the correct response can be selected and executed. This addition is needed to account for experimental findings. Firstly, one of the most robust findings within the choice RT literature is that response repeats yield faster RTs than response switches (Bertelson, 1965). Secondly, PRP experiments also suggest that responses must be selected sequentially, and cannot be selected in parallel (e.g. Pashler, 1994, but see Oberauer & Kliegl, 2004 for an exception). One notable feature of both the response focus and the bridge is that representations held within these components will remain active until they are replaced (instead of being removed after implementation of a task-set or execution of a response).

Finally, one important assumption inherent to Oberauer's model of WM is that declarative and procedural WM operate according to analogous principles. Most of Oberauer's experiments have investigated the properties of DWM. In such experiments, participants are typically required to remember lists of items (words or digits). Both list size and the number of lists have been manipulated. For example, Oberauer has previously shown that the time required to retrieve an item from a list is increased when that list is different from the list retrieved on a previous trial – these so called "list switch costs" are analogous to task switch costs, and are thought to reflect the time required to upload the relevant list into a limited capacity component of DWM, the focus of attention. Very few of Oberauer's experiments directly concern PWM, hence many of Oberauer's predictions about the properties of PWM are derived from by analogy from findings for DWM.

To summarise, then, Oberauer's PWM consists of three components: the activated part of LTM (which does not have a strict capacity limit), the capacity-limited "bridge" and the response focus. Whereas task-switching theories' stipulation of what constitutes task-set reconfiguration (TSR) have often been vague, and thus rarely enable empirical tests of their validity, Oberauer's (2009) model of PWM yields some explicit predictions with regards to the processes that support task-set control. Two of these predictions are addressed by the research in this thesis: firstly, in a task switching experiment all potentially relevant task-sets are represented in the activated part of LTM. Because task-sets represented at this level are not subject to a capacity limit, Oberauer predicts that increasing the number of tasks in play should not affect the switch cost. This prediction is addressed in the experiments reported in Chapter 2, as summarised below in section 1.2.1. Secondly, because the currently relevant task-set is held in a capacity limited component of PWM (the "bridge"), Oberauer expects RT to increase as a function of the complexity of the currently operative task-set. An experiment reported in Chapter 3 tests this assumption. For the same reason, Oberauer expects switch costs to increase as a function of the complexity of the currently operative task-set. This prediction is addressed in Chapter 4, and the relevant evidence is discussed in more detail in section 1.2.2

The third question addressed in this thesis concerns the basic distinction between declarative and procedural WM. Although this distinction appears reasonable, the assumption that DWM and PWM operate according to analogous principles has not sufficiently been supported by experimental evidence. Among the predictions one might make is that if DWM and PWM are completely distinct, one would not expect linguistic representations in DWM to contribute directly to task-set control, though this claim has been made, or at least implied, by several theorists (as reviewed in section 1.2.3). Chapter 5 investigates this possibility.

I now summarise briefly how these three kinds of question are addressed in the thesis. Relevant literature will be reviewed in detail in the later chapters.

### 1.2.1 Retrieving a task-set into the capacity-limited component of PWM

The first part of this thesis is concerned with the process of selecting one task-set from among alternatives. As mentioned above, this process has been conceptualised both in terms of “retrieval” and in terms of “activation”. In Oberauer’s theory of procedural working memory, task-set retrieval is described as updating the contents of the limited capacity component of PWM, the “bridge”: when the task changes, the currently relevant task-set must be promoted from the activated part of LTM into the bridge, and the now irrelevant task-set must be replaced, or demoted into the activated part of LTM. The first question addressed in this thesis asks whether the time required for task-set retrieval increases as a function of the number of tasks the participant is required to switch among.

Intuition might suggest that it would. But because the activated part of LTM does not have a strict capacity limit, Oberauer (2009, 2010) predicts that the switch cost should not increase as a function of the number of task-sets in play. One reason this prediction appears counterintuitive is because there are several examples in the memory literature where the time it takes to retrieve an item from memory increases as a function of the number of competing items. Examples include the logarithmic increase in reaction time with the number of S-R mappings (Hick’s law), and the decrease in free recall performance with list length (Murdock, 1962).

Task switching studies that have manipulated the number of tasks participants are required to switch among have yielded mixed results: some found a greater switch cost with more tasks (Kray, Li & Lindenberger, 2002; Buchler, Hoyer & Cerella, 2008) others found a smaller switch cost with more tasks (Emerson & Miyake, 2003, Experiment 4), yet other found no difference (Allport & Styles, 1994; da Silva Souza, Oberauer, Gade & Druey, in press; Rubin & Meiran, 2005; Kessler & Meiran, 2010).

Furthermore, in many of these experiments the number of tasks manipulation was confounded with the recency and frequency with which a task was performed: the frequency (and hence average recency) of a task substantially decreases when the total number of tasks is increased. It is not unreasonable to assume the availability of a task-set is influenced by its recency and frequency of usage (cf. Anderson, 1976).

We therefore manipulated the number of tasks participants were required to switch among (three or five, within-subjects) whilst matching recency and frequency for a subset of those tasks. These experiments also varied CSI: if task-set retrieval is influenced by the number of tasks in play then the switch cost should be larger with more tasks, particularly when there is no opportunity for advance preparation. To pre-empt the results, these experiments demonstrated that retrieval of a task-set is not influenced by the number of alternatives when the recency and frequency of tasks was matched. For the other (recency and frequency confounded) tasks there was an apparent effect of the number of tasks. This led us to wonder whether task-set retrieval is special, or whether other effects (Hick's law in particular) are in fact also confounds of stimulus recency and frequency. This question was addressed in Experiment 3 (Chapter 3).

### 1.2.2 Selecting a stimulus-response mapping within the capacity-limited component of PWM

Following on from the question of whether the process of promoting a task-set from the activated part of LTM into the limited capacity component of PWM (Oberauer's "bridge") is influenced by the number tasks in play, the second part of this thesis (Chapters 3 and 4) is concerned with the properties of the limited capacity "bridge". In contrast to the capacity-limits of declarative working memory, very few experiments have directly investigated the capacity limits of PWM. Oberauer has presumed the content of the bridge is restricted to a single task-set, but he simultaneously questions the unit of measurement, and asks whether "there [is] a limit to how many stimulus-response bindings as part of one task set can be maintained at the same time in the bridge? Can a prepared reflex be established for task sets with any number of stimulus-response bindings?" (Oberauer, 2010, p. 289). These are the kinds of questions which motivated the second part of this thesis.

Some evidence for a limited capacity component of PWM has been provided by a phenomenon outside the task-switching domain: set-size effects in choice RT experiments (Hick's law) have been interpreted by Oberauer as key piece of evidence in support of his "bridge". The idea that the most active task-set is held in capacity-limited state is also apparent in Duncan's research on goal neglect (e.g. Duncan, Emslie, Williams, Johnson & Freer, 1996; Duncan, Parr, Woolgar, Thompson, Bright, Cox, Bishop & Nimmo-Smith, 2008). This research has demonstrated that, sometimes, participants fail to implement task instructions despite being able to reproduce the instructions (Duncan et al., 2008). As will be described in Chapter 4, this "goal neglect" can be modulated by the complexity of the task-set, suggesting that the problem may be a limited PWM capacity. Together, these findings point towards an account of task-set control in which the currently operative task-set is held in a severely capacity limited component of PWM.

To my knowledge, only two previous task switching studies have investigated the effect of task-set complexity on the switch cost (Rubinstein et al., 2001; Hübner, Kluwe, Luna-Rodriguez & Peters, 2004). Despite this, the idea of a structural limit of some sort as a determinant of the switch cost is not new: Lien et al. (2005, reviewed above) have also suggested that the residual switch cost reflects partial preparation. Although de Jong's (2000) original account opposed this assumption (assuming that preparation is all or none), he has more recently suggested that the "all-or-none" hypothesis applies to task-sets of a simpler description, and that "if the number of S-R rules within a task-set exceeds short-term memory capacity, advance preparation of the complete task-set is not feasible" (Lindsen & de Jong, 2010, p. 1222).

But the two experiments that have manipulated task-set complexity (described in more detail in Chapter 4) have yielded inconclusive results: Rubinstein et al. (2001) found that the mixing cost increased with task-set complexity, consistent with their theory that the switch cost reflects (in part) the time required for rule activation. However, Hübner et al. (2004, Experiment 4) found that the switch cost is larger when switching towards the less complex task.

Before moving on to a description of how complexity was manipulated in this thesis (Chapter 4), two final issues requires consideration. Firstly, it is worth noting that I do not propose that the capacity limit of the bridge is necessarily a structural limit (according to this view, the bridge might be a "box" with a limited number of "slots" available). Instead, it is possible that there is a limit on the processing capacity of PWM (for example, a capacity limit might arise because of interference between competing items). The research reported in this thesis does not aim to distinguish between these possibilities.

Secondly, whether or not the complexity of a task-set affects the switch cost depends not only on the capacity of the bridge, but also on whether task-set complexity affects the time required to upload (or remove) a task-set into (or from) the bridge. For example, the switch cost might be larger with more complex task-sets because it takes longer to retrieve such a task-set into the bridge, not because the capacity limit of the bridge is as such that is cannot maintain all S-R mappings simultaneously. If an effect of complexity on the switch cost is the result of increased retrieval time, then any such effect should be eliminated with plenty of opportunity for preparation (cf. Mayr & Kliegl, 2000, who demonstrated that task-set retrieval can occur prior to stimulus onset). However, based on an analogy with declarative working memory, Oberauer expects that the time required to upload a task-set into the bridge should not depend on its complexity.

In his experiments on DWM, Oberauer (2005, also see Wickens, Moody & Dow, 1981; Conway & Engle, 1995) presented participants with two lists. One red list appeared in the top half of the screen, and a blue list appeared in the bottom. Each list consisted of either one or three words (words were drawn randomly without replacement)[3]. These lists remained on the screen for (number of words * 1300) ms, followed by a 700 ms blank, after which a coloured square was presented. This square indicated the relevant list colour (thus serving as a cue), and after a CSI of 100 or 2000 ms, a probe word was presented. This probe was either positive (50%, i.e. from the relevant list, to be accepted) or negative (i.e. to be rejected, with 25% intrusion probes from the irrelevant list, and 25% new probes). After a response, feedback, a 200 ms blank, a second cue appeared (50% list switch); and after a variable (100 ms or 2000 ms) CSI, a second probe occurred.

The most important comparison with regards to the question at hand (is the duration of task-set retrieval influenced by the complexity of the task-set?) is whether the cost of switching to a new list is influenced by the size of that list. As it turns out, this is not the case. List switch costs are influenced by the size of the irrelevant list (i.e. the list to be switched away from) and this list switch cost is eliminated after 2s, suggesting that within that time, participants are able to completely remove the irrelevant list from the capacity limited part of DWM. But list switch costs are not

---

[3] One might ask whether one versus three is an appropriate comparison, as in the less complex condition a positive probe will always be the most recently encoded or rehearsed probe, whereas that is not the case in the more complex (three word) list condition.

influenced by the size of the list to be retrieved, suggesting it is retrieved as a chunk. This finding is somewhat puzzling: why are we able to retrieve lists in chunk form but must we remove them item by item? Oberauer speculates that perhaps one cannot delete an item from the capacity limited part of WM without first having identified it (personal communication).

To summarise, there are studies both within and outside the task switching literature that appear to suggest that the currently operative task-set is held in a capacity-limited component of procedural working memory. But the effect of task-set complexity on the switch cost remains unclear: Rubinstein et al. (2001) found larger switch costs for the more complex task-set whereas Hübner et al. (2004) found smaller switch costs. Experiments (4-6) reported in Chapter 4 therefore attempted to clarify the effect of task-set complexity on the switch cost. In Experiment 4 participants switched between identifying the most prevalent colour or letter within a string of letters. In the less complex condition there were two S-R mappings for the colour and letter task, in the more complex condition there were three. In Experiments 5 and 6 complexity was manipulated in a different manner: in both the more and the less complex condition eight stimulus names (per task) mapped onto two responses, but in the less complex condition these S-R mappings could be summarised using a global category rule, whereas in the more complex condition they could not.

In all three experiments we also varied CSI: if the bridge can accommodate even the more complex task-set, then any effect of complexity on the switch cost should be eliminated with a sufficient preparation interval. Alternatively, if task-sets are retrieved as chunks, but the bridge has a limited capacity that cannot accommodate all of the more complex task-set, then switch costs should remain larger for the more complex task-set at a long CSI.

### 1.2.3 The representation of S-R rules

The previous sections have summarised how this thesis addresses the process by which a task-set is uploaded into procedural working memory buffer (1.2.1), and the properties (notably the capacity limit) of that buffer (1.2.2). This final section is concerned with the nature of representation of S-R rules, and the distinction between declarative and procedural WM. Specifically, Experiments 7-9 (Chapter 5) asked whether linguistic (specifically phonological) representations contribute to task-set control.

Linguistic representations have long been thought to play an important role in action control (Vygotsky & Luria, 1994). A number of findings in the task switching literature are consistent with the idea (most explicitly stated by Goschke, 2000) that linguistic self-instruction plays a role in task-set control. For example, task switching experiments often find that task-set preparation is better with transparent verbal cues than with transparent pictorial cues (Elchlepp et al., 2012; Lavric, Mizon & Monsell, 2008; Monsell & Mizon, 2006). Furthermore, Mecklinger, von Cramon, Springer and Matthes-von Cramon (1999) found that increased switch costs in patients with left prefrontal damage were associated with language problems, suggesting that impaired verbal mediation at least contributes to the switching deficit.

These and other findings (reviewed in more detail in Chapter 5) suggest that *linguistic* representations might contribute to task-set control. The results of a number of studies that used the concurrent articulation of irrelevant material ("articulatory suppression" or AS) have led to more specific claims about the involvement of *phonological* representations in particular. The typical finding is that switch costs increase when participants are required to engage in AS whilst switching between tasks. To explain this result a number of different theories have been proposed with regards to the specific role of phonological representations in task switching, including a role for such representations in aiding LTM retrieval (Miyake et al., 2004), representing S-R rules (Liefooghe et al., 2005) and keeping track of the task sequence (Bryck & Mayr, 2005). Although the precise role of phonological representations in task-set preparation remains unclear, each theory assumes that such representations contribute *in some manner* to task switching performance.

Alternatively, others have proposed that the S-R rules are represented in a PWM buffer (e.g. Rubinstein et al., 2001; Oberauer, 2009). Procedural and declarative working memory have long been thought to be separate systems (James, 1890), with distinct neuroanatomical substrates (Milner, 1962). In Oberauer's account of WM, which also distinguishes between PWM and DWM, declarative representations cannot directly control action, though they might help to establish procedural representations in the bridge.

Oberauer's assertion (also see Kray, Eber & Karbach, 2008; Monsell, 2003) that declarative representations may aid the formation of procedures is consistent with theories of skill acquisition (Fitts, 1964; Anderson, 1982) and production systems like ACT-R (Anderson, Bothell, Byrne, Douglas, Lebiere & Qin, 2004), according to which the acquisition of procedural skill progresses through three different stages. Firstly, a declarative representation is generated. This representation is essentially “a set of facts about the skill” (Anderson, 1982, p. 370) that can be used to guide the appropriate behaviour. This stage is often accompanied by verbal mediation. Anderson (1982) argues this is the case because during this stage, the facts have to be maintained in working memory through the use of rehearsal. The second stage is a transitional stage and serves to transform the declarative representation into a set of procedures (Anderson, 1982). Any initial errors in understanding are corrected during this stage. Throughout this stage, the role of verbal mediation is reduced. Note that although declarative information might continue to exist (i.e. the procedural form does not necessarily replace the declarative form), it is now the procedural form that governs action (Anderson, 1980). During the final autonomous stage, the implementation of procedures becomes increasingly automatic, and hence reaction time decreases. Anderson (1980, p. 226) notes that there is no “sharp distinction” between the associative and the autonomous stage, but rather that the autonomous stage can be considered an “extension” of the associative stage. During this stage, the original declarative representation is sometimes lost, as is evident in the case of extremely well practiced and automatic procedures such as driving a car, where one is able to perform the skill without being able to describe exactly the rules required for accurate performance.

To summarise, the increased switch costs obtained under AS suggest that phonological representations contribute to task-set control, though the precise contribution of such representations remains unclear. Other accounts of task-set control propose that S-R rules are represented in a procedural WM buffer which is separate from declarative WM. Experiments 7-9 (Chapter 5) investigate whether phonological representations contribute to task-set control, or not. This was achieved through the manipulation of two effects that are thought to reflect phonological buffer involvement: the word length effect (Experiment 7) and the phonological similarity effect (Experiments 8 and 9). If phonological representations are used to represent the S-R rules of the most active task-set, then task switching performance should be influenced by the articulatory duration or the phonological similarity of the stimulus names. However, if non-linguistic representations in PWM control action, no such effects would be expected (except perhaps early in practice).

### 1.2.4 Overview of thesis

To summarise, this thesis investigated the process of promoting one of several task-sets in play into a most activate state. The experiments reported in Chapters 2, 4 and 5 used the reduction in switch cost (RISC) with preparation observed in the task-cueing paradigm as an index of this process. Through the manipulation of the number tasks in play (Chapter 2) and the complexity of the task-set (Chapters 3 and 4) these experiments aimed to distinguish between two levels of task-set representation: the level at which recently exercised (but currently irrelevant) task-sets are represented, and the level at which the S-R rules of the currently relevant task-set are maintained in a privileged state of activation. Experiments 7-9 (Chapter 5) manipulated the phonological properties of the stimulus terms, to investigate whether a phonological representation of S-R rules contributes to task-set control. The implications of these findings for theories of task-set control and procedural working memory are discussed in Chapter 6, which also offers some directions for future research.

# Chapter 2: Is it harder to switch among a larger set of tasks?

## Abstract

When stimuli afford multiple tasks, switching among them involves promoting one of several task-sets in play into a most-active state. This process, often conceptualised as retrieving task parameters and S-R rules into a procedural working memory buffer, is a likely source of the RT cost of a task-switch, especially when no time is available for task preparation before the stimulus. We report two task-cuing experiments that ask whether the time consumed by task-set retrieval increases with the number of task-sets in play. Previous studies varying the number of tasks have generally confounded number of tasks with their frequency and recency of usage. Our participants were required to switch among 3 or 5 tasks: orthogonal classifications of perceptual attributes of an object (Experiment 1) or of phonological/semantic attributes of a word (Experiment 2), with a 100 or 1300 ms cue-stimulus interval. For two of the tasks recency and frequency were identical in the 3- and 5-task conditions; for these tasks there was no effect of number of tasks on the switch cost. For the other tasks, there did appear to be a greater switch cost in the 5-task condition, with no time for preparation, attributable to effects of frequency/recency. Thus retrieval time for active task-sets is not influenced by the number of alternatives per se (unlike several other kinds of memory retrieval) but is influenced by recency or frequency of usage.

## Introduction

To perform any cognitive task, an appropriate organisation of the mind, or "task-set", is required, specifying what to attend to and when, and what procedures or rules to apply to the object(s) of attention to generate internal or external actions. Moreover, this task-set must temporarily dominate all the other potential task-sets that could be instantiated, regardless of their familiarity or recency of use (Norman & Shallice, 1986). This requires a process that promotes (activates, retrieves, loads) an existing task-set into *procedural working memory (PWM)*, or creates a new representation in PWM from instructions (Duncan, Emslie, Williams, Johnson & Freer, 1996; Duncan, Parr, Woolgar, Thompson, Bright, Cox, Bishop & Nimmo-Smith, 2008; Monsell, 1984, 1996, 2003; Oberauer, 2009). One way to engage the promotion process for existing task-sets is a task-switching experiment, in which participants are required to classify stimuli according to one of several sets of stimulus-response (S-R) rules on which they have been trained. For example, participants might have to classify a digit with left or right key-presses based on its parity (odd or even) on some trials, and on its numerical size (bigger or smaller than 5) on other trials. To specify the task, a task cue can be used (Shaffer, 1965; Sudevan & Taylor, 1987; Meiran, 1996), each digit being preceded by a cue indicating which task to perform. In this way, the task required may change or repeat from trial to trial, other things being kept equal. The time available for preparation can be manipulated by varying the cue-stimulus interval (CSI), while keeping the time since the previous response constant to control for any potential decay of task-set activation from the previous trial (Meiran, 1996).

As is well known (see Kiesel, Steinhauser, Wendt, Falkenstein, Jost, Philipp & Koch, 2010; Monsell, 2003; Vandierendonck, Liefooghe & Verbruggen, 2010 for reviews), there is a substantial task-switch cost: reaction times (RTs) and error rates are greater when the task changes, compared to when it repeats (e.g. Meiran, 1996; Rogers & Monsell, 1995; Allport, Styles & Hsieh, 1994). When the interval between the cue onset and the stimulus onset is increased to allow preparation for the upcoming task, the switch cost is found to decrease (e.g. Meiran, 1996; Monsell & Mizon, 2006). This reduction in switch cost (RISC effect) is thought by many to index active reconfiguration of task-set (e.g. Rogers & Monsell, 1995; Mayr & Kliegl, 2000; Rubinstein, Meyer & Evans, 2001; but see Schneider & Logan, 2005). Hypothesised sub-processes of task-set reconfiguration include reorienting attention to the relevant stimulus dimension, suppressing the previous (now irrelevant) task-set, and retrieving the currently relevant task-set rules into procedural working memory (Monsell, 2003). In this paper, we focus on the last of these, asking whether the difficulty of making the required task-set the most active depends on the number of potential task-sets in play.

Procedural working memory, and the process of promoting a task-set into a most-active state, are theoretical constructs in need of clarification. The task-switching literature contains both "retrieval" and "activation" conceptions of this promotion process. According to the former, sets of S-R rules (e.g. Mayr & Kliegl, 2000; Rubinstein, Meyer & Evans, 2001) or incomplete sets (Lien, Ruthruff, Remington & Johnston, 2005) are discretely retrieved or "loaded" into a PWM store. According to the second, illustrated by connectionist accounts of task-set control (e.g. Gilbert & Shallice, 2002; Brown, Reyolds & Braver, 2007) task-sets (represented in these models by task-nodes) have continuous degrees of activation which are determined by a combination of top-down control input, bottom-up activation from associated contexts and stimuli, and by their previous level of activation. In these accounts, there is no distinct PWM, only a most-active task-set.

One explicit conception of PWM that incorporates elements of both “retrieval” and “activation” accounts is found in Oberauer’s (2009) recent general account of working memory. This model distinguishes, first, between declarative working memory and procedural working memory (PWM). The former holds representations in an active state for operating upon, the latter makes processing possible by holding the sets of operations or S-R rules (task-sets) currently in play. According to Oberauer (2009), PWM consists of three components: the activated part of long-term memory (LTM), a component he calls the “bridge” (representing the S-R rules constituting the current task-set), and a response focus (essentially the response code most strongly activated by applying the current task-set rules to the contents of declarative working memory — i.e. the current perceptual or mental content). Procedures or task-sets that one has practiced are represented in the activated part of LTM. In a task-switching experiment, the several task-sets in play will be in an active state, competing for control over the bridge. The bridge represents the S-R rules currently governing performance. For effective performance, and specifically to avoid response ambiguity, it is assumed that the content of the bridge must be restricted to a single task-set.[4] Hence task-switching requires updating the contents of the bridge with an alternative task-set retrieved from

---

[4] Note that it is assumed that S-R rules in the active part of LTM can still prime responses even though they are not in the bridge – hence the response congruence effects observed in numerous paradigms, including task-switching.

the activated part of LTM. This retrieval process is assumed to be at least one source of task-switch costs.

If the retrieval of a task-set into PWM contributes to the cost of a task-switch, then if we make retrieval more difficult, this should increase the cost. To test this idea, Mayr and Kliegl (2000) manipulated the difficulty of retrieval. In their Experiment 1, participants were required to classify a set of centrally presented words, switching predictably either between a size and a living/nonliving judgement (“low retrieval demand”), or between classification of the words’ colour and screen location during a practice session (“high retrieval demand”). The switch cost was indeed found to be larger in the latter case. Furthermore, when the task cue explicitly displayed the S-R mappings (Experiment 3), the effect of retrieval demand was eliminated, suggesting that it is the time required to retrieve the S-R rules in particular that contributes to the switch cost.

Further articulation of the properties of PWM requires more data on what determines the difficulty of promoting one of several active task-sets to become the dominant task-set – retrieval into the PWM "bridge" in Oberauer's terms. In this paper we ask whether the number of currently active task-sets determines retrieval time (and hence switch costs). In a parallel paper (van 't Wout, Monsell & Lavric, in preparation) we consider the effect of the complexity of task-set on the switch cost and its implications for PWM capacity.

One argument for expecting the number of tasks in play to affect the difficulty of TSR is that there are a number of other cases of human memory retrieval where retrieval becomes more difficult the more candidates for retrieval are in contention. Examples include the decrease in free recall performance with list length (Murdock, 1962), the "fan" effect (Lewis & Anderson, 1976; Anderson, 1974), and Hick's law (Hick, 1952; see Schneider & Anderson, 2011, for a recent treatment in terms of memory retrieval). Of course these examples differ in various ways from retrieval of a task set. The fan effect – the increase in the time to retrieve a fact associated with a concept as a function of the number of facts learned about that concept – involves competition among memory elements associated with a single cue. Hick's law, the logarithmic increase in choice reaction time with the number of S-R mappings may in Oberauer's (2009) terms be seen as competition between the S-R rules represented *within* the PWM "bridge", whereas we are interested here in competition among sets of rules for *access to* the bridge. There is no a priori requirement that the set-size effects in these other forms of memory retrieval should apply to task-set activation. Indeed, Oberauer (2009) explicitly asserts that while the Hick's law set-size effect – RT increasing with number of S-R rules for a task-set held in the bridge – is expected because of competition for the response 'focus', there should be no effect of the number of activated task sets in LTM on retrieval time into the bridge.

Da Silva Souza, Oberauer, Gade and Druey (in press) recently acquired some evidence in support of this latter assumption. Participants were cued to perform (a subset of) three classification tasks (inner/outer, odd/even, high/low) on a digit, and the switch cost was not larger with 3 than with 2 tasks. Several other task-switching studies have also manipulated the number of tasks in play, with mixed results. Rubin and Meiran (2005; Experiment 2) used two perceptual tasks (colour and shape classification) and two spatial tasks (upper/lower and left/right classification); the two sets of tasks used separate responses. When participants switched between 2 or 3 tasks in a task-cuing paradigm (CSI was 100 ms or 1000 ms), the switch cost did not increase with the number of tasks. However, the comparison between the 2-task and 3-task conditions was imperfect because in the 3-task condition one of the tasks required separate responses. In another experiment (Kessler & Meiran, 2010), participants switched between a colour and a shape task on 75% of all trials. On the remaining 25% of trials participants either performed one or three other perceptual tasks (judging the size, thickness or fill of a clover shape), so that participants were essentially switching between 3 or 5 tasks. Because only the colour and shape task trials were included in the analysis, task frequency was matched between the two conditions. Again, no effect of number of tasks on the switch cost was found. But again, the additional tasks always required a set of different stimuli and responses, complicating the interpretation of this null result. Kray, Li and Lindenberger (2002) required participants to switch between 2, 3 or 4 word classification tasks in a task-cuing paradigm (CSI was 1000 ms); the switch cost was substantially larger with more tasks. However, this effect could be attributed almost entirely to faster RTs on repeat trials in the 4-task condition, which does not sit easily with a retrieval account. Buchler, Hoyer and Cerella (2008), had participants switch unpredictably between 2, 3 or 4 mathematical tasks, with no opportunity for advance preparation. Although the frequency with which each task was performed across set sizes was controlled task recency was not (see below). And the outcome was equivocal: the switch cost was nearly twice as large in the 4-task condition as in the 2-task condition, but this difference was not statistically reliable. Emerson and Miyake (2003; Experiment 4) required participants to switch between 2 or 3 mathematical tasks, using a list paradigm. The authors predicted that switching between 2 tasks might in fact be harder than switching between 3 tasks, based on the "backwards-inhibition" or "n-2 task repetition" effect (e.g. Arbuthnott & Frank, 2000; Mayr & Keele, 2000) – slower RTs for a task that was more recently switched away from (i.e. a longer RT on the last trial of an ABA than of a CBA task sequence). Emerson and Miyake (2003) found no effect of number of tasks on the RTs but the error switch cost was larger in the 2-task condition. The authors speculate that a counteracting 'fan effect' may have caused the lack of interaction in the RTs. Finally, Kleinsorge and Apitzsch (2012) found that switch costs were larger with 4 tasks than with 2 (particularly at short CSI) in cued switching but not predictable switching.

Taken together, the results are clearly inconsistent, with some experiments showing a larger switch cost with more tasks (Kray et al., 2002; Buchler et al., 2008, Kleinsorge & Apitzsch, 2012), others the opposite (Emerson & Miyake, 2003); yet others have found no effect of number of tasks on the switch cost (Buchler & Meiran, 2005; Meiran & Kessler, 2010; Da Silva Souza et al., in press). Moreover, the number of tasks manipulation in these studies is generally confounded with their recency and/or frequency. Simply increasing the number of tasks performed in a fixed number of trials necessarily decreases both the frequency and the average recency with which each task has been performed. Over a typical one-session experiment, performance on a given task generally improves with practice, so frequency cannot be ignored. And the ease with which a task-set is retrieved or activated is very likely to reflect the recency with which that task was last performed.

The purpose of the two task-cuing experiments reported in this paper was to manipulate the number of tasks in play while controlling for recency and frequency. To this end, participants were, in separate sessions, required to switch among either 3 tasks or 5 tasks, in a task-cuing paradigm. In Experiment 1, participants switched among perceptual classifications of objects; in Experiment 2, participants switched among semantic/phonological classifications of words. In both experiments, for two of these tasks – which we will refer to as the "probe" tasks – recency and frequency of task performance was matched across the 5-task and 3-task conditions. Both experiments also manipulated preparation interval. With a sufficiently long cue-stimulus interval and motivation to prepare, task-set retrieval should be accomplished prior to stimulus onset (Mayr & Kliegl, 2000). Hence we would expect any effect of number of tasks to be seen when the participant has no time to prepare (a CSI of 100 ms in our experiments), but not when they have ample time (a CSI of 1300 ms).

In both experiments, the tasks were orthogonal classifications of the stimulus objects, all using the same pair of left and right key-press responses. Another manifestation of competition among the task-sets currently in play is the effect of response congruency. In experiments with just two tasks mapped to the same response set, it is usually found that congruent stimuli (mapped to the same response by both tasks) are responded to more rapidly and accurately than incongruent stimuli (mapped to different responses), and that this congruence effect is amplified by a task switch (see Kiesel et al., 2010 for review). This indicates that the S-R rules comprising the irrelevant task-set are not completely inactive, and that they are more active if they were used on the previous trial. With more than two tasks mapped to the same response set, response congruency becomes a more complex variable: all, some or none of the currently irrelevant stimulus attributes can be mapped to a competing response. Meiran and Kessler (2010), described above, found no effect of number of tasks on the effect of response congruency, but their “additional tasks” involved separate stimuli and required different responses. In contrast to Meiran and Kessler (2010), in the experiments reported here each stimulus afforded all (3 or 5) tasks, which allowed us to compare response congruency effects as a function of degree of congruence (the number of dimensions specifying the same response) and the number of tasks they afforded, all other things being equal. To anticipate, this analysis was more revealing in the second experiment, in which a much larger response congruence effect was obtained.

## Experiment 1

On each trial, following a task cue, participants classified one of a set of 32 visual objects (see Figure 2.1 for examples). The set was designed so that each object afforded a binary classification on 5 orthogonal dimensions. In one session, the participant could be cued to perform any of the 5 corresponding classification tasks. In the other session, participants performed only 3 out of 5 tasks. Different sets of stimuli (32 creatures, and 32 trees) were used in the two sessions, so that the number of tasks (3 or 5) could be manipulated within subjects, without performance in the second session being affected by exposure in the previous session to a sub- or super-set of the same classifications. For two of the tasks in each session, which we refer to as the *probe* tasks, recency and frequency were matched across sessions, to allow an estimate of the effect of number of tasks in play uncontaminated by differences in the recency and frequency with which the current or previous task had been performed. To test our expectation that any effect of the number of tasks in play would be on task-set preparation rather than the execution of a prepared task-set, we compared blocks of trials in which the CSI was short (100 ms) to blocks in which the CSI was long (1300 ms), and encouraged preparation. To control for any carryover of task-set activation or competition from the previous trial, the response-stimulus interval (RSI) was the same with both CSIs, and we limited the critical analysis to two-trial sequences on which both tasks were probe tasks.

## Method

*Participants* Twenty four participants (aged between 18 and 22 [M = 19.5]; 4 male and 20 female) provided informed consent before taking part in two sessions on consecutive days. All participants were paid between £12.80 and £16.00, depending on the speed and accuracy of their performance.

| | | | **responses** | |
|---|---|---|---|---|
| **stimulus set** | **tasks** | **auditory cues** | **left** | **right** |
| creature stimuli | pattern | "pattern" and "texture" | dots | stripes |
| | head | "head" and "face" | big | small |
| | body | "body" and "torso" | round | rectangular |
| | legs | "legs" and "limbs" | long | short |
| | tail | "tail" and "rear" | down | up |
| tree stimuli | trunk | "trunk" and "stem" | narrow | wide |
| | number | "number" and "count" | one | three |
| | ground | "ground" and "base" | flat | grass |
| | leaves | "leaves" and "branches" | vertical | horizontal |
| | fruit | "fruit" and "crop" | pear | triangular |

**Table 2.1 (Above)** Tasks, task-cues and response assignments for the 2 sets of stimuli (creatures and trees) used in Experiment 1.

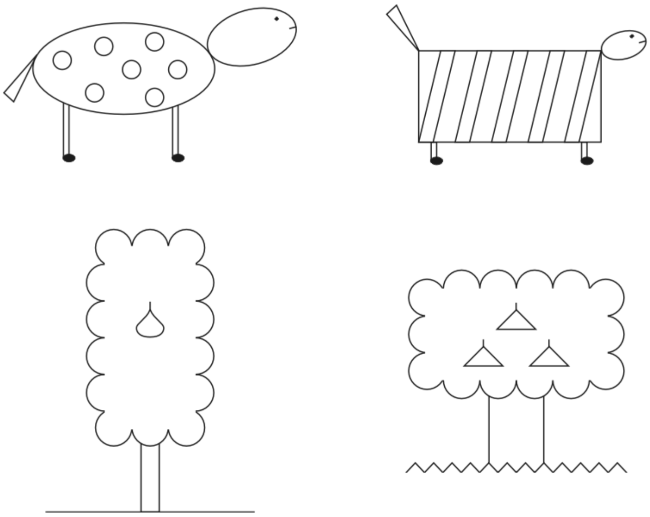


**Figure 2.1 (Left)** Examples of stimuli used in Experiment 1. The stimuli on the left require a left response in all tasks, whilst the stimuli on the right require a right response in all tasks.

*Stimuli and cues* The stimuli were schematic line drawings (see Figure 2.1 for examples) of imaginary animals or trees. The set of animal stimuli was composed of the 32 combinations of binary values of five attributes: body pattern (dots or stripes); head (big or small); legs (long or short); tail (up or down) and body (square or round). The set of tree stimuli was similarly composed of the 32 combinations of: trunk (narrow or wide); number of fruit (1 or 3); ground (grassy or flat); orientation of leaves (horizontal or vertical) and type of fruit (pear-shaped or triangle-shaped).

For each set, two auditory words could be used to cue each of these five classifications. The cue always changed from one trial to the next, to avoid confounding the effects of cue- and task-repetition (Logan & Bundesen, 2003; Monsell & Mizon, 2006). The two cue words used for each attribute are shown in Table 2.1, as are the assignments of the two values of that attribute to left and right response keys.

*Design and procedure* Participants classified the tree stimuli in one session and the creature stimuli in the other. Half the participants performed the 5-task condition with the tree stimuli, and the 3-task condition with the creature stimuli; the assignment was reversed for the other half. Within each group of participants, half performed the 3-task condition in the first session and half in the 5-task condition.

On each trial, an auditory cue word (duration 350 ms) was presented, indicating which classification task to perform. In the long CSI blocks, the trial began with a 500 ms blank screen, followed by presentation of a fixation dot: 100 ms later the cue word was presented, and the stimulus was displayed 1300 ms after the cue's onset, replacing the fixation dot. In the short CSI blocks, the cue onset was 1300 ms after the onset of the fixation dot, and the stimulus was presented just 100 ms after the onset of the cue word. (Hence the response-stimulus interval – and hence the recency of the previous task performance – was the same in long and short CSI blocks.) The stimulus was displayed until a response was detected. Responses were made using the left or right arrow key of a PC keyboard, pressed by the index finger of the left or right hand. If the wrong key was pressed, “Error!” was displayed on the screen for an extra 1000 ms. Otherwise a new trial began immediately.

After practice (see below), a session consisted of 594 trials divided into 18 blocks, each of 1+32 trials. (The first trial of a block was a warm-up trial needed to make the second trial a task-switch or -repeat, but was not otherwise included in the design or analysis.) Per session, each of the 32 stimuli occurred once for each of the 9 possible task-task transitions for each of the 2 CSIs. Long and short CSI blocks alternated, with the starting CSI counterbalanced between subjects. At the end of each block, participants were presented with a score (the mean correct RT in ms divided by 10 + 5 for each error). When participants improved on this score in a subsequent block of the same CSI, a bonus (£0.20) was awarded.

Of the five tasks of each set, two were chosen to serve as *probe* tasks matched for recency and frequency across the 3-task and 5-task conditions: the pattern and head classification tasks (for the animal stimuli) and the number and trunk classification tasks (for the tree stimuli). The two tasks of each pair were approximately matched for difficulty based on pilot data. Which of the other 3 tasks served as the non-probe task in the 3-task condition was varied over participants so that each was used equally often. New trial sequences were generated for each participant, according to the following principle. In order to match the recency and frequency of the probe tasks, every participant ($P_1$) was yoked to another participant ($P_2$). First, a 3-task sequence containing a random sequence of equal numbers of trials with tasks A and B (probe tasks) and C (non-probe task) was generated for participant $P_1$. From this a 5-task sequence was created for participant $P_2$ by randomly replacing all instances of task C with equal numbers of instances of task C, D and E.[5] As a result, probe task trials were in exactly the same positions in the trial sequence for the 3-task and the 5-task participant with this stimulus set. Consequently, for AA and BB task repetitions and AB and BA task switches, the comparison between the 3-task and 5-task condition was unconfounded by effects of the frequency and recency of the task used on either the current or previous trial. (Recency/frequency might affect not just the difficulty of retrieving the next task-set but also the difficulty of suppressing the previous task-set.)

Before the experimental blocks of each session, there were practice blocks. First there was one block of 32 practice trials per task for each of the three or five tasks to be performed in that session. During these single task blocks, the two auditory cues for that task were also presented, alternating from trial to trial, with a CSI of 1300 ms. Half the participants practiced the tasks in one order, and the other half in the reverse order. These 160 (or 96 in the 3-task condition) practice trials were followed by 2 additional practice blocks (one of 20 long CSI trials, the other of 20 short CSI trials) which required participants to switch between tasks in the same way as in the experimental

---

[5] As an unavoidable corollary of this, the proportion of switch trials was slightly higher in the 5-task condition (74.3% in Exp 1; 74.1% in Exp 2) than in the 3-condition (66.7% in Exp 1 and 2). However, only probe task trials followed by probe task trials (AA, BB, AB and BA transitions) were analysed for the frequency and recency matched comparison.

blocks. In total, this practice phase lasted approximately 15 minutes, and the experimental blocks about 45 minutes.

## Results and Discussion

Trials following an error and correct responses with very short (< 200 ms) or very long (> 3000) reaction times (RTs) were excluded from analysis (0.02% of the correct responses).

In each of the analyses described below, task assignment was entered as a between subjects variable with two levels (5 trees and 3 creatures versus 5 creatures and 3 trees), to remove any variance attributable to differences in performance between the tasks per se from the error term. Because we were not interested in the effects of task per se, interactions with task assignment are not reported here.

*Effect of number of tasks for the probe tasks* Mean correct RTs and error rates for two trial sequences involving only the probe tasks (i.e. controlled for recency and frequency of the present and previous task) are shown in Figure 2.2 as a function of number of tasks and CSI. Switch costs are shown in Table 2.2. A 2 (3 or 5 tasks) x 2 (long or short CSI) x 2 (switch or repeat) repeated measures ANOVA was conducted on the mean correct RTs and error rates, averaging over the two tasks.

Figure 2.2 shows that there was no detectable effect of number of tasks on RT, $F<1$. In fact, overall mean RT was identical in the 5 and 3-task conditions (531 ms). There were reliable main effects of a task switch, $F(1,22) = 53.76$, $p < 0.001$, and CSI, $F(1,22) = 573.41$, $p < 0.001$. These interacted so that the switch cost reduced from 51 ms at the short CSI to 22 ms at the long CSI, a RISC effect of 29 ms or 56%, $F(1,22) = 38.76$, $p < 0.001$. This switch cost and its reduction with preparation were very similar in the 5-task condition (from 50 ms to 24 ms; a RISC of 26 ms) and in the 3-task condition (from 52 ms to 20 ms; a RISC of 32 ms), $F<1$ for the 2- and 3-way interaction. Error rates showed a reliable switch cost of 1.2%, $F(1,22) = 13.89$, $p = 0.001$, and non-significant CSI effect of 0.7%, $F(1,22) = 2.89$, $p = 0.103$. The error switch cost reduced from 2.1% in the short CSI to 0.2% in the long CSI, $F(1,22) = 7.74$, $p = 0.011$. The error rate ANOVA showed no reliable interactions involving the number of tasks, $Fs < 1.13$, but there was a near-reliable main effect, with a greater error rate in the 3-task (3.5%) than in the 5-task condition (2.5%), $F(1,21) = 3.88$, $p = 0.062$.

Hence, when recency and frequency are controlled there is no evidence that retrieving the appropriate task-set is harder when there are more tasks. If anything, there were slightly fewer errors with 5 tasks, but this did not interact with switching or preparation.

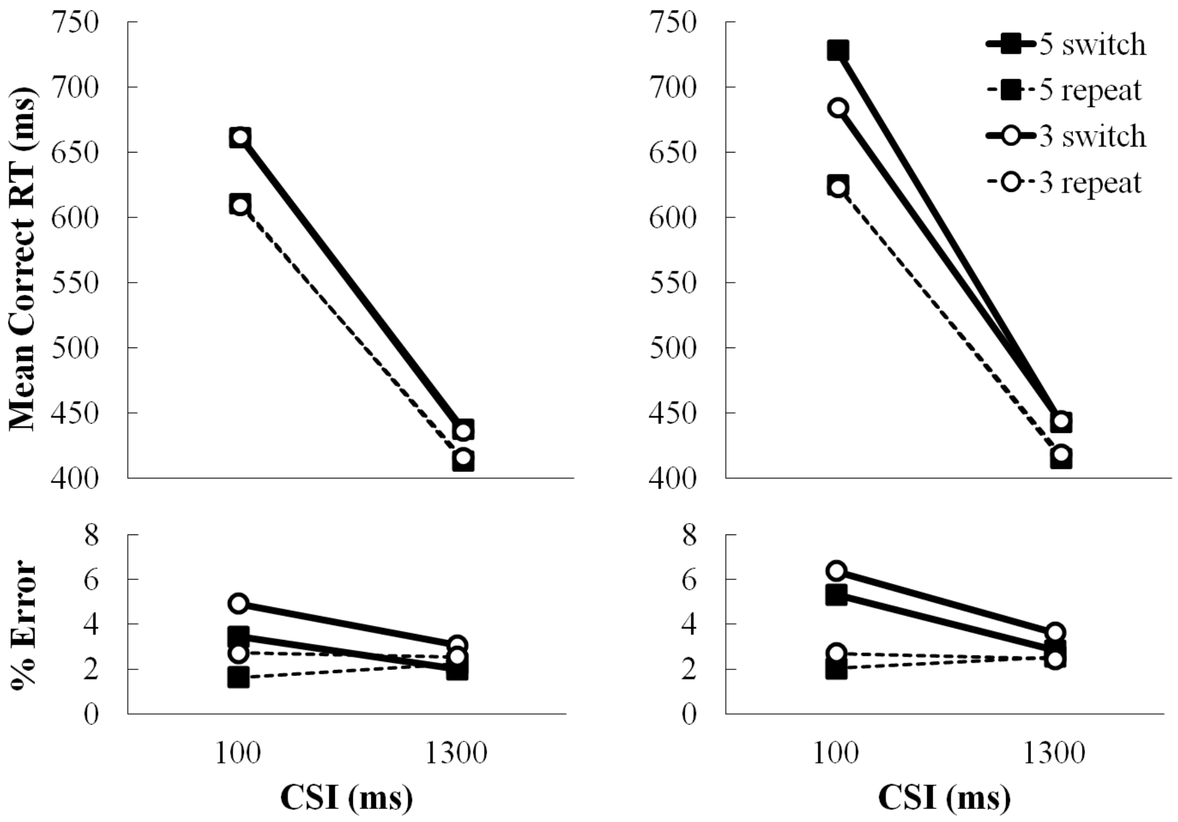


**Figure 2.2** Mean correct RTs (top) and % error (bottom) data in Experiment 1 with three and five tasks, on switch and repeat trials, as a function of CSI; plotted separately for probe task trials (left) and all task trials (right).

| | | **Probe tasks** | | **All tasks** | |
|---|---|---|---|---|---|
| | | **5 tasks** | **3 tasks** | **5 tasks** | **3 tasks** |
| **100 ms CSI** | **repeat** | 611 (1.7) | 610 (2.8) | 625 (2.0) | 624 (2.7) |
| | **switch** | 662 (3.6) | 662 (5.3) | 729 (5.3) | 685 (6.4) |
| | **switch cost** | 51 (1.9) | 52 (2.5) | 103 (3.3) | 61 (3.7) |
| **1300 ms CSI** | **repeat** | 414 (2.4) | 417 (2.6) | 416 (2.5) | 420 (2.5) |
| | **switch** | 438 (2.1) | 437 (3.3) | 444 (2.8) | 445 (3.6) |
| | **switch cost** | 24 (-0.3) | 20 (0.8) | 27 (0.3) | 25 (1.2) |
| | **RISC** | 26 (2.2) | 32 (1.8) | 76 (3.0) | 36 (2.5) |

**Table 2.2** Mean correct RTs (ms), with the % error data between brackets, for the different conditions in Experiment 1.

*Effect of number of tasks for all trials* The above analysis was restricted to trial sequences controlled for task recency and frequency. To explore the importance of this control, we repeated the analysis, now pooling over all trials for each combination of CSI, switch/repeat and number of tasks (see Figure 2.2), so that the analysis was no longer restricted to the 4 in 9 trials on which a probe task trial followed a probe task trial. Now the outcome looks quite different: for task-switch trials at the short CSI, RT was longer for 5 tasks (729 ms) than for 3 tasks (685 ms). The relevant interactions were reliable: switch costs were larger in the 5-task (65 ms) than in the 3-task (43 ms) condition, F(1,22) = 19.27, $p < 0.001$; the effect of CSI was greater in the 5-task (247 ms) than in the 3-task condition (222 ms), F(1,22) = 10.09, p = 0.004; and the switch cost was reduced to a greater extent, F(1,22) = 23.64, $p < 0.001$, in the 5-task condition (from 103 ms to 27 ms; a RISC of 76 ms) than in the 3-task condition (from 61 to 25 ms, a RISC of 36 ms).

Consistent with the RT analysis, the reduction in error rate with preparation was slightly larger in the 5-task (3%) than in the 3-task condition (2.5%), though not reliably so, $F<1$. Error rates also showed a marginally reliable main effect of number of tasks in the opposite direction to RTs, — slightly more errors with 3 tasks (3.8%) than with 5 (3.2%), F(1,22) = 4.07, p = 0.056.

This analysis suggests that, had we not controlled for frequency and recency, we might have mistakenly interpreted the longer RT for unprepared switch trials in the 5-task than in the 3-task condition as evidence that task-set retrieval is harder the more tasks there are in play. They also suggest that performance is in part determined by task recency and/or frequency, especially on switch trials

*Effects of task recency* The effect of number of tasks found in the uncontrolled data (right panel of Figure 2.2), suggests that task recency (the lag in trials since the previous performance of the same task) and/or frequency may (in part) determine performance. To examine the effects of recency more directly, we sorted the data for each number of tasks condition into three categories according to the lag since the last trial on which the same task had been cued: 2-3 trials back, 4-5 trials back, and 6-7 trials back, separately for short and long CSI trials, pooling over task. (For longer lags, too few trials were available.) Figure 2.3 shows this analysis restricted to probe-probe task trial sequences, and Figure 2.3 to an analysis pooling over all tasks. Only effects of and interactions with the linear component of the effect of recency are reported below.

If the time consumed by task-set retrieval is influenced by the task's recency of usage, RTs should increase with lag specifically on unprepared (short CSI) switch trials. The probe task trials, for which task frequency is unconfounded with number of tasks, should provide a "pure" estimate of the effect of recency. Indeed the data in Figure 2.2 show just such a small recency effect in the short CSI condition (slope 4 ms per intervening trial), and none in the long CSI condition (slope -4 ms); the slopes differed reliably, F(1,22) = 5.56, p = 0.028, though separate ANOVAs showed that the main effect of lag was not reliable in either the long CSI, F(1,22) = 3.20, p = 0.087, or the short CSI, F(1,22) = 2.83, p = 0.106, separately. No interaction with number of tasks was significant, F<1. For the % error data, there were no reliable effects or interactions involving task recency (F<1 except for recency x number for which approached significance, F(1,22) = 4.06, p = 0.056).

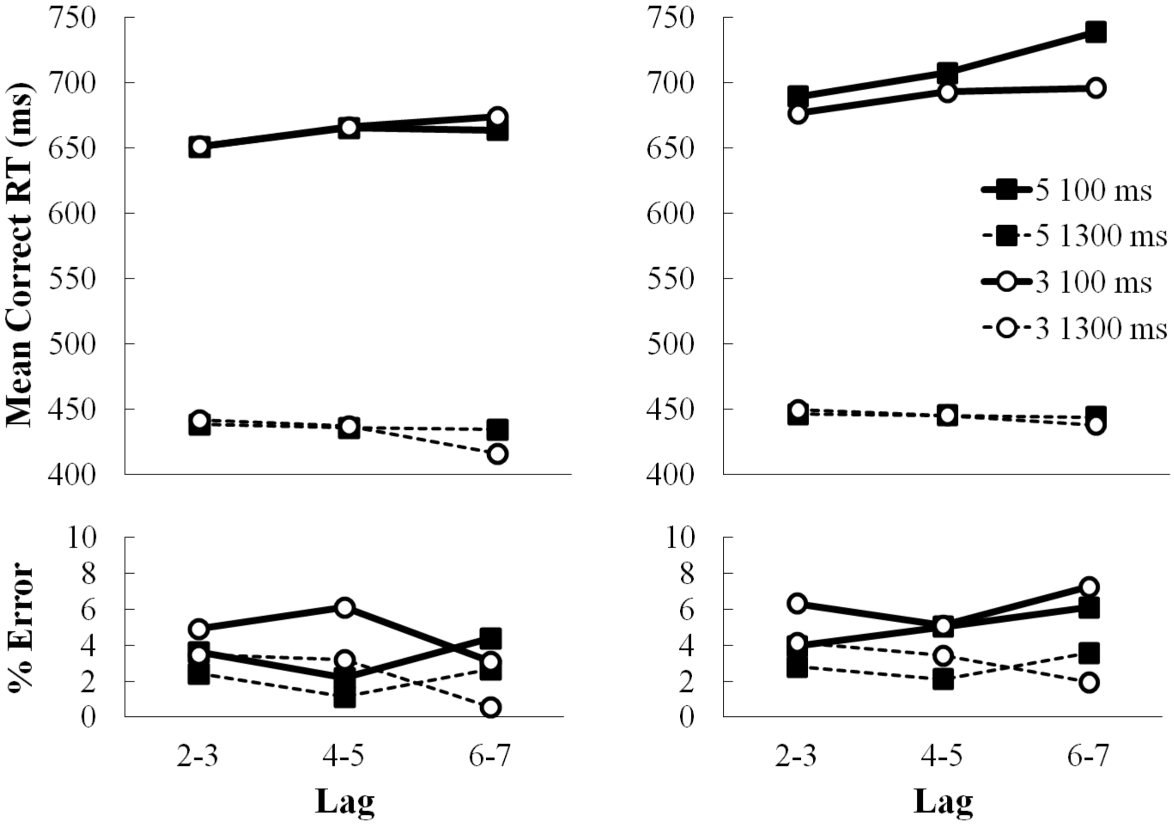


**Figure 2.3** Mean correct RTs (top) and % error (bottom) data in Experiment 1 with three and five tasks, on long and short CSI trials, as a function of lag; plotted separately for probe task trials (left) and all task trials (right).

The data for the uncontrolled tasks in Figure 2.3 are similar: the main effect of task recency was reliable, F(1,22) = 5.73, p = 0.026, but as for the probe task trials, this recency effect originated entirely from short CSI blocks (slope 9 ms per intervening trial), and was slightly negative in long CSI blocks (slope -2 ms), F(1,22) = 10.97, p = 0.003. However, unlike the controlled data, RTs for unprepared trials appear longer, and the recency slope somewhat greater, for the 5- than for the 3-task condition. As the controlled data indicate there is no real effect of number of tasks per se, we can only attribute this additional difference to the lower average frequency with which some tasks are performed. The recency slope was larger in the 5-task condition (6 ms) than the 3-task condition (1 ms), F(1,22) = 8.04, p = 0.01. This difference in recency slope between the 5- and 3-task conditions was larger in the short CSI (slopes 12 ms and 5 ms, respectively) than in the long CSI (-1 ms and -3 ms), though the 3-way interaction was not reliable, F(1,22) = 1.41, p = 0.247. However, separate 2 (3 vs. 5) by 3 (lag) ANOVAs for each CSI revealed a significant interaction between number of tasks and lag in the short CSI, F(1,22) = 5.15, p = 0.033, but not in the long CSI, F (1,22) = 1 (ns). The error rate analysis yielded similar trends, i.e. larger recency slopes for the 5-task (0.4%) than for the 3-task (-0.2%) condition, F(1,22) = 2.95, p = 0.1; and a larger recency slope in short CSI (0.4%) than long CSI blocks (-0.2%), F(1,22) = 3.63, p = 0.07. For the linear trend of recency and its other interactions, F<1.

These analyses provide some evidence that the difficulty of reactivating a task-set (but not of performing it, once prepared) is modulated by the recency with which it was last performed and, additionally, by the frequency with which it is performed, but not by the number of active tasks in play.

*N – 2 Repetition effect* A one-way ANOVA (restricted to switch trials from the 3-task session) comparing trials on which participants performed a task they had more (ABA) or less (ABC) recently switched away from found no evidence of an n-2 repetition effect. Participants were faster on ABA (560 ms) than ABC (566 ms) sequences, F(1,23) = 1.11 (ns), but they made slightly more errors on ABA (5.3%) than ABC sequences (4.3%), F(1,23) = 3.31 (ns).

*Effects of response congruence* In the 5-task condition, there were 5 possible levels of response congruence (defined by the number of irrelevant attributes that are mapped to the same/opposite response), ranging from completely congruent (all four irrelevant attributes mapped to the same response) to completely incongruent (all four irrelevant attributes mapped to the opposite response). In the 3-task condition, there were only 3 levels of congruency (2, 1 or 0 irrelevant attributes mapped to the same response). As explained in the introduction, the experiment provides an opportunity to ask whether and how the effect of congruence depends on the proportion of attributes that are congruent/incongruent. Is there more detectable conflict when four irrelevant attributes are mapped to the wrong response than when only two of four are? In order to compare the congruency effect with 3 and 5 tasks independent of the recency and frequency of the tasks performed on the current and immediately preceding trial the analysis was restricted to probe-probe sequences.

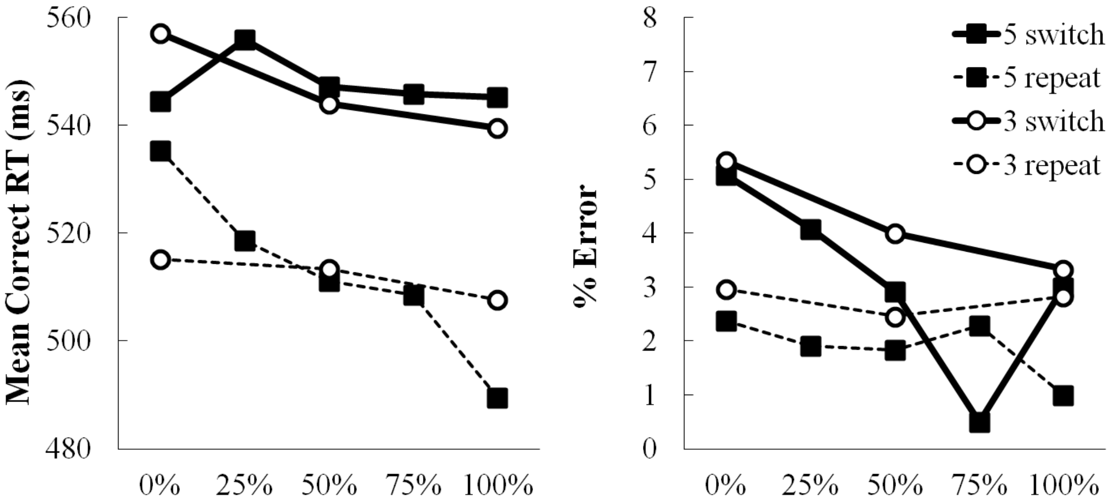


**Figure 2.4** Mean correct RTs (left) and % error (right) data in Experiment 1 with three and five tasks, on switch and repeat trials, as a function of response congruence.

Figure 2.4 shows the effect of degree of congruence for each combination of switch/repeat and number of tasks, with the abscissa representing the proportion of attributes that are incongruent (thus aligning the extreme values for the 3- and 5-task conditions). Allowing for noise, the data suggest a graded effect of degrees of response congruence, reasonably well described by regression lines fitted to these functions. An ANOVA on the 3-tasks RTs showed a reliable effect of degree of congruence, F(2,44) = 3.72, p = 0.032[6], with a significant linear trend: F(1,22) = 6.27, p = 0.02, and no quadratic trend, F<1. An ANOVA on the 5-tasks RTs showed a reliable effect of degree of congruence, F(4,88) = 3.14, p = 0.03, with a reliable linear trend: F(1,22) = 10.14, p = 0.004, and no other reliable trend components, F<1. An ANOVA on the 3-task error

---

[6] Huyn-Feldt corrected

rates found no reliable main effect of congruency, F(2,44) = 1.23 , p = 0.303 (nor any linear trend). An ANOVA on the 5-task error rates yielded no reliable main effect of congruency, F(4,88) = 2.13, p = 0.083, but the linear trend was significant, F(1,22) = 5.57, p = 0.028.

To compare these effects between the 3- and 5- task conditions, we express the congruence effect as what we will call the "congruence slope": the difference between the effect of 0% congruent and 100% congruent estimated from the regression line. For RTs the slope was slightly larger for 5 tasks (22 ms) than for 3 tasks (13 ms), but this difference was not significant, F(1,22) = 1.13. Unusually, the congruence effect was slightly larger on repeat (24 ms) than on switch (-11 ms) trials, although the slope difference was not reliable, F(1,22) = 2, p = 0.171). The interaction between effects of number and switch approached significance, F(1,22) = 3.6, p = 0.071, reflecting a larger congruency slope on repeat trials for the 5-task (41 ms) than for the 3-task condition (8 ms), but not on switch trials, for which the 5-task slope was in fact smaller (3 ms) than the 3-task slope (18 ms).

For error rates there was no significant difference between the 5-task slope (2.0%) and the 3-task slope (1.1%), F<1. In contrast to the RT analysis, congruency slopes were slightly larger on switch (2.6%) than on repeat trials (0.5%), although this difference was also not significant, F(1,22) = 2.03 p = 0.169. Finally, there was no reliable interaction between number and switch, F<1, and error rate pattern did not support the near-significant interaction in the RTs (i.e. congruency slopes for switch and repeat trials were 3.1% and 1.0%, respectively, in the 5-task condition; and 2.0% and 0.1%, respectively, in the 3-task condition).

Hence we obtained a graded effect of the number of incongruent attributes, but there was no clear evidence that this effect was stronger for the 5-task condition, when a greater number of attributes could be congruent/incongruent.

*Summary* Participants were required to switch among 3 or 5 tasks. For two of those tasks (probe tasks), recency and frequency were matched between the 3- and 5-task conditions. When analysis was restricted to probe task trials preceded by probe task trials, there was no evidence that having a greater number of tasks in play made performance more difficult, or increased the difficulty of switching or preparing to switch to a new task. (In fact the overall error rate was slightly higher overall for 3 tasks, but we do not dwell on this as it was not replicated in Experiment 2). The importance of the potential confound of number of tasks with recency or frequency was demonstrated by an analysis including trial sequences in which number of tasks was confounded with recency and (in)frequency. Now RTs were reliably longer for unprepared switch trials in the 5-task than in the 3-task condition, giving the appearance of a greater difficulty of task-set retrieval with more tasks in play.

That recency has an impact on task-set preparation was confirmed by a trend in the probe task data indicating that on short CSI task-switch trials, RTs were longer when the task had been less recently exercised, suggesting that it is harder to retrieve a less recently performed task-set. Data from the uncontrolled tasks suggested that it was also harder to retrieve a less frequently performed task set. Finally, a modest graded congruence effect could be seen, such that performance deteriorated the more irrelevant attributes were mapped to the wrong response, but there was no evidence that such effects were more extreme in the 5-task condition (when four irrelevant attributes can be congruent/incongruent) than in the 3-task condition (when only two can) — consistent with the lack of effect of the number of competing tasks on preparation. For these sets of tasks at least, we conclude that the difficulty of promoting an active but currently unselected task-set into procedural working memory does not increase with the number of currently active task-sets, but is influenced by frequency and recency.

## Experiment 2

Does this conclusion generalise to other pairs of tasks? One special property of the tasks used in Experiment 1 is that they involved attention to different perceptual attributes of an object. Moreover the attributes were spatially separate (head, tail, etc.) The response congruence effects were quite small, suggesting that participants were good at filtering out the irrelevant attributes, even on task-switch trials. Hence it could be argued that this is a situation where competition was not so much in the retrieval of the relevant task-set rules, or in attending to the relevant dimension, as in the potential targets for spatial attention — and we know that Hick's law does not apply to eye movements to a target (e.g. Kveraga, Boucher & Hughes, 2002). Experiment 2 therefore used a similar design, but with a word stimulus and an orthogonal set of five lexical classification tasks, for each of which attention was required to a different semantic/phonological property of the word, not a spatially distinct attribute of an object.

There were two other minor changes. In Experiment 1, the two tasks we selected as recency- and frequency-controlled probe tasks were the same for all participants. To avoid the possibility that there was something special about the controlled tasks, in

Experiment 2 each pair of tasks served equally often as probe tasks. The other change was that only one set of tasks was used in Experiment 2 (because we found it hard to generate more than one set of five orthogonal task pairs applicable to the same words). Participants performed five tasks in one session, and a subset of three in the other session, with order counterbalanced. To minimise the effect in the second session of earlier exposure to (a sub-set of) the tasks, the sessions were about two weeks apart.

### Method

*Participants* Twenty participants aged between 18 and 41 [M = 22.3], 5 male and 15 female, provided informed consent before taking part in two sessions separated by at least 11 days (M 14 days) days. Participants were paid between £12.80 and £16.00, depending on the speed and accuracy of their performance.

*Design* The stimuli were a set of 32 object names, each classifiable along 5 dimensions, four being semantic properties of the object and one a phonological property of the word. The five possible classifications, and the associated pair of task cues[7] used for each were:

- Is it bigger or smaller than a bread box? (Cue: “big?” or “small?”.)
- Is it a normally an occupant of the air or the ground? (Cue: “ground?” or “sky?”.)
- Is it common or rare in the local environment? (Cue: “common?” or “rare?”.)
- Is it an animal or an object? (Cue: “animal?” or “object?”.)
- Does the word have one or two syllables? (Cue: "one?" or "two?".)

Table 2.3 shows the words and classifications. Frequency and recency were matched between the 3- and 5-task conditions in the same way as in Experiment 1, but with a different pair of probe tasks used for each pair of participants. All 10 possible pairs of recency-controlled probe tasks were used equally often, and each task was used equally often as the third task in the 3-task condition.

Ten participants completed the 3-task condition in the first session and the 5-task condition two weeks later; the other ten were tested in the opposite order. If

---

[7] Using one of the two potential response category labels as the cue results in faster responses when the cue matches the response category (e.g. Monsell & Mizon, 2006; Schneider & Logan, 2005) but this effect is orthogonal to those of interest here, and does not appear to modulate the effectiveness of preparation.

participants had completed the 5-task session first it was made clear in the second session that only 3 of those 5 tasks would ever be performed in that session. In addition, in both sessions, before the practice blocks, participants were shown the list of 32 words and the tasks applicable in that session only, to familiarise them with the words and their categorisations.

Otherwise, Experiment 2 was identical to Experiment 1, with the same trial and practice trial numbers, and the same method of constructing trial sequences. As in Experiment 1, long (1300 ms) and short (100 ms) CSI blocks alternated.

| | | | | | |
|---|---|---|---|---|---|
| **ground** | **object** | **big** | **one** | **common** | car |
| | | | | **rare** | throne |
| | | | **two** | **common** | sofa |
| | | | | **rare** | tandem |
| | | **small** | **one** | **common** | spoon |
| | | | | **rare** | gem |
| | | | **two** | **common** | pencil |
| | | | | **rare** | ruby |
| | **animal** | **big** | **one** | **common** | horse |
| | | | | **rare** | moose |
| | | | **two** | **common** | donkey |
| | | | | **rare** | camel |
| | | **small** | **one** | **common** | rat |
| | | | | **rare** | stoat |
| | | | **two** | **common** | hamster |
| | | | | **rare** | gecko |
| **sky** | **object** | **big** | **one** | **common** | plane |
| | | | | **rare** | blimp |
| | | | **two** | **common** | chopper |
| | | | | **rare** | glider |
| | | **small** | **one** | **common** | dart |
| | | | | **rare** | flare |
| | | | **two** | **common** | frisbee |
| | | | | **rare** | discus |
| | **animal** | **big** | **one** | **common** | swan |
| | | | | **rare** | stork |
| | | | **two** | **common** | buzzard |
| | | | | **rare** | condor |
| | | **small** | **one** | **common** | moth |
| | | | | **rare** | chough |
| | | | **two** | **common** | sparrow |
| | | | | **rare** | goldfinch |

**Table 2.3** All 32 word stimuli used in Experiment 2, and their respective classifications along the 5 task dimensions.

## Results and Discussion

The same exclusion criteria used in Experiment 1 were applied, resulting in the exclusion of 0.4% of the correct trials from the computation of RT means. The data were analysed in the same was as for Experiment 1.

*Effect of number of tasks for the probe tasks* Mean RT and error rates for the controlled task sequences are shown in Figure 2.5. They largely replicate the result obtained in Experiment 1 (see Figure 2.2). There was little sign of any overall effect of number of tasks in play and, more important, no evidence for RTs on unprepared switch trials being selectively prolonged by an increase in the number of tasks, as would be expected if promoting a task-set to most-active state (after stimulus onset) was more difficult with more active task-sets in play. The mean RT difference between the 5-task (852 ms) and the 3-task condition (843 ms) was not reliable, F<1. The overall reduction in switch cost with opportunity for preparation (from 66 to 32 ms) was a reliable 34 ms or 52%, F(1,19) = 7.99, p = 0.011. This RISC effect did not differ reliably between the 5-task (27 ms) and 3-task conditions (42 ms), F<1 for the 3-way interaction. Indeed the difference was in the opposite direction to what would be expected if TSR were slower with more tasks to select from.

There were reliable main effects of CSI, F(1, 19) = 97.86, p < 0.001, and switching, F(1,19) = 20.12, p < 0.001, The effect of a switch did not differ reliably between the 5-task (56 ms) and 3-task conditions (42 ms), F<1. There was no reliable interaction between the effects of CSI and number of tasks, F<1.

In the % error analysis, only the switch cost (1.9%) was reliable, F(1,19) = 13.31, p = 0.002. For the effect of, and all interactions with, number of tasks, F<1.

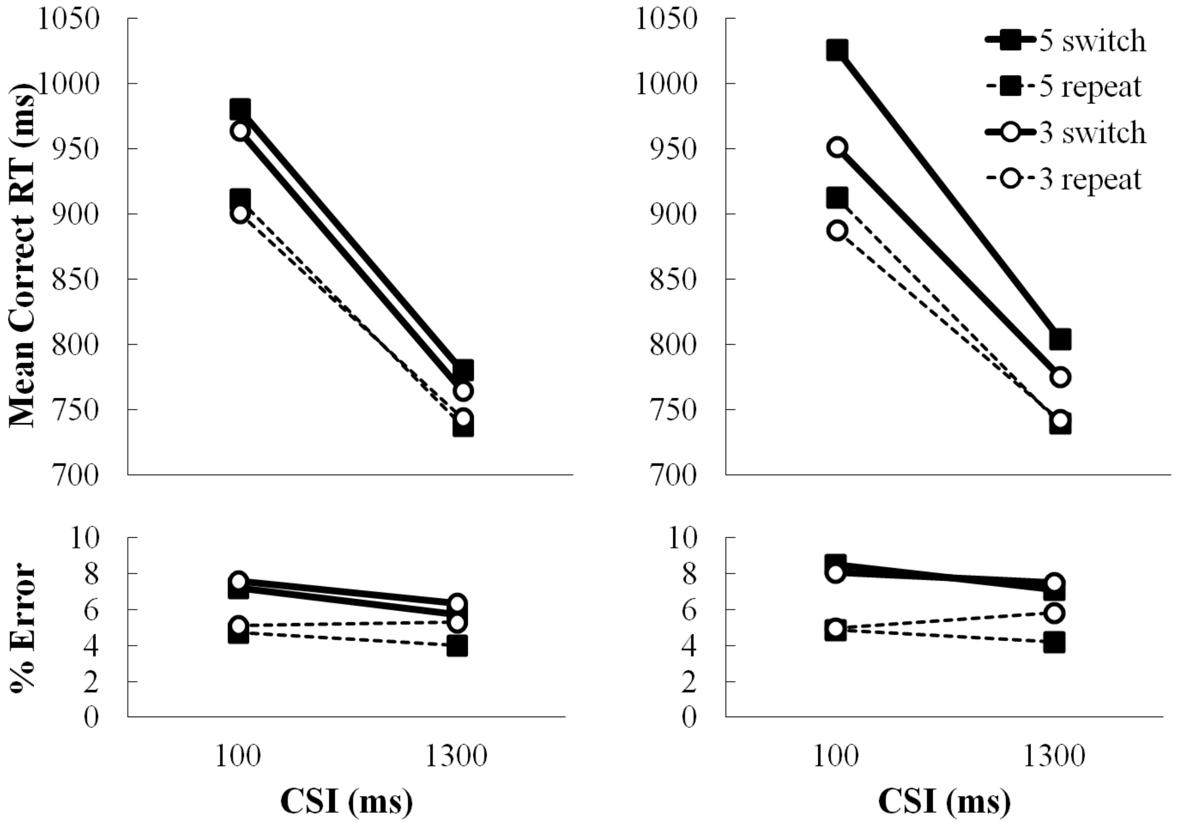


**Figure 2.5** Mean correct RTs (top) and % error (bottom) data in Experiment 2 with three and five tasks, on switch and repeat trials, as a function of CSI; plotted separately for probe task trials (left) and all task trials (right).

| | | **Probe tasks** | | **All tasks** | |
|---|---|---|---|---|---|
| | | **5 tasks** | **3 tasks** | **5 tasks** | **3 tasks** |
| **100 ms CSI** | **repeat** | 909 (4.7) | 898 (5.2) | 912 (4.9%) | 887 (5.0) |
| | **switch** | 980 (7.3) | 963 (7.7) | 1026 (8.5) | 951 (8.1) |
| | **switch cost** | 71 (2.6) | 66 (2.5) | 113 (3.7) | 64 (3.2) |
| **1300 ms CSI** | **repeat** | 980 (7.3) | 963 (7.7) | 739 (4.2) | 742 (5.9) |
| | **switch** | 737 (4.1) | 743 (5.3) | 804 (7.1) | 775 (7.5) |
| | **switch cost** | 71 (2.6) | 66 (2.5) | 65 (2.9) | 33 (1.7) |
| | **RISC** | 28 (0.7) | 43 (1.2) | 49 (0.8) | 30 (1.5) |

**Table 2.4** Mean correct RTs (ms), with the % error data between brackets, for the different conditions in Experiment 2.

*Effect of number of tasks for all trials* As for Experiment 1, we repeated the above analysis, but now including all trials. As may be seen from Figure 2.5, the pattern of results closely resembled the result of Experiment 1. Mean RTs were now substantially longer in the 5-task (870 ms) compared to the 3-task condition (839 ms), although this difference was not significant. The analysis also revealed a reliable switch cost of 69 ms, F(1,19) = 46.31, $p < 0.001$; a CSI effect of 179 ms, F(1,19) = 90.13, $p < 0.001$, and a reduction in switch cost of 40 ms or 45%, F(1,19) = 22.47, $p < 0.001$. As was the case in Experiment 1, the largest RT difference between the 5-task and the 3-task condition was on short CSI switch trials. This resulted in a larger switch cost in the 5-task (89 ms) compared to the 3-task condition (49 ms), F(1,19) = 12.91, p = 0.002, and a marginally larger CSI effect, F(1,19) = 3.566, p = 0.074. The RISC effect was also larger in the 5-task condition (49 ms) compared to the 3-task condition (30 ms), though the 3-way interaction was not reliable, F(1,19) = 0.81, p = 0.379. Comparing RT for each combination of long and short CSI and task switch and repeat trials between the 5-task and 3-task conditions, only the 75 ms difference for short CSI switch trials was reliable, F(1,19) = 5.26, p = 0.033.

The error data analysis only revealed a reliable switch cost of 2.8%, F(1,19) = 42.33, $p < 0.001$, and a marginally reliable RISC of 1.1%, F(1,19) = 3.88, p = 0.064 (for main effect and all interactions with number of tasks, $F < 2.67$).

*Effects of task recency* When this analysis was restricted to probe task sequences (Figure 2.6), the linear trend of recency (slope of 9 ms), approached significance, F(1,19) = 3.70, p = 0.070. Although the recency slope was larger in the short CSI (11 ms) than in the long CSI (7 ms), the interaction was not reliable, $F<1$. The difference between the 5-task (8 ms) and 3-task (10 ms) recency slopes was also not reliable, $F<1$, and neither was the 3-way interaction, $F<1$. (When the recency slopes for short and long CSI trials with 3 and 5 tasks were analysed individually, none of the linear trends were significant, all Fs < 2.45, ns.) In the error analysis, neither the linear trend of recency nor any of its interactions were reliable, all Fs < 1.

When all switch trials were included in the analysis (see Figure 2.6), the linear trend of recency (slope 5 ms) was reliable, F(1,19) = 5.00, p = 0.038. As in Experiment 1, the recency slope was larger in the short CSI (9 ms) than in the long CSI (2 ms)*;* and larger with 5 tasks (7 ms) than with 3 (4 ms), though neither interaction was reliable, F(1,19) = 2.78, p = 0.112 and $F<1$, respectively. The 3-way interaction was also not reliable F(1,19) = 2.79, p = 0.112. However, separate 2 (5 or 3) x 3 (lag) ANOVAs for each CSI showed that the recency slope was (almost) reliably larger with 5 tasks (14 ms) than with 3 tasks (4 ms) in the short CSI, F(1,19) = 4.32, p = 0.051, but not in the long CSI, F<1. In the error data, the linear trend of recency and all interactions with this factor were not reliable, F<1.

Thus, though the effects are somewhat noisy in this analysis, the data are consistent with the trend seen in Experiment 1, namely a modest trend towards shorter RTs when the task has been performed more recently at a short CSI in the controlled data, though the relevant interaction between the linear component of recency and CSI was not reliable. Furthermore, as in Experiment 1, an additional difference between the 3- and 5-task condition appeared when the uncontrolled data was plotted as a function of recency, suggesting again that it is harder to activate a less frequently performed task set.

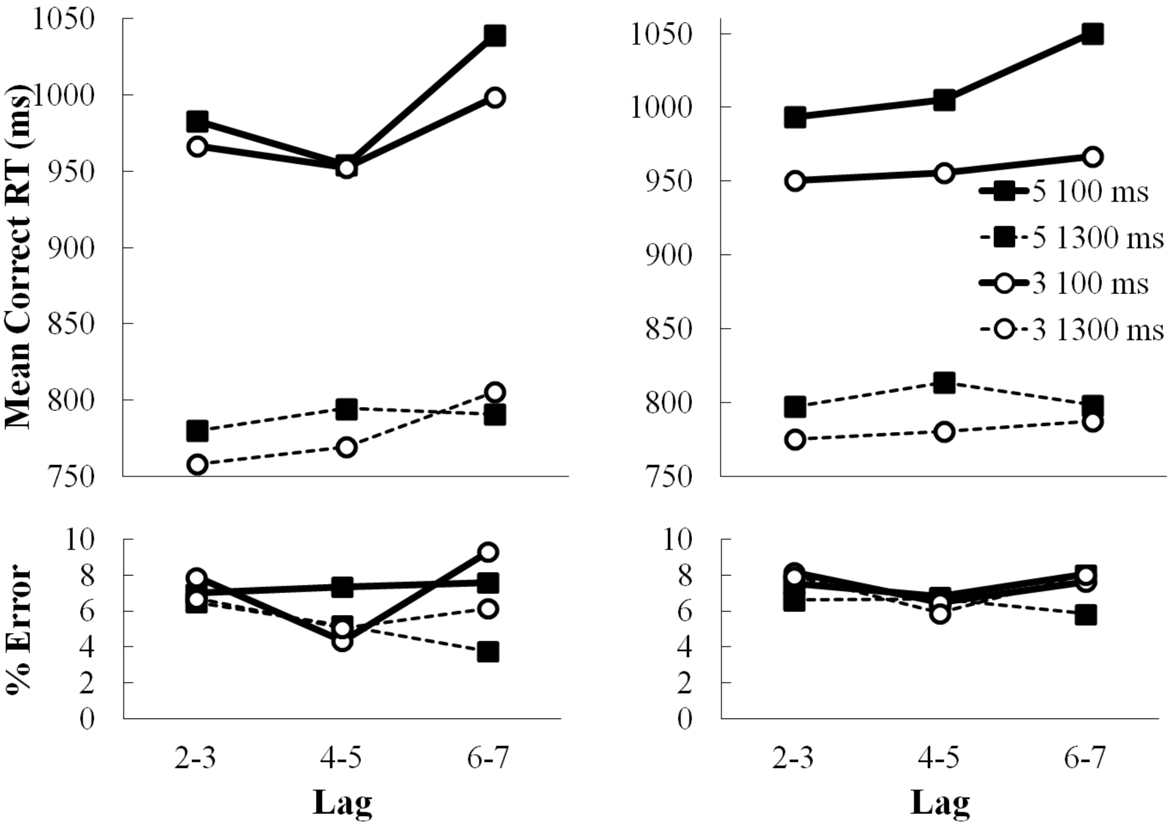


**Figure 2.6** Mean correct RTs (top) and % error (bottom) data in Experiment 2 with three and five tasks, in long and short CSI blocks, as a function of lag; plotted separately for probe task trials (left) and all task trials (right).

*N – 2 repetition effect* As in Experiment 1, there was no evidence for an n – 2 task repetition effect. Participants were faster on (854 ms) ABA than on CBA sequences (871 ms), F(1,23) = 2.60 (ns), but they made slightly more errors on ABA (8.3%) than CBA sequences (7.3%), F(1,23) = 1.98 (ns).

*Effects of response congruence* It can be seen in Figure 2.7 that the effects of response congruence were much stronger than in Experiment 1. The 3-task RT ANOVA showed a reliable main effect of congruency (slope 92 ms), F(2,38) = 21.30, $p < 0.001$[8], with a linear trend component, F(1,19) = 33.31, $p < 0.001$ (quadratic trend F<1). The 5-task RT ANOVA revealed a congruency slope of similar size (104 ms), F(4,76) = 9.91, $p < 0.001$, with a linear trend component, F(1,19) = 23.78, $p < 0.001$, and no other reliable trend components (F < 2.30, ns). The 3-task error ANOVA showed a reliable main effect of congruence (slope 10%), F(2,38) = 35.27, $p < 0.001$, with a linear trend, F(1,19) = 51.88, $p < 0.001$, and no quadratic trend, F = 3.21 (n.s.). A much smaller congruence slope was found in the 5-task error data (slope 3.6%), though the main effect was still significant, F(4,76) = 2.89, p = 0.036, as was the linear trend, F(1,19) = 9.83, p = 0.005, (all other trend components, F<1).

To compare congruence slopes between the 3-task and the 5-task conditions, 2 (3 or 5) x 2 (switch or repeat) ANOVAs were run on the RT and error congruence slopes. The RT congruence slope did not differ reliably between the 3-task (92 ms) and 5-task condition (104 ms), F<1, or between switch (84 ms) and repeat trials (108 ms), F<1. The interaction between effects of number of tasks and a task switch was also not significant, F = 1.23 (ns). The error congruence slope was much larger in the 3-task (10%) than in the 5-task condition (3.6%), F(1,19) = 28.21, $p < 0.001$, and it was of similar size for switch (7.8%) and repeat trials (5.9%), F<1. The interaction was also not significant, F<1.

To summarise, Experiment 2 yielded graded congruence effects much larger than in Experiment 1. Moreover, for the errors (but not for the RTs) the congruency slope was much larger in the 3-task condition.

---

[8] Huyn-Feldt corrected

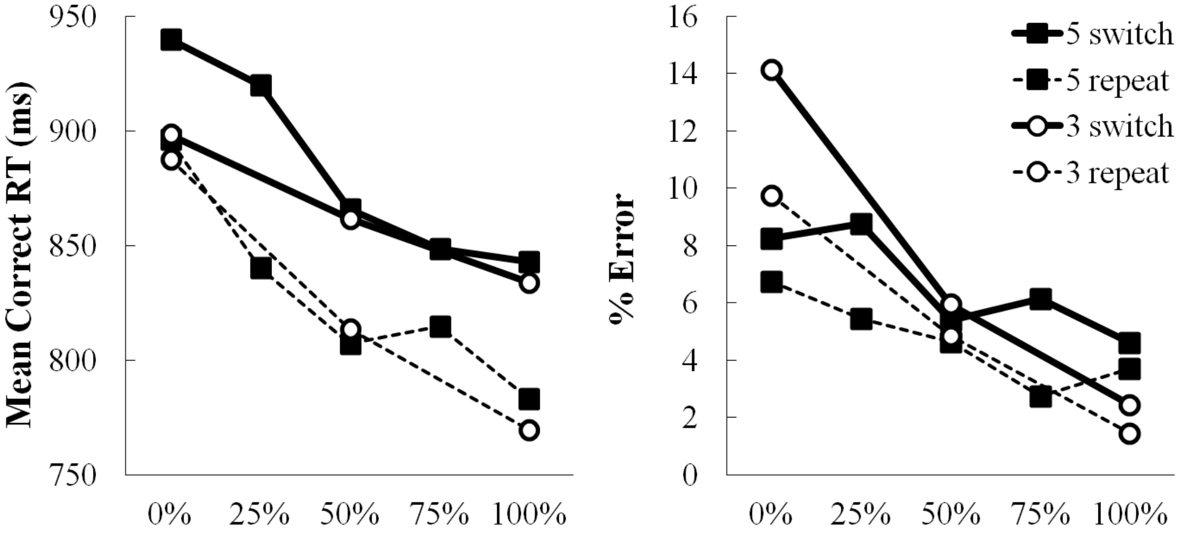
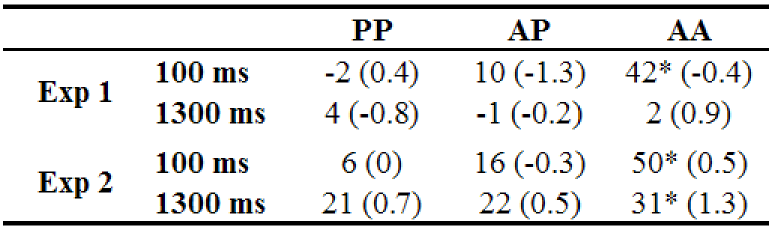


**Figure 2.7** Mean correct RTs (top) and % error (bottom) data in Experiment 2 with three and five tasks, on switch and repeat trials, as a function of response congruence.

*Summary* Using a quite different set of tasks, requiring classification of semantic/phonological attributes of a word rather than perceptual attributes of an object, Experiment 2 confirmed the results obtained in Experiment 1. When task recency and frequency were uncontrolled, RTs were longer in the 5-task condition, particularly on a task switch trial with no opportunity for advance preparation. But when two-trial sequences of tasks matched for frequency and recency were examined, there was again no effect of number of tasks to suggest that task-set reconfiguration is influenced by the number of alternative tasks in play. Again, a trend in the controlled data was consistent with the idea that recency has a modest impact on task-set preparation in addition to any effects of task frequency. And again, differences between the 3- and 5-task condition remained in the uncontrolled data when it was plotted as a function of recency, suggesting that task frequency must have also contributed to apparent number of tasks effect in this data.

In both experiments the probe task analysis was restricted to trials for which the preceding trial was also a probe task, on the assumption that the recency/frequency of the preceding task might also affect the switch cost. To test this assumption we compared the difference in switch costs with 3 and 5 tasks for the following three trial types: probe-probe sequences (4/9 of all trials), all trial-probe sequences (2/3 of all trials) and all trial-all trial sequences (see Table 2.5). Switch costs were only substantially and reliably larger with 5 tasks for all trial-all trial sequences; there was not much difference between the probe-probe and all trial-probe sequences. Apparently it is the frequency/recency of the *current* task that matters, not that of the *preceding* task.



| | | PP | AP | AA |
|---|---|---|---|---|
| **Exp 1** | **100 ms** | -2 (0.4) | 10 (-1.3) | 42* (-0.4) |
| | **1300 ms** | 4 (-0.8) | -1 (-0.2) | 2 (0.9) |
| **Exp 2** | **100 ms** | 6 (0) | 16 (-0.3) | 50* (0.5) |
| | **1300 ms** | 21 (0.7) | 22 (0.5) | 31* (1.3) |

**Table 2.5** Differences between the 5-task and 3-task switch costs in ms (% errors between brackets) for the following trial types: probe trials following probe trials (PP), probe trials following all trials (AP) and all trials following all trials (AA). * = $p < 0.05$.

In both experiments, we obtained a graded response congruence effect reflecting the proportion of incongruent attributes, but this effect was much bigger (by about a factor of five) in Experiment 2 than in Experiment 1. It was clearly harder to ignore the values of the irrelevant dimensions for the lexical classifications of Experiment 2, presumably because spatial attention could no longer be used to filter out irrelevant attributes. Moreover, in Experiment 2, the congruency effect for error rates was found to be much larger in the 3-task than in the 5-task condition. One explanation for this difference is the strength of the S-R associations of the other (currently irrelevant) tasks. Specifically, although recency and frequency of the current task was matched between the 3- and 5-task conditions (in the analysis restricted to probe trials), the recency and frequency of other tasks in play was not. And whereas the one non-probe task occurred as often as the probe tasks in the 3-task condition, in the 5-task condition 3 out of 4 of the currently irrelevant tasks occurred much less frequently (and hence on average, less recently) than the probe tasks. Hence with 3 tasks the S-R mappings of the irrelevant tasks (represented in the activated part of LTM) were “stronger”, resulting in more response priming (facilitation on congruent trials, and interference on incongruent trials).

## General Discussion

The aim of the experiments reported here was to elucidate the process of promoting a task-set to the most-active state, which in many accounts is conceptualised as retrieving it (or its S-R rules) into working memory (Mayr & Kliegl, 2000) or, more specifically, a protected current-task buffer in procedural working memory (Oberauer's, 2009, "bridge"). In particular, we asked whether this process is influenced by the number of alternative, currently irrelevant task-sets. Previous experiments have yielded inconclusive results, in part because most did not control for the recency and frequency with which a task was performed, or the "extra" tasks were not afforded by the same stimuli. In the present experiments, participants switched among three and five tasks (in separate sessions), and for two of those tasks (“probe tasks”), recency and frequency were matched between sessions. A natural expectation, based on other cases of memory retrieval, would be that retrieval time should increase as a function of the number of task-sets in play. This should be reflected in a larger switch cost when switching among five tasks, particularly when the cue-stimulus interval allows no opportunity for advance preparation.

In both experiments, when the analysis was restricted to sequences of two tasks matched for recency and frequency, number of tasks was found to have no effect on overall RT, the switch cost, or reduction in switch cost with preparation; more specifically there was no sign that having to select from a larger set of tasks caused extra difficulty on unprepared switch trials. Clearly, the difficulty of retrieving a task-set does not increase as a function of the number of available competitor task-sets. However, when pooling over all trials, and hence including tasks not matched for recency and frequency, a different result was obtained: now participants were slower with five tasks, particularly when there was no opportunity for advance preparation. Although the pattern of results in this analysis was slightly different between Experiment 1 (participants were exclusively slower on short CSI switch trials, reliable 3-way interaction) and Experiment 2 (participants were substantially slower on short CSI switch trials, but also somewhat slower on short CSI repeat trials, hence no reliable 3-way interaction), the contrast to the analysis restricted to probe tasks is marked. Thus, had we not controlled for recency and frequency, we would have likely have been led to conclude that the time required to upload a task-set into PWM increases as a function of the number of alternative task-sets. Indeed, the same confound may have led others to conclude that it is harder to switch among a larger set of tasks (Kray et al., 2002;

Buchler et al., 2008).

So what caused participants to be slower with five tasks in the uncontrolled data? Overall, the probability of a task switch was slightly higher with five tasks, but this is unlikely to have caused the result, as a higher proportion of switch trials normally results in smaller switch costs (Monsell & Mizon, 2006). To explore the possibility that differences in task recency caused the apparent effect of number of tasks, task switch trials were analysed as a function of the task's recency. For tasks matched for frequency (probe tasks) RTs increased (marginally) as a function of lag when there was no opportunity for preparation, suggesting that task recency does have a small effect on the difficulty of retrieval. For data including uncontrolled tasks, there was an additional effect for unprepared trials that we can attribute only to the effect of task frequency. Hence as well as demonstrating that there is no effect of number of tasks in play when recency and frequency are controlled, the data provide evidence that task-set activation on a switch trial is easier when the task has been more recently and frequently performed. Moreover, there was very little difference between the switch costs for probe task trials preceded by probe task trials, and probe task trials preceded by all task trials, demonstrating that it was the recency/frequency of the current task (not the preceding task) that mattered.

We also examined response congruency effects as another index of competition among task sets. In both experiments, a graded congruency effect was obtained, i.e. RTs increased as a function with the proportion of incongruent dimensions (also see Meiran & Kessler, 2010). In Experiment 1, the congruency slope did not differ between the 3- and 5-task conditions. However, in Experiment 2, the error congruence slope was much larger with three tasks though a slight trend in the opposite direction was present in the RTs. One possible explanation is that in the 5-task condition, the other (irrelevant) task-sets were practiced on average less frequently (and hence less recently), resulting in weaker S-R associations for the currently irrelevant tasks, resulting in less facilitation on congruent trials, and interference on incongruent trials. At least we can say that interference did not *increase* with the number of tasks, consistent with our finding that switch costs and its reduction with preparation is not larger with five tasks than with three.

The demonstration that promoting a task-set into the most-active state is no harder when there are more tasks in play is consistent with Oberauer’s (2009) theory of procedural working memory, or at least with his speculation that, because the activated part of LTM which holds the currently irrelevant task-sets is capacity-unlimited, promoting one to the bridge should not take longer the greater the number of tasks in play. Oberauer made this prediction on the basis of his assumption that declarative and procedural WM operate by analogous principles, but recent data from his own lab on declarative working memory (Da Souza Silva et al., in press) appear to violate this principle! In this experiment participants were cued to perform (a subset of) 3 mathematical tasks (big/small, inner/outer, high/low) on (a subset of) 3 lists (presented on the top, centre and bottom of the screen) of 3 digits each. The number of lists and the number of tasks was varied independently, between blocks of 9 trials. The position of the task cue on the screen indicated the relevant list (horizontal axis) and the relevant item within that list (vertical axis). Task switch costs and list switch costs were underadditive, supporting Oberauer's assertion that procedural and declarative representations are selected in parallel. However, contrary to expectations, whereas task-switch costs were not larger with three tasks than with two, list-switch costs were largest with three lists (only at a short CSI – like the uncontrolled data in our experiments), suggesting a difference between declarative and procedural WM processes. Da Silva Souza et al. (in press) explain this difference by suggesting that whereas the capacity-limited declarative buffer may sometimes hold more than one list (hence the increase in list switch costs with the number of lists), the capacity-limited PWM buffer only ever holds one task at a time (hence no effect of the number tasks on the switch cost). But as Da Silva Souza et al. did not match the recency and frequency with which each list (or item within each list) was retrieved, it is possible that their number of lists effect results from a confound of the recency and frequency with which each list was accessed.

Finally, although we have interpreted the lack of a number of tasks effect in the controlled data as evidence that the number of tasks in play does not affect task-set retrieval time, one other explanation deserves consideration: it is possible that the difficulty of task-set retrieval *does* increase task-set retrieval time, but that this effect is 'cancelled out' by an effect of task-set competition in the opposite direction. (In the five task condition, the other tasks are performed less often than the probe tasks, and hence have less competitive strength). However, this seems unlikely, and the current experiments are unable to assess this possibility.

To conclude, our finding of no effect of number of tasks in play on task-set retrieval provides an additional reason for distinguishing two levels of action control: the level of selection of a task-set, and the level of selection of responses within a task-set, for the latter does appear to show set-size effects, as reflected in Hick's law. One might, of course, wonder whether Hick's law too could arise from a confound between recency/frequency and set size. This seems unlikely from early and more recent manipulations of probability and repetition in choice RT (Hyman, 1952, Kornblum, 1968, Scheider & Anderson, 2011). However, in the next chapter (Chapter 3) we demonstrate that the Hick’s law effect of number of S-R rules on choice RT within one task survives the control of recency and frequency that we applied in the present experiments.

# Chapter 3: The contribution of stimulus recency and frequency to Hick's law

---

## Introduction

The experiments described in Chapter 2 were designed to elucidate the process by which the relevant task-set is retrieved into a limited capacity procedural working memory (PWM) buffer. More specifically, these experiments asked whether this process of task-set retrieval is influenced by the number of alternative task-sets. If this were the case, the switch cost should increase as a function of the number of potentially relevant task-sets. But Experiments 1 and 2 demonstrated that it does not, provided that recency and frequency of task usage are matched. This finding is inconsistent with what we know about other kinds of memory retrieval, where retrieval time appears to be influenced by the number of competing alternatives. For example, according to Hick's law, reaction time (RT) increases as a function of the number of stimulus-response (S-R) mappings (Hick, 1952; Hyman, 1953). Another example, the "fan effect", refers to the observation that as the number of unrelated facts learned in association with a concept increases, retrieval time for one of these facts increases (Anderson, 1974).

Schneider and Anderson (2011) have recently argued that both Hick's law and the fan effect arise as a consequence of memory retrieval. In support of this proposal, they point out that when memory retrieval is not needed, as in experiments requiring saccades to a target location (Kveraga, Boucher & Hughes, 2002), or aimed arm movements (Wright, Marino, Belovsky & Chubb, 2007), effects of set-size (number of alternatives) are not found. In Experiments 1 and 2 also, the switch cost did not differ when the task had to be selected from among 3 versus 5 alternatives. But if we are right to interpret the switch cost, especially with no opportunity to prepare, as a measure of the time taken to retrieve a task-set, this is a case where memory retrieval is not influenced by the number of alternatives. There are two possible explanations: 1) The process of retrieving a task-set in a task-cueing environment is somehow "special", and differs from other kinds of memory retrieval in this respect; or 2) The set-size effects observed for these other kinds of memory retrieval actually result from confounds with the recency and/or frequency of the retrieval, not from set-size per se. The experiment reported here investigated the latter explanation, focusing in particular on the set-size effect described by Hick's law.

Since the publication of Hick (1952), several factors which may modulate or even eliminate set-size effects have been identified (e.g. Morin & Forrin, 1965; Mowbray, 1960; Theios, 1973). S-R compatibility is one: the cases of directed arm movements and saccades have already been mentioned above; as another classic example, Leonard (1959) varied set size between 1 and 10 for a finger flexion in response to vibro-tactile stimulation of the responding finger, but found no effect beyond an increase from simple to two-choice RTs. Other factors known to modulate Hick's law include response mode (set-size effects are normally smaller with verbal responses, see Longstreth, Zahhar & Alcorn, 1985), familiarity with the stimuli or S-R associations (no set-size effects are found when participants are required to name familiar stimuli such as letters, Morin, Konick, Troxell & McPherson, 1965, or digits, Brainard, Irby, Fitts & Alluisi, 1962). The effects of familiarity and S-R compatibility might be attributed, as above, to their eliminating the need for retrieval of an arbitrarily associated response (or to such retrieval becoming so automatic that no set size effect can be detected, e.g. Schneider & Anderson, 2011). Another – perhaps more parsimonious – explanation for the absence of set-size effects is that in all the examples listed above, there was actually only one S-R rule (e.g. "name the stimulus" or "point/saccade towards the stimulus").

With regard to the contribution of stimulus recency and frequency effects to Hick's law, perhaps the most relevant dispute in the literature concerns response repetitions. The importance of response repetitions in set-size effects was first recognised by Kornblum (1968), who noted that there is an inverse correlation between the number of S-R mappings and the probability of an S-R repeat. Because RTs are typically faster for an S-R repeat than for an S-R switch, decreasing the number of S-R mappings inflates the proportion of fast(er) responses in the RT distribution. To test this prediction, Kornblum contrasted conditions which were equal in terms of information (i.e. the number of S-R mappings), but differed with respect to the proportion of S-R repetitions. When there was a high probability of an S-R repetition, a large effect of set-size was found. However, when the probability of an S-R repetition was low, RT did not increase as a function of set-size, supporting Kornblum's theory that Hick's law is actually a response repetition effect.

However, Hyman and Umiltà (1969) argued that the conditions set out in Kornblum's experiment were not optimal. They pointed out that, according to Hyman

(1953), ideal conditions should include plenty of practice, motivated subjects and, most notably, a relatively long response-stimulus interval (RSI) of 9 to 10 seconds. According to Hyman and Umiltà, the RSI of 140 ms used by Kornblum did not allow participants sufficient time to prepare for the upcoming trial. In Hyman and Umiltà's replication of Kornblum's experiment, participants were given plenty of time to prepare (RSI of 7.5 seconds). Indeed, under these circumstances, a large effect of set-size was obtained.

Although Kornblum's theory has not gone unchallenged, the underlying principle is relevant to the current research because a similar argument can be made with regards to the contribution of frequency and recency effects to Hick's law. Specifically, provided that the total number of trials is equal, and that each stimulus occurs equally often, with a smaller stimulus set each stimulus will occur on average more recently, and also more frequently, than in the case of a larger stimulus set. As more recently and frequently performed S-R associations normally yield faster responses, increasing the proportion of such trials will lead to a decrease in overall RT. The contribution of recency and frequency effects to set-size effects was also addressed in Hyman's (1953) original paper. Specifically, as well as varying the number of S-R mappings, Hyman also manipulated the frequency (Experiment 2) and recency (Experiment 3) with which stimuli occurred. In these experiments, 4 participants responded to (a subset of) 8 lights by naming them ("bun", "boo", "bee", etc.). Hyman (1953) claimed that the correlations between RT and information computed on the basis of frequency or recency were similar to the correlation between RT and information computed on the basis of the number of S-R mappings. However, the interaction between these variables was not further investigated. In other words, Hyman's experiments cannot tell us whether, if recency and frequency are matched, the set-size effect would disappear. The experiment reported here was aimed at investigating whether this was indeed the case.

Some insight into the role of recency and frequency in memory retrieval can also be derived from fan effect experiments. In such experiments it is possible to independently manipulate the number of facts associated with a concept on the one hand, and the frequency of those facts on the other hand (Anderson, 1983). To illustrate this, Anderson (1983) contrasted a two-fan condition in which both facts were studied three times to a four-fan condition in which one fact was studied three times, and the three other facts were each studied once. According to Anderson (1983), there was practically no difference between the two. In a description of this experiment elsewhere, Anderson (1990) concluded that "it was probability, and not fan, that was the controlling variable" (p. 69). In the light of this conclusion it may seem astonishing that this same author has continued to attribute the fan effect to the number of competing alternatives rather than frequency (Anderson & Reder, 1999; Schneider & Anderson, 2011, 2012). However, closer inspection of Anderson's (1983) data reveals that Anderson's (1990) conclusion, apparently based on the observation that the frequency-matched fan RT effect was non-significant, may not have been warranted. Although there was a small opposite trend in the errors, the RT fan effect was not entirely eliminated: a 44 ms difference remained, which was in fact significant one-tailed. So from this experiment it is not clear whether matching for frequency eliminates the fan effect, or not.

At the same time, it is interesting to note that the ACT-R (Adaptive Control of Thought – Rational) cognitive architecture – whose development was heavily influenced by the existence of fan effects – does incorporate parameters that reflect recency and frequency of stimulus usage. In ACT-R, the retrieval or "availability" of a chunk depends on its total activation, which is the sum of a chunk's associative activation (which decreases with increasing fan), and the chunk's base-level activation (which reflects recency and frequency of usage, Anderson, Bothell, Byrne, Douglass, Lebiere & Qin, 2004).

A final note concerns the role of practice. With regards to the effect of practice on Hick's law, Longstreth et al. (1985) have summarised the methods and results of a number of experiments in which set-size effects were (almost) absent, and concluded that practice was an important factor. Others have also asserted that with vast amounts of practice, the set-size effect would disappear (Teichner & Krebs, 1974). However, set-size effects are rarely completely eliminated, and often considerable effects remain even after extensive practice (for example, both Hale, 1968, and Hyman, 1953, reported sizeable set-size effects after 5000 and 15000 trials, respectively). The role of practice has also been investigated in the context of the fan effect: Anderson (1983) has described an experiment in which he compared RT as a function of practice in a fan and a no-fan condition. Each subject completed 25 sessions, and by the end of the experiment each sentence had been encountered 600 times. Even after the twenty-fifth session, a small fan effect remained. Although Anderson (1983) has argued that it is likely that additional practice would eventually eliminate the fan effect, to my knowledge there are no studies in which the fan effect was eliminated as a function of practice. In any case, the results of both Hick's law and fan experiments seem to suggest that although practice can massively reduce the effect of the number of alternatives on memory retrieval, it is unlikely to eliminate such effects altogether.

In summary then, it appears at least possible that the set-size effects captured by Hick’s law and the fan effect are (in part) the result of the frequency and recency with which the specific S-R retrieval has been exercised. However, the exact contribution of these factors remains unknown. The effect of practice also remains somewhat ambiguous, with some suggestions that extensive practice should eliminate set-size effects (Anderson, 1983; Teichner & Krebs, 1974), yet there are no experimental findings confirming this. In order to address these unresolved questions, this experiment manipulated stimulus frequency and set-size in a Hick’s law experiment. To achieve this, participants were required to classify a set of 4 stimuli, and a set of 6 stimuli. The number of stimulus-response (S-R) rules was manipulated between blocks, and two different sets of stimuli were used: a set of colours, and a set of shapes, so that half of the participants classified 4 colours, and 6 shapes, and vice versa for the other half of participants. For 2 of those stimuli (“probe” stimuli), recency and frequency were matched between the 4 S-R and the 6 S-R conditions in the same way as in Experiments 1 and 2. If Hick’s law really results from a confound between set-size and recency or frequency, then a set-size effect should be found for the nonprobe stimuli, but not for the probe stimuli for which recency and frequency did not differ between the 4 and 6 S-R conditions.

## Experiment 3

### Method

*Participants* Twenty four participants (aged between 18 and 45 [M = 21.6], 22 female and 2 male) took part in this experiment. All provided informed consent prior to participating, and the experiment was approved by the University of Exeter School of Psychology Ethics Committee. Participants were paid between £5.40 and £7.00, depending on the speed and accuracy of their performance.

**Figure 3.1** Stimuli used in Experiment 3. Half of the participants were assigned 4 of the shapes and all 6 colours, and vice versa for the other half of participants (procedure for assigning 4 out of 6 S-R mappings is described below).

*Design and procedure* In order to manipulate the number of S-R mappings within subjects between the two halves of the experiment, whilst avoiding any impact of previous exposure to the same stimuli on performance in the second half, 2 sets of 6 stimuli were used (see Figure 3.1): a set of 6 colours (green, red, light blue, purple, yellow and dark blue), and a set of 6 shapes (circle, cross, drop, square, star and triangle). Responses were made using (four or all of) the X, C, V, B, N and M keys on a computer key board, pressed with the ring, middle and index fingers of the left and right hand. Half the participants completed the shape task with 4 S-R mappings, and the colour task with 6 S-R mappings, and vice versa for the other half of participants. In this way, the experiment was split into 2 parts (a 6 S-R and a 4 S-R part), and the order of parts (and tasks) was balanced between participants.

| | **A** | **B** | C | D | E | F |
|---|---|---|---|---|---|---|
| **A** | **36** | **36** | 9 | 9 | 9 | 9 |
| **B** | **36** | **36** | 9 | 9 | 9 | 9 |
| C | 9 | 9 | 9 | 9 | 9 | 9 |
| D | 9 | 9 | 9 | 9 | 9 | 9 |
| E | 9 | 9 | 9 | 9 | 9 | 9 |
| F | 9 | 9 | 9 | 9 | 9 | 9 |

→

| | **A** | **B** | C | D |
|---|---|---|---|---|
| **A** | **36** | **36** | 18 | 18 |
| **B** | **36** | **36** | 18 | 18 |
| C | 18 | 18 | 36 | 36 |
| D | 18 | 18 | 36 | 36 |

**Figure 3.2** Trial matrix displaying the frequency of all transition types in the 6 S-R condition (left) and the 4 S-R condition (right) in Experiment 3, with the probe transitions highlighted in bold.

Each part was split up into 9 blocks of 48 trials each, plus 1 warm-up trial. For 2 of the S-R mappings (the "probe" mappings), recency and frequency were matched between the 4 S-R and 6 S-R conditions. As in Experiments 1 and 2, this was achieved by yoking one participant ($P_1$) with another participant ($P_2$). Firstly, a 6 S-R sequence (for participant $P_1$) was created, in which the probe transitions (AA, BB, AB and BA) occurred 4 times as often as all the other (nonprobe) transitions (see Figure 3.2). In order to create the 4 S-R sequence (for participant $P_2$), all E's and F's were replaced with C's and D's. In this manner, in the 4 S-R sequence, all 4 probe transitions and the CC, DD, CD and DC nonprobe transitions occurred twice as often as the other nonprobe transitions. So, stimulus recency was matched for probe stimuli between (yoked) subjects, and frequency was also matched within subjects (see Figure 3.2).

The order of number of S-R mappings (in the two halves of the session) and the order of stimulus set used for each half was manipulated between subjects. Furthermore, in the 4 S-R condition each participant was assigned one of 3 response sets, consisting of 2 responses of the left and right hand: 1) ring and middle finger; 2) middle and index finger; and 3) index and ring finger. This was done so that between subjects, each response was made equally often in the 4 and 6 S-R condition.

Prior to the start of the 4 S-R and 6 S-R parts, participants completed 1 practice block of 48 trials (plus one warm-up trial). In both the practice and the experimental session, the trial sequence was as follows: a 500 ms blank, followed by a 500 ms fixation dot, after which the stimulus appeared. The stimulus remained on the screen until a response was made. On incorrect trials, an error message remained on the screen for 1000 ms. The 4 S-R and 6 S-R part both consisted of 9 blocks of 48 trials (plus 1 warm-up trial) each, and the parts were separated by a 5 minute break. At the end of each block, participants were presented with a score that was based on the speed and accuracy of their responses. When participants improved on this score, a bonus point (£0.10 each) was awarded. In total, the session lasted approximately 50 minutes.

### Results

Very long (>2000 ms) and short (<200 ms) reaction times (RTs), trials following an error and warm-up trials were excluded from the data set (0.5% of the correct responses). Furthermore, response repetitions were excluded in all of the analyses (Kornblum, 1968), except for the recency analysis. As a result, the probe task trials were either an AB or a BA transition. As both of these trial types were also a hand switch, for the nonprobe data, hand repeats were also excluded from the analysis.

*Set-size effects* Set-size effects for probe and nonprobe stimuli are shown in Figure 3.3. A 2 (4 S-R or 6 S-R) x 2 (probe or nonprobe) repeated measures ANOVA was run on the mean correct RT data, in order to compare set-size effects for probe trials (following probe trials) and nonprobe trials (following nonprobe trials). Overall, this analysis revealed a reliable set-size effect of 99 ms, F(1,23) = 36.94, p < 0.001. A reliable two-way interaction demonstrated that the set-size effect was much larger for the nonprobe stimuli (150 ms) compared to the probe stimuli (52 ms), F(1,23) = 13.48, p = 0.001. RTs were also faster for probe stimuli (622 ms) compared to nonprobe stimuli (686 ms), F(1,23) = 19.65, p < 0.001. Additional ANOVAs revealed that the set-size effect was reliable both for probe stimuli, F(1,23) = 5.56, p = 0.027, and for nonprobe stimuli, F(1,23) = 53.74, p < 0.001. Hence, even though matching for recency and frequency reduced the set-size effect considerably (by 65%), a significant set-size effect still remained for probe stimuli, suggesting that the effect cannot be entirely attributed to a stimulus recency or frequency. [9]

Participants also made slightly more errors in the 6 S-R (5.9%) compared to the 4 S-R condition (5.0%), although this difference was not significant, F(1,23) = 1.33, p = 0.261. They made reliably fewer errors on probe (4.1%) compared to nonprobe (6.8%) trials, F(1,23) = 12.32, p = 0.002. Although the set-size effect was larger for the nonprobe (1.9%) compared to the probe stimuli (-0.1%) – as for RTs , the two-way interaction was not reliable, F(1,23) = 1.64, p = 0.213, and further ANOVAs revealed that the set-size effect was not reliable for probe and nonprobe stimuli separately, F(1,23) = 0.01, p = 0.935 and F(1,23) = 2.57, p = 0.122, respectively.

---

[9] This analysis was restricted to probe-probe and nonprobe-nonprobe sequences in case performance was also influenced by the recency/frequency of the immediately preceding trial. Without restrictions on the preceding trial type, the set-size effect was still much larger for the nonprobe stimuli (142 ms, F(1,23) = 48.31, p < 0.001) than for the probe stimuli (67 ms, F(1,23) = 10.91, p = 0.003), F(1,23) = 8.19, p = 0.009, though this difference is somewhat smaller (the effect for non probe exceeds that for probe stimuli by a factor of two rather than three). This suggests frequency/recency of the preceding trial may matter; hence the main analysis was restricted to probe-probe and nonprobe-nonprobe sequences.

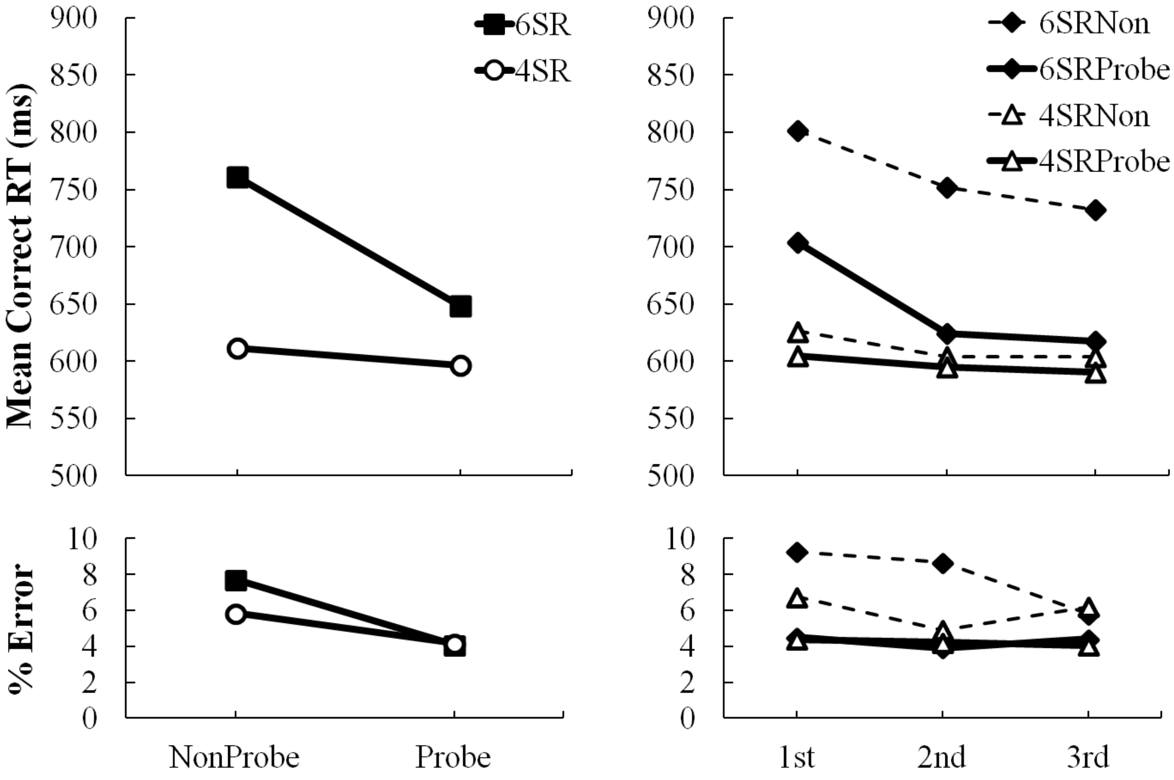


**Figure 3.3** Mean correct RT (top) and % error (bottom) data in Experiment 3, for 6 S-R and 4 S-R trials, plotted as a function of probe/nonprobe stimuli (left), and as a function of practice (right).

*Modulation of set-size effect by practice* In order to investigate potential effects of practice on the set-size effect, the data were split up into 3 parts of 144 trials each (first 3 blocks versus second 3 blocks versus third 3 blocks. Note that the sequential constraints described in the Method section were applied per block trio, so that each part contained equal numbers of all stimuli, transition types, etc.). A 3 (practice) x 2 (4 S-R or 6 S-R) x 2 (probe or nonprobe) repeated measures ANOVA was run on the mean RT and % error data. Only significant interactions with the linear component of the effect of practice are reported below.

Overall, the effect of set-size was significantly reduced with practice (from 137 ms in the first part to 78 ms in the third part, a reduction of 59 ms or 43%), F(1,13) = 19.04, $p < 0.001$. This reduction in set-size effect with practice was greater for the probe stimuli (from 99 ms to 27 ms, a reduction of 72 ms or 73%) than for the nonprobe stimuli (from 176 ms to 128 ms, a reduction of 47 ms or 27%). When the interaction between practice and number is tested separately for probe and nonprobe trials, a reliable reduction in set-size effect with practice is found for probe trials, F(1,23) = 14, p = 0.001, but not for the nonprobe trials, F(1,23) = 2.60, p = 1.21. However, as the three-way interaction between probe, number and practice is not reliable, this result should be interpreted with caution.

Still, a reduction in set-size effect with practice is interesting because it might suggest that, perhaps with more practice, the set-size effect might disappear altogether, especially for probe trials, for which the effect of practice on the set-size effect was largest. Indeed, for probe trials, when the first part was excluded from the analysis, the set-size effect was no longer reliable, F(1,23) = 1.68, p = 0.208. However, it was non-trivial in magnitude (28 ms) and appears asymptotic.

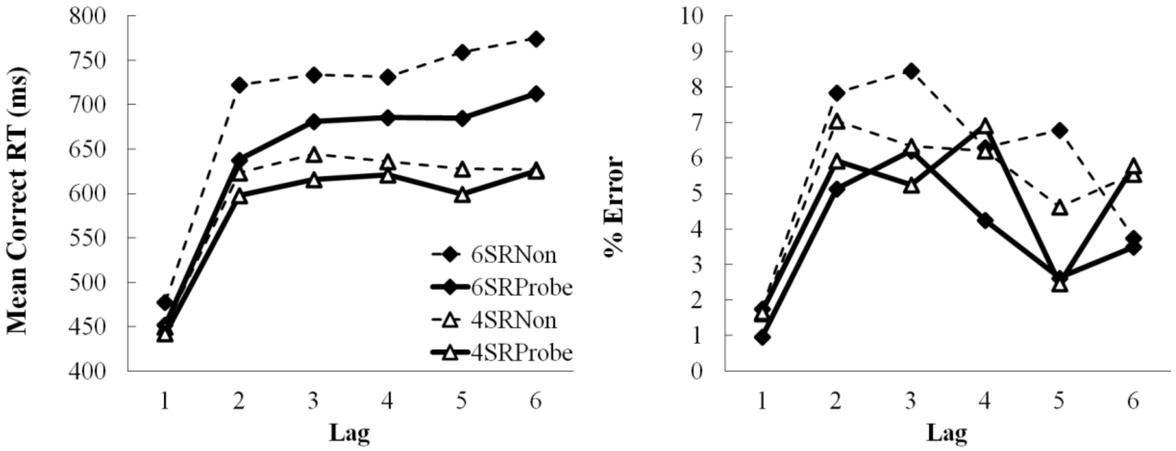


**Figure 3.4** Mean correct RT (left) and % error (right) data in Experiment 3, for 6 S-R and 4 S-R trials, plotted separately for probe/nonprobe stimuli as a function of lag.

*Recency analysis* RT and errors were also analysed as a function of lag. This analysis was restricted to trials with a lag up to 6 (with lag 1 being an immediate S-R repeat). The analysis included all probe and all nonprobe trials (there was not enough data to restrict this analysis to probe-probe and nonprobe-nonprobe sequences). The first thing to notice from Figure 3.4 is the massive (190 ms, F(1,23) = 267.64, p < 0.001) increase in RT from lag 1 to lag 2. That is, most of the effect of lag was due to immediate repetitions. This increase was also reliably larger in the 6 S-R condition (216 ms) than in the 4 S-R condition (164 ms), F(1,23) = 9.77, p = 0.005.

Of interest, though, is whether RTs continued to increase beyond lag 2. In a further analysis, lag 1 trials were excluded from the analysis. Only effects of and interactions with the linear component of recency will be reported below. There was a small (slope 8 ms) but significant recency effect beyond lag 2, F(1,23) = 9.50, p = 0.005. But this recency effect is only present in the 6 S-R condition (slope 14 ms, F(1,23) = 10.37, p = 0.004), and not in the 4 S-R condition, (slope 1 ms, F < 1) as reflected in a reliable two-way interaction, F(1,23) = 6.80, p = 0.016.

This result (an effect of lag beyond lag 2 in the 6 S-R but not the 4 S-R condition) is complicated by a speed-accuracy trade-off. As for RTs, error rates increased substantially (by 5.0%) from lag 1 to lag 2, F(1,23) = 65.87, p < 0.001, although in contrast to the RTs, this increase was not significantly larger in the 6 S-R (5.1%) than in the 4 S-R (4.8%) condition, F < 1. The outcome of the analysis investigating recency effects beyond lag 2 is less clear. The error lag effect opposed the RT lag effect: participants made fewer errors (slope -0.6%) with an increase in lag, F(1,23) = 12.17, p = 0.002. Though the interaction between number and lag was not reliable, F(1,23) = 1.83, separate ANOVAs revealed that this "inverse" recency effect was only reliable in the 6 S-R condition (slope -0.8%), F(1,23) = 28.04, p < 0.001, not in the 4 S-R condition (slope -0.4%), F(1,23) = 2.08, p = 0.162.

To summarise, in the 4 S-R condition, there was no reliable effect of lag beyond immediate S-R repetitions. The results were less clear for the 6 S-R condition: a reliable increase in RT beyond immediate stimulus repetitions was accompanied by a reliable decrease in error rates. Furthermore, the difference in slope between the 6 S-R and 4 S-R condition demonstrates that unequal proportions of more recent trials alone cannot explain the occurrence of set-size effects: had this been the case, there should have been no difference whatsoever between the 4 S-R and 6 S-R conditions once the data is plotted as a function of recency. So, frequency must also be important.

*Frequency analysis* The evidence just mentioned suggests frequency is important. The effect of frequency can also be assessed independently of the number of S-R mappings: half of the 4 S-R stimuli were responded to twice as often as the others (see Figure 3.2). The more frequent stimuli (638 ms) were responded to 34 ms faster than less frequent stimuli (604 ms), F(1,23) = 19.27, p < 0.001. The same trend was apparent in the errors (5.6% less frequent stimuli; 5.0% more frequent stimuli), though this difference was not significant, F < 1. As this manipulation of frequency effect was overall confounded with stimulus recency, the same analysis was also run averaging over lag position (restricted to lag positions 2 to 6, hence not including immediate stimulus repetitions) – this estimate of the frequency effect could not be affected by an inflated proportion of more recent stimuli for the more frequent stimuli. The result was very similar: RTs were 33 ms faster for the more frequent (610 ms) than the less frequent (643 ms) stimuli, F(1,23) = 14.43, p = 0.001. Again, the same trend was apparent in the errors (0.5% fewer errors for more frequent stimuli), though it was not reliable, F < 1. This frequency analysis confirms that frequency determines RT independently of set size.

*RT distribution analysis* The difference in set-size effects for probe and nonprobe trials described above suggests that stimulus recency and frequency contribute substantially to set-size effects. One other way in which the effects of stimulus recency in particular can be assessed involves examining the RT distribution. By analogy with Kornblum's (1968) theory which considers set-size effects to be caused by an inflated proportion of stimulus repeats with smaller set-sizes, the following prediction can be made: if set-size effects are caused by a larger proportion of more recent trials (resulting in faster responses) in the 4 S-R condition, then this should mostly be reflected in a change in the *shape* of the distribution. Specifically, the 4 and 6 S-R functions should then share the same minimum. In contrast, from the shape of the nonprobe functions (Figure 3.5) below it is clear that there is mostly a *general shift* in the 6 S-R function (demonstrating that participants are slower in the 6 S-R compared to the 4 S-R condition throughout the distribution); and the 4 S-R and 6 S-R nonprobe functions do not share the same minimum. Thus, the distribution analysis confirms the result of mean RT analysis of the recency matched probe trials: an increased proportion of more recent trials is not the sole cause of set-size effects.

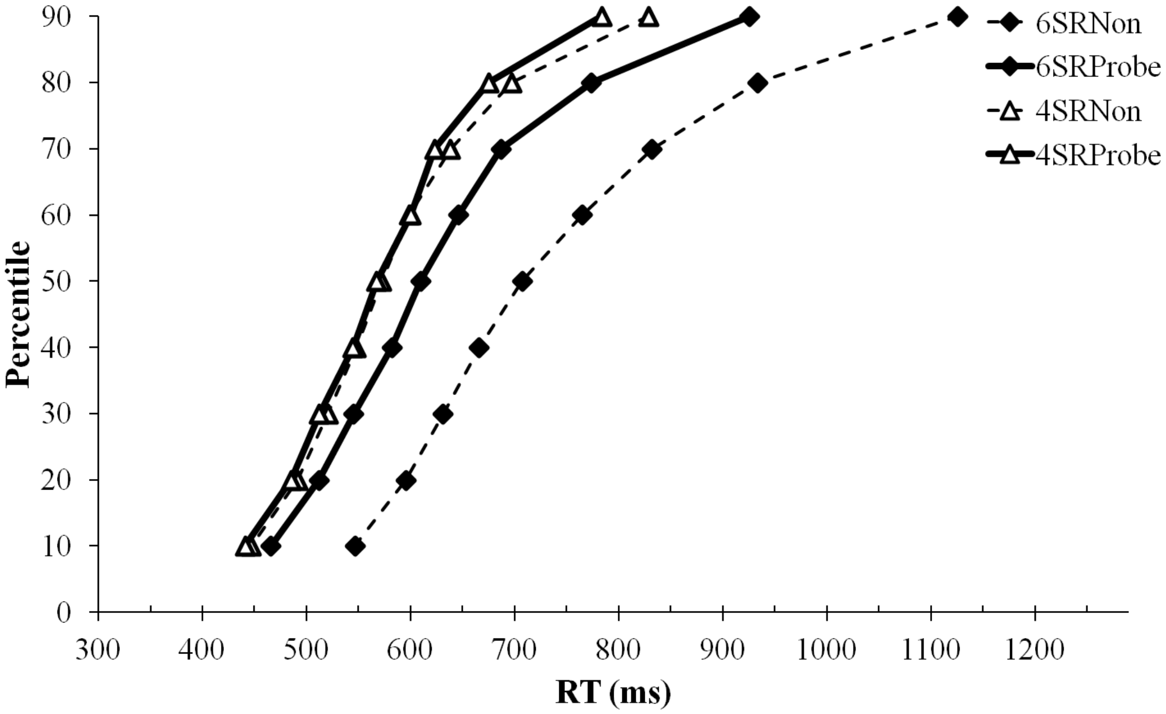


**Figure 3.5** Percentiles for probe stimuli and nonprobe stimuli in Experiment 3, in the 6 S-R and 4 S-R conditions.

## Discussion

This experiment set out to determine whether the set-size effect described by Hick's law is caused by the number of alternative responses, or whether it is the result of the effects of frequency and/or recency instead. To this end, participants were required to identify 6 colours and 4 shapes (or vice versa), in two separate parts of the experiment. For two of the stimuli in each part (the "probe" stimuli), frequency and recency of usage were matched between parts. It was predicted that if Hick's law describes an effect of recency and/or frequency, and not of set-size, then no set-size effect should be observed for the probe stimuli. But the results did not confirm this prediction: even though the set-size effect was approximately 3 times larger for nonprobe stimuli compared to probe stimuli, a reliable overall set-size effect of 52 ms remained for probe stimuli. This finding suggests that although stimulus frequency and recency are a source of set-size effects in 'uncontrolled' data sets, they are not the only source. Further analyses showed that immediate stimulus repetitions substantially contribute to set-size effects, as does stimulus frequency. Stimulus recency beyond immediate repetitions did not appear to have much of an effect. Finally, the RT distribution analysis showed that the 4 S-R and 6 S-R nonprobe functions did not share the same minimum, consistent with the idea that inflated proportions of more frequent/recent trials alone cannot explain set-size effects. Altogether, the analyses clearly demonstrate that even when recency and frequency of usage are matched, retrieving an S-R mapping from among alternatives is influenced by the number of competitors.

As it has previously been demonstrated that practice modulates set-size effects (Hale, 1968; also see Longstreth et al., 1985), the data were also analysed as a function of practice. Consistent with previous findings, this analysis revealed that overall, the set-size effect decreased as a function of practice. For the nonprobe stimuli, the set-size effect decreased by 27% in the last third of the half-session, compared to the first third. For the probe stimuli, however, this reduction was much larger (73%). Indeed, when the first third was removed from the analysis, the set-size effect for probe trials was no longer reliable. However, at 28 ms, the set-size effect had still not disappeared entirely, and appeared asymptotic.

The set-size effect obtained for probe trials in this experiment can be contrasted with the results of Experiments 1 and 2, in which the effect of the number of alternative task-sets on task-set retrieval disappeared when the tasks were matched for recency and frequency. This discrepancy demonstrates that the process of selecting one S-R mapping amongst alternatives (Experiment 3), and the process of retrieving one task-set amongst others (Experiments 1 and 2) are not subject to the same constraints or capacity limits. This difference is consistent with an account of procedural memory which distinguishes between a component holding all potentially relevant task-sets, some in an active state, with no limit on the total pool of activation (or none that affects retrieval time), and a capacity limited component, holding only the currently operative task-set (e.g. Oberauer, 2009). Such a theory makes two predictions: 1) The time consumed by a task change is not influenced by the number of other task-sets, because these are represented in the (capacity unlimited) part of LTM; this prediction was confirmed by Experiments 1 and 2; 2) The time required the activate the appropriate S-R mapping should increase as a function of the number of competitors (within the same task-set), because the currently operative task-set is held in a capacity limited buffer; this prediction is confirmed by Experiment 3.

Finally, the results obtained in this experiment argue against Kornblum's (1968) suggestion that set-size effects are caused by the inverse correlation between the proportion of immediate S-R repetitions and set-size. Specifically, response repetitions were excluded from all the analyses reported here, and still significant set-size effects remained. These results are also relevant with regards to Hyman and Umiltà's analysis of Kornblum's experiment. According to Hyman and Umiltà, the most likely reason Kornblum did not obtain a set-size effect was because of the short RSI (140 ms) employed in that study. However, the results obtained in the present experiment (using an RSI of 1000 ms) show that set-size effects may still be obtained using RSIs much shorter than the 9 to 10 second interval proposed by Hyman and Umiltà.

To summarise, Experiment 3 found that a reliable set-size effect remained even when stimuli were matched for recency and frequency. This result can be contrasted with that of Experiments 1 and 2, in which the number of tasks effect was eliminated by matching for recency and frequency. Together, these results show that the process of retrieving a task-set (Experiments 1 and 2) or selecting an S-R mapping (Experiment 3) from among competitors are not subject to the same capacity constraints. Hence these results provide an additional reason for distinguishing between the level at which task-sets (packages of S-R rules) are represented, and the level at which S-R rules are represented.

# Chapter 4: The effect of task-set complexity on preparation for a task switch: Evidence for a limited capacity component of procedural working memory.

## Introduction

The experiments reported in Chapter 2 found that the time consumed by task-set retrieval is not influenced by the number of alternative task-sets in play, provided that frequency and recency of usage are matched. The discrepancy between this result and other types of memory retrieval (which do appear to be influenced by the number of alternatives) led us to consider two opposing explanations: either task-set retrieval is somehow special, or set-size effects in other kinds of memory retrieval can also be explained by the recency or frequency with which a stimulus is retrieved. Chapter 3 examined the latter explanation as applied to Hick’s law, but rejected it: in Experiment 3, the effect of number of S-R rules (though much reduced) was not eliminated for stimuli matched for frequency and recency. This contrast between the effect of number of alternatives on retrieval of one response within a task-set (Hick's law), and the lack of an effect of number of alternative task-sets on retrieval of a task set, provides a new and additional reason for taking seriously a model of procedural working memory (PWM) which is capacity-limited with respect to the current task-set, but not with respect to the ability to maintain task-sets "in play". Oberauer's (2009) model has this property. Multiple task-sets in play are represented without limit in the activated part of LTM, but only the currently operative task-set is represented in the limited-capacity “bridge”. In the present chapter I return to the task-cueing paradigm, and report three experiments that manipulated the complexity of the upcoming task by varying the number of S-R rules required to specify it. Broadly, these experiments explore two possibly conflicting intuitions. On the one hand, a more complex task should benefit more from advance preparation. On the other hand, preparation can only be effective to the extent to which there is capacity to accommodate the S-R rules for the upcoming task. These experiments may be seen as exploring the process of loading Oberauer's "bridge" and its capacity limits.

There is quite a history of research trying to quantify the capacity limited component of *declarative* working memory – in terms of chunks (Miller, 1956; Cowan, Chen & Rouder, 2004), or (for verbal WM) articulatory complexity (Caplan & Waters, 1994) or duration (Baddeley, Thomson & Buchanan, 1975), or (for visual short term memory) number of visual objects (Luck & Vogel, 1997, see Cowan et al., 2004 for review). Much less is known about the capacity limits of *procedural* working memory — or indeed about its other properties. One key question concerns the unit of capacity: how should one assess the complexity of a task in relation to its demand on PWM capacity. For the sort of one-step choice reaction time tasks typical of research on task-switching, the number of S-R rules required to specify the task seems an obvious measure of complexity. Hick’s law suggests that even increasing the complexity from two to three or four rules is pushing up against a functional capacity limit, i.e., that three rules cannot be represented as accessibly as two.

The idea that the currently operative task-set is held in a capacity-limited state of representation is also supported by Duncan and colleagues’ research on goal neglect (e.g. Duncan, Emslie, Williams, Johnson & Freer, 1996; Duncan, Parr, Woolgar, Thompson, Bright, Cox, Bishop & Nimmo-Smith, 2008). Duncan et al. (1996) presented participants with two adjacent streams of stimuli and instructed them to respond to one stream (e.g. the one on the left) until a cue instructed them to change side. Goal neglect occurred when participants ignored the cue and continued to respond to stimuli on the left. But despite apparently ignoring this task requirement, participants were still able to describe what they were supposed to do. This discrepancy between adequate declarative memory for the instruction and failure to implement it as an effective procedure provides further evidence for a declarative/procedural WM distinction. Moreover, goal neglect can be modulated by the complexity of the instructions: in Experiment 3, Duncan et al. (2008) manipulated the complexity of the instructions given to two groups of participants. Both groups were instructed to do two tasks, one in a first, another in a second block of trials. In the “full instructions” condition, participants were informed of the task requirements for both the first and the second block before beginning the first. In the “reduced instructions” conditions, participants were informed of the task requirements for the second block only after completion of the first block. It was found that participants with the more complex task description (“full instructions”) more frequently exhibited goal neglect in the first block. According to Duncan et al. (2008), the more complex “task model” exceeds capacity limits, resulting in the loss of some (vulnerable) components of the task model.

In task switching experiments, only the currently operative task-set is thought to be held in the bridge. Hence, when the task changes, the contents of the bridge must be updated, and it is this process which is captured by the switch cost (Oberauer, 2009). Adopting this view, one might expect the switch cost to be affected by the complexity of the task set. Two previous studies provide preliminary support for this assumption.

Rubinstein, Meyer and Evans (2001) were among the first to explicitly consider the role of PWM in task switching. The authors distinguish between two distinct stages: goal shifting and rule activation. Goal shifting involves identifying the current task goal and uploading it into working memory, and this stage can, they assume, occur prior to stimulus onset. The rule activation stage involves enabling the current (and disabling the previously relevant) set of SR mappings, and is assumed to be initiated only after stimulus identification.[10] Rubinstein et al. (2001) reasoned that if the switch cost (in part) reflects the time consumed by rule activation, then increasing the complexity of the task-set should increase the switch cost. In Experiment 1, participants were required to sort stimulus cards according to the shape, size, shading or numerosity of geometric objects (each task had four values). In the low complexity condition, participants alternated predictably between unidimensional sorts (e.g. IF shape = circle THEN response = 1). In the high complexity (bidimensional) condition required participants to sort the cards according to conjunctions of two rules (e.g. IF shape = star AND numerosity = two THEN response = 1). Stimulus discriminability was also manipulated between tasks: stimuli were easier to discriminate based on numerosity and shape than on size and shading. Rubinstein et al. (2001) found that mixing cost (greater mean RT on task-repeat trials in alternating than in single task blocks) increased with rule complexity, but not with a decrease in stimulus discriminability. In Experiment 2, the same procedure was used but, instead of perceptual classifications, arithmetic tasks were used: participants switched either between addition and subtraction or between multiplication and division, and the mixing cost was larger for the multiplication-division condition. The latter condition was deemed more complex because multiplication and division require the rules of addition and subtracting plus “other supplementary ones” (Rubinstein et al., 2001, p. 778). This would be true in a randomly drawn pool of operations, because multiplication and division require carrying digits over more often. But Rubinstein et al. (2001), fixed the proportion of problems that required carrying over to 83%, so that one might think that performance in both

---

[10] The rationale for this assumption is unclear, except that it provides an account of the "residual cost".

conditions would require (on average) the same number of steps or production rules. Thus it is not obvious why the mixing cost should have been larger for the multiplication-division condition.

Moreover, a subsequent attempt by Hübner, Kluwe, Luna-Rodriguez and Peters (2004) at a conceptual replication of Rubinstein et al.'s (2001) Experiment 1 failed. In this experiment, participants switched between a unidimensional letter identification task and a bidimensional colour-numerosity identification. No difference in switch cost was found, except for when stimulus repetitions were included, causing Hübner et al. to argue that Rubinstein’s finding resulted from a stimulus repetition confound. Although this may explain the result of Rubinstein et al.’s Experiment 1 it cannot explain the results of Experiment 2, in which many different stimuli were used, so that stimulus repetitions were rare. Another difference between these experiments which may explain the discrepancy in results is that Hübner et al. (2004) had participants switch between tasks of different complexity (within-block manipulation), whereas Rubinstein et al. (2001) varied complexity between blocks.

In the same article, Hübner et al. (2004) manipulated complexity in a different manner by varying the number of S-R rules. In Experiment 1A a word cue presented prior to the onset of a letter instructed participants to respond to either its identity or its colour. One task had four S-R mappings (e.g. A, B, C and D mapped to different keys) and the other had two (e.g. yellow and green mapped to two keys) so that task-set complexity was manipulated within-block. They also manipulated CSI. They found that switch costs were larger when switching towards the less complex (2 S-R mapping) task, and this effect was not reliably modulated by CSI. But as the 2 S-R and 4 S-R tasks were performed equally often, the individual values were responded to twice as often for the 2 S-R task. There are a number of studies showing asymmetrical switch costs between tasks of different strength (Allport, Styles & Hsieh, 1994; Monsell, Yeung & Azuma, 2000; Rubinstein et al, 2001; Yeung & Monsell, 2003a,b), including a demonstration that such strength can be manipulated merely via recency of practice (Yeung & Monsell, 2003a). Hübner et al. attempted to eliminate the confound with recency/frequency but were unable to do so without introducing other confounds: in Experiment 2A the 4 S-R task occurred twice as often as the 2 S-R task so that each S-R rule was exercised with equal frequency. But this is also not ideal because now one task (the 4 S-R task) was performed twice as often, and presumably some preparation also takes place at the task-set level (e.g. Miyake, Emerson, Padilla & Ahn, 2004). Moreover, this manipulation resulted in the stimulus values of the 2 S-R task occurring much more frequently as the irrelevant value. Hübner et al. (2004) addressed the latter problem in Experiment 2b, in which 75% of 4 S-R task trials were univalent stimuli. But, although only bivalent stimuli were included in the analysis, the inclusion of the univalent stimuli might affect the perceived difficulty of that task. Finally, a third experiment contrasted a 3 S-R task-set with a 6 S-R task-set. Although this experiment was somewhat consistent with their previous findings (larger % error switch cost for the less complex condition), the RT data looked strange: a very small switch cost (20 ms) was not reduced with preparation, nor was it modulated by the number of S-R rules.

Although the finding of a larger switch cost for the less complex condition was relatively consistent, then, Hübner et al.'s results were totally unclear with regards to the CSI manipulation: in Experiment 1 this switch cost difference appeared to manifest itself at the long CSI and in Experiment 2 at the short CSI. Furthermore, the interpretation of this effect is also problematic: it remains unclear whether it is easier to switch towards that task, or harder to switch away from the other tasks. Finally, it also remains unclear what caused the discrepancy between Hübner et al.'s (2004, Experiment 4) and Rubinstein et al.'s (2001) results.

The experiments reported in this chapter (Experiments 4 to 6) attempted to manipulate the complexity of a task-set while avoiding the confounds inherent to the manipulations used in the studies reviewed so far. Complexity was manipulated by varying the number of S-R rules required to *specify* a task-set, while keeping the number of stimuli and the number of responses to individual stimuli constant. Moreover, in any block of trials, participants switched between two tasks of the same complexity, so the issue of asymmetrical costs did not arise and the number of times that participants responded to each value of the relevant stimulus dimension was identical for the different levels of complexity. In Experiment 4, this was achieved by manipulating whether all of three S-R rules for each task could apply in the current block, or a subset of two. Experiments 5 and 6 manipulated number of rules in a different way. Each stimulus was a composite of two elements, one for each task. The eight elements for each task came from two familiar categories. Four were assigned to each of two responses either according to their category (so that the task was specified by only two categorical S-R rules[11]) or arbitrarily (so that the task required eight stimulus-specific S-R rules[12] to specify it).

We also varied CSI to explore the two contrasting intuitions mentioned at the

---

[11] or one rule if we assume that half the stimuli are categorised by default

[12] or four rules if we assume that half the stimuli are categorised by default

beginning. On the one hand, one might expect that the more S-R rules have to be "loaded" into the bridge the longer it will take, and the greater the reduction in switch cost one would see between unprepared and fully prepared switching; with a sufficient interval, however, any effect of complexity should be eliminated (cf. Mayr & Kliegl, 2000). But this prediction makes two assumptions: (a) that there is "room" for all the S-R rules in the bridge; and (b) that loading more rules takes longer. Suppose that neither of these is the case. If the capacity of the bridge is smaller than the number of rules (so that some have to remain in the active part of LTM until called for) then no amount of preparation will eliminate the effect of complexity. And if as many S-R rules as can be loaded can be loaded as a single chunk rather than seriatim — as has been claimed for retrieving a list into declarative WM (Wickens, Moody & Dow, 1981; Conway & Engle, 1995; Oberauer, 2005) then the switch cost on unprepared trials should be of similar size for small and large S-R sets. Under the first pair of assumptions, we would predict large complexity effects at a short CSI and none at a long one. Under the second pair, we would predict the opposite.

## Experiment 4

Participants were cued with a spoken word to perform one of two tasks (described below) requiring attention to colour or letter identity when a string of coloured letters was presented. The cue-stimulus interval (CSI) was 150 ms (allowing little or no time for preparation) or 1200 ms (ample for preparation). Task-set complexity was manipulated by varying the number of stimulus-response mappings between two (2-choice condition) or three (3-choice condition). For reasons mentioned above, the number of S-R mappings was manipulated between blocks, and the experiment was designed so that the total number of responses to each value of the relevant stimulus dimension was identical in the 2-choice and 3-choice conditions.

To achieve these goals, the stimulus was a string of six coloured letters (see Figure 4.1 for examples). Every string contained four tokens of one colour, and two tokens of another. It also contained (independently of their colours) four tokens of one letter and two of another. One task was to identify the predominant colour in the string (green in the top-left example); the other task was to identify the predominant letter (T in the same example). In the 3-choice blocks of trials, any pair of two colours from the three colours could occur on a trial, each assigned to one of three response keys, and any pair of letters from the three, likewise. Hence, although each stimulus contained only two colours and two letters, the participant did not know in advance which they would be, and should prepare all 3 S-R rules when the relevant dimension was cued. In the 2-choice condition, only two of the three colours and two of the three letters were used on every trial in a block, and the participant knew which two they would be for that block. Hence they needed to prepare only those two S-R mappings for each task.

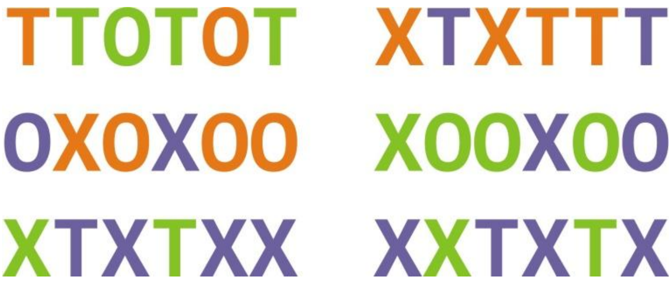


**Figure 4.1** Example of stimuli used in Experiment 4. Congruent stimuli are illustrated on the left (dominant features: green and T; orange and O; purple and X); incongruent stimuli on the right (orange and T; green and O; purple and X).

## Method

*Participants* 24 participants (aged between 18 and 32 [M = 23]; 8 male and 16 female) provided informed consent prior to taking part. All participants were paid between £10 and £16, depending on the speed and accuracy of their performance.

*Stimuli and cues* Each six-letter stimulus contained two of three colours (green, orange and purple) and two of three letters (T, O and X) (see Figure 4.1 for examples). We used 12 different 'combinations' of letters or colours. An example of a letter combination is 2 Xs and 4 Os. For each combination the letter (or colour) tokens could be ordered in 72 different ways, resulting in (12 x 72 =) 864 unique stimuli.[13] Participants responded to green and T with the index finger, to orange and O with the middle finger and to X and purple with the ring finger, of the right hand, using the V, B and N keys on a standard PC keyboard. Before the onset of each stimulus, an auditory cue instructed participants to perform the colour or the letter task. The auditory cues were the words “colour” or

---

[13] In total there were 36 different combination (3 dominant colour x 2 subordinate colour x 3 dominant letter x 2 subordinate letter = 36) but we only used the 12 stimulus types featuring in the 2-choice block types (as described below): 3 block types x 4 stimulus types = 12 stimulus types. For each combination 72 different colour/letter distributions were used: with 2 letters of the dominant value and 4 letters of the subordinate value, all 6 letters within each stimulus type could be ordered in 15 unique ways. Out of 15, 6 instances with immediately adjacent subordinate values (e.g. 2 Xs followed by 4 Ts) were removed (too easy). The same applied to colour distributions, resulting in 9 different letter and 9 different colour distributions. Within these 81 combinations the 9 instances of an identical distribution of colours and letters were removed resulting in 9 colour x 8 letter distributions = 72. 72 distributions x 12 stimulus types = 864 unique stimuli.

"paint" for the colour task, and "letter" or "symbol" for the letter task. The cue-stimulus interval (onset to onset) was varied between blocks, and was either short (150 ms) or long (1200 ms).

*Design and procedure* Task-set complexity was manipulated between blocks, by varying the number of potentially relevant S-R mappings. In the more complex 3-choice condition, stimuli could contain any pair of the three colours, and any pair of the three letters. Hence, whatever the task, participants should prepare all three of its S-R mappings. In the less complex 2-choice condition, only two letters and two colours could appear in any one block. As a result, participants had to prepare only two S-R mappings in this condition.

All participants completed two sessions on consecutive days. Sessions were identical on each day: both were split into two parts (a 2-choice part and a 3-choice part) consisting of six blocks each. The order of parts was counterbalanced between participants. In the 3-choice condition, the response requirements were identical in each block. In the 2-choice condition, there were three different block types, corresponding to the three potential subsets of two out of three S-R mappings. The S-R mappings required in block type 1 were purple, orange, X and O (i.e. middle and ring fingers only); orange, green, O and T in block type 2 (i.e. index and middle fingers only); and green, purple, T and X in block type 3 (index and ring fingers only). In this manner, participants responded to each colour and each letter equally often in the 2-choice and 3-choice condition. The order of 2-choice block types was balanced between participants.

At the start of each session, participants completed four single task practice blocks (two for each task, using all 3 S-R rules) of 18 trials each, and four switching practice blocks (two for each CSI) of 36 trials each. In the experimental sequence, each block pair was also preceded by a switching practice block of 24 trials, to enable the participant to adjust to the subset of responses to be used in the next block pair. To equate for practice, participants also completed a similar practice block prior to each 3-choice block pair. All twelve experimental blocks (6 3-choice blocks and 2 blocks of each 2-choice block type) consisted of 36 trials, plus a filler trial at the start of each block (432 trials per session: 3 colour values x 3 letter values x 3 switch/repeat x 2 task x 2 cue x 2 CSI x 2 2-choice/3-choice = 432).

The trial sequence was as follows: A blank screen (presented for 200 ms in long CSI blocks and 1250 ms in short CSI blocks) was followed by the 500 ms presentation of a fixation cross. The onset of the auditory cue (whose duration was ~350 ms) preceded the stimulus onset either by 150 ms (short CSI) or 1200 ms (long CSI). In this way, the interval between the response and the subsequent stimulus onset was kept constant at 1900 ms (to avoid confounding preparation time with time since the previous task performance). When the wrong key was pressed, the word "Error!" was displayed on the screen for an additional 1000 ms. At the end of each block, participants were presented with a score (the mean correct RT in ms divided by 10 + 5 for each error). When participants improved on this score on the next block with the same CSI, bonus points (£0.20 each) were awarded.

## Results

Trials following an error were excluded from analysis, as were the 0.5% of correct responses with very short (< 200 ms) or very long (> 3000) reaction times (RTs).

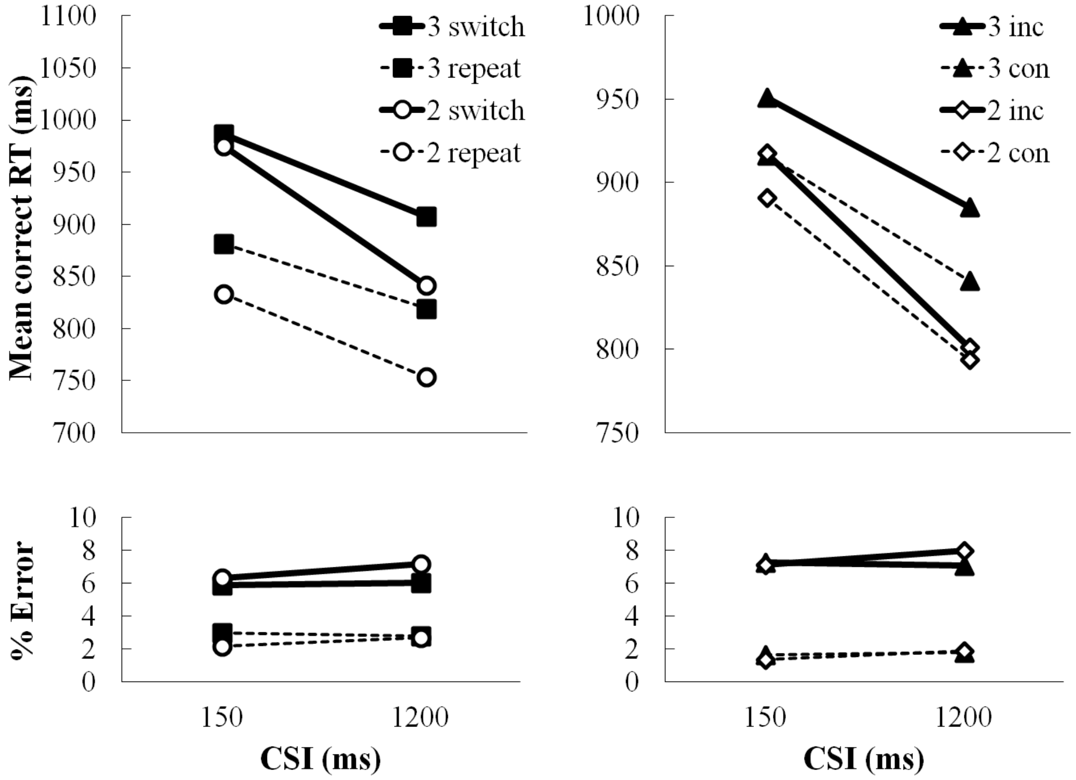


**Figure 4.2** Mean correct RTs (top) and % error (bottom) data in Experiment 4, plotted separately as a function of switch/repeat (left) and congruency (right).

### *Effect of task-set complexity on the switch cost and its reduction with preparation*

Figure 4.2 shows the mean correct RTs and error rates for 2-choice and 3-choice switch and repeat trials, as a function of CSI. The most important finding was that the reduction in switch cost (RISC) with preparation was larger in 2-choice choice blocks than in 3-choice blocks. A 2 (2-choice or 3-choice) x 2 (100 ms or 1200 ms) x 2 (switch or repeat) x 2 (congruent or incongruent) repeated measures ANOVA showed that the relevant interactions were reliable.

Overall, participants were 48 ms faster in 2-choice than 3-choice blocks, F(1,23) = 10.62, p = 0.003, 106 ms faster on repeat than switch trials, F(1,23) = 156.66, p < 0.001, and 89 ms faster in long CSI than short CSI blocks, F(1,23) = 26.42, p < 0.001. The CSI effect was larger in 2-choice (107 ms) than in 3-choice blocks (71 ms), F(1,23) = 8.78, p = 0.007, and on average the switch cost reduced from 124 ms in the short CSI to 88 ms in the long CSI, a highly reliable RISC of 36 ms or 29%, F(1,23) = 16.28, p = 0.001, indicating successful task-set preparation.

As mentioned above, the RISC was larger in 2-choice blocks, in which the switch cost reduced from 142 ms in the short CSI to 88 ms in the long CSI (RISC of 54 ms or 38%) than in 3-choice blocks, in which the switch cost reduced from 105 ms in the short CSI to 88 ms in the long CSI (a RISC of 17 ms or 16%), F(1,23) = 4.90, p = 0.037 (three-way interaction).

For the error rates, the same ANOVA found a reliable switch cost of 3.7 %, F(1,23) = 37.38, p < 0.001. The switch cost was, on average, also larger in 2-choice (4.3%) than in 3-choice blocks (3.1%), F(1,23) = 5.65, p = 0.026. Although the error graph might suggest a speed accuracy trade-off (slightly more errors in 2-choice blocks in the long CSI – perhaps suggesting a slight lack of caution in that condition), the three-way interaction did not approach significance, F < 1.

### *Effect of task-set complexity on the congruency effect and its reduction with preparation*

The larger RISC for the less complex task-set suggests that task-set preparation is more effective in that condition. Another index of the successful preparation of a task-set is its susceptibility to interference, as measured by response congruency effects (slower RTs for stimuli which afford a different response in both tasks). If the less complex task-set is better prepared, this might also be reflected in smaller congruency effects at a long CSI in that condition.

Figure 4.2 shows the mean correct RTs and error rates for 2-choice and 3-choice congruent and incongruent trials, as a function of CSI. The RT ANOVA described above found a congruency effect of 28 ms, F(1,23) = 18.81, $p < 0.001$. The congruency effect was also larger on switch (41 ms) than on repeat trials (15 ms), F(1,23) = 9.04, p = 0.006, and marginally smaller in 2-choice (17 ms) than in 3-choice blocks (39 ms), F(1,23) = 3.93, p = 0.059. As shown in Figure 4.2, the congruency effect was reduced with preparation only in 2-choice blocks (from 27 ms in the short CSI to 8 ms in the long CSI), and not in 3-choice blocks (35 ms in the short CSI and 44 ms in the long CSI). Although the three-way interaction was not reliable, F(1,23) = 1.52, p = 0.229, separate ANOVAs for each CSI showed the congruency effect was reliably smaller for 2-choice trials in the long CSI, F(1,23) = 6.39, p = 0.019, but not in the short CSI, $F < 1$. The four-way interaction was not significant, $F < 1$.

For the errors, there was a congruency effect of 5.7%, F(1,23) = 35.12, $p < 0.001$, which was also larger on switch (8.7%) than on repeat trials (2.7%), F(1,23) = 23.44, p = 0.001. The four-way interaction was just reliable, F(1,23) = 4.28, p = 0.050, but as the three-way interaction (complexity x CSI x switch) was not reliable for either congruent, F = 1.55, or incongruent trials, F = 1.03, this interaction is difficult to interpret.

*RT distribution analysis* Examination of RT distributions in task-switching experiments typically shows a larger RT variance on switch than on repeat trials, with an increasing difference between switch and repeat trials at longer RTs; the fastest RTs can show little difference between switch and repeats. This suggests that task-set preparation fails on a proportion of trials (De Jong, 2000), or at least that there is a large range of preparation effectiveness on task-switch trials.

The larger RISC effect we found in 2-choice blocks is consistent with a capacity so limited that it cannot accommodate even 3 S-R rules in as prepared a state as two. One possible realisation of this state of affairs is that only 2 S-R mappings can be fully prepared on each trial. Hence on a subset of trials in the 3-choice condition the stimulus-relevant rule would not be in this state of preparation (leading to slow RTs), while on the remaining trials it would (leading to RTs just as fast as in the 2-choice condition). To test this possibility we examined the RT distributions for the switch/repeat and short/long CSI conditions in the 2-choice and 3-choice blocks, pooling over congruency. CDFs were computed by rank ordering the correct RTs (for each switch/repeat x complexity condition separately) and computing deciles ($10^{th}$-$90^{th}$ percentiles in steps of 10). In the interest of clarity I have captured the difference between the highest and lowest percentiles in a fast-slow analysis, which averages over the 10th-40th percentiles (fast RTs) and the 60th-90th percentiles (slow responses). RT switch costs for fast and slow responses in 2-choice and 3-choice blocks are plotted as a function of CSI in Figure 4.3.

For the faster responses, the difference between the 3-choice RISC (42 ms or 48%) and the 2-choice RISC (51 ms or 49%) is negligible, $F<1$, suggesting that on this proportion of trials, preparation was equally effective in the 2-choice and 3-choice conditions. But for slower responses, task-set preparation was only effective in 2-choice blocks (RISC of 54 ms or 30%) and not in 3-choice blocks (RISC of 2 ms or 1%), $F(1,23) = 4.75$, $p = 0.040$. The four-way interaction approached significance, $F(1,23) = 3.54$, $p = 0.073$.

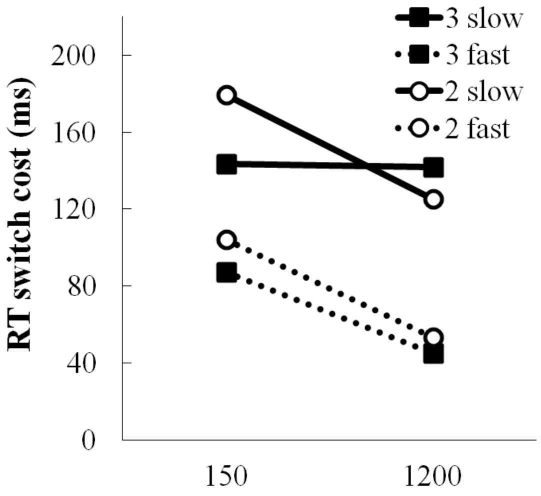


**Figure 4.3** RT switch cost (ms) in the 3-choice and 2-choice conditions in Experiment 4, for fast (average of 10th-40th percentile) and for slow responses (average of 60th-90th percentile), as a function of CSI.

*Summary* This experiment sought to determine how task-set complexity (as defined by the number of S-R mappings) affects the ability to prepare for an upcoming change of task. We found that the reduction in switch cost with preparation was substantially larger in 2-choice blocks, suggesting that more is gained by preparing or “activating” a task-set of simpler description. More effective preparation in 2-choice blocks was also reflected in a smaller congruency effect on prepared trials, suggesting reduced susceptibility to interference in that condition.

The general reduction in switch cost is consistent with theories (e.g. Rubinstein et al., 2001; Oberauer, 2009), which assume the currently relevant task-set must be held in a limited capacity component of PWM (Oberauer's “bridge”), and that updating the contents of the bridge can be done in advance of the stimulus. The results suggest that the effectiveness of task-set preparation is compromised when the complexity of a task-set is increased from 2 to 3 S-R mappings. One interpretation is that PWM can accommodate only 2 S-R mappings – at least in a fully active state – so that preparation in the 3-choice condition is only partial, in that one of the S-R mappings cannot be represented in such an available state. The distribution analysis is consistent with this. Switch costs for the faster responses show equivalent benefits of task-set preparation in the 2-choice and 3-choice condition, while the slower responses suggest inability to benefit from preparation in the 3-choice condition. The equivalence of the RISC effect for fast responses also argues against participants just not trying so hard to prepare in the 3-choice blocks.

However, the way we manipulated the number of S-R rules in this experiment is open to some objections: in the 2-choice condition, participants may have been able to prepare for a specific visual discrimination (e.g. purple versus green; T versus O), whereas this was not possible in the 3-choice condition; this might explain the greater effectiveness of preparation in the 2-choice condition (if not the fast-slow results). Also, although the overall frequency of responses was the same in the 2- and 3-choice conditions, within a block it is necessarily the case that the recency and frequency of responses was greater in a 2-choice than in a 3-choce block. Moreover, although there were 864 stimuli in the whole experiment, so that the recency of a particular string is not an issue, there were only 12 stimulus combinations (dominant colour x subordinate colour x dominant letter x subordinate letter), and they repeated with shorter lags in the 2-choice condition.

Hence in the remaining experiments (Experiments 5 and 6), we devised a paradigm in which task-set complexity could be varied while keeping the number of responses and the stimulus set size constant.

## Experiment 5

Experiment 4 found the effectiveness of task-set preparation was much reduced when task-set complexity was increased from 2 to 3 S-R mappings per task. Our interpretation was that the limited capacity component of PWM (Oberauer’s “bridge”) can only maintain 2 S-R rules in an optimal state of accessibility. But varying task-set complexity via the number of response rules resulted in some (potential) confounds, as described above.

To avoid such confounds, we now varied task-set complexity in such a way as to keep the number of stimuli and responses constant. In Experiments 5 and 6, for each task four stimuli were mapped onto each of two response keys, and these mappings were either arbitrary or governed by a categorical rule. For example, in the category-rule condition of a country classification task, some participants responded to four European countries (France, Italy, Spain and Germany) with the left response key, and four Asian countries (China, Japan, India and Thailand) with the right response key. These participants could this prepare for this task by activating just two[14] S-R rules (Europe → left; Asia → right). In the arbitrary-rule condition, four country names were drawn from a pool of sixteen to map onto each response key in such a way as to evade description by a simple category rule, so that participants had to prepare all eight[15] S-R rules instead (see Figure 4.4 for examples).

Our interpretation of the results of Experiment 4 was that task-set preparation was more successful for a less complex task-set, because for the more complex task-set the bridge cannot maintain all the rules in the same active state of representation simultaneously. Hence here too we expected a larger reduction in switch cost with preparation for the less complex category-rule condition than for the more complex arbitrary-rule condition. One limitation of Experiment 4 was that with only two CSIs, we could not be sure that 1200 ms was adequate for asymptotic preparation in this particular case (although RISC functions are normally asymptotic well before then – see Monsell & Mizon, 2006). Hence in the present experiment we used four CSIs (150 ms, 400 ms, 800 ms and 1300 ms), in the expectation that we would see little difference between the last two. In Experiment 4 more effective preparation for the less complex task-set also manifested itself in a reduced susceptibility to interference at a long CSI; hence we also predicted a larger reduction in congruency effect in the less complex category rule condition.

## Method

*Participants* 24 participants (aged between 18 and 31 [M = 20]; 5 male and 19 female) provided informed consent prior to taking part. All participants were paid between £12 and £17, depending on the speed and accuracy of their performance.

---

[14] Or one, if one hand is mapped by default (e.g. "if Europe, left, otherwise right").

[15] Or four, if one hand is mapped by default (e.g. "China, Spain, India, Brazil → left, otherwise right").

*Stimuli and cues* Two task pairs were used. In the country-shape task pair, the stimulus consisted of one of sixteen country names (4 European, 4 Asian, 4 African and 4 South-American) surrounded by one of sixteen shapes (4 shapes with no corners, 4 with two corners, 4 with three corners and 4 with four corners), resulting in 256 unique stimuli (see Figure 4.4 for examples). Each participant saw only 64 of these combinations (8 different country names surrounded by 8 different shapes). In the category-rule condition, the 4 country names mapping onto the left and right responses belonged to the same continent (e.g. European countries map onto the left response, and Asian countries map onto the right response) and the 4 shapes mapped onto each response would have the same number of corners (shapes with two corners map onto the left response, and shapes with three corners map onto the right response). In the arbitrary-rule condition, participants were unable to use a categorical rule to decide whether a country or a shape should be responded to with the left or right key (one country from each continent and one shape from each corner subset were mapped onto each response).

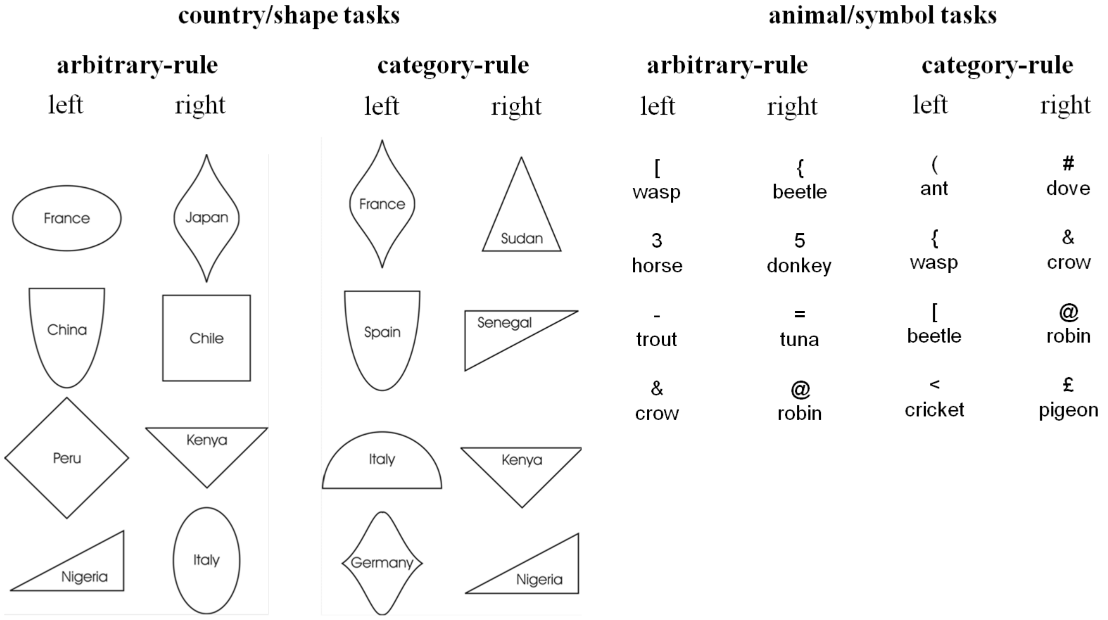


**Figure 4.4** Examples of stimuli used in Experiments 5 and 6. In each experiment, half of the participants switched between the country and shape tasks in the category-rule condition and between the animal and symbol tasks in the arbitrary-rule condition, and vice versa for the other half of participants.

The assignment of stimuli to conditions (rule/arbitrary) and response (left/right) was balanced between participants. In the arbitrary condition stimuli were assigned as to minimise the chance of participants coming up with alternative task rules to aid performance (e.g. “All warm countries map onto the left”; or “All laterally extended shapes map onto the right”).

The animal-symbol task pair was structurally identical to the country-shape pair. The stimuli in this task combined one of 16 animal names (4 mammals, 4 fish, 4 birds and 4 insects) with one of 16 symbols (4 brackets, 4 numbers, 4 mathematical signs and 4 symbols) placed above the name (see Figure 4.4 for examples).

The auditory task cues (~350 ms long) were “country” and “land” (country task), “shape” and “form” (shape task), “animal” and “living” (animal task) and “symbol” and “graphic” (symbol task).

*Design and procedure* Twelve participants performed the country-shape task pair in the category-rule condition and the animal-symbol task pair in the arbitrary-rule condition, and twelve participants did the opposite. Each participant completed the two sessions on consecutive days, and both the order of complexity (category- or arbitrary-rule) and assignment of task-pair to arbitrary- versus category-rule condition was balanced over participants. Prior to completing each experimental session, all participants completed a practice session. The practice sessions for the category- and arbitrary-rule conditions were slightly different. Specifically, in the category-rule condition, participants firstly completed 6 single-task training blocks of 40 trials each (3 blocks of each task). In total, each of the 8 S-R mappings of each task was practiced 15 times in the single task blocks. These single task blocks were then followed by 4 switching blocks of 20 trials each (1 block for each CSI).

In the training blocks for the arbitrary-rule condition, S-R mappings were introduced incrementally to facilitate acquisition and to minimise the chance of participants detecting some common thread among the stimulus elements assigned to each response (see Dreisbach & Haider, 2008, for a similar procedure). Every participant started out with a mini practice block for one task, with one stimulus element requiring a left response and one requiring a right response (5 trials of each, 10 trials in total). In the next block, 2 new stimuli (one left, right, same task) were introduced (5 trials each), and this block also contained one of each stimulus from the previous block (12 trials in total). In the third block, two new stimuli were again introduced (5 trials each), with 4 trials containing stimuli from the previous two blocks

(14 trials in total). In the final block, the last 2 stimuli were introduced (5 trials each), with 6 stimuli from the previous 3 blocks (16 trials in total). This procedure was then repeated for the other task. Then, the whole procedure was repeated but in reversed order, so that the two stimuli that were introduced first in the first two blocks were now introduced last, and vice versa, resulting in a total of 208 trials in which each response mapping was presented 13 times. This was followed by two mini blocks (one for each task) in which the 8 S-R mappings for each task were presented another two times, so that in total, each S-R mapping was practiced 15 times (as in the category-rule condition).

In both the arbitrary- and the category-rule condition, the practice session was followed by an experimental session containing 768 trials (8 of each of 96 trial types: 2 task x 2 cue x 3 switch x 2 congruency x 4 CSI = 96; each of the 64 stimuli occurred 3 times per CSI) plus 24 filler trials (1 filler at the start of each block), resulting in a total of 792 trials. The switch : repeat ratio was 1:2. In each session, all participants completed 24 blocks of 33 trials each.

On each trial, a fixation cross was presented for 500 ms, followed by the auditory cue (the fixation cross remained on the screen for the duration of the CSI). This cue preceded the stimulus onset by 150 ms, 400 ms, 800 ms or 1300 ms. The stimulus then remained on the screen until the participant responded. In the event of an error, the word “Error!” was displayed on the screen for 1000 ms. Each response was followed by a blank interval, the duration of which depended on the duration of the CSI (blank duration = 1400 – CSI), so that the time elapsing between the response and the subsequent stimulus onset was kept constant at 1900 ms.

At the end of each block, participants were presented with a score, computed as before based on mean RT and error rate. When participants improved on this score in the next block, bonus points (worth £0.10 each) were awarded.

## Results

The same exclusion criteria used in Experiment 4 were applied, resulting in the exclusion of 0.05% of the mean correct RTs.

### *Effect of task-set complexity on the switch cost and its reduction with preparation*

Figure 4.5 shows the mean correct RTs and error rates for category-rule and arbitrary-rule switch and repeat trials, as a function of CSI. A 2 (category- or arbitrary-rule) x 4

(150, 400, 800 or 1300 ms) x 2 (switch or repeat) x 2 (congruent or incongruent) repeated measures ANOVA was then run on these RTs and errors.

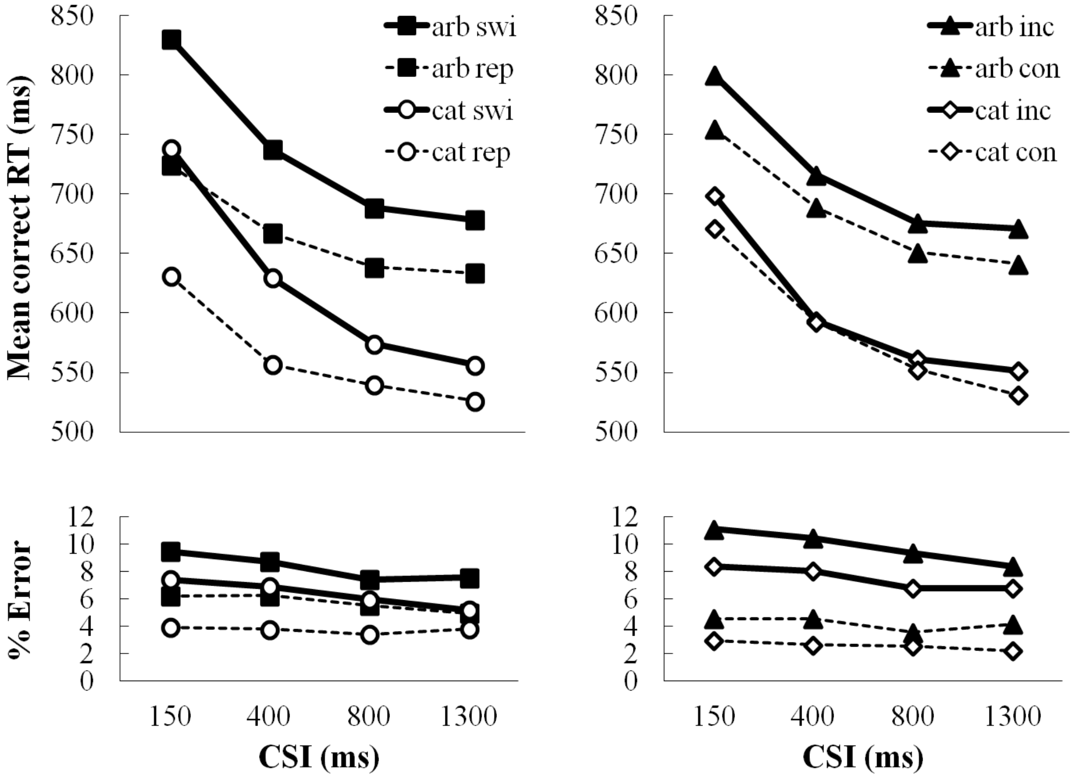


**Figure 4.5** RTs and errors in Experiment 5 for switch and repeat trials (left) and congruent and incongruent trials (right), plotted separately for the category-rule and the arbitrary-rule condition, as a function of CSI.

There was a main effect of complexity: RTs were 106 ms slower in the arbitrary-rule blocks than in the category-rule blocks, F(1,23) = 23.06, p < 0.001. Overall, the switch cost was 65 ms, F(1,23) = 111.13, p < 0.001. Overall mean RT decreased as a function of increasing CSI (150 ms = 731 ms, 400 ms = 648 ms, 800 ms = 610 ms, 1300 ms = 599 ms), F(3,69) = 111.88, p < 0.001 (H-F).

Overall, the switch cost also reliably reduced as a function of increasing CSI (150 ms = 107 ms, 400 ms = 72 ms, 800 ms = 42 ms, 1300 ms= 38 ms – a reduction from shortest to longest CSI of 69 ms or 65%), F(3,69) = 22.02, p < 0.001. When the same analysis was restricted to the two longest CSIs (800 ms and 1300 ms) only, the RISC was not reliable (F < 1): preparation for a task switch appears asymptotic by 1300 ms.

The switch cost was not reliably larger in arbitrary-rule (68 ms) than in the category-rule (61 ms) conditions, F < 1. As predicted, the RISC was larger in the category- (77 ms or 72%) than in the arbitrary-rule (61 ms or 58%), but the three-way interaction was not significant, F < 1. (But the difference in RISCs between the arbitrary- and category-rule condition was reliably modulated by congruency, as described below.)

Error rates showed a reliable switch cost of 2.6%, F(1,23) = 33.57, p < 0.001. Participants also made 2% more errors in arbitrary- than in category-rule blocks, F(1,23) = 6.25, p = 0.020, and error rates decreased as a function of CSI (150 ms = 6.8%, 400 ms = 6.4%, 800 ms = 5.6 %, 1300 ms = 5.4%), F(3,69) = 5.24, p = 0.006 (H-F).

*Effect of task-set complexity on the congruency effect and its reduction with preparation* The ANOVA described above also revealed a congruency effect of 23 ms, F(1,23) = 23.32, p < 0.001. The congruency effect reduced reliably with preparation (150 ms = 36 ms, 400 ms = 14 ms, 800 ms = 17 ms, 1300 ms = 25 ms), F(3,69) = 4.04, p = 0.010 (H-F). The congruency effect was also approximately twice as large in the arbitrary-rule (32 ms) as in the category-rule (14 ms) condition, F(1,23) = 5.17, p = 0.033, suggesting that performance in the more complex arbitrary-rule condition is more susceptible to interference.

For the errors, participants made 5.3% more errors on incongruent than on congruent trials, F(1,23) = 53.67, p < 0.001, and the congruency effect was larger on switch (8%) than repeat trials (2.6%), F(1,23) = 48.99, p < 0.001.

*Interaction of effects of task-set complexity and position in run* Several task-switching studies have found that RTs continue to decrease beyond the second trial in a run of repeats of the same task (the first trial being a switch, see Monsell, Sumner & Waters, 2003, for a review). For unpredictable task switching, Monsell et al. (2003, Experiment 2) also found that this decrease in RT with an increase in position in run was more pronounced when there was no opportunity for advance preparation, suggesting that the position in run effect reflects the effectiveness of task-set preparation to some extent.

If, as we hypothesise, preparation is less effective (or less complete) for the more complex task, then one might also expect this task to benefit from further repetitions of the same task. This is indeed what we found: for prepared repeat trials, the approach to asymptotic performance was more gradual for the more complex task.

Figure 4.6 shows performance on prepared and unprepared repeat trials as a function of position in a run. This analysis averaged over the 400 ms, 800 ms and 1300 ms CSIs (long CSI condition) to ensure enough trials per cell. A 2 (arbitrary- or category-rule) x 2 (long or short CSI) x 4 (position in run) repeated measures ANOVA was run on the mean correct RTs and error rates. Only effects of and interaction with the linear component of position in run are reported here.

RTs decreased with an increase in position in run of task repeats (slope -19 ms per intervening trial), F(1,23) = 64.58, p < 0.001. This slope was steeper on unprepared (-29 ms) than prepared (-10 ms) trials, F(1,23) = 26.19, p < 0.001. Moreover, slopes were similar in arbitrary-rule (-27 ms) and category-rule blocks (-30 ms) on unprepared trials, F < 1, but the slope was twice as steep for prepared trials in arbitrary blocks (-13) than in category-rule blocks (-6 ms), F(1,23) = 4.30, p = 0.039 (three-way interaction, F(1,23) = 3.49, p = 0.074). Clearly, further repetition of the same task improves performance more when (1) there was no time for advance preparation, or (2) there is a larger set of S-R rules. In the error rates only the linear component of position in run (slope -0.4%) was reliable, F(1,23) = 4.42, p = 0.047.

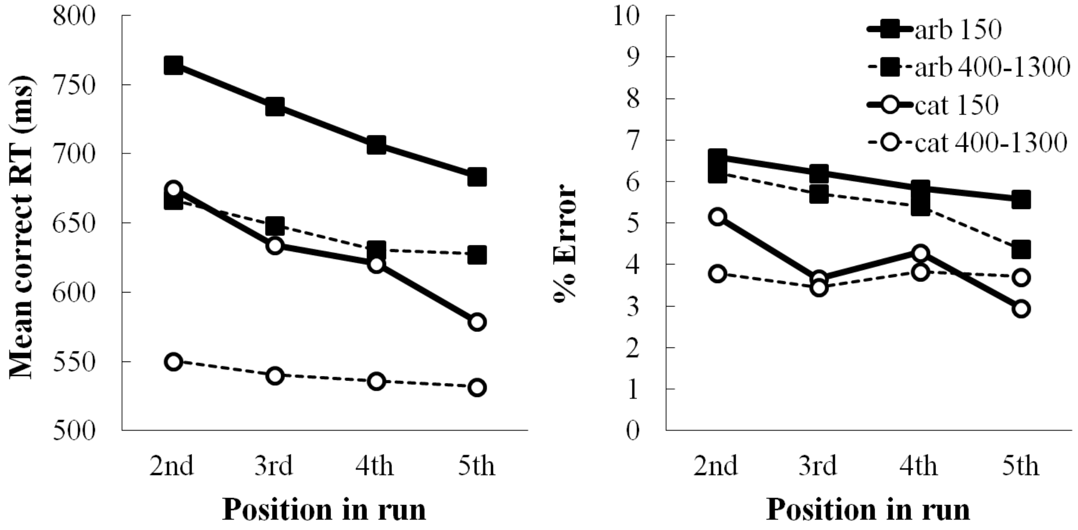


**Figure 4.6** RTs and errors in Experiment 5 for repeat trials plotted as a function of position in run.

Considering these findings, one might argue that the switch cost calculations reported above underestimate the true cost of switching, because mostly longer RTs for the second run position contribute to this calculation (see Mayr & Kliegl, 2000, for a similar argument). This would apply disproportionally to the arbitrary-rule condition in which the position in run effect was larger for prepared trials.

For this reason, the 2 (category- or arbitrary-rule) x 4 (150, 400, 800 or 1300 ms) x 2 (switch or repeat) x 2 (congruent or incongruent) repeated measures ANOVA reported above was repeated, excluding run positions 2 and 3. This analysis found a reliable four-way interaction between complexity, CSI, switch and congruency, F(1,23) = 2.89, p = 0.042 (H-F). Table 4.1 shows that this interaction reflects that preparation proved particularly ineffective (indexed by a larger residual switch cost) in the more complex arbitrary-rule condition when the stimulus happened to be incongruent: the more complex task-set appears less well prepared, and hence more susceptible to interference.

| | **congruent** | | **incongruent** | |
|---|---|---|---|---|
| | **150 ms** | **400-1300 ms** | **150 ms** | **400-1300 ms** |
| **arbitrary** | 131 | 65 | 167 | 86 |
| **category** | 142 | 53 | 151 | 54 |
| **arb - cat difference** | -10 | 11 | 17 | 32* |

**Table 4.1** Switch costs (ms) for the arbitrary- and category-rule conditions (and the difference between them) in Experiment 5 as a function of CSI and congruency (* = $p < 0.05$).

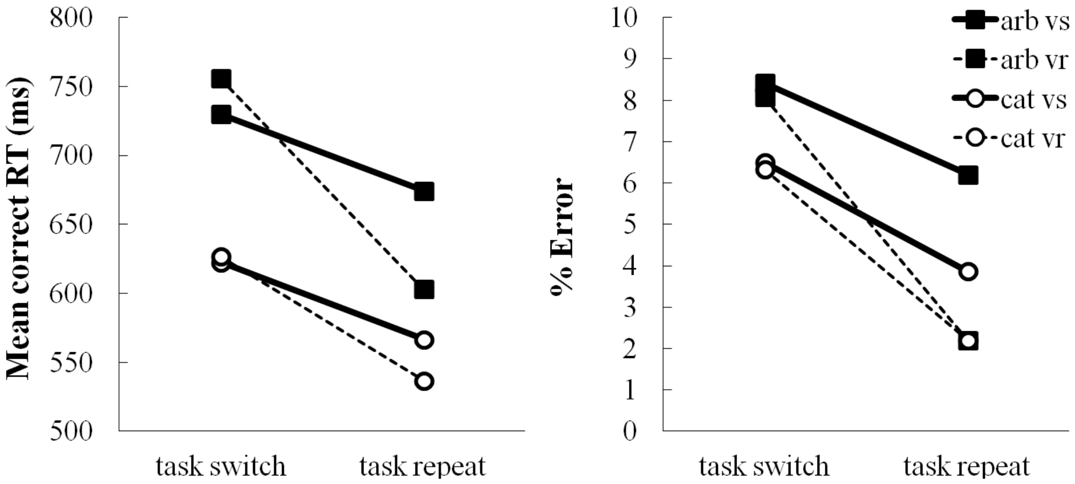


**Figure 4.7** RTs and errors in Experiment 5 for now relevant value switch (vs) and now relevant value repeat (vr) trials as a function of task switch/task repeat.

*Effect of task-set complexity on a (relevant or irrelevant) value repeat* Our interpretation of a smaller RISC with the more complex task-set is that only a subset of its S-R mappings can be represented in the bridge. Hence, on a proportion of task switch trials, the relevant S-R mapping will not be fully prepared for the more complex task-set. This proportion will include some of the trials on which the task repeats, but the relevant stimulus value changes (e.g. in the country task: Kenya surrounded by a circle, preceded by France surrounded by a square). However, when the more complex task repeats, and the relevant stimulus also repeats (e.g. in the country task: Kenya surrounded by a circle, preceded by Kenya surrounded by a square), then the required S-R rule (Kenya → left response) is likely still to be in the bridge from the previous trial. Consequently, one might expect performance in the more complex arbitrary-rule blocks to benefit more on a task repeat trial when the relevant value repeats from the previous trial. Whether the relevant value repeats or changes should matter much less in the less complex category-rule condition because the bridge can accommodate both of the S-R rules for that task.

It is somewhat less clear what one would predict on a task switch trial. But we have already seen that the more complex condition is more susceptible to interference. We might additionally assume that S-R mappings that have been more recently exercised (e.g. on the previous trial) interfere more than less recently exercised S-R mappings. If these assumptions are correct, then performance should be worse on a task switch trial when the now relevant value repeats (because it required a response on the previous trial) – particularly in the arbitrary-rule condition.

The effect of a now relevant value repeat (as a function of switch/repeat x arbitrary/rule) was assessed in a 2 x 2 x 2 ANOVA. Complete repetitions of a composite stimulus were not included in this analysis as they entail a repeat of both the relevant and the irrelevant value. Figure 4.7 shows RTs and error as a function of the now relevant value switch/repeat only (there were no effects of the now irrelevant value switch/repeat, as below).

The benefit of repeating the relevant value (for RTs and errors) on a task repeat trial was more than twice as large in the arbitrary (71 ms and 4%) than in the category-rule condition (30 ms and 1.7%), $F(1,23) = 13.52$, $p = 0.001$ (RTs) and $F(1,23) = 6.54$, $p = 0.018$ (errors). Moreover, RTs were slightly slower for a relevant value repeat on a task switch trial, but much more so in arbitrary (-26 ms), $F(1,23) = 4.61$, $p = 0.043$, than in category-rule blocks (-4 ms), $F < 1$ (interaction $F(1,23) = 1.96$, $p = 0.175$).

We ran the same analysis considering the effect of a change or repeat of the now irrelevant value. Whether the now irrelevant value had changed or repeated from the previous trial seemed to have very little effect on performance: there was a slight benefit for repeating the now irrelevant value (5 ms or 0.5%), but this effect was not reliable, nor was it modulated by complexity or task switch (all F's < 1.82).

*Summary* In Experiment 5 participants were required to switch between two tasks of greater or lesser complexity. The results were generally consistent with those of Experiment 4: task-set preparation, as indexed by the reduction in switch cost was again more effective (and less susceptible to interference) for the less complex task-set, though the critical 3-way interaction was not reliable.

Moreover, the position in run analysis showed that with the opportunity for advance preparation, further repetition of the same task resulted in a greater performance benefit for the more complex arbitrary-rule condition. This finding fits our previous interpretation that in the more complex condition, the relevant S-R rule has on a proportion of the trials not been uploaded into the bridge. For the same reason, repetition of the relevant value benefits performance more on a task repeat trial in the more complex condition (when the relevant S-R rule is still in the bridge from the previous trial).

Finally, when the first two task repetitions were excluded from the analysis, preparation was found to be particularly ineffective in the arbitrary-rule condition when the stimulus was incongruent, again suggesting a greater susceptibility to interference for that task-set. These results are consistent with the findings of Experiment 4, and the notion that as much of the operative task-set as possible is held in a capacity limited component of PWM. However the critical analysis (interaction between complexity, CSI and switch) lacked the statistical power needed for an unequivocal conclusion. For this reason, another very similar experiment was conducted in an attempt to replicate the outcome of Experiment 5.

Experiment 5 did establish, however, that a CSI of 1300 ms provides ample time for asymptotic preparation with even the more complex tasks. There was very little improvement in overall performance or switch cost as CSI increased from 800 to 1300 ms; given the shape of the preparation function it is highly unlikely that an even longer CSI would achieve more. Hence in the next experiment we could use just the two CSIs (150 and 1300 ms) to increase the power, confident that they span the extremes of task-set preparation.

## Experiment 6

Experiment 6 was essentially identical to Experiment 5, except that we used only two CSIs (150 ms and 1300 ms) with the same number of trials overall.

### Method

*Participants* Twenty four participants (aged between 18 and 35 [M = 20.5]; 5 male and 19 female) provided informed consent before taking part in two sessions on consecutive days. All participants received course credit in return for participation, and an additional sum of money (maximum £4.40), depending on the speed and accuracy of their performance.

*Stimuli and procedure* There were now only two CSIs (150 ms and 1300 ms), each with twice as many trials as each CSI in Experiment 5. Blocks with short and long CSIs alternated, and the starting CSI was balanced over participants. Otherwise all details were as in Experiment 5.

### Results

The same exclusion criteria used in Experiment 4 were applied, resulting in the exclusion of 0.06% of the mean correct RTs.

*Effect of task-set complexity on the switch cost and its reduction with preparation* Mean correct RTs and error rates in Figure 4.8 show that the results of Experiment 6 essentially replicate those of Experiment 5: the RISC effect was smaller for the arbitrary-rule condition, particularly (in this experiment: exclusively) on incongruent trials. A 2 (complexity) x 2 (CSI) x 2 (switch) x 2 (congruency) repeated measures ANOVA was conducted on these RTs and error rates.

RTs were 89 ms slower in arbitrary blocks than in category-rule blocks, $F(1,23) = 46.29$, $p < 0.001$. There was a switch cost of 76 ms, $F(1,23) = 156.32$, $p < 0.001$, and a CSI effect of 104 ms, $F(1,23) = 146.93$, $p < 0.001$. Switch and CSI also interacted producing an overall RISC of 56 ms or 53%, $F(1,23) = 36.17$, $p < 0.001$. The RISC was again larger in the category-rule condition (from 107 ms to 43 ms, a RISC of 63 ms or 59%) than in the arbitrary-rule condition (from 102 ms to 53 ms, a RISC of 49 ms or

48%), though the three-way interaction was again not reliable, F = 1.87. (As in Experiment 5, this difference in RISCs was reliably modulated by congruency, as described below).

Although participants made slightly more errors in arbitrary (7.7%) than in category-rule (6.4%) blocks, this difference was not reliable, F = 2.12. Overall, there was a reliable switch cost of 3.7%, F(1,23) = 53.22, $p < 0.001$, a CSI effect of 1.3%, F(1,23) = 9.06, p = 0.006, and a RISC of 2.6%, F(1,23) = 9.24, p = 0.006.

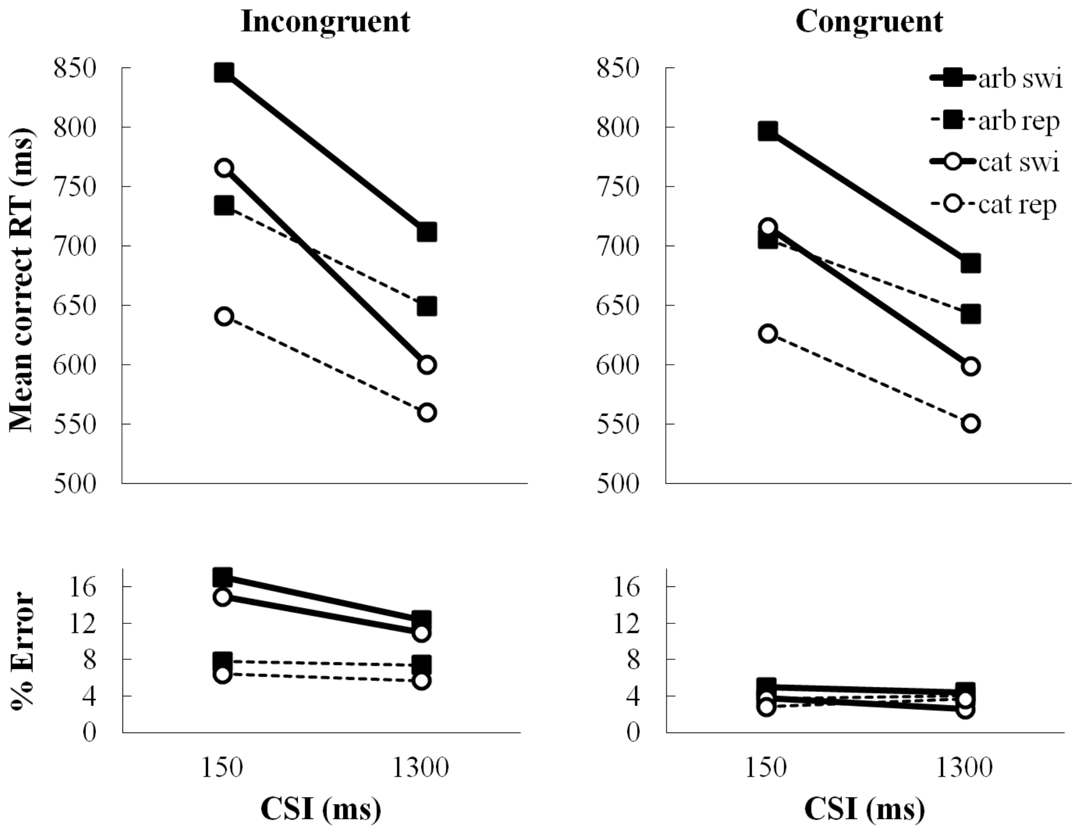


**Figure 4.8** RTs and errors in Experiment 6 as a function of switch/repeat, arbitrary-/category-rule and CSI plotted separately for incongruent (left) and congruent (right) trials.

| | congruent | | incongruent | |
|---|---|---|---|---|
| | 150 ms | 1300 ms | 150 ms | 1300 ms |
| **arbitrary** | 91 | 43 | 112 | 62 |
| **category** | 89 | 48 | 125 | 40 |
| **arb - cat difference** | 2 | -6 | -13 | 21* |

**Table 4.2** Switch costs (ms) for the arbitrary- and category-rule conditions (and the difference between them) in Experiment 6 as a function of CSI and congruency (* = $p < 0.05$).

*Effect of task-set complexity on the congruency effect and its reduction with preparation* The ANOVAs described above also found effects of and interaction with congruency. For the RTs there was a congruency effect of 23 ms, F(1,23) = 70.25, p < 0.001, which was larger on switch (31 ms) than repeat (14 ms) trials, F(1,23) = 15.67, p = 0.001, and larger in the short (35 ms) than in the long CSI (11 ms), F(1,23) = 23.23, p < 0.001. The congruency effect was larger on switch than on repeat trials, but more so in the short CSI (28 ms difference) than in the long CSI (6 ms difference), F(1,23) = 5.10, p = 0.034. F(1,23) = 5.18, p = 0.034.

Importantly, the RISC was larger in the category-rule (84 ms or 68%) than in the arbitrary-rule condition (50 ms or 44%) on incongruent trials, F(1,23) = 7.93, p = 0.010, but not on congruent trials (category-rule RISC 41 ms or 46%; arbitrary RISC 48 ms or 53%), F < 1 (four-way interaction: F(1,23) = 5.20, p = 0.032). A comparison between Tables 1 and 2 shows that in both Experiments 5 and 6, the switch cost was larger in the arbitrary-rule condition particularly on prepared incongruent trials.

For the errors, there was a congruency effect of 6.6%, F(1,23) = 72.79, p < 0.001, which was larger on switch (9.9%) than on repeat (3.3%) trials, F(1,23) = 42.22, p < 0.001 and larger in the short CSI (7.8%) than in the long CSI (5.4%), F(1,23) = 10.69, p = 0.003.

*Effect of task complexity on position in run* In Experiment 6 repetition of the same task improved performance more in the arbitrary-rule condition, particularly at a long CSI. To assess whether this was also true in this experiment, RTs and errors were analysed as a function of position in run. As in Experiment 6, this analysis was restricted to the first four repeats in a run. Only the effect of and interactions with the linear component of position in run are reported below.

As shown in Figure 4.9, there was a general improvement in RT through a run (slope -13 ms), F(1,23) = 31.24, p < 0.001, and a reliable interaction with CSI: as in Experiment 6, the effect of position in run was much greater in the short CSI (-22 ms) than in the long CSI (-4 ms), F(1,23) = 23.23, p < 0.001. But unlike Experiment 6, the linear trend of position in run did not differ reliably in the category-rule (slope -16 ms) and arbitrary-rule conditions (slope -6 ms), F < 1 (three-way interaction F < 1).

The error data also showed a more positive slope for the short CSI (slope -0.7%) than for long CSI blocks (slope 0.3%), F(1,23) = 6.54, p < 0.001, though again this effect was not modulated by complexity, F < 1.

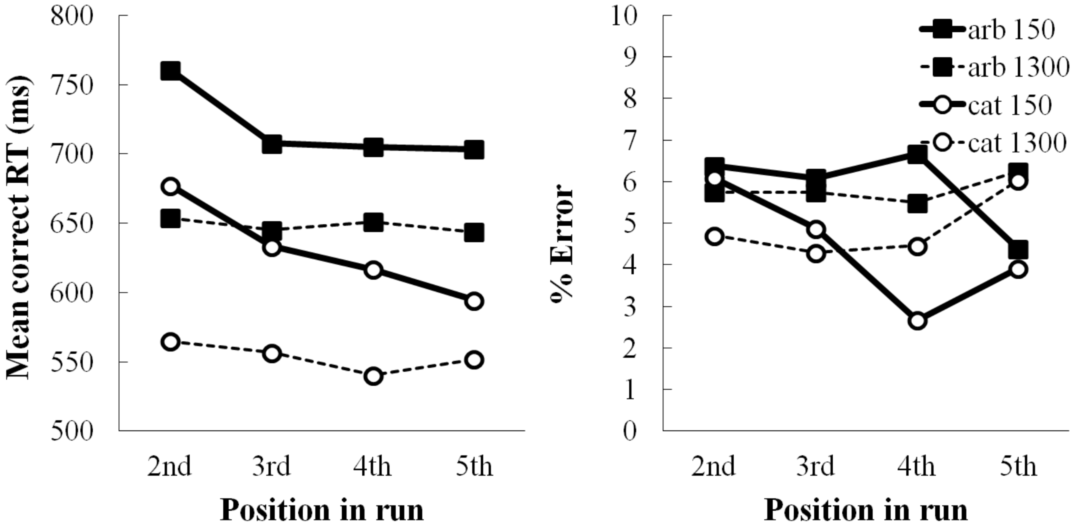


**Figure 4.9** RTs and errors in Experiment 6 for repeat trials plotted as a function of position in run.

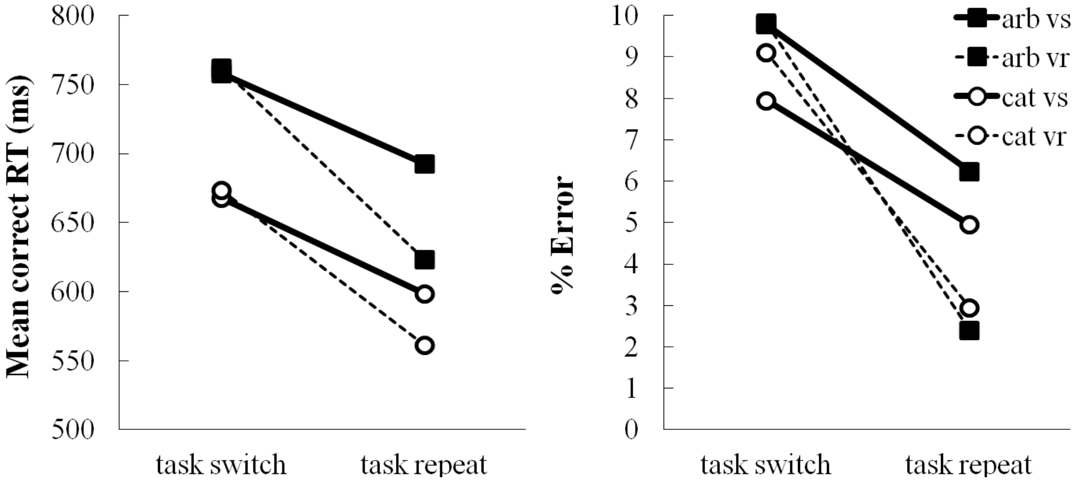


**Figure 4.10** RTs and errors in Experiment 6 for now relevant value switch (vs) and now relevant value repeat (vr) trials as a function of task switch/task repeat.

*Effect of task-set complexity on a (relevant or irrelevant) value repeat* RTs and error rates for arbitrary- and category-rule task-switch and -repeat trials are shown in Figure 4.10 as a function of the relevant value repeat/switch. As in Experiment 5, the RT and error benefit of repeating the relevant value was confined to task repeat trials, for which this benefit was almost twice as large in arbitrary- (69 ms and 3.8%) than in category-rule (37 ms and 2.0%) blocks, F(1,23) = 6.91, p = 0.015 (RTs) and F(1,23) = 5.60, p = 0.022 (errors). Though there was a slight disadvantage of repeating the now relevant value on a task switch trial (-5 ms and -0.6%), this effect was not reliable or reliably modulated by complexity, all F's < 1.

As for irrelevant value repeats, there was a small overall benefit of repeating the irrelevant value – 4 ms in the RTs, F < 1 and 0.8% in the errors, F(1,23) = 3.21, p = 0.087 – but there were no interactions with task switch/repeat or complexity, all F's < 2.8.

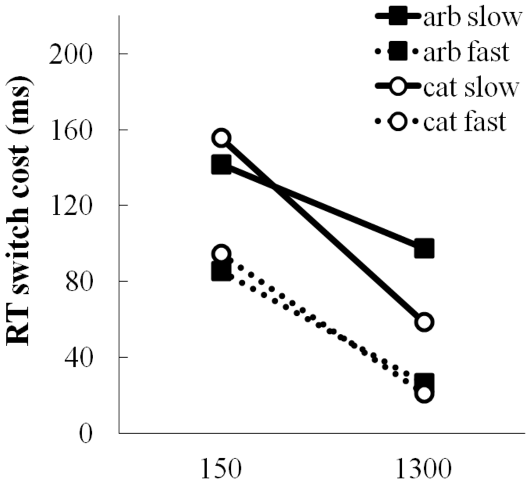


**Figure 4.11** RT switch cost (ms) in Experiment 6 in the arbitrary and category-rule conditions, for fast (average of 10th-40th percentile) and for slow responses (average of 60th-90th percentile), as a function of CSI.

*RT distribution analysis* Correct RTs were rank ordered and portioned into deciles in the same way as in Experiment 4[16]. This analysis was restricted to incongruent trials as the effect of task-set complexity manifested itself selective on those trials in this experiment. RT switch costs for fast (10th-40th percentiles) and slow responses (60th-90th percentiles) in category- and arbitrary-rule blocks are plotted as a function of CSI in Figure 4.11.

As in Experiment 4, the difference between the category-rule RISC (73 ms or 78%) and the arbitrary-rule RISC (59 ms or 69%) was similar for fast responses, F = 2.84, n.s., but for slow responses the category-rule RISC (97 ms or 62%) was much larger than the arbitrary-rule RISC (44 ms or 31%), F(1,23) = 7.18, p = 0.013 (four-way interaction F(1,23) = 3.62, p = 0.070).

---

[16] I did not include an RT distribution analysis for Experiment 5 because in that experiment, there were too few trials per cell (per decile) per CSI, because that experiment used four CSIs instead of two.

*Summary* Experiment 6 replicated the result obtained in Experiment 5: task-set preparation was less effective for the more complex task-set when the stimulus turned out to be incongruent. Though the result of the position in run analysis did not replicate Experiment 5 (task-set complexity did not modulate the position in run effect in this Experiment), the results from the value repeat analysis showed that the benefit of repeating the relevant value on a task repeat trial was again twice as large in the arbitrary condition. This finding is consistent with our interpretation that with a greater number of S-R rules, not all S-R rules can be maintained in the bridge simultaneously. But when the relevant value repeats (on a task repeat trial) the S-R mapping is already in the bridge, resulting in faster RTs for value repeats, particularly in the more complex condition.

The results of a fast-slow analysis (restricted to incongruent trials) were also consistent with this interpretation: category- and arbitrary-rule RISCs were of similar size for fast responses, suggesting that preparation was equally effective in both conditions at this end of the distribution. But for the slow responses, the RISC was much smaller for the arbitrary condition, consistent with the idea that the capacity-limited component of PWM can only accommodate a subset of the S-R mappings: on a subset of trials (slow responses), the relevant value has not been prepared in the more complex condition, consistent with the result of Experiment 4.

Altogether, these findings support the results of Experiments 4 and 5. The finding that task-set preparation is affected by the complexity of the task-set is consistent with models in which the most active task-set is held in capacity-limited mode of representation (Oberauer, 2009).

## General Discussion

The experiments reported in Chapter 2 and Chapter 3 suggest that only the currently operative task-set is represented in a capacity limited “bridge”, and that the currently irrelevant task-sets are represented in a capacity-unlimited part of activated LTM. The three experiments reported in this chapter were designed to further define the process of uploading a task-set into the bridge, and its capacity limit. To achieve this task-set complexity was manipulated: participants either switched between task-sets specified by 2 S-R rules or between task-sets specified by 3 (Experiment 4) or 8[17] (Experiments 5 and 6) S-R rules. The predicted outcome depended both on the effective capacity of the bridge, and on the time consumed by retrieving a task-set into the bridge. Two of the many potential outcomes were outlined in the Introduction: (1) if task-set retrieval is more time consuming for the more complex task-set, and the bridge can accommodate all of the more complex task-set, then the reduction in switch cost should be greater for the more complex task-set; (2) but if task-sets are uploaded into the bridge as a “chunk” (cf. Oberauer, 2005) and the bridge can only accommodate a subset of the more complex task-set, then the reduction in switch cost with preparation would be greater for the less complex task-set.

In all three experiments preparation was found to be more effective for the less complex task-set, as indexed by a larger RISC in that condition. In two experiments this effect was particularly apparent on (Experiment 5) or exclusively restricted to (Experiment 6) incongruent trials, suggesting that incomplete preparation of the more complex task-set also made it more vulnerable to interference. Consistent with this, in Experiment 4 the less complex task-set was also less susceptible to interference on prepared trials, as reflected by smaller congruency effects in that condition.

These results imply that the bridge can only represent very few S-R mappings in their most active state. When a task-set exceeds the capacity of the bridge, effective task-set preparation can only be achieved for a sub-set of the S-R mappings. This interpretation is further supported by the results of fast-slow analyses (Experiments 4 and 6), a position in run analysis (Experiment 5) and value repeat analyses (Experiments 5 and 6).

Firstly, the fast-slow analyses showed a much reduced RISC for the more complex condition at the slow end of the RT distribution and a very similar RISC effect

---

[17] or 4 if we assume half the stimuli were classified by default.

at the fast end – suggesting that with the more complex task-set, some S-R rules were as well prepared as for the less complex set, but a subset were less well prepared.

Secondly, the position in run analysis showed that repetition of the same task improved performance more (for prepared trials) in the more complex arbitrary-rule condition. Although this finding was not replicated in Experiment 6, it is at least consistent with the idea that on some prepared trials, the relevant S-R rule is not represented in the bridge in the more complex condition.

Thirdly, Experiments 5 and 6 showed that the benefit of repeating a stimulus value on a task repeat trial was considerably greater in the arbitrary- than in the category-rule condition. This result is again consistent with the idea that the bridge can maintain only a subset of the S-R mappings for the more complex task; but when the relevant value repeats on a task-repeat trial, it is likely that the relevant S-R mapping is already in the bridge (from the previous trial), resulting in improved performance on those trials. The benefit of repeating the relevant value on a task-repeat trial is much smaller in the category-rule condition because the bridge can accommodate all the S-R rules for that task.

However, in two experiments this larger RISC for the more complex task was found mostly (Experiment 5) or only (Experiment 6) on incongruent trials. Although it makes sense that the effect of complexity is *more* pronounced on incongruent trials (reflecting the notion that a more complex task-set is also more susceptible to interference, cf. Dreisbach and colleagues), it is less obvious why complexity should have *no* effect on congruent trials, as we found in Experiment 6: on a congruent trial the tasks' rules must also be loaded into the bridge, and when the task-set exceeds the capacity of the bridge, this should be problematic. It is therefore difficult to explain why we obtained no effect of complexity on task-set preparation on congruent trials in Experiment 6.

Though the larger RISC for the less complex condition found in all three experiments supports a limited capacity PWM buffer, as proposed by Oberauer, they are not entirely conclusive with regards to the effect of complexity on retrieval. Based on the results of analogous manipulations in declarative WM, which show that the time it takes to make one of several lists available again for operating upon does not depend on the size of that list (Wickens, Moody & Dow, 1981; Conway & Engle, 1995; Oberauer, 2005), Oberauer predicts that the time it takes to upload a task-set into the bridge should not depend on the complexity of that task-set. If this were the case, then a difference in switch costs should remain at a long CSI (according to the alternative, effects of complexity reflect increased retrieval time for the more complex task-set, and hence any effect of complexity should manifest itself at a short CSI, and should be eliminated at a long CSI). This is indeed what was found in Experiments 5 and 6, in which a difference in switch costs (between the category- and arbitrary-rule conditions) was found for prepared trials. But the data looked slightly different for Experiment 4, in which the switch cost was larger for the more complex 3-choice condition only for unprepared trials. But on closer inspection this result does not argue *against* the notion that a task-set is retrieved as a chunk: if Oberauer was wrong, and task-set retrieval takes longer for the more complex task-set, then surely one would expect the difference between (2-choice and 3-choice) task switch RTs to *decrease* with preparation (cf. Mayr & Kliegl, 2000)? But this was not what was found: in Experiment 4, as in Experiments 5 and 6, the switch trials in the less complex condition benefited more from preparation – consistent with our conclusion that preparation is most effective when all S-R mappings can be maintained simultaneously in the bridge.

So what happens then when a task-set exceeds the capacity of the bridge? The two most likely options are: 1) the available activation is shared so that all S-R rules are represented in a lesser state of activation; 2) only a subset of the task's S-R mappings can be represented in a state of full activation[18]. I have assumed the latter, because the distribution analysis is not consistent with the former: if all of the more complex task-set is represented in the bridge in a "degraded" state, then one would not expect preparation to be equally effective for the more and less complex condition on a proportion of trials.

Assuming then that only a subset of the more complex task's S-R mappings are represented in the bridge, what happens when the stimulus arrives and its rule has not been loaded? One possibility is that participants must load this rule after stimulus onset (perhaps the stimulus retrieves the S-R rule automatically if it is not already in the bridge on stimulus presentation). The observation that RTs were substantially slower in the more complex condition when the required rule changed on a task repeat trial is

---

[18] Within this second possibility (only a subset of S-R rules can be maintained in the bridge), one can further distinguish between two possibilities: the S-R rules not in the bridge either don't get into the bridge in the first place, or they are displaced from the bridge by rules loaded subsequently (which leaves them in a higher state of activation in the active part of long term procedural memory than they would be if they had not be loaded at all). This second possibility actually appears more likely, if one assumes that if a subset of S-R rules was never in the bridge, then when such a rule is required on a task repeat trial, performance should be similar to performance on switch trials, but this is not the case: in Experiments 5 and 6 we found that a relevant value switch (on a task repeat trial) is still faster than a task switch trial.

consistent with the view that, even on a (some) task repeat trial(s), the relevant S-R rule is not (yet) in the bridge.

On the whole then, these results suggest that the currently operative task-set is held in a very capacity-limited mode of representation. These results are partially consistent with Rubinstein et al. (2001), who also found larger switch costs with more complex tasks. But Rubinstein et al.'s (2001) conclusion – that it takes longer to "load" a more complex task-set – would predict that the effect of complexity on the switch cost should be eliminated with plenty of opportunity for preparation, which is not what was found in Experiments 5 and 6. One might argue that Rubinstein et al. (2001) proposed that rule activation cannot commence prior to stimulus onset (which could also explain larger switch costs for the more complex condition at a long CSI), but the reduction in the congruency effect with preparation found in this thesis and elsewhere (see Kiesel et al., 2010 for review) rejects this proposal. The results of Hübner et al. (2004, larger switch costs for the less complex task-set) are also not easily reconciled with our result, but then Hübner manipulated complexity within a block, whereas here complexity was manipulated between blocks. Establishing whether there are any differences between the effect of task-set complexity on uploading (switching towards) and removing (switching away from) a task-set in/from the bridge would require an orthogonal manipulation of the complexity of the previous and current task-set within the same experiment.

# Chapter 5: Are stimulus-response rules represented phonologically for task-set preparation and maintenance?

## Abstract

Accounts of task-set control generally assume that the current task's stimulus-response (S-R) rules must be elevated to and maintained in a privileged state of activation. How are they represented? In three task-cuing experiments we tested the hypothesis that phonological working memory is used to represent S-R rules for task-set control, by getting participants to switch between two sets of arbitrary S-R rules, and manipulating the articulatory duration (Experiment 7) or phonological similarity (Experiments 8 and 9) of the names of the sets of stimulus terms. The task cue specified which of two objects (Experiment 7) or consonants (Experiment 8) in a display to identify with a key press. In Experiment 9, participants switched between identifying an object/consonant and its colour/visual texture. Neither the duration nor the similarity of the stimulus terms had any detectable effect on overall performance, task-switch cost, or its reduction with preparation in the task-cueing blocks. Only in the initial single-task training blocks was there any effect of phonological similarity. Hence beyond a very transient role there is no evidence that (declarative) phonological working memory makes a functional contribution to representing S-R rules for task-set control, arguably because once learned they are represented in non-linguistic procedural working memory.

## Introduction

Our ability to switch flexibly among tasks requires a process by which the current task-set is selected and maintained in a privileged state of activation. The task-cueing paradigm exercises this process: on each trial participants are presented with a stimulus which must be classified according to one of several sets of task-rules, the relevant task being signalled by a task cue preceding each stimulus (Sudevan & Taylor, 1987; Meiran, 1996). When the task switches, reaction times (RTs) are longer and error rates are greater than when the task repeats (e.g. Meiran, 1996; Rogers & Monsell, 1995; Allport, Styles & Hsieh, 1994). One interpretation of this switch cost is that it reflects the amount of time needed by the cognitive system to reconfigure task-set. This account has been supported by the observation that the switch cost reduces (up to a point) the greater the time available for advance preparation, suggesting that participants are able to engage the appropriate task-set in advance, and providing a way of measuring the difficulty of engaging or maintaining a set of S-R rules (e.g. Meiran, 1996; Monsell & Mizon, 2006). The sub-processes that constitute task-set reconfiguration and the nature of representation of task-set remain poorly specified. This paper focuses on two linked issues: the role of language, specifically of phonological representations, in task-set control, and the distinction between declarative and procedural working memory.

Goschke (2000) argued from a Vygotskian perspective that linguistic self-instruction plays a crucial role in task set control. He gave participants two-trial sequences in which they predictably switched between classifying the colour and identity of letters, or repeated one of these tasks, with a short (14 ms) or long (1500 ms) response-stimulus interval (RSI) between the two letters. During the long RSI participants were required to say the task name ("letter" or "colour), an irrelevant word (“Monday” or “Tuesday”), or neither. Saying an irrelevant word eliminated the reduction in switch costs observed in the other RSI conditions because, Goschke argued, the irrelevant articulation blocked the use of the task-name as a self-instruction to activate the appropriate task-goal. Consistent with the idea that a verbal representation at least facilitates task-set activation, work in our lab has found that transparent verbal cues naming the dimension to be attended to ("colour", "shape") produce more effective preparation than a transparent pictorial cue (a collage of the colours or shapes) (Elchlepp, Lavric, Mizon & Monsell, 2012; Lavric, Mizon & Monsell, 2008; Monsell & Mizon, 2006). Moreover a non-verbal cue triggered lefthemisphere activation suggestive of the participant generating their own internal verbal representation, where a verbal cue did not (Lavric et al., 2008). The observation that increased switch costs in patients with left prefrontal damage are associated with language problems (Mecklinger, von Cramon, Springer & Matthes-von Cramon, 1999) lends further support to the role of verbal mediation in task switching.

There is, of course, more to language than speech. But experiments requiring concurrent articulation of irrelevant material ("articulatory suppression" or AS) have focused attention specifically on the role of phonological coding. Some of these studies found that AS increased alternation costs in Jersild's (1927) list paradigm (Baddeley, Chincotta & Adlam, 2001; Emerson & Miyake, 2003), but subsequent studies suggested that this was largely because of the use of the phonological buffer to maintain a representation of the order in which the tasks were to be performed ("letter, digit, letter..."), and where the participant was in the sequence. For example Bryck and Mayr (2005) found that when they eliminated the need to keep track of task sequence this paradigm by using univalent stimuli (i.e. stimuli that are only associated with one task) or explicit task cues, AS no longer increased the switch cost. And Saeki and Saito (2004) demonstrated that AS only increased the alternation cost in a list paradigm when external cues indicating which stimulus to respond to were absent (see also Saeki & Saito, 2006, 2009).

A role of phonological representations in task-set activation, rather than remembering the order of tasks, was suggested by Miyake, Emerson, Padilla and Ahn, (2004), who required participants to switch unpredictably between colour and shape classifications (CSI = 200 ms or 1200 ms), using either (transparent) word cues (“colour” and “shape”) or (less transparent) letter cues (“C” and “S”). Participants also engaged in either AS (saying “Tuesday” or Thursday”) or a control foot tapping task (both tasks were performed to the beat of a metronome – once every 750 ms). Switch costs were found to be substantially larger in the AS condition compared to the foot tapping or silent control condition, but only with letter cues. The authors conclude that in the absence of an explicit verbal cue, inner speech is used to activate and retrieve the relevant task goal.

Liefooghe, Vandierendonck, Muyllaert, Verbruggen and Vanneste (2005) argued for a third role of phonological representations in task-set preparation: maintaining the S-R rules in working memory – but also that such maintenance was needed only on task-repeat trials. When participants switched predictably among three tasks in an AS or silent control condition they found increased reaction times (RTs) under AS limited largely to repeat trials. In a second experiment they had participants classify a target based on its shape (square or circle) or colour (red or blue). The task-cue either specifically displayed the S-R rule on the screen (e.g. a square on the left and a circle on the right) or was an arbitrary colour cue (a yellow hexagon for the shape task and a green hexagon for the colour task). The switch cost was larger with arbitrary cueing, but then disappeared under AS. The authors argued that (when the S-R mapping was not on the screen) participants use phonological working memory to maintain the rules on task-repeat trials, but are presumably retrieving the rules from long term memory (LTM) on switch trials (cf. Mayr & Kliegl, 2000).

The present experiments were designed to examine further the possible role of phonological representations in representing the S-R rules for task-set control. Effects of articulatory suppression are one classic diagnostic of the use of phonological representation, though an effect of AS does not tell us which elements of the task-set are represented phonologically. We used two other diagnostics which can specifically target the representation of S-R rules: the effects of phonological similarity and word length.

It has been known for a long time that immediate serial recall is much worse for lists of letters (Conrad & Hull, 1964) or words (Baddeley, 1966) with phonologically similar names than for lists with dissimilar names. This observation was one of the main drivers of the proposal that an important component of working memory is a buffer capable of holding a sequence of several words or syllables in a phonological code (Conrad, 1964; Morton, 1969; Baddeley & Hitch, 1974). There remains some disagreement with regards to the locus of the effect, with Baddeley and colleagues ascribing it to processes that take place during retrieval, whilst others attribute it to processes that take place during encoding (Oberauer & Kliegl, 2006; Farrell & Lewandowsky, 2003). But the presence of the effect clearly demonstrates the use of a phonological representation (see Camos, Mora & Oberauer, 2011). A second driver was the word length effect: immediate serial recall of a word list is inversely correlated with the articulatory complexity/duration of the constituent words, even when they are matched for visual length (e.g. Baddeley, Thomson & Buchanan, 1975; Avons, Wright & Pammer, 1994; Ellis & Henneley, 1980; Longoni, Richerdson & Aiello, 1993; Caplan, Rochon & Waters, 1992). The most popular interpretation of the word-length effect has been that it reflects the use of a covert articulation process to maintain representations of the items in a phonological buffer in working memory – the "articulatory loop" hypothesis of Baddeley et al. (1975). Longer words take longer to rehearse, and therefore fewer can simultaneously be maintained within the phonological store. The elimination of the word length and phonological similarity effects under AS (Baddeley et al., 1975) supported the assumption that AS blocks the generation and maintenance of a phonological code. As a rule of thumb, Baddeley et al. proposed that the effective capacity of the phonological buffer is the amount of speech that can be articulated in about two seconds. However, others (e.g. Caplan, Rochon & Waters, 1992; Service, 1998) have argued that the effect of word length is an effect not of articulatory duration, implicating rehearsal rate, but of phonological complexity, implicating the capacity limits of the buffer. For our purposes, the effect is diagnostic of the involvement of a phonological code whether it is due to complexity or articulatory duration.

The task-cueing experiments reported here required participants to switch between two tasks specified by sets of four arbitrary S-R rules, with or without the opportunity for advance preparation, and manipulated the articulatory complexity/duration (Experiment 7) or the phonological similarity (Experiments 8 and 9) of the four stimulus terms in a set. Bringing a task-set into a state where it dominates another recently performed task-set still afforded by the stimuli is commonly thought to involve retrieving its S-R rules into working memory from longer-term memory (e.g. Mayr & Kliegl, 2000; Rubinstein, Meyer & Evans, 2001). If those rules are represented in a phonological code, then one might expect phonological complexity or similarity to create problems. Exactly what problems depends on one's theory. If, for example, maintenance in a phonological code is important only on task-repeat trials (Liefooghe et al., 2005) then the effects should presumably be restricted to such trials and larger with longer preparation times. If one sees task-set activation as a process of serially activating one S-R rule at a time (e.g. Lien, Ruthruff, Remington & Johnston, 2005), and if this is done phonologically, then one might expect the largest effects of articulatory duration and similarity on RT on switch trials, especially when there is no time for preparation.

However, such predictions assume that the representations used for activating or maintaining S-R rules for task-set control are the same as those used in the phonological buffer. An alternative possibility is that task-set control involves the representation of S-R rules in a procedural working memory buffer (e.g. Meyer & Kieras, 1997; Rubinstein et al., 2001) distinct from declarative working memory (including the phonological buffer). Oberauer's (2009) recent account distinguishes between declarative and procedural working memory in just this way. The task-sets in play are represented in the activated part of procedural long term memory, and only the currently operative task-set is held in, and must be retrieved into, a capacity-limited procedure buffer that Oberauer calls the “bridge”. Task-switch RT costs include the time it takes to upload the relevant task-set into the bridge when the task changes. Elsewhere in this thesis we report explorations of the effect of the number of tasks in play (Chapter 2) and the complexity of the task on this process (Chapter 4).

Oberauer's (2009) account assumes, as have many others (James, 1890; Anderson 1976) that declarative and procedural memory are parallel but separate systems, with separate capacity limits. Evidence for an anatomical distinction began to accumulate in the 1960s (Milner, 1962). A long tradition of research on skill acquisition (Fitts, 1954; Anderson, 1982) assumes that acquisition of a new task-set begins with “declarative stage”, which is often accompanied by verbal mediation. In a transitional stage this declarative representation is transformed into a set of procedures which then become progressively more automated, and the role of verbal mediation drops away. Indeed, in the practice blocks of task-switching experiments it is common for participants to overtly vocalise both goals ("colour") and rules ("red-left") or report covertly doing so. Oberauer similarly notes that although declarative representations in working memory cannot control action directly, they can help establish procedural representations (also see Monsell, 2003; Kray, Kipp & Karbach, 2009). And although the declarative representations might continue to exist thereafter (our participants may continue to rehearse the S-R rules verbally) they no longer play a functional role in task-set control. One might then expect effects diagnostic of phonological representation to be detectable early in practice but perhaps not later.

This idea has received some support. In one experiment (Kray, Eber & Karbach, 2008), participants alternated predictably between a semantic categorisation task and a perceptual colour task. As expected, articulatory suppression was found to increase the switch cost, but this effect steadily decreased with practice. However, a large effect of AS on the switch cost remained, even after participants had performed over a thousand trials. But as this study did not include an appropriate secondary task control condition (such as the foot tapping control task used by Miyake et al., 2004), it is possible that at least some of the effect of AS found by Kray et al. (2008) can be attributed to general secondary task demands. Logan and Schneider (2006) present evidence that participants are likely to generate verbal task-mediators when the task-cues are non transparent, but that the use of such mediators decreases with practice.

In summary, if phonological codes are used to represent the set of S-R rules currently in play, we expect to observe effects of articulatory complexity/duration and phonological similarity of the set of stimulus terms, although the exact interaction of these effects with switching and preparation is hard to predict. If, however, task-set control is mediated by non-linguistic representations of S-R rules in procedural working memory, then we would expect no effects, except perhaps early in practice.

### Experiment 7

Experiment 7 manipulated the word length of the stimulus terms in the S-R rules of a task-set under preparation. A task-cueing paradigm was used, requiring participants to switch between a task requiring arbitrary key-press responses to pictured objects with monosyllabic names, and an equivalent task in which the objects had trisyllabic names and the same responses were used. If task-set preparation involves reinstating and maintaining S-R rules in phonological working memory, then the monosyllabic task-set should be easier to reinstate, and we could expect to see smaller switch costs and/or a greater benefit from the opportunity to prepare, and hence a larger reduction in switch cost. Picture rather than word stimuli were used to avoid confounding phonological with visual complexity. The difference between the two sets of names in articulatory duration was checked in Experiment 7a.

| Pair number | 1 syllable | | 3 syllable | |
|---|---|---|---|---|
| **1** | dog | 545 | elephant | 540 |
| **2** | arm | 695 | banana | 695 |
| **3** | ring | 900 | parachute | 888 |
| **4** | shelf | 855 | strawberry | 845 |
| **5** | tap | 565 | typewriter | 565 |
| **6** | stamp | 615 | kangaroo | 617 |
| **7** | leg | 766 | pyramid | 795 |
| **8** | goat | 883 | cucumber | 880 |
| **9** | bed | 530 | butterfly | 528 |
| **10** | fork | 585 | umbrella | 595 |
| **11** | dart | 810 | caravan | 795 |
| **12** | skirt | 895 | microscope | 880 |
| **Mean naming latency (ms)** | | **720** | | **719** |

**Table 5.1** Monosyllabic and threesyllabic words (and normative naming latencies) descriptive of pictures used in Experiment 7.

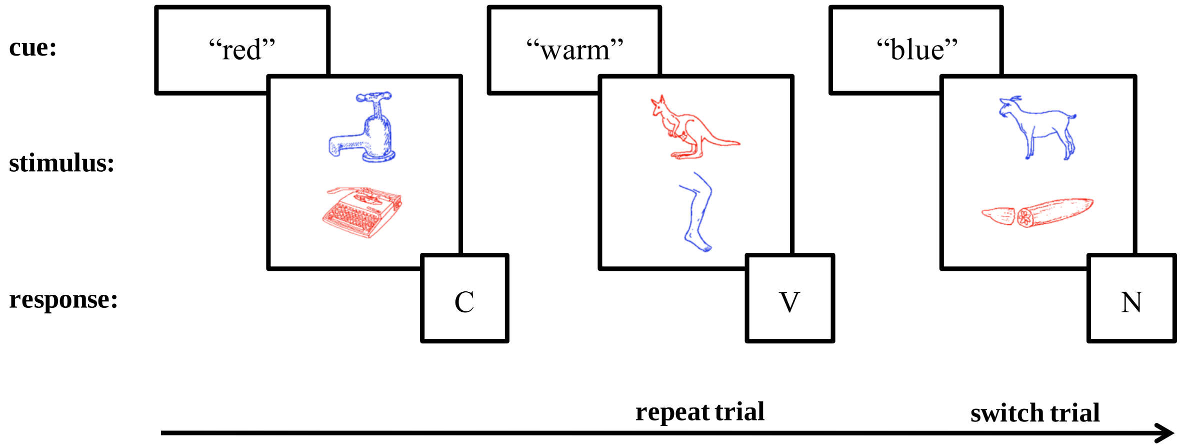


**Figure 5.1** Example trial sequence and stimuli for Experiment 7.

## Method

*Participants* Twenty-four participants (aged between 19 and 34 [M =21.4]; 4 male and 20 female) provided informed consent before taking part. All participants were paid between £5 and £7.20, depending on the speed and accuracy of their performance.

*Procedure and design* For each task four object pictures were each mapped to one of four response keys (C, V, B and N) on the PC keyboard, pressed with the middle and index fingers of left and right hands. For half the participants, objects with short names were presented in red, and objects with long names in blue, and vice versa for the other half. On each trial, participants saw both a red and a blue object one above the other, equidistant from the fixation point in the centre of the screen. The visual angle between the fixation point and the centre of each image was approximately 1.4°. The red object was above the blue object on a half of the trials and below it on the other half, so that the location of the task-relevant picture on each trial was unpredictable. Task switches were also unpredictable, with a 1:2 ratio of switch to repeat trials. There were no immediate stimulus repetitions. An auditory cue word told participant which object to respond to; the cue was ‘red’ or ‘warm’ for the red object, ‘blue’ or ‘cold’ for the blue object. The cue was selected from alternating pairs so that there were no immediate cue repetitions. Its spoken duration was approximately 350 ms. On each trial, a blank screen (500 ms) was followed by a fixation cross preceding the cue onset by either 1300 ms or 100 ms; the cue's onset then preceded each pair of pictures by either 100 ms (short CSI) or 1300 ms (long CSI), so that the response-stimulus interval (RSI) was kept constant at 1900 ms. Long CSI blocks and short CSI blocks alternated, and the order of CSI was counterbalanced between participants. Each block consisted of 1 warmup and 32 experimental trials, and there were 24 experimental blocks in total; in total each instance of 2 tasks x 2 cues x 3 switch/repeat x 2 CSI x 32 stimuli (= 768) occurred once.

Prior to the experimental blocks, participants completed a practice session consisting of 2 single task blocks (16 trials each) and 2 task switching blocks (16 trials each, one long CSI and one short CSI block). With regards to the single task practice, half of the participants practiced the trisyllabic task-set first, and half of the participants practiced the monosyllabic task-set first. Each of the four responses was made four times in both single task practice blocks. At the end of the session, participants were asked if they noticed any differences between the red and the blue objects, to determine whether they were aware of the difference in name length. All but one participant reported not having noticed this difference. The exceptional participant was replaced.

*Selection of images* The images were chosen from a database of line drawings of objects developed and normed for naming latency and accuracy by Linda Wheeldon (e.g. Wheeldon & Monsell, 1994). Although each participant was exposed only to four objects with monosyllabic names and four with trisyllabic names, in total 24 different pictures were used. This was to minimise the chance that any differences in task switching performance between the two sets were due exclusively to some property of particular pictures. The two sets were matched for naming latency as this would indirectly provide an approximate match for frequency and any other property influencing naming latencies. Oldfield and Wingfield (1965) found a high inverse correlation between object naming latency and frequency. The average naming latencies for the 12 monosyllabic and the 12 trisyllabic items were 720 ms and 719 ms, respectively (see Table 5.1).

Each monosyllabic item was approximately matched in naming latency with a trisyllabic item to form one of 12 pairs. From the complete set, 6 sets of 4 pairs were created. The compilation of these sets was subject to the following constraints: each set contained 4 items (2 monosyllabic and 2 trisyllabic) of low naming latency (< 700 ms), and 2 items (2 1 syllable and 2 3 syllable items) of relatively high naming latency (> 700 ms). In addition to that, attempts were made to ensure that pictures that were either semantically (leg and arm) or perceptually (cucumber and banana) similar were not part of the same set. In this way, the following 6 sets were created (see Table 5.1 for pair numbers): set 1 contained pairs 1 to 4, set 2 pairs 2 to 6, set 3 pairs 4 to 8, set 4 pairs 6 to 10, set 5 pairs 8 to 12, and set 6 pairs 11, 12, 1 and 2. With 24 participants taking part, each set was used 4 times.

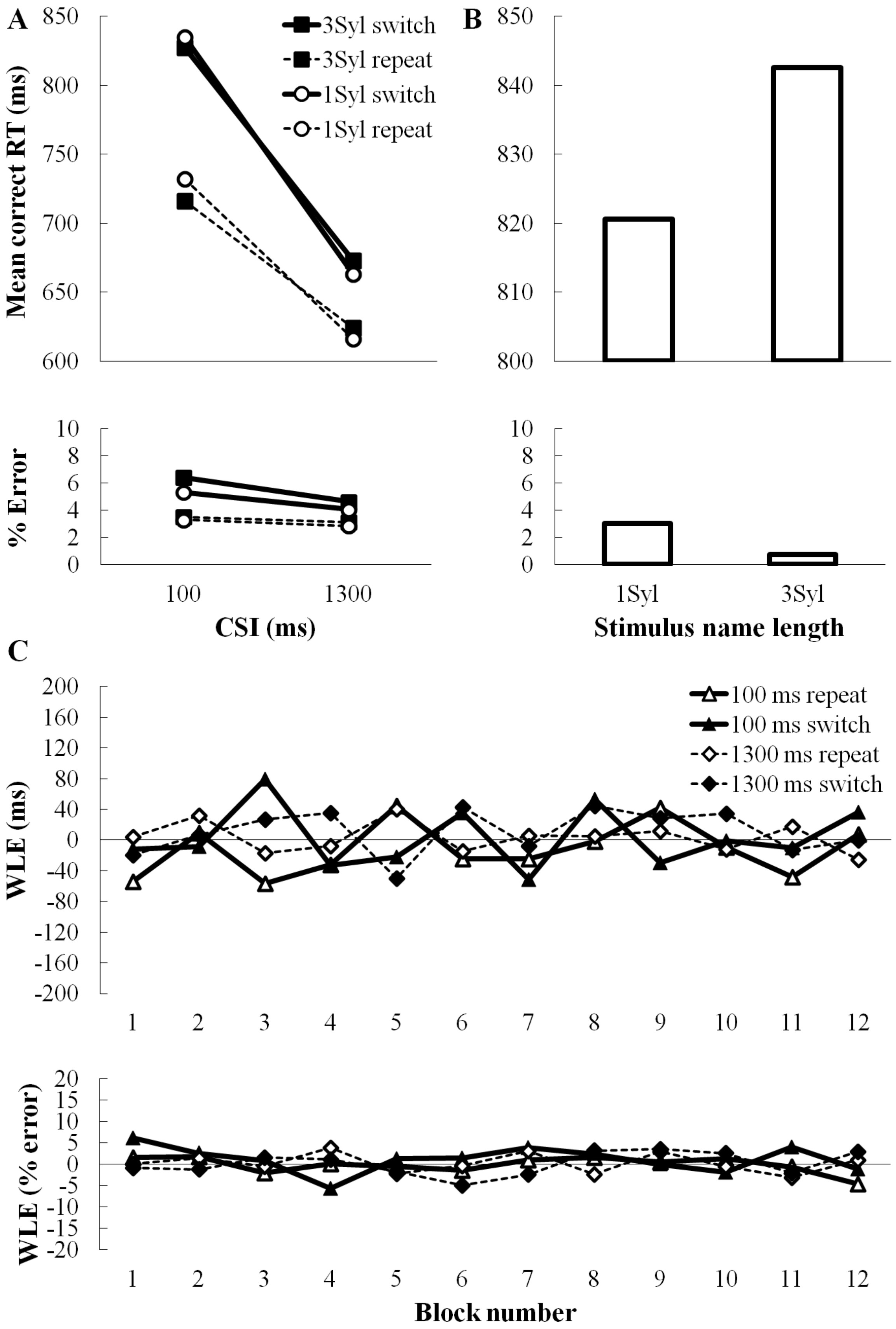


**Figure 5.2** Mean correct RTs and error data for Experiment 7. **A** RTs and errors for the one- and three-syllable stimulus names as a function of switch/repeat and CSI. **B** Single task practice data. **C** Word-length effects (WLE) plotted for switch and repeat trials, by CSI (100 ms or 1300 ms), as a function of block number.

## Results

The first trial of each block and trials following an error were excluded from analysis. Trials with very short (< 200 ms) or very long (> 3000) RTs were also excluded from the computation of mean correct RTs (0.02% of correct responses).

*Effect of stimulus name length in the task switching blocks* The effects of a task switch, preparation (100 ms or 1300 ms CSI) and task (i.e. name length) are summarised in Figure 5.2A. Repeated measures ANOVAs with the within-subject variables name length (monosyllabic or trisyllabic), CSI (100 ms or 1300 ms), congruency (congruent or incongruent) and switch (switch or repeat) were run on the mean correct RT and % error data. Mean correct RTs were almost identical for the trisyllables task (710 ms) and the monosyllables task (712 ms), $F<1$, SE of the mean difference = 7 ms. As usual, RTs were faster on long CSI trials (644 ms) than on short CSI trials (778 ms), $F(1,23) = 163.05$, $p < 0.001$. Participants were also faster on repeat (672 ms) than on switch trials (750 ms), $F(1,23) = 106.89$, $p < 0.001$. This substantial switch cost was smaller on long CSI trials (48 ms) compared to short CSI trials (107 ms), resulting in a significant 55% reduction in switch costs with preparation of 59 ms $F(1,23) = 23.23$, $p < 0.001$.

The analysis also revealed a small but reliable congruency effect of 17 ms, $F(1,23) = 13.55$, $p < 0.01$; and a near reliable interaction between congruency and switch, $F(1,23) = 4.08$, $p = 0.055$. With regards to the latter interaction, congruency effects were slightly smaller on switch (8 ms) than on repeat trials (26 ms). This unusual pattern of results[19] was mostly caused by faster RTs on congruent repeat compared to incongruent repeat trials.

With regards to the interactions with name length, the following was found: switch costs were very similar for the monosyllable (80 ms) and trisyllable (75 ms) tasks, $F<1$, and so were the RISC effects (62 ms and 56 ms, respectively), $F<1$. However, name length did interact significantly with CSI, $F(1,23) = 5.54$, $p = 0.028$: compared to monosyllable items, RTs for trisyllable items were slightly faster (by 12 ms) for the short CSI, and slightly slower for the long CSI (by 9 ms), but these differences were not separately reliable. This unexpected finding was further qualified

---

[19] Interactions found between switch and congruency normally reflect larger congruency effects on switch trials, caused by slower RTs on incongruent switch trials particularly.

by a reliable three-way interaction between task-set length, CSI and congruency, F(1,23) = 5.341, p = 0.03: RTs were 27 ms faster for trisyllabic incongruent stimuli at a short CSI, but 13 ms slower at a long CSI (for congruent stimuli, trisyllabic RTs were slightly slower at both short (3 ms) and long (5 ms) CSIs). Although this result is statistically reliable, there is no obvious explanation for it. The error graph perhaps reveals some evidence of a speed-accuracy trade-off (participants made more errors on three-syllabic items in the short CSI), although the same three-way interaction is not reliable in the % error data.

Error rates were similar for trisyllable (4.4%) and monosyllable (3.9%) tasks, F<1, and the effect of name length interacted reliably with no other variables (all Fs≤1). The error cost was slightly larger when switching to the trisyllable (2.2%) than to the monosyllable (1.6%) task, but the difference did not approach significance, F<1. There were more errors on switch (5.1%) than on repeat trials (3.2%), F(1,23) = 17.11, p < 0.001; on short CSI (4.6%) than on long CSI (3.7%) trials, F(1,23) = 4.33, p = 0.049; and on incongruent (5.1%) than on congruent (3.2%) trials, F(1,23) = 36.30, p < 0.001. The only reliable interaction in the error data was between congruency and switch: unlike the RT data, the congruency effect was larger on switch (3.5%) than on repeat (0.3%) trials, F(1,23) = 23.77, p < 0.001. The error switch cost did reduce with preparation from 2.4% in the short CSI to 1.4% in the long CSI, this RISC effect was only marginally reliable, F(1,23) = 2.99, p = 0.096.

*Effect of stimulus name length during practice* Given the possibility, discussed in the Introduction, that phonological mediation of S-R rule representation might play a role only early in practice, three additional analyses were conducted: (1) A 2 (name length) x 2 (CSI) x 2 (switch) repeated measures ANOVA was run on the data from the first two experimental blocks only; (2) The complete data set was also analysed with block pair as a factor in a 2 (name length) x 2 (CSI) x 2 (switch) x 6 (block pair) analysis; and (3) the single task practice data was analysed. For all three analyses data was pooled over congruency due to small trial numbers. Given the very small number of trials available in the two single-task practice blocks, to avoid empty cells trials following an error were included in this analysis. Only effects of or interactions with name length are reported below.

Figure 5.2C shows performance as a function of name length and switch/repeat over the course of the experimental blocks. There is no hint of an early name-length effect disappearing with practice. And indeed, neither the first nor the second analysis revealed any significant main effects of or interactions with task-set length.[20] The single task practice data were also analysed in order to reveal any potential effects of task-set length on performance (see Figure 5.2B). Although participants were on average 22 ms slower in categorising trisyllable items, the main effect of task-set length did not approach significance, $F<1$. Furthermore, participants actually made more errors on monosyllable items (3.0%) than on trisyllable items (0.8%), $F(1,23) = 6.677$, $p = 0.017$. In sum, there is no convincing evidence that name length had an effect on task switching performance or even a detectable effect during the learning of the S-R pairs.

### Experiment 7a: Measurement of articulatory duration

If it is articulatory duration rather than articulatory complexity that matters (see Introduction), the null effect of name length argues against a role for phonological codes only if there was an effective difference in articulatory duration between the monosyllabic and trisyllabic items. To check this, the participants who took part in Experiment 7 were asked to return for measurement of their overt and covert articulation rates. Sixteen (of 24) were able to do so. These participants were presented with 2 lists of 4 nouns, which were the nouns depicted by the images the participant had seen during the experiment in the order of the corresponding keys, left to right. For example, for one participant one list contained the words LEG, GOAT, BED and FORK; and the other list contained the words PYRAMID, CUCUMBER, BUTTERFLY and UMBRELLA.

Prior to the start of measuring articulation rates, participants first practiced saying each list five times, alternating between lists. During the last two rehearsals of each list, participants also timed their performance with a stop watch. The participants pressed the ‘start’ and ‘go’ buttons on the stop watch, and then showed the timing to the experimenter. These timings were not recorded, but familiarised them with the use of the stop watch. If participants did forget (any of) the items in a list, they were reminded verbally of the list by the experimenter. After this practice they alternated between the two lists (with the name-length of the first list balanced), saying each list aloud five times whilst measuring their speed. All participants were instructed to go through the

---

[20] The % error analysis did reveal a marginally reliable interaction between CSI, task and block pair, $F(5, 115) = 2.36$, $p = 0.052$ Huynh-Feldt. However, this interaction was only reliable as an order 4 interaction, and is therefore difficult to interpret.

list as quickly as possible (cf. Baddeley & Andrade, 1994, on the importance of speeded rehearsal) but without ‘cheating’, i.e. chopping off (parts of) words. The experimenter recorded the completion times in centiseconds.

Covert rehearsal rates were then also collected. The procedure was identical to that for overt rehearsal except that for each measurement the participant was asked to cycle through the list four times, instead of once. In total, this procedure lasted no longer than 10 minutes, after which participants were thanked, paid £1 and debriefed.

From Figure 5.3 it can be seen that completion times were longer for the trisyllable list than for the monosyllable list for both overt and covert articulation. The increase in completion time for trisyllable lists was very similar for overt (43%) and covert (42%), articulation. A 2 (overt/covert articulation) x 2 (mono-/trisyllable) repeated measures ANOVA yielded reliable main effects of articulation type, F(1,15) = 168.28, p < 0.001 and name length, F(1,15) = 146.66, p < 0.001, and a reliable interaction F(1,15) = 43.56, p < 0.001. Furthermore, longer articulation times for trisyllabic lists were observed for each participant.

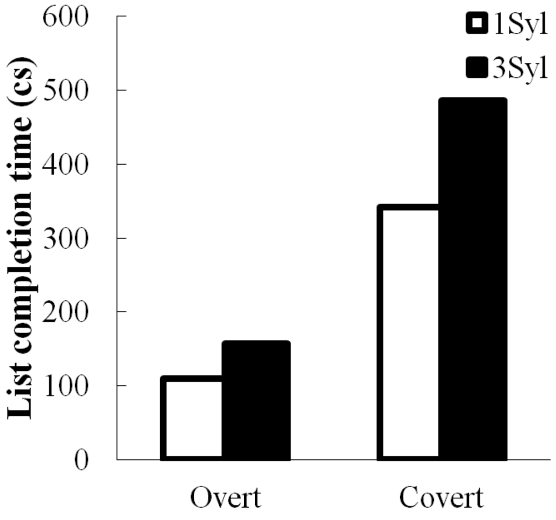


**Figure 5.3** List completion times in centiseconds (cs) for mono and trisyllable stimulus names, in the overt condition (participants cycled through list once) and the covert condition (participants cycled through list four times) in Experiment 7a.

### Discussion

Experiment 7 set out to investigate whether or not task-set preparation makes use of phonological representations of S-R rules. It exploited the well-known effect of word length on phonological working memory and required participants to switch between classifying two sets of pictures whose names differed in both phonological complexity (one syllable versus three) and articulatory duration (the one-syllable words took about 40% shorter to articulate both overtly and covertly, as shown in Experiment 7a). If phonological representations of S-R rules are rehearsed and/or maintained in the phonological buffer during task-set preparation, and this contributes to task-set preparation, preparation should be less effective for the task with the longer names. There was no evidence that this was the case. Neither the switch cost nor its reduction with preparation were influenced by the spoken word duration, or phonological complexity, of the stimulus terms. Moreover, overall RTs and error rates were also very similar for the monosyllable and trisyllable task-sets. As phonological representations might be particularly important at the start of the experiment, when the task rules are not well-established, effects of practice were also analysed, but provided no evidence that name length was important even early in practice.

## Experiment 8

The remaining two experiments exploited, in just the same way as word length in Experiment 7, another factor known to have a major impact on retrieval from or maintenance in phonological working memory, the phonological similarity of the stimulus terms in the S-R rules. In Experiment 8, instead of pictured objects, participants switched between identifying two sets of consonants with key presses. If representation in phonological working memory is critical to task-set reconfiguration and preparation, we might expect better performance and smaller switch costs, and/or a greater benefit of preparation when switching to the task with phonologically dissimilar stimulus terms.

### Method

*Participants* Thirty-two participants (aged between 18 and 29 [M = 19.1]; 6 male and 26 female) provided informed consent prior to taking part. In return for participation, all participants received 1 research credit and up to £2.50, depending on the speed and accuracy of their performance.

*Design and procedure* This experiment was identical to Experiment 7, except for the stimuli used. Instead of identifying one of two pictures on the screen, participants responded to one of two letters, one above the other (see Figure 5.4 for example stimuli). The visual angle between the fixation point and the centre of each letter was approximately 0.7°. Participants switched between identifying with a key press members of a set of similar sounding consonants (either D, P, B and T; or X, S, F and Z) and members of a set of dissimilar sounding consonants (Q, J, R and ?) in which the fourth letter "?" was manipulated between participants to be one of the letters from the phonologically similar set not presented to that participant. This was done so that if an effect of phonological similarity were obtained, it would be possible to compare performance (between participants) for the same consonant when it was a member of a similar and a dissimilar set, and thus test the possibility that the effect is due to some other property of the specific sets of stimuli used. For example, for one participant, the similar set might be D, P, B and T; and the dissimilar set would be Q, J, R and X, and their performance on X could be compared to the performance of other participants for whom X was a member of the similar set X, S, F and Z. For half the participants, letters from the similar set were presented in blue and letters from the dissimilar set in red, and vice versa for the other half of the participants. Participants responded to the letters with the left and right hands' index and middle fingers, using the four left most keys on a five keyed response box (with the order of response assignments reversed for half of the participants). The composition of both the practice and experimental session blocks was otherwise identical to those of Experiment 7.

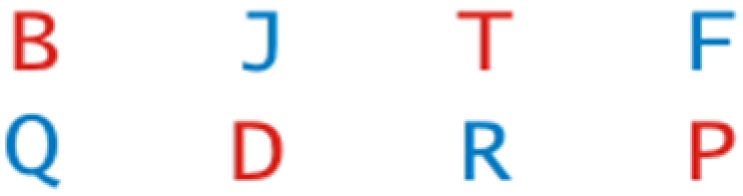


**Figure 5.4** The stimulus sets used in Experiment 8 (dissimilar set in blue: Q, J, R and F; similar set in red: B, D, T and P), arranged to provide illustrations of stimulus pairs.

## Results

The same data exclusion and analysis strategy was applied as in Experiment 7, resulting in the exclusion of 1% of the correct RTs. The results are summarised in Figure 5.5.

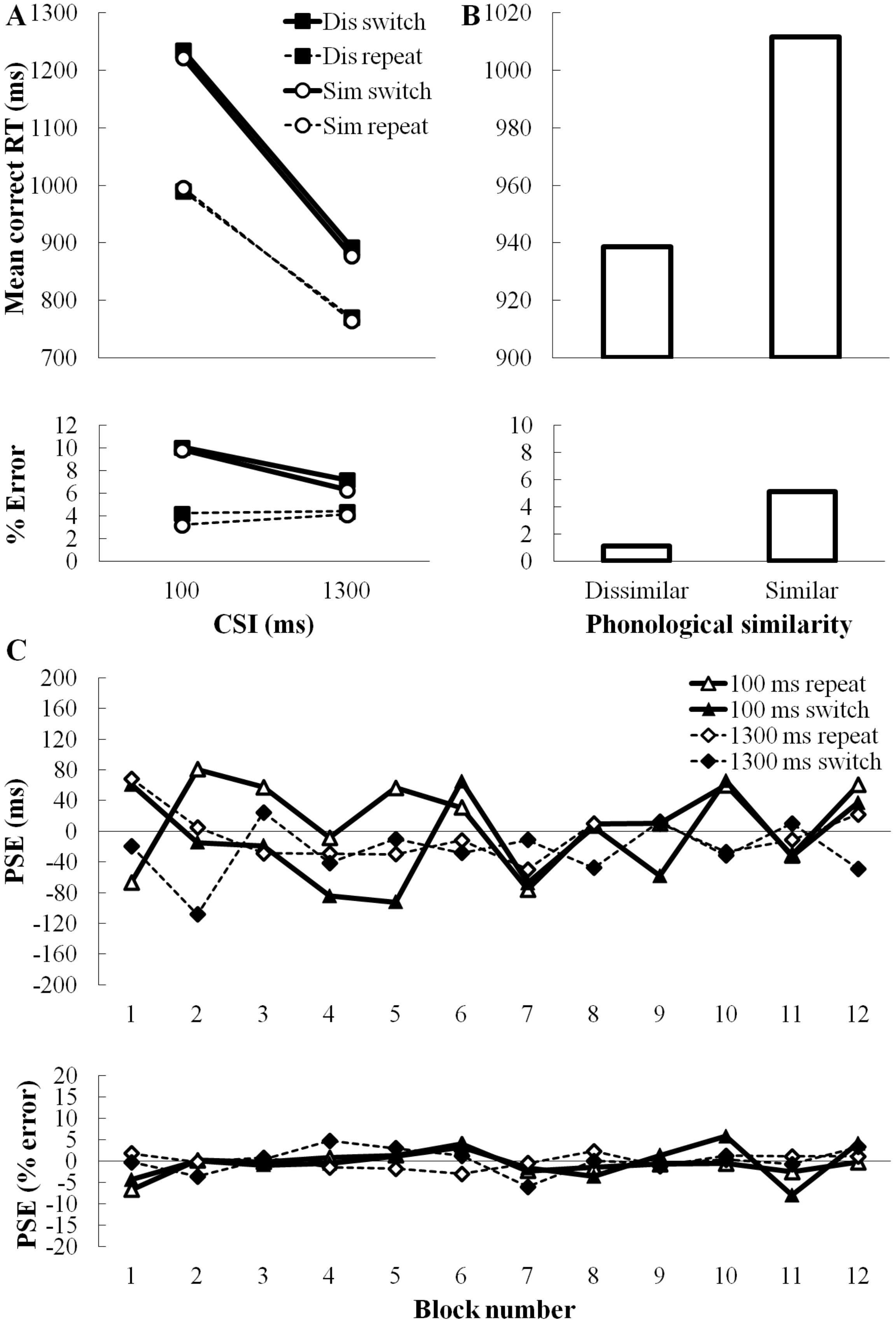


**Figure 5.5** Mean correct RTs and error data for Experiment 8. **A** RTs and errors for the similar and dissimilar stimulus names as a function of switch/repeat and CSI. **B** Single task practice data. **C** Phonological similarity effects (PSE) plotted for switch and repeat trials, by CSI (100 ms or 1300 ms), as a function of block number.

*Effect of phonological similarity in the task switching blocks* With regards to the mean correct RTs, neither the main effect of phonological similarity nor any of its interactions approached significance (F<1 for all except the four-way interaction, for which F(1,31) = 1.60). Mean correct RTs were very similar for the tasks with phonologically similar (965 ms) and dissimilar (972 ms) items, SE of the mean difference = 17 ms. As can be seen from Figure 5.5A, mean switch costs were similar for phonologically similar and dissimilar conditions (183 ms and 169 ms, respectively), as were RISC effects (114 ms and 124 ms, respectively). Otherwise, participants were faster on repeat (880 ms) than on switch (1056 ms) trials, a substantial switch cost of 176 ms; F(1,31) = 350, p < 0.001. There was a substantial improvement in performance at the long CSI (826 ms) relative to the short CSI (1111 ms), F(1,31) = 381.38, p < 0.001, and the switch cost reduced from 235 ms at the short CSI to 117 ms at the long CSI; a reduction of 119 ms or 50%, F(1.31) = 34.64, p < 0.001. A marginally reliable congruency effect of 15 ms was obtained, with RTs being faster on congruent (961 ms) than on incongruent (976) trials, F(1,31) = 3.7, p = 0.064. In addition to this, the switch cost was larger on incongruent (191 ms) than on congruent trials (161 ms), F(1,31) = 4.9, p = 0.034.

The % error data largely concurred with the mean correct RT data. Participants committed slightly fewer errors in the phonologically similar (5.9%) than in the dissimilar (6.5%) condition, F(1,31)= 1.78. Participants made fewer errors on repeat (4%) than on switch trials (8.4%), F(1,31) = 76.87, p < 0.001; and fewer errors on long CSI (5.5%) than on short CSI trials (6.8%), F(1,31) = 9.67, p = 0.004. The error switch cost was reduced reliably from 6.2% in the short CSI to 2.5% in the long CSI, F(1,31) = 13.3, p = 0.001. Participants also made fewer errors on congruent (4.7%) than on incongruent trials (7.7%), F (1,31) = 36.95, p < 0.001; and the congruency effect also reduced with preparation, from 3.8% in the short CSI to 2.2% in the long CSI, F(1,31) = 4.68, p = 0.038.

As explained in the method, for each participant, the dissimilar task-set contained one item (which we call the “common” stimulus) from the similar task-set used by other participants. In order to determine whether the null effect of phonological similarity would still hold when the stimuli were identical (between subjects), a 2 (similar or dissimilar) x 2 (CSI) x 2 (switch) repeated measures ANOVA was run on a restricted data set, which only contained dissimilar task trials which required a response to the common stimulus (i.e. excluding responses to the letters Q, J and R). This analysis pooled over congruency to avoid empty cells and only responses from the left and right middle finger were included, as the common stimulus was never responded to with the left or right index finger. With regards to the mean correct RTs, RTs were identical for similar and dissimilar stimulus sets (917 ms). As can be seen from Figure 5.6, switch costs (169 and 183 ms) and RISC effects (183 and 159) were also alike for the similar and dissimilar stimulus sets, respectively. Indeed, for the main effect of, and all interactions with, phonological similarity, F<1. With regards to the % error data, error rates were comparable for the similar (6.8%) and dissimilar (5.9%) task-sets, F(1,31) = 1.537, p = 0.224. For all interactions with phonological similarity, F<1. It would appear that the null result of phonological similarity in the complete data set was not due to an effect being masked by an accidental difference in the opposite direction in the difficulty of discriminating the two sets of items, or associating responses to them.

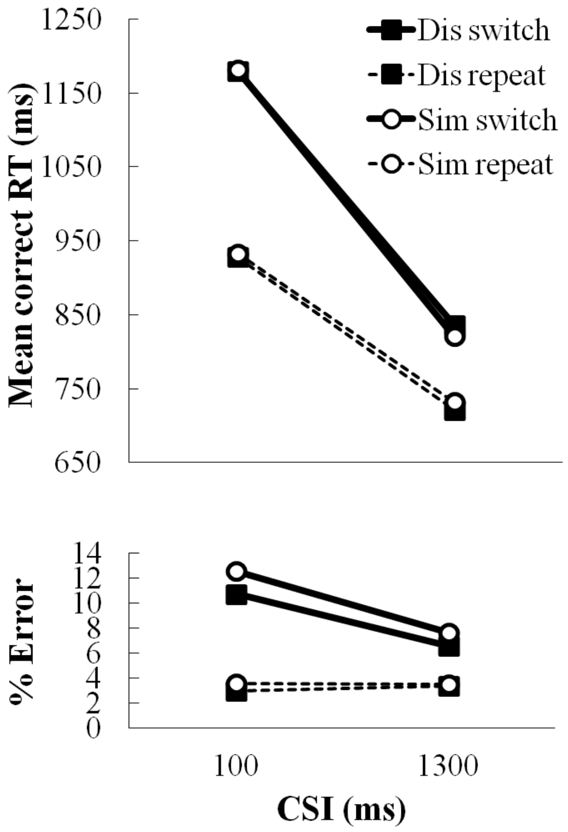


**Figure 5.6** Mean correct RTs and errors for the “common” stimulus in Experiment 8.

*Effect of phonological similarity during practice* To investigate any potential effects of practice, data are plotted as a function of block in Figure 5.5C in the same way as for Experiment 7, and the same three analyses were run. There is no suggestion of an effect of similarity present in the initial blocks and then disappearing. For the first analysis, limited to the first two experimental blocks, some participants were excluded (4 participants for the mean RT analysis and 2 participants for the % error analysis) due to empty cells. Only effects of or interactions with phonological similarity are reported.

The first analysis did not reveal any significant effects of phonological similarity during the first two blocks. The three-way interaction between phonological similarity, CSI and switch did approach significance, F(1,27) = 2.99; p = 0.095 in the RT analysis, but no simple main effects of similarity were reliable. The second analysis also did not reveal any interaction of phonological similarity with blocks.

However, as can be seen from Figure 5.5B, an effect of phonological similarity was found in the single task practice data: participants were on average 73 ms slower classifying phonologically similar compared to dissimilar items, F(1,31) = 4.86; p = 0.035, and they also made more errors with similar (6.3%) than dissimilar items (4.6%), though this difference was not reliable, F = 1.06.

### Discussion

Experiment 8 further investigated the possibility that S-R rules are represented phonologically for task-set preparation and/or maintenance by requiring participants to switch between sets of task-rules for which the stimulus terms were phonologically similar or dissimilar consonants. Consistent with the findings on word length in Experiment 7, this manipulation had no detectable effect on overall performance, the switch cost or its reduction with preparation. The additional analysis run on the common stimuli (i.e. consonants that were part of a similar set for some participants, and a dissimilar set for other participants) ruled out the possibility that the null effect resulted from an effect of phonological similarity being cancelled out by an unlucky confound between similarity and some other difference in difficulty between the similar and dissimilar sets of items.

Interestingly, the analysis of the two single task practice blocks did reveal a significant phonological similarity effect: in spite of the small amount of data (only 16 trials each), RTs were reliably greater for phonologically similar compared to dissimilar items, and the error rates differed in the same direction. This finding is important for two reasons: firstly, it demonstrates that the phonological similarity manipulation used in this experiment can have an effect on performance in these participants. Secondly, it suggests that the contribution of phonologically represented S-R rules in this paradigm is extremely short-lived. Indeed, no phonological similarity effect was obtained at the start of the experimental blocks, which began after only 32 single-task practice trials and 32 task-cueing trials. These findings certainly agree with previous suggestions that the role of linguistic representations might be confined to the earliest stages of practice, when the task rules are not yet well established (e.g. Oberauer, 2009; Kray et al., 2008).

One potential limitation of Experiments 7 and 8 is that the tasks used in these experiments were not typical of tasks ordinarily used in task-cueing experiments (including those claimed to reveal a contribution of linguistic representations). In most such studies, the tasks between which participants switch have clearly distinct task goals. For example, participants might be required to switch between colour and letter identification (Goschke, 2000), or between a colour and shape classification (Miyake et al., 2004; Liefooghe et al., 2005). In contrast to this, in Experiments 7 and 8, participants switched between two tasks with identical goal descriptions – i.e. press the key corresponding to the identity of the object (Experiment 7) or letter (Experiment 8). All that changed was the particular set of items that the participant had to identify. The benefit of this design was that it facilitated the manipulation of spoken word duration and phonological similarity without introducing other confounds, such as task difficulty. The disadvantage is that it remains possible that effects diagnostic of phonological representation might still be found in more commonly used task cueing paradigms.

Another feature of Experiments 7 and 8 – displays consisting of compounds of two univalent stimuli each mapped to bivalent responses – has another possible limitation: because the object or consonant constituents were univalent it might be suggested[21] that participants maintained all 8 S-R rules in procedural working memory on all trials, rather than uploading only the task-relevant S-R rules when the task changes. Of course this would require some other basis for selecting the appropriate consonant for response, perhaps rapidly focusing the spatial spotlight on the object with the cued colour. We note the following three points. First, the large switch costs and RISC effects obtained in Experiments 7 and 8 would have to be attributed to something other than activating S-R rules – perhaps shifts of spatial attention? Second, the capacity of procedural working memory required to represent with equal availability 8 distinct S-R rules would have to be considerably larger than our findings in Chapter 4 suggest. And thirdly, even if all 8 S-R rules were maintained in working memory on all trials, then if this representation were phonological we would still expect effects of phonological similarity: effects on immediate serial recall are still obtained when phonologically similar items are part of a “mixed list”, which also contains dissimilar

---

[21] We thank Frederick Verbruggen and Stephen Lewandowsky for this suggestion

items (e.g. Farrell, 2006). These points notwithstanding, it is desirable to test the impact of phonological similarity in a more standard task-cueing paradigm, where the stimulus attribute attended to changes with the task, and this we did in Experiment 9.

## Experiment 9

Each participant performed two pairs of tasks in a task-cueing paradigm, one pair in each of two sessions. In one session, the stimulus was one of four consonants displayed in one of four visual textures (see Figure 5.7). Participants were cued to identify the consonant or its texture by pressing one of four keys. The set of four consonants had phonologically similar or dissimilar names, as in Experiment 8. In the other session, the stimulus was one of four line drawings of an object displayed in one of four colours. Participants were cued to identify with a key press either the object or the colour in which it was drawn. The object names were phonologically similar or dissimilar.

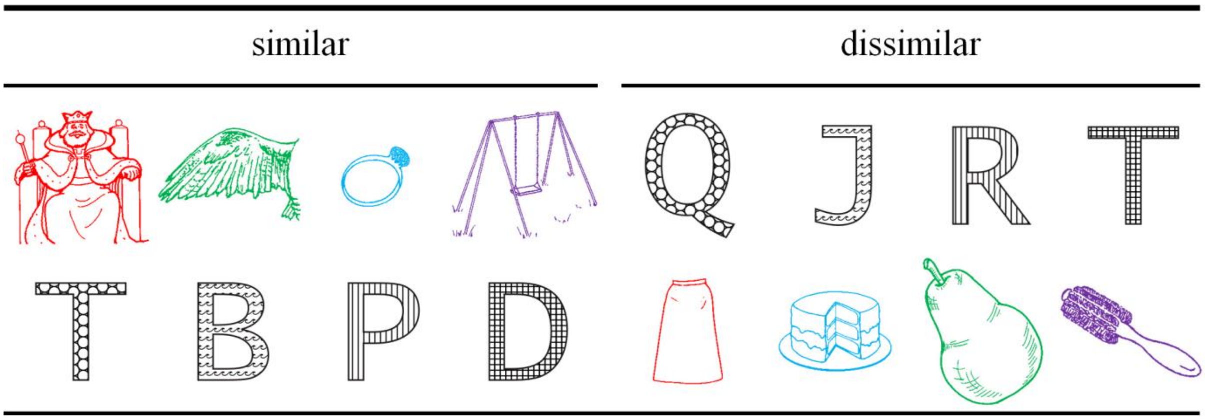
Half the participants were given phonologically similar letter names to identify in the letter task, and phonologically dissimilar object names in the object task. The other half were given phonologically dissimilar letter names, and phonologically similar object names. Hence although phonological similarity was manipulated between subjects for one task pair, it was manipulated in the opposite direction between the same set of subjects for the other task pair, providing overall a within-subjects (or within-subject-pairs) contrast.

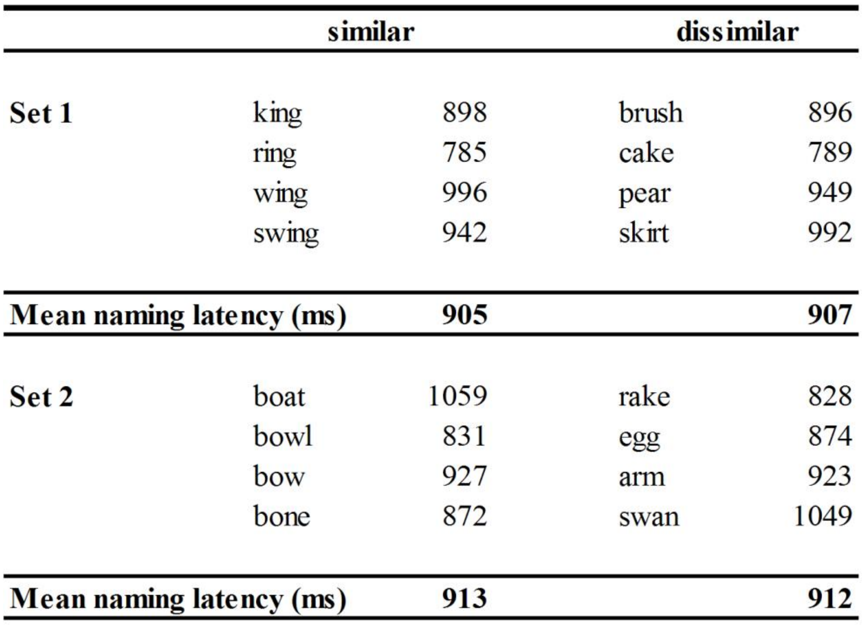
### Method

*Participants* Sixteen participants (aged between 18 and 23 [M = 20.1]; 6 male and 10 female) provided informed consent prior to taking part. In return for participation, all participants received between £10.60 and £15, depending on the speed and accuracy of their performance.

*Procedure and design* In order to manipulate phonological similarity within subjects, two sets of stimuli (see Table 5.2) were used: coloured pictures and patterned letters (see Figure 5.7 for examples). Each participant completed two sessions on consecutive days, switching between the colour task and the picture task in one session, and the letter task and the pattern task in the other session. The order of task pairs was balanced orthogonally to the manipulation of similarity. The phonological similarity of the pictures and letters was manipulated as described above, so that for half of the participants, the letters were phonologically similar, and the pictures dissimilar, and vice versa for the other half of participants. The letters used in this experiment were identical to those used in Experiment 8. That is, for half the participants given a similar set, it was D, P, B T, and for the other half F, S, X and Z. The dissimilar set was always Q, J, R and one letter from the similar set the participant did not see, using each one equally often. The letters could be filled with 4 different patterns (waves, dots, stripes and squares, see Figure 5.7), resulting in a total of 16 stimuli per participant (visual angle was approximately 5.7°). For the object task pair, all pictures were line drawings of monosyllabic nouns selected from the International Picture Naming Project (IPNP, Bates, D’Amico, Jacobsen, Székely, Andonova, Devescovi, Herron et al., 2003). For this task pair four sets of pictures (2 similar sets and 2 dissimilar sets) were used. One similar set comprised four objects with rhyming names (king, ring, swing and wing), and the other comprised four objects whose names shared onset and vowel nucleus (boat, bone, bowl, bow). The items in the similar sets were matched for naming latency with the items in the dissimilar sets (visual angle was approximately 5.7°). The auditory cue words instructing participants which task to perform were: “letter” and “symbol” (letter task), “pattern” and “texture” (pattern task), “picture” and “image” (picture task) and “colour” and “paint” (colour task).

Apart from these details, and the fact that the stimulus replacing the fixation point was a single central letter (in one session) and object drawing (in the other), the design and procedure for each session of Experiment 9 were identical to those of Experiments 7 and 8. Each session consisted of 792 trials (split up into 24 blocks of 33 trials), preceded by single task practice trials, and two blocks of 16 task-cueing practice trials.





| | similar | | dissimilar | |
|---|---|---|---|---|
| **Set 1** | king | 898 | brush | 896 |
| | ring | 785 | cake | 789 |
| | wing | 996 | pear | 949 |
| | swing | 942 | skirt | 992 |
| **Mean naming latency (ms)** | | **905** | | **907** |
| **Set 2** | boat | 1059 | rake | 828 |
| | bowl | 831 | egg | 874 |
| | bow | 927 | arm | 923 |
| | bone | 872 | swan | 1049 |
| **Mean naming latency (ms)** | | **913** | | **912** |

**Figure 5.7** Examples of stimuli used in Experiment 9 (top) assigned to participants for whom the picture stimuli are similar, and the letter stimuli dissimilar, and (bottom) vice versa.

**Table 5.2** Pictures used in Experiment 9 and their normative naming latencies.

### Results

The same exclusion criteria were used as in Experiments 7 and 8, resulting in the exclusion of 0.06% of correct RTs. As before a 2 (similar or dissimilar) x 2 (100 ms or 1300 ms CSI) x 2 (switch or repeat) x 2 (congruent or incongruent) ANOVA was run on the mean correct RTs and errors. An additional between-subjects variable (task assignment) distinguished between participants who had identified similar letters and dissimilar picture names, and participants who had done the reverse.

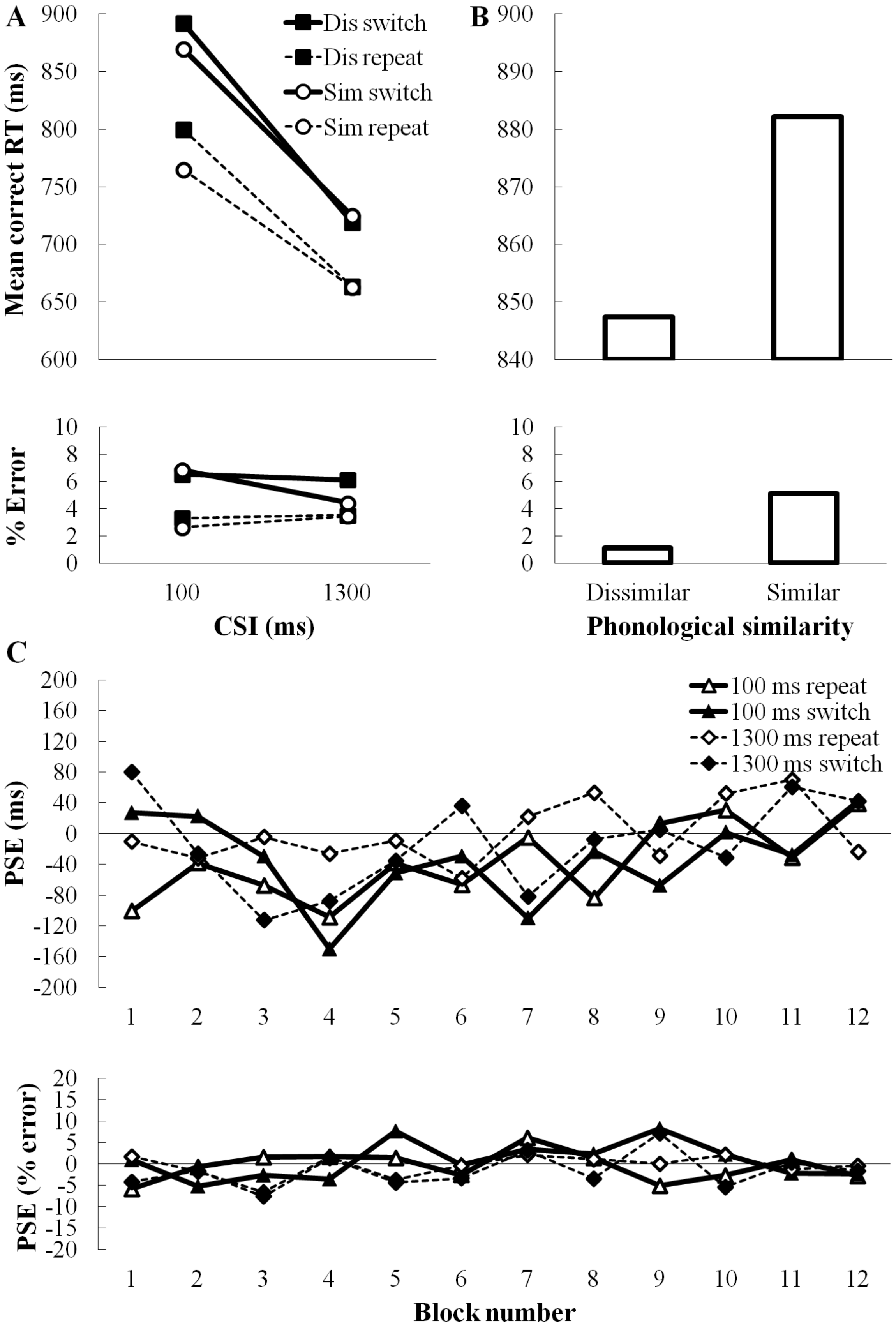


**Figure 5.8** Mean correct RTs and error data for Experiment 9. **A** RTs and errors for the similar and dissimilar stimulus names as a function of switch/repeat and CSI. **B** Single task practice data. **C** Phonological similarity effects (PSE) plotted for switch and repeat trials, by CSI (100 ms or 1300 ms), as a function of block number.

*Effect of phonological similarity length in the task switching blocks* As in Experiment 8, mean correct RTs were not shorter for dissimilar (769 ms) than for similar (755 ms) stimulus sets, F<1 (mean difference 14 ms, SE 45 ms). Similar and dissimilar task-sets also yielded comparable switch costs (84 ms and 74 ms, respectively), F < 1, and RISC effects (42 ms and 37 ms, respectively), F<1.

Overall, there was a reliable switch cost of 79 ms, F(1,14) = 40.69, p < 0.001, a reliable CSI effect of 139 ms, F(1,14) = 46.11, p < 0.001, and a reliable congruency effect of 53 ms, F(1,14) = 83.05. The switch cost reduced from 99 ms in the short CSI to 59 ms in the long CSI, F(1,14) = 8.94, p = 0.010, and the switch cost was also larger on incongruent (72 ms) than congruent (34 ms) trials, F(1,14) = 11.78, p = 0.004.

Although there were no effects of similarity (as above), there were (near) reliable interactions involving similarity and task assignment (similarity x task assignment, F(1,14) = 33.13, p < 0.001; CSI x task assignment, F(1,4) = 36.34, p < 0.001; similarity x CSI x switch x congruency x task assignment, F(1,14) = 4.46, p = 0.053). Interactions including task assignment and phonological similarity are relatively uninteresting because they arise from a difference in task difficulty (rather than a genuine effect of phonological similarity): RTs and errors were greater for the letter task (835 ms and 6.2%) than for the picture task (698 ms and 3%), hence in one group of subjects the phonologically similar stimulus names specified the more difficult task, and in the other group they specified the easier task.

As a different (but less powerful) test for an effect of phonological similarity, separate ANOVAs (CSI x switch x congruency) were run for each task with phonological similarity as a between-subjects factor. Though the means suggest a small phonological similarity effect in the picture task (47 ms), and a larger reverse similarity effect (RTs were 73 ms slower for dissimilar stimulus sets) for the picture task, neither the effect of similarity nor any of its interactions approached significance in either task, Fs<2.67 for the picture task, and Fs<1.75 for the letter task.

Overall error rates were comparable for similar (4.3%) and dissimilar stimulus sets (4.9%), as were switch costs (2.9% and 2.6%, respectively), F<1. The RISC effect was somewhat larger for similar (3.3%) than dissimilar (0.6%) stimulus sets but not reliably so, F(1,14) = 1.94, ns. In the errors there was an overall switch cost of 2.8%, F(1,14) = 22.09, p < 0.001, and a congruency effect of 5.2%, F(1,14) = 14.11, p = 0.002. Switch costs were also larger on incongruent (6.6%) than congruent (3.8%) trials, F(1,14) = 5.77, p = 0.031, and marginally larger in the short CSI (3.7%) than in the long CSI (1.8%), F(1,14) = 3.52, p = 0.082.

Again there were some reliable interactions involving similarity and task assignment (similarity x task assignment, F(1,14) = 7.6, p = 0.015; similarity x task assignment x switch, F(1,14) = 5.6, p = 0.040; similarity x task assignment x congruency, F(1,14) = 6.54, p = 0.023), but separate ANOVAs for the letter and picture task (as for the RTs) revealed that neither the effect of similarity (-0.5% in the picture task and -0.6 in the letter task) nor any of its interactions approached significance in either task, Fs < 2.23 for the picture task, and Fs < 1.99 for the letter task.

*Effect of phonological similarity during practice* Effects of practice were analysed in the same way as in Experiments 7 and 8. The first analysis was a 2 (similarity) x 2 (CSI) x 2 (switch) repeated measures ANOVA on the data from the first two experimental blocks, with task assignment as a between-subjects variable (pooled across congruency). One participant was excluded from the analysis because of an empty cell. In the RTs, there was no main effect of similarity, F<1, though the switch cost was larger for the phonologically similar items (109 ms) than for the dissimilar items (10 ms), F(1,13) = 3.54, p = 0.083 (all other interactions with similarity, F<1). For the errors, for the main effect of similarity and its interactions, F<1, except for the three-way interaction between similarity x CSI x switch, F(1,13) = 3.08, p = 0.103 (for similar items, switch cost reduced from 7.9% in the short CSI to -1.6% in the long CSI, RISC of 9.5%; for dissimilar items, there was a switch cost of 1.7% in the short CSI and of 4.7% in the long CSI, RISC of -3%).

The second analysis (includes block pair as a factor in a 2 (task-set length) x 2 (CSI) x 2 (switch) x 6 (block pair) analysis) revealed no effect of or interaction with phonological similarity in the RTs or the errors. Figure 5.8C does not suggest any modulation of the lack of effect of phonological similarity by practice.

Finally, the single task practice data was analysed. As in Experiment 8, RTs were slower for phonologically similar items (882 ms) than for dissimilar items (847 ms), although this difference of 35 ms was not significant, F<1. However, participants also committed more errors on phonologically similar (5.1%) than dissimilar trials (1.1%), and this phonological similarity effect of 4.0% was significant, F(1,14) = 6.17, p = 0.026 (see Figure 5.8B). Altogether, the practice analyses replicated the pattern of results obtained in Experiment 8: there was no effect of phonological similarity, except during the initial single task practice blocks.

## Discussion

The aim of Experiment 9 was to extend the result obtained in Experiment 8 to the case of switching between tasks requiring attention to different attributes of a single stimulus, and hence more closely resembling the paradigms used in other experiments on the role of phonology. The results were entirely consistent with the findings of Experiment 8. Phonological similarity of the stimulus terms in the S-R rules had no detectable effect on overall performance, task-switch costs, or task-set preparation as indexed by the reduction in switch cost with preparation. However, as in Experiment 8, a transient but reliable effect of phonological similarity effect was found in the initial single-task practice blocks.

## General discussion

Three experiments investigated whether S-R rules are represented phonologically for task-set preparation and maintenance. Previous studies suggest that linguistic (e.g. Mecklinger et al., 1999) and specifically phonological (e.g. Liefooghe et al., 2005) representations might contribute to task-set control. However the precise role remains unclear: roles proposed include keeping track of a task sequence when it must be held in memory (Bryck & Mayr, 2005), aiding task goal retrieval (Goschke, 2000; Miyake et al., 2004) and maintaining the S-R mappings in the phonological buffer (Liefooghe et al., 2005). The experiments reported here investigated the role of phonological codes in activating or maintaining S-R rules. Unlike previous experiments, which have used the effect of articulatory suppression (AS) as diagnostic of the involvement of phonological representations, we used word length (Experiment 7) and phonological similarity (Experiments 8 and 9).

On the whole we found no effect of either word length or phonological similarity on task-cueing performance, the switch cost or its reduction with preparation. These results argue against the idea that S-R rules are represented phonologically for task-set preparation and maintenance. Hence our results are inconsistent with Liefooghe et al.’s (2005) idea that the phonological buffer is used to maintain S-R mappings on repeat trials. They base their conclusions on two key findings: (1) smaller switch costs (due to inflated RTs on repeat trials) under AS, and (2) the disappearance of this effect when the task cues explicitly represent the S-R mappings. But it is possible that either or both of their manipulations (AS and cue type) affected elements of task-set control other than the representation of S-R mappings. It is possible that AS interfered with phonological representation at the task-set level rather than at the S-R level (also see below). The use of task-cues that explicitly provide the S-R rules eliminates not only the need of S-R maintenance in WM, but also the need for task-goal activation/retrieval and cue interpretation.

Indeed Miyake et al.'s (2004) findings indicate that a phonological representation of the task label *does* aid task-set retrieval, particularly when the task cues are not transparent. Our results are not inconsistent with those of Miyake et al. (2004), as it is possible that phonological representations contribute to task-set control at the task-set level (e.g. activating the task-goal through verbal self-instruction), but are not helpful in subsequent activation or maintenance of the S-R rules. Previous chapters presented evidence consistent with the idea that selecting a task from among alternatives (Chapter 2) and activating a S-R rule within a task-set (Chapters 3 and 4) are not subject to the same capacity limit, supporting distinct levels of action control for task-sets on the one hand, and the S-R mappings within a task-set on the other hand. It is possible that phonological representations are helpful at one level (task-set level, Miyake et al., 2004) but play no role at another (S-R level, this chapter).

How then are S-R rules represented, if not phonologically? Our results are consistent with the idea that the currently operative task-set is represented in a procedural working memory buffer (Rubinstein et al., 2001; Oberauer, 2009), separate from declarative WM. Declarative (possibly phonological) representations may aid the formation of novel task-sets but only compiled procedures can control action. Though the results reported in this chapter do not directly provide evidence for the declarative/procedural distinction, they are least consistent with it. And the alternatives appear unlikely: S-R mappings might be stored in declarative working memory in a non-phonological form, but we know that people readily use phonological representations for short term maintenance even when alternative options (e.g. visual imagery) are available (see Camos, Mora and Oberauer, 2010). Another alternative is that the S-R mappings are maintained phonologically in procedural working memory but in a way that is somehow immune to the manipulations of word length and phonological similarity that have large effects on declarative working memory. This seems both implausible, and at odds with the basic rationale for the existence of procedural working memory: the advantage of compiling perceptuo-motor translation procedures in a code appropriate to their performance, not their communication.

Finally, in Experiments 8 and 9 we did find an effect of phonological similarity in the single task practice blocks (the first 32 trials). This finding demonstrates that our phonological similarity manipulation can affect performance and supports previous suggestions (e.g. Fitts, 1954; Kray et al., 2008; Oberauer, 2009) that linguistic representations might be particularly important in the early stages of practice. Our results suggest that, at least for these small sets of S-R rules, the proceduralisation of a declarative representation proceeds rapidly: after 64 trials, there was no detectable effect of phonological similarity.

# Chapter 6: General Discussion

The aim of this thesis was to further define the properties of procedural working memory (PWM), and its contribution to task-set control. Of particular interest was the process by which one task-set is promoted into a most activate state (“uploaded” into the PWM buffer). As an index of this process we used the reduction in switch cost (RISC) with preparation that is typically found in task-cueing experiments.

The first two experiments asked whether the time required to upload a task-set into a PWM buffer is influenced by the number of tasks in play. Experiments 1 and 2 found that there was no effect of the number of tasks on the switch cost, or its reduction with preparation, provided that recency and frequency of usage were matched. Experiment 3 showed that the time required to select an S-R rule from within a task-set *does* increase with the number of alternative rules, even when recency and frequency are matched. Moreover, the finding that task-set preparation is more effective for the less complex task-set (Experiments 4-7, Chapter 4) provided further support for the notion that the most active task-set is represented in a limited capacity PWM buffer. Together, these findings support a distinction between two levels of action control: the level at which a task-set is selected from among other task-sets, and the level at which S-R rules are represented within a task-set.

Experiments 7-9 (Chapter 5) investigated the possibility that a phonological representation of S-R rules contributes directly to task-set control, and rejected it: the phonological properties of the stimulus terms in the S-R rules did not affect performance in a task-cueing paradigm. These results are also consistent with the idea that S-R rules are represented in a procedural WM buffer that is separate from declarative WM. I will now summarise the main findings of each chapter (in italics) and give some directions for future research.

### The effect of the number of tasks on task-cueing performance (Chapter 2).

*The time required to select and upload a task-set from the activated part of LTM into a capacity-limited PWM buffer is not influenced by the number of other potentially relevant task-sets – provided that recency and frequency of usage are matched.*

The successful preparation and implementation of a task-set requires a process by which that task-set-set temporarily dominates all other task-sets. If the duration of this process is influenced by the number of task-sets in play, then the switch cost should be larger with more tasks, particularly when there is no opportunity for advance preparation. Results of previous task switching studies yielded mixed results, not least in part because many of those studies confounded number of tasks with the recency and frequency of their usage.

In Chapter 2 participants switched among three or five tasks requiring the identification of a perceptual attribute of an object (Experiment 1) or the categorisation of a phonological/semantic attribute of a word (Experiment 2). Both experiments found no effect of number of tasks on overall RT, the switch cost or its reduction with preparation when the recency and frequency of task usage was matched. But in the uncontrolled data, the switch cost was larger with five tasks, particularly when there was no opportunity for advance preparation. Further analyses suggested effects of task recency and implied effects of task frequency, suggesting that it is not the number of tasks per se but the recency and frequency of usage that determines the time required to upload a task-set into a PWM buffer. This may explain why other studies did find an effect of the number of tasks on the switch cost, as they did not control for both recency and frequency.

The lack of a number of tasks effect in the frequency- and recency-matched data is consistent with theories of task switching according to which only the currently relevant task-set is maintained in a limited capacity buffer, whilst other (currently irrelevant task-sets) are held in a relatively capacity unlimited activated part of LTM (Oberauer, 2009). Recent data from Oberauer’s lab is consistent with this, too: switch costs were not larger when switching among three compared to two tasks (Da Silva Souza, Oberauer, Gade & Druey, in press). But at the same time these authors did find that list switch costs (longer RTs for items from a different list than the previous item) increased with the number of potentially relevant lists. The difference in the effect of the number of tasks and lists on their respective switch costs is inconsistent with Oberauer’s assumption that declarative and procedural WM operate according to analogous principles. But Da Silva Souza et al. did not match the frequency and recency of the tasks or lists, so before concluding that task- and list-representations are subject to different constraints in DWM and PWM (as Da Silva Souza et al. imply) we must first see if their number of lists effect survives matching for frequency and recency.

Within PWM itself, some questions remain with regards to the contributions of recency and frequency to task-set retrieval. Experiments 1 and 2 found somewhat steeper recency slope for unprepared switch trials in the frequency-matched probe tasks, but the effect of recency (its linear component) was not reliable in either experiment;

hence it remains unclear to what extent task recency contributes to retrieval time. Furthermore, our results suggested that task frequency contributes to retrieval time, but Experiments 1 and 2 were unable to address whether it is the relative or absolute frequency of task usage that matters. In relation to the fan effect, Anderson (1990) has argued that it is relative frequency that is important, but evidence for this would require an experiment which independently varies both factors.

### The effect of task-set complexity on single task performance (Chapter 3).

*The time required to select a response within a task-set is influenced by the number of alternative responses – even when recency and frequency of usage or matched.*

The finding that the switch cost (and overall RT) did not increase as a function of the number of tasks when these were matched for recency and frequency can be contrasted with typical set-size effects in choice RT experiments: RT increases when the number of S-R rules within a task-set is increased (Hick's law). At first sight this contrast suggests task-set control operates at two levels: at the level of task-sets, and at the level of S-R rules that constitute a task-set.

But in light of the apparent number of tasks effect found in the uncontrolled data of Experiments 1 and 2, it seemed necessary to investigate whether set-size effects within a task-set can also be eliminated by matching for the recency and frequency with which an S-R rule is selected. This turned out not to be the case: set-size effects were much reduced but not eliminated for recency and frequency matched stimuli. Clearly, the currently operative task-set is maintained in a capacity limited mode of representation, which appears distinct from the level at which task-sets (packages of S-R rules) are represented.

### The effect of task-set complexity on task-cueing performance (Chapter 4).

*Task-set preparation is more effective for the less complex task-set, suggesting that a capacity-limited PWM buffer can only maintain a very limited number of S-R rules in a "most active" state of representation.*

The experiments reported in Chapter 4 further investigated the properties of the limited-capacity PWM buffer (Oberauer's "bridge"), and the process of loading it, focusing on the effect of task-set complexity on the switch cost and its reduction with preparation. In all three experiments task-set preparation was more effective for the less complex task-set, as indexed by a larger RISC with preparation in that condition (in Experiments 5 & 6 this was mostly so when the stimulus was incongruent – this interaction will be discussed later). This finding suggests that the capacity of the bridge is extremely limited: it can only represent very few S-R mappings in a most active state. The results of Experiment 4 best reflect the severity of this capacity limit: a task-set specified by as few as 3 S-R mappings already appears to exceed the effective capacity of the bridge. In Chapter 4 I argued that when this capacity limit is exceeded, only a subset of the S-R mappings can be effectively prepared and maintained. Examination of the RT distributions (Experiments 4 and 6) supports that interpretation: task-set preparation was not affected by complexity for the fast RTs, but for slow RTs preparation was much less effective for the more complex condition, suggesting that for those trials the relevant S-R mapping had not been prepared.

Though these are the first experiments to convincingly demonstrate that complexity affects task-set preparation, the idea that task-sets can only be partially prepared (hence resulting in larger residual switch costs) because of capacity limitations is not new. According to Lien et al.’s (2005) partial mapping hypothesis, residual switch costs occur because preparation is only effective for a subset of the S-R mappings. The authors also claim that preparation of S-R mappings proceeds in a serial manner (from left to right). In support of this they found smaller or even absent switch costs for the leftmost S-R pairs, though the observation of zero costs for the leftmost in a row of responses with one stimulus per response was not replicated by Monsell and Mizon (2006). The findings reported in Chapter 4 agree that residual switch costs arise because of incomplete preparation though they make no assertions about the serial nature of this process.

In all three experiments, task-set complexity affected not only the switch cost but also the congruency effect: in Experiment 4, the congruency effect was larger for the more complex 3-choice condition, particularly with a long preparation interval. In Experiments 5 and 6, task-set preparation was less effective (as indexed by a smaller RISC) for the more complex condition, mostly (Experiment 5) or exclusively (Experiment 6) so when the stimulus turned out to be incongruent. Together these findings imply that the representation of the more complex task-set is more susceptible to interference. Why would this be the case? One possibility is that the more complex task-set is more susceptible to interference due to less effective task-set shielding (Dreisbach & Haider, 2008, 2009, Dreisbach & Wenke, 2011) in that condition.

According to Dreisbach and colleagues (2008, 2009, 2011), the purpose of task-sets is to shield the task representation from irrelevant information. Dreisbach and Haider (2009) required participants to classify eight words (items of clothing). Participants were either informed of the task rule (covers top or bottom half of body) or not – in the latter case they had to remember all eight S-R mappings. The S-R group group's performance was much more susceptible than that of the task-rule group to interference from semantically unrelated distractors (each word was superimposed on a line drawing), consistent with their theory task-sets shield performance from irrelevant information.

Dreisbach and Wenke (2011) have since shown that task-set shielding is less effective on a task switch trial (this must necessarily be so to allow for a change of task): participants were presented with a letter (requiring a vowel/consonant classification) or a digit (requiring an odd/even classification). Each stimulus also had an irrelevant feature (a colour in Experiment 1, a font in Experiment 2) which could change or repeat from trial to trial. Consistent with a shielding function of task-sets, RTs on switch (but not repeat) trials were slower when the irrelevant feature changed as well. The notion that task-sets are more vulnerable to interference on a task switch trial is also supported by larger response congruency effects on those trials (Rogers & Monsell, 1995, also see this thesis).

Why would task-set shielding be less effective in the more complex condition? Dreisbach and Haider (2009, Experiment 3) have shown that the effectiveness of task-set shielding is dependent on how well the task-set has been established. Presumably, the task-set is less well-established in the S-R condition. But when participants were given the opportunity to practice the S-R mappings in the absence of the distractors (1 practice block of 128 trials), performance in the S-R condition was no more susceptible to interference from the distractors than performance in the task condition. Decreased task-set shielding for the less well-established or less stable task-set can explain the results of Experiments 4 and 5, in which the effect of complexity was more pronounced on incongruent trials, though it remains unclear why task-set preparation was not at all affected by complexity on congruent trials in Experiment 6.

In any case, the findings appear to suggest that the effectiveness of task-set preparation is subject to the limited capacity of the bridge. However, one question about the effect of complexity on the switch cost was not conclusively answered in Chapter 4: is the effect of complexity on the switch cost caused entirely by the capacity limit on the bridge, or does it (also) take longer to upload a more complex task-set into the bridge?

Based on analogous findings from experiments on declarative working memory, Oberauer (2009) predicts that the latter should not be the case, and that task-sets (PWM) or lists (DWM) of any complexity are retrieved as a package or chunk. The results of Experiments 5 and 6, which showed that switch costs remained larger for the more complex condition even with plenty of opportunity for preparation, are consistent with this. If the effect of complexity on the switch cost were due solely to increased retrieval time then this effect should be eliminated with a long preparation interval.

There is another and perhaps more obvious reason as to why set-size effects are unlikely to be restricted to task-set retrieval time alone. If it were true that complexity only affects the time it takes to retrieve a task-set, and that the representation of this task-set is not restricted further by a capacity limit imposed on this representation, then complexity should have no effect on repeat trials, or indeed on performance in single-task blocks. But these predictions were disproved by the experiments of Chapters 4 and 3, respectively, demonstrating that (at least part of) the effect of complexity on the switch cost is due to a capacity limited bridge.

### The effect of the phonological properties of the stimulus terms on task-cueing performance (Chapter 5).

*Performance in a task-cueing paradigm is not influenced by the phonological properties of the stimulus terms, arguing against a direct role for phonological representations in preparation and maintenance of S-R rules.*

The final part of this thesis was concerned with how task-sets are represented, focusing on the representation of S-R rules. The results of previous studies have suggested that linguistic representations may contribute to task-set control. Some of these studies found that articulatory suppression (AS) affects the switch cost, suggesting that task-sets may be represented phonologically. The experiments reported in Chapter 5 further investigated this possibility by varying the spoken word duration (Experiment 7) or phonological similarity (Experiments 8 and 9) of the stimulus terms in the S-R rules. These experiments found no effect of name length or phonological similarity on overall RT, the switch cost or its reduction with preparation in the task-cueing blocks, suggesting that phonological representations of the S-R rules do not contribute to task-set control. They are consistent with the idea that task-sets are maintained in a procedural working memory buffer (Rubinstein et al., 2001; Oberauer, 2009), which is separate from (phonological representations in) declarative memory.

However, despite very small trial numbers (16 per condition) we did find a reliable effect of phonological similarity in the single task practice blocks. This finding suggests phonological representations are used when the task-sets are not well-established yet, but only very briefly, as this effect was no longer found in the first experimental switching block. Though it has previously been suggested phonological representations of task rules may aid performance particularly early on in practice, Experiments 8 and 9 provided the first unambiguous support for this idea. Kray, Eber and Karbach (2008) previously showed that the effect of AS decreased with practice, but their experiment did not include an appropriate dual task control condition.

The ability to transform a set of (declarative) instructions into a set of procedures is one of the fundamental cornerstones of human cognition: it allows us to rapidly acquire skills that take other animals months or years, if they are able to learn them at all (Stoet & Snyder's, 2003, "task switching monkeys" received 100.000 training trials before data collection even started). Not many studies have directly investigated the contribution of phonological representations to the proceduralisation of skill – it may be difficult to capture this process *because* it proceeds so rapidly (at least in the case of relatively simple task-sets). In this thesis I was only able to distinguish between the absolute presence (single task practice blocks) and absence (task-cueing blocks) of phonological mediation (Experiments 6 and 7). In future, it would be interesting to plot the demise of phonological representations in more detail: with a greater number of trials one would expect to see the effect of phonological similarity disappear gradually, as predicted by theories of skill acquisition. In addition, one might also expect the duration of phonological mediation to depend on the complexity of the task-set – presumably, a less complex task-set is more rapidly "proceduralised" than a more complex task-set.

That leaves open the question of why previous experiments that have used AS to investigate the role of linguistic representations in task-set control have obtained such mixed results. Some studies found larger switch costs with AS, others found smaller switch costs and yet others found that AS affected switch and repeat trials equally. The results reported in this thesis (no effect of the phonological properties of the stimulus terms on task-cueing performance) and those of Miyake et al. (2004, larger switch costs with AS, particularly when the cues were not transparent) are not necessarily incompatible: it is possible that linguistic representations facilitate task-set control at one level (selecting a task-set from among alternatives) but not another (selecting an S-R rule within a task-set).

However, the idea that linguistic representations contribute to task-set control at one level but not another cannot explain why some experiments using very similar or almost identical paradigms have obtained such different results. For example, Miyake et al. (2004) found the switch cost to increase with AS, whereas Liefooghe et al. (2005) found the reverse (both used a task-cueing paradigm with bivalent stimuli). Moreover, an attempt by Bryck and Mayr (2005) to replicate Miyake et al.'s (2004) result using an almost identical procedure was not successful. It is possible that the use of linguistic representations in task-set control is strategic and differs widely between participants and tasks.

Previous studies provide some support for the idea that the use of phonological representations and rehearsal in declarative WM tasks might be subject to the task demands, and strategic modulation: Salamé and Baddeley (1982, also see Camos, Mora & Oberauer, 2010) have previously suggested that when lists contain many phonological similar items, participants might use a strategy other than phonological rehearsal because phonological encoding of the items is likely to be difficult. Hanley and Bakopoulou (2003) have provided evidence consistent with this proposal. In their Experiment 2, 7 similar (B, C, D, G, P, T, V) or dissimilar (H, J, R S, L, Y, Z) consonants were presented sequentially on a computer screen for 1 s (ISI 0.5 s). One group of participants were instructed to use a semantic strategy (turn each consonant into a word starting with that letter and create a sentence from those words). Participants in another group were instructed to use a phonological strategy (subvocal rehearsal). Participants in a third group were not given any instructions. Effects of phonological similarity were found in the control and phonological strategy group, but not in the semantic strategy group.

Further evidence that the use of a rehearsal strategy is subject to strategic modulation can be inferred from Baddeley, Thomson and Buchanan's (1975) observation that a word length effect (for words matched for the number of syllables but differing in terms of articulatory duration) was found only for participants who reported using a rehearsal strategy, and not for participants who used visual imagery. Hence it is possible that participants in Experiment 8 and 9 abandoned a phonological strategy because the stimulus names were similar. Moreover, it then remains possible that phonological representations may be used for task-set preparation when participants are explicitly instructed to use such representations (as was indeed the case in Camos et al., 2010, Experiment 2). A repeat of Experiment 8 or 9 with an additional between subjects manipulation (instructed to rehearse the S-R mappings, or not) could address this possibility.

Finally, if S-R rules are not represented phonologically (possibly out of strategic considerations), as the results of Experiments 7-9 suggest, then how are they represented? The results of Experiments 7-9 are consistent with the notion that S-R rules are represented in a PWM buffer which is separate from declarative WM (e.g. Oberauer, 2009). In any case, the alternative (a phonological representation somehow immune to the effect of articulatory duration or similarity) appears unlikely.

One other mechanism for maintaining representations in a heightened state of accessibility is attentional refreshing (Camos, Lagner & Barrouillet, 2009, Camos, Mora & Oberauer, 2011, Johnson, 1992): thinking briefly of a representation serves to reactivate it and hence protect it from decay or interference. In contrast to rehearsal, which can only be used for the maintenance of phonological representations, attentional refreshing does not make any explicit assumptions about the nature of the representation. It is possible to distinguish behaviourally between refreshing and rehearsal: whereas refreshing is thought to be attentionally demanding (e.g. Camos et al., 2011) and hence its effectiveness depends on the amount of attention available, the effectiveness of rehearsal depends on the phonological properties of the representation. For example, Camos et al. (2011, Experiment 1) used a complex span task which required participants to recall a list of six words. The words were phonologically similar or dissimilar. In between the presentation of each word a series of black squares was presented, to which participants had either to respond with a single key press (SRT condition), or by judging its location (CRT condition). Camos et al. (2011) reasoned that participants would prefer to use refreshing because the material was phonologically similar. But their ability to refresh would be impaired in the (more attentionally demanding) CRT condition, forcing participants to rely on rehearsal. Hence the authors expected a phonological similarity effect in the CRT but not the SRT condition, which is precisely what they found. They furthermore found that the use of refreshing or rehearsal could be modulated through explicit instruction: the phonological similarity of list items only affected performance when participants were instructed to use rehearsal (Experiment 2). Hence attentional refreshing is an alternative to rehearsal, and it may be the case participants used this strategy in our experiments. However, following the logic of Camos et al. (2011), who reasoned that participants are not likely to use attentional refreshing when the task is attentionally demanding, this seems unlikely considering that cued-switching is attentionally very demanding.

## Conclusions

In a nutshell, the experiments reported in this thesis have shown that we must distinguish between two levels of action control and representation in active procedural working memory: the level of selection among active task-sets, whose availability is not subject to capacity limitation (Chapter 2); and the level of selection of S-R mappings within a task-set –represented in a procedural working memory buffer of very limited capacity (Chapters 3 and 4). Furthermore, Chapter 5 showed the S-R rules are not represented phonologically (or at least that such representations do not contribute to task-set control) – a finding that is furthermore consistent with the notion that task-sets are represented in a procedural working memory.

Although previous accounts of task-set control have emphasised working memory involvement, they have rarely specified in any detail the properties of this working memory buffer. This thesis made some progress in further defining the properties of PWM, and its contribution to task-set control. In doing so Oberauer's theory of WM proved a useful framework: some of the properties of PWM proposed by this theory received support in this thesis – others await future research.

# References

Allport, D. A., & Styles, E. A. (unpublished manuscript). Multiple executive functions, multiple resources? Experiments in shifting attentional control of tasks.

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press.

Altmann, E. M. (2002). Functional decay of memory for tasks. *Psychological Research, 66 (4),* 287–297.

Altmann, E. M. (2004). The preparation effect in task switching: Carryover of SOA. *Memory & Cognition, 32,* 153-163.

Altmann, E. M. (2007). Comparing switch costs: Alternating runs and explicit cuing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 475-483.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13,* 27-33.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review, 115,* 602-639.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451-474.

Anderson, J. (1976). *Language, Memory and Thought.* Hillsdale, NJ: Erlbaum Associates.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220,* 25-30.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Anderson, J. R. & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General, 128,* 186-197.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111 (4),* 1036-1060.

Arbuthnott, K., & Frank, J. (2000). Executive control in set switching: Residual switch costs and task-set inhibition. *Canadian Journal of Experimental Psychology, 54,* 33–41.

Aron, A. R., Monsell, S., Sahakian, B. J., & Robbins, T. W. (2004). A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain, 127,* 1561-1573.

Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science, 15,* 610-615.

Arrington, C. M., & Logan, G. D. (2005). Voluntary task switching: Chasing the elusive homunculus. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 683-702.

Avons, S. E., Wright, K. L., & Palmer, K. (1994). The word-length effect in probed and serial recall. The Quarterly Journal of Experimental Psychology, 47A, 207-231.

Baddeley, A.D. (1966). The influence of acoustic and semantic similarities on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18,* 302-309.

Baddeley, A. D., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47(A),* 1047-1054.

Baddeley, A.D., Chincotta, D. M. & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130,* 641-657.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G.A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975).Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Bates, E., D’Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10 (2),* 344-380.

Bertelson, P. (1965). Serial choice reaction-time as a function of response versus signal- and response repetition. *Nature, 206,* 217-218.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. (1962). Some variables influencing the rate of gain of information. *Journal of Experimental Psychology, 63,* 105-110.

Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology, 55,* 37-85.

Bryck, R. L., & Mayr, U. (2005). On the role of verbalization during task-set selection: Switching or serial order control? *Memory & Cognition, 33,* 611-623.

Bryck, R. L., & Mayr, U. (2008). Task selection cost asymmetry without task switching. *Psychonomic Bulletin & Review, 15 (1),* 128-134.

Buchler, N. G., Hoyer, W. J., & Cerella, J. (2008). Rules and more rules: The effect of multiple tasks, extensive training, and aging on task-switching performance. *Memory & Cognition, 36 (4),* 735-748.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61,* 457-469.

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39,* 231-244.

Caplan, D., Rochon, E., Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology, 45 (2),* 177-92.

Caplan, D., & Waters, G. S (1994). Articulatory length and phonological similarity in span tasks: a reply to Baddeley and Andrade. *Quarterly Journal of Experimental psychology A, 47 (4),* 1055-1062.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75-84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55,* 429-432.

Conway, A. R. A., & Engle, R. W. (1995). Working memory and retrieval: a resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123 (4),* 354-373.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science, 15,* 634-640.

Da Silva Souza, A., Oberauer, K., Gade, M., & Druey, M. D. (in press). Processing of representations in declarative and procedural working memory. *Quarterly Journal of Experimental Psychology.*

De Jong, R. (2000). An intention-activation account of residual switch costs. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention and performance XVIII* (pp. 357-376). Cambridge, MA: MIT Press.

Dreisbach, G., & Haider, H. (2008). That's what task-sets are for: shielding against irrelevant information. *Psychological Research, 72,* 355-361.

Dreisbach, G., & Haider, H. (2009). How task representations guide attention: further evidence for the shielding function of task-sets. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35 (2),* 477-486.

Dreisbach, G., & Wenke, D. (2011). The shielding function of task-sets and its relaxation during tasks switching. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37 (6),* 1540-1546.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, J., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General, 137 (1),* 131-148.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257-303.

Elchlepp, H. (2011). *The temporal dynamics of switching tasks.* Unpublished doctoral thesis, University of Exeter, Exeter, UK.

Elchlepp, H., Lavric, A., Mizon, G. A., & Monsell, S. (2012). A brain-potential study of preparation for and execution of a task-switch with stimuli that afford only the relevant task. *Human Brain Mapping, 33 (5)* 1137-1154

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71,* 43-51.

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48,* 148-168

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55,* 587-600.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47,* 381-391.

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

Gopher, D. (1996). Attention control: Explorations of the work of an executive controller. *Cognitive Brain Research, 5,* 23-38.

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* (pp. 331-355). Cambridge, MA: MIT Press.

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review, 10,* 435-444.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11-26.

Hübner, M., Kluwe, R. H., Luna-Rodriguez, A., & Peters, A. (2004). Response selection difficulty and asymmetrical costs of switching between tasks and stimuli: No evidence for an exogenous component of task-set reconfiguration. Journal of Experimental Psychology: *Human Perception and Performance, 30 (6),* 1043-1063

Hale, D. (1968). The relation of correct and error responses in a serial choice reaction task. *Psychonomic Science, 13,* 299-300.

Hommel, B. (1998). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of Cognitive Processes.* Cambridge, MA: MIT Press.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 53,* 188-196.

Hyman, R., & Umiltà, C. (1969). The information hypothesis and non-repetitions. *Acta Psychologica, 30,* 37-53.

James, W. (1890). The principles of psychology. New York: Henry Holt.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, whole nr. 89.*

Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience, 4,* 268-280.

Karayanidis, F., Jamadar, S., Ruge, H., Phillips, N., Heathcote, A., & Forstmann, B. U. (2010). Advance preparation in task-switching: converging evidence from behavioral, brain activation, and model-based approaches. *Frontiers in Psychology, 25,* 1-13.

Keele, S. W., & Rafal, R. (2000). Deficits of task-set in patients with left prefrontal cortex lesions. In S. Monsell & J. S. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 627-651). Cambridge, MA: MIT Press.

Kessler, Y. & Meiran, N. (2010). The reaction-time task-rule congruency effect is not affected by working memory load: Further support for the activated long-term memory hypothesis. *Psychological Research, 74,* 388-399.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin, 136,* 5, 849-874.

Kleinsorge, T., & Apitzsch, N. (2012). Task preparation based on precues versus memory: Precues lead to superior performance with two tasks but not with four tasks. *Journal of Cognitive Psychology, 24 (2),* 140-156.

Kornblum, S. (1968). Serial-choice reaction time: Inadequacies of the information hypothesis. *Science, 159,* 432-434.

Kray, J., Li, K. Z., H., & Lindenberger, U. (2002). Age-related changes in task switching components: The role of task uncertainty. *Brain and Cognition, 49,* 363-381.

Kray, J. & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15 (1),* 126-147.

Kray, J., Eber, J., & Karbach, J. (2008) Verbal self-instructions in task switching: a compensatory tool for action-control deficits in childhood and old age? *Developmental Science, 11 (2),* 2236-236.

Kray, J., Kipp, K. H., & Karbach, J. (2009). The selective development of inhibitory control: The influence of verbal labeling. *Acta Psychologica, 130,* 48-57.

Kveraga, K., Berryhill, M., & Hughes, H. C. (2006). Directional uncertainty in visually guided pointing. *Perceptual and Motor Skills, 102,* 125-132.

Lavric, A., Mizon, G. A., & Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: a task-switching investigation. *European Journal of Neuroscience, 28 (5),* 1016-1029.

Leonard, J. A. (1959). Tactual choice reactions: I. *Quarterly Journal of Experimental Psychology, 11 (2),* 76-83.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). Response to Altmann: Adaptive forgetting by decay or removal of STM contents? *Trends in Cognitive Sciences, 13,* 280-281.

Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology, 8,* 311–335.

Liefooghe, B., Vandierendonck, A., Muyllaert, I., Verbruggen, F., & Vanneste, W. (2005). The phonological loop in task alternation and task repetition. *Memory, 13 (5),* 550-560.

Lien, M. C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task switch: Do people prepare all the task some of the time or some of the task all the time? *Journal of Experimental Psychology: Human Perception and Performance, 31 (2),* 299-315.

Lindsen, J. P., & de Jong, R. (2010). Distinguishing between de partial-mapping preparation hypothesis and the failure to engage hypothesis of residual switch costs. *Journal of Experimental Psychology: Human Perception and Performance, 36 (5),* 1207-1226.

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., & Schneider, D. W. (2006). Interpreting instructional cues in task switching procedures: The role of mediator retrieval. *Journal of Experimental psychology-learning memory and cognition, 32 (3),* 347-363.

Longman C., Lavric A., & Monsell, S. (under revision). More attention to attention? An eye tracking investigation of selection of perceptual attributes during a task switch. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Longoni, A. M., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition, 21,* 11-22.

Longstreth, L. E., El-Zahhar, N., & Alcorn, M. B. (1985). Exceptions to Hick’s law: Explorations with a response duration measure. *Journal of Experimental Psychology, General, 114,* 417-434.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology & Aging, 16,* 96-109.

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1124-1140.

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,* 362–372.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General,* 129, 4-26.

Mecklinger, A., von Cramon, D. Y., Springer, A., & Matthes-von Cramon, G. (1999). Executive control functions in task switching: Evidence from brain injured patients. *Journal of Clinical and Experimental Neuropsychology, 21 (5),* 606-619.

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. (2000). Reconfiguration of stimulus task sets and response task sets during task switching. In J. Driver & S. Monsell (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII* (pp. 377-399). Cambridge, MA: MIT Press.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41,* 211-253.

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.

Meyer, D. E., & Kieras, D. E. (1997). EPIC - A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review,* 104, 749-791.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review, 63,* 81-97.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. C. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica, 115 (2-3),* 123-142.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7 (3),* 134-140.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind* (pp. 93-148). Hove, E. Sussex: Erlbaum (UK) Taylor and Francis.

Monsell, S. (1984). Components of working memory underlying verbal skills: A 'distributed capacities' view – A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 327-350). Hillsdale, NJ: Erlbaum.

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., & Mizon G.A (under revision). Stimulus-task associations and the task-switch cost. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31*, 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*, 250-264.

Morin, R. E., & Forrin, B. (1965). Information-processing: Choice reaction times of first- and third-grade students for two types of associations. *Child Development, 36,* 713-720.

Morin, R. E., Konick, A., Troxell, N., & McPherson, S. (1965). Information and reaction time for "naming" responses. *Journal of Experimental Psychology, 70,* 309-314.

Mowbray, G. H. (1960). Choice reaction times for skilled responses. *Quarterly Journal of Experimental Psychology, 12,* 193-202.

Murdock, B. B. Jr. (1962). The serial position effect in free recall. *Journal of Experimental Psychology, 64,* 482-488.

Nieuwenhuis, S., & Monsell, S. (2002). Residual costs in task switching: Testing the failure to-engage hypothesis. *Psychonomic Bulletin & Review, 9,* 86-92.

Norman, D. A. and Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In Davidson, R. J., Schwartz, G. E., and Shapiro, D., (Eds), *Consciousness and Self-Regulation: Advances in Research and Theory*. Plenum Press.

Oberauer, K. (2005). Control of the contents of working memory: a comparison of two paradigms and two age groups. Journal of Experimental Psychology: Learning, *Memory & Cognition, 31 (4),* 714-728.

Oberauer, K. (2009). Design for a working memory. *Journal of learning and motivation, 51,* 45-100.

Oberauer, K. (2010). Declarative and procedural working memory: common principles, common capacities? *Psychologica Belgica, 50* (3-4), 277-308.

Oberauer, K., & Kliegl, R. (2004). Simultaneous execution of two cognitive operations: Evidence from a continuous updating paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 689-707.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601-626.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 4,* 272-281.

Philipp, A. M., & Koch, I. (2006). Task inhibition and task repetition in task switching. *European Journal of Cognitive Psychology, 18,* 624-639.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology-General, 124 (2),* 207-231.

Rubin, O. & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task switching paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1477-1491.

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.

Ruthruff, E., Remington, R. W., & Johnston, J. C. (2001). Switching between simple cognitive tasks: The interaction of top-down and bottom-up factors. *Journal of Experimental Psychology: Human Perception & Performance, 27,* 1404-1419.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150-164.

Saeki, E., & Saito, S. (2004). Effect of articulatory suppression on task-switching performance: Implications for models of working memory. *Memory, 12,* 257-271.

Saeki, E., Saito, S., & Kawaguchi, J. (2006). Effect of response-stimulus interval manipulation and articulatory suppression on task switching. *Memory, 14,* 965-976.

Saeki, E., & Saito, S. (2009). Verbal representation in task order control: An examination with transition and task cues in random task switching. *Memory & Cognition, 37 (2),* 1040-1050.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology & Aging, 13,* 445-461.

Schneider, D. W., & Anderson, J. R. (2011). A memory-based model of Hick's law. *Cognitive Psychology, 62,* 193-222.

Schneider, D. W., & Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognitive Psychology, 64,* 127-160.

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,* 343-367.

Schneider, D. W., & Logan, G. D. (2007). Defining task-set reconfiguration: The case of reference point switching. *Psychonomic Bulletin & Review, 14,* 118-125.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51 (2),* 283-304.

Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology, 70,* 284-288.

Steinhauser, M., & Hübner, R. (2007). Automatic activation of task-related representations in task shifting. *Memory & Cognition, 35 (1),* 138-155.

Stoet, G., & Snyder, L. H. (2003). Executive control and task switching in monkeys. *Neuropsychologia, 41 (10),* 1357-1364.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 89-103.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review, 81,* 75-98.

Theios, J. (1973). Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory.* New York: Academic Press.

Van ‘t Wout, F., Monsell, S., & Lavric, A. (in preparation a). Is it harder to switch among a larger set of tasks?

Van ‘t Wout, F., Monsell, S., & Lavric, A. (in preparation b). The effect of task-set complexity on preparation for a task switch: Evidence for a limited capacity component of procedural working memory.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychological Bulletin, 136 (4),* 601-626.

Verbruggen, F., Liefooghe, B., Vandierendonck, A., & Demanet, J. (2007). Short cue presentations encourage advance task preparation: A recipe to diminish the residual switch costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 342-356.

Vygotsky, L., & Luria, A. R. (1994). Tool and symbol in child development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 99-174). Oxford, England: Blackwell.

Wheeldon, L. R., & Monsell, M. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language, 33 (3),* 332-356.

Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General, 10 (1),* 1-20.

Wright, C. E., Marino, V. F., Belovsky, S. A., & Chubb, C. (2007). Visually guided, aimed movements can be unaffected by stimulus-response uncertainty. *Experimental Brain Research, 179,* 475-496.

Yeung, N., & Monsell, S. (2003a). The effects of recent practice on task switching. *Journal of Experimental Psychology: Human Perception and Performance, 29 (5),* 919-936.

Yeung, N., & Monsell, S. (2003b). 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.