

ORIGINAL ARTICLE

Glenn Wylie · Alan Allport

Task switching and the measurement of “switch costs”

Received: 31 March 1999 / Accepted: 23 July 1999

Abstract The measurement of “switch costs” is held to be of interest because, as is widely believed, they may reflect the control processes that are engaged when subjects switch between two (or more) competing tasks. [In task-switching experiments, the reaction time (RT) switch cost is typically measured as the difference in RT between switch and non-switch (repeat) trials.] In this report we focus on the RT switch costs that remain even after the subject has had some time to prepare for the shift of task, when the switch cost may be approximately asymptotic (so-called residual switch costs). Three experiments are presented. All three experiments used Stroop colour/word, and neutral stimuli. Participants performed the two tasks of word-reading and colour-naming in a regular, double alternation, using the “alternating runs” paradigm (R. D. Rogers & S. Monsell, 1995). The experiments were designed to test the hypothesis that RT switch costs depend on a form of proactive interference (PI) arising from the performance of a prior, competing task. A. Allport, E. A. Styles and S. Hsieh (1994) suggested that these PI effects resulted from “task-set inertia”, that is, the persisting activation-suppression of competing task-sets, or competing task-processing pathways. The results confirmed the existence of long-lasting PI from the competing task as a major contributor to switch costs. Non-switch trials, used as the baseline in the measurement of switch costs, were also shown to be strongly affected by similar PI effects. However, task-set inertia was not sufficient to account for these results. The results appeared inconsistent also with all other previous models of task switching. A new hypothesis to explain these between-task interference effects was developed, based on the stimulus-triggered retrieval of competing stimulus-response (S-R) associations, acquired (or strengthened) in earlier trials.

Consistent with this retrieval hypothesis, switch costs were shown to depend primarily on the S-R characteristics of the preceding task (the task that was switched from) rather than the upcoming task. Further, the effects of the other, competing task were found to persist over many successive switching trials, affecting switch costs long after the stimulus overlap (and hence the principal S-R competition) between the current tasks had been removed. Switch costs were also found to be affected by recent, item-specific experience with a given stimulus, in either the same or the competing task. Finally, the results showed that switch costs were massively affected by the ratio of the number of prior trials, in response to the same stimuli, that had implemented either the currently intended or the competing S-R mappings. None of these effects are predicted by current models of residual switch costs, which appeal to the differences in control processes assumed to be engaged in switch versus non-switch trials.

Introduction

During the past 70 years, researchers have attempted to study “executive” control processes through the behavioural effects of task switching (e.g., Allport, Styles & Hsieh, 1994; De Jong, in press; Goshke, in press; Jersild, 1927; Mayr & Keele, in press; Meiran, 1996; Meiran, Chorev, & Sapir, in press; Rogers & Monsell, 1995; Spector & Biederman, 1976; and R. De Jong, B. Emans, R. Eenshuistra, & E. Wagenmakers, *Strategies and intrinsic limitations in intentional task control. Experimental and Occupational Psychology*, submitted for publication). In such studies, participants are asked to switch between two (or more) tasks, and performance in two conditions is compared: (1) a condition in which participants switch between the tasks on successive trials, and (2) a condition in which participants perform the same task on successive trials. A common assumption is that participants must use top-down, voluntary,

control processes when they have to switch between competing tasks, and that the same switching control processes are not needed on non-switch (repeat) trials. Therefore, it is widely assumed, if performance on Switch trials is compared to performance on the same task on Repeat trials, slower performance on the Switch trials can be attributed to the additional cost of the control processes involved in the switch.

This is a highly intuitive supposition. Furthermore, the work done by Jersild (1927) and by Spector and Biederman (1976) provided some boundary conditions to this supposition, which appear readily interpretable within the framework of popular models of executive control. These researchers found a performance difference between switch and repeat conditions only when the same stimulus-set was used for both tasks. When one task (Task A) was to add three to a given number, and the other task (Task B) was to provide the antonym for a given word – and thus the stimuli used for Task A and B were different (numbers for Task A and words for Task B) – performance in the switch condition was no slower or less accurate than performance in the repeat condition. However, when the two tasks used the same stimuli, e.g., if Task B was to subtract two from a given number, and Task A was to add three as before – thus requiring numbers to be used as stimuli for both tasks, participants' response times (RTs) were substantially longer in the switch condition than in the repeat condition.

This finding fits well into Baddeley's model of working memory (Baddeley, 1986) (c.f., Norman & Shallice, 1986). In these models, top-down, control mechanisms – the "central executive" or "supervisory attentional system" (SAS) – are called into action only in certain circumstances, among which is the circumstance in which a stimulus is compatible with more than one currently relevant task. According to these models, stimuli (along with their associated context) become associated with the action sequence normally performed on or with them. When these stimuli are encountered, the system performs the action sequence most strongly associated with them, and this is done without the intervention of any control processes. This corresponds to an experiment in which the stimuli used for Task A and Task B are of different types. However, if the system is presented with a stimulus that is strongly associated with two (or more) incompatible actions, then the intervention of some other mechanism – the central executive or SAS – is required, to ensure that the system performs the action sequence that is currently appropriate. This corresponds to an experiment in which the same stimuli are used both for Task A and Task B.

Perhaps because these early experiments fit so snugly into currently popular models of executive control, it is tempting to suppose that the RT difference between switch and repeat conditions (the so called "switch cost") reflects, more or less directly, the time that control processes need to switch the system from a readiness to perform Task A to a readiness to perform Task B. If

viable, this simple measurement would be an invaluable tool in the study of control processes.

However, there are several grounds for doubting this simple view. (For a recent review of some of the arguments see Allport and Wylie, in press-a.) Three lines of argument will be developed in this present publication. The first concerns the states of the baseline (the repeat condition) against which the performance cost on Switch trials (the switch cost) is conventionally measured. The second (and related) line of argument concerns the occurrence of between-task interference effects (in our experiments, "reverse Stroop" interference) affecting Repeat trials, as well as Switch trials. The third line of argument, linking both of the above, concerns the long-lasting nature of these between-task interference effects, such that switch costs can reflect the relative proportion of competing Task A-Task B trials performed in the course of an experiment.

Our discussion (and the experiments to be described here) refer specifically to the so-called residual switch costs, observed even after a relatively long (e.g., 1 s) preparation interval (c.f., R. De Jong, B. Emans, R. Eenshuistra, & E. Wagenmakers, submitted for publication; and Meiran, 1996; Meiran et al., in press; Rogers and Monsell, 1995). The residual switch cost in these conditions (i.e., after a long preparation interval) is the difference between RT (or error rate or both) on Switch trials and Repeat trials. An implicit (sometimes explicit) assumption underlying this subtractive methodology is that the repeat condition represents a state in which "task-set reconfiguration" is already fully completed (during and/or after the Switch trial), *before* the task stimulus for the Repeat trial is presented; and hence that, on the Repeat trial, the system is "fully prepared" for the relevant task. This is, of course, a fundamental assumption, if RT switch costs are to be taken as a direct measure of the executive control processes involved in the switching or reconfiguration of task-set. If it can be shown that performance on Repeat trials also depends on the recent execution of a competing task, from which the system was supposed to have switched on a prior Switch trial, then the rationale for this subtractive methodology, and indeed the whole enterprise of inferring executive function (reconfiguration of task-set) from switch costs will be called into question.

In most of the experiments published to date, however, making the comparison between Repeat trial RTs and RTs in a baseline condition is not possible. This is because most studies have used the Repeat condition as the baseline condition. [A major exception is in the work of De Jong and colleagues (e.g., R. De Jong et al., submitted for publication).] In the experiments by Jersild (1927), Spector and Biederman (1976), and Allport et al. (1994), participants worked through lists of stimuli. In one condition, they were instructed to regularly alternate between the two tasks as they worked through the list (the switch condition); in another they were asked to do one or other of the two tasks throughout the

list (the non-switch or repeat condition). The dependent measure was list completion time, and switch costs were calculated by subtracting the mean time taken to complete the two pure task (repeat) lists from the time taken to complete the switch list. Thus, in this paradigm, the pure task (repeat) list completion time was taken as the baseline.

Rogers and Monsell (1995) criticised this methodology. They pointed out that this method potentially confounded factors such as working memory load, and arousal, in the critical comparison. They argued that, in the switch condition, participants had to "hold in mind" two tasks while, in the repeat condition, they had to hold in mind only one task. This, they felt, might have led to the use of different strategies or different levels of arousal, etc., in the switch and repeat conditions. They, therefore, introduced their "alternating runs" paradigm, in which Switch and Repeat trials both occurred within a single block. In the alternating runs paradigm, participants switch from one task to the other on every second trial (i.e., AABBAABBA...), or on every fourth trial, etc. The first trial of a given task is, therefore, the Switch trial and the second (and later) trials are the Repeat trials. This, they hoped, equates the working memory demands on the Switch and Repeat trials, as well as the level of arousal, etc., because both conditions are presented within the same block. Others (e.g., R. De Jong et al., submitted; and Meiran, 1996; Meiran et al., in press) have controlled for these variables, instead, by using a random cueing paradigm. In such a paradigm, a cue is presented prior to every trial, to inform the participant of the task that is relevant on that trial. Sometimes the cues on two successive trials are the same (making the second trial a Repeat trial), and sometimes they differ (making the second trial a Switch trial). In both the alternating runs paradigm, and the random cueing paradigm, the Repeat trials are used as the baseline, and are assumed to represent a condition in which participants are "fully prepared". However, there is some evidence to suggest that participants are not fully prepared on these Repeat trials.

If participants were fully prepared on Repeat trials, we would expect their performance on a given task to be at least as fast as (or faster than) in any other, less well prepared, condition. Thus, if it is demonstrated that RTs in another condition are faster than RTs on the Repeat trials, we would have some grounds for doubting that participants were fully prepared for a given task on Repeat trials. Recently, this has been demonstrated: RTs in "fixed-task" conditions – in which participants expected to have to perform only one task – were significantly shorter than RTs on Repeat trials – when participants expected to have to switch to another task on forthcoming trials (Allport & Wylie, in press-b; and R. De Jong et al., submitted; see also Los, 1996, for a review). This, it might be argued, only demonstrates that holding two tasks in mind impairs participants' performance. However, Allport and Wylie (in press-b, Experiment 3) found that participants' performance on a given

very familiar task (word reading) continued to be substantially slower, even over a large number of Repeat trials, in pure task conditions, after having performed a competing task. This was found despite the fact that the participants had no expectation of ever having to perform the competing task again. This was relative to a baseline condition in which they had not yet been told about the competing task. Thus, in both conditions (before and after the competing task), participants had to hold in mind only one task, yet for more than 30 trials after performing a competing task, their performance was slowed. This suggests that trials immediately following Switch trials do not necessarily represent a fully prepared state.

A second line of evidence that suggests that participants may not be fully prepared on Repeat trials, following a switch of tasks, derives from task congruity (or task-interference) effects. These effects represent one of the most stable, and, perhaps, under-appreciated, effects in the task-switching literature. Nearly all task-switching studies include Congruity as a dependent variable because, as Jersild (1927) and Spector and Biederman (1976) showed, participants do not appear to exhibit switch costs when responding to univalent (i.e., non-overlapping) stimulus-sets (with non-overlapping response sets). As a result, most studies have used divergent (bivalent) stimuli, and divergent stimuli (especially with forced-choice key-press tasks) can be either congruent or incongruent in their response mappings, in the two tasks. If participants had attained a fully prepared state during the course of the Switch trial, we would not expect to find incongruity effects on the Repeat trial that were different from any incongruity effects found in pure task conditions. If such effects were found – that is, if stimulus congruity affected RTs on Repeat trials more than in pure tasks – we would have to conclude that participants were less than fully prepared for the task at hand on Repeat trials. In fact, R. De Jong et al. (submitted; Experiment 3) found incongruity effects were smaller in pure tasks than in Repeat trials in switching blocks.

One of the experiments reported by Allport et al. (1994) provides another, even more striking, example. In Experiment 5, Allport et al. used Stroop colour/word stimuli (e.g., the word "red" printed in *blue* ink). These stimuli afford the two tasks of word reading and colour naming. While participants typically read such Stroop stimuli with little difficulty, they find colour-naming an especially challenging task (Stroop, 1935; see MacLeod, 1991, for a review). Thus, by asking participants to switch between colour-naming and word-reading, Allport et al. required them to switch between an asymmetrically "difficult" pair of tasks. Neutral stimuli were included as well – to provide a baseline against which to measure RTs to the Stroop stimuli. The five colour words printed in black served as the neutral word-reading stimuli, and a string of xs printed in the five colours served as the neutral colour-naming stimuli. Thus, Allport et al. were able to compare the effect of

performing both tasks in response to *incongruent* Stroop stimuli to the effect of performing them in response to *neutral* stimuli, and hence to estimate the incongruity effect.

Allport et al. (1994) found there to be a large incongruity effect on both switch and repeat colour-naming RTs. It is not surprising to find incongruity effects on Stroop colour-naming, as a large body of research on the "Stroop effect" has shown (MacLeod, 1991). More surprisingly, however, incongruity effects were found on word-reading RTs as well. This is surprising because it is very unusual to find incongruity effects on word-reading using Stroop stimuli (so called "reverse Stroop" interference effects; MacLeod, 1991). Furthermore, incongruity effects were found on both the Switch and the Repeat trials of word-reading, suggesting that the participants were not fully prepared for the dominant word-reading task, even on Repeat trials.

This experiment also yielded an unexpected pattern of RT results. The switch costs associated with word-reading were far larger than those associated with colour-naming (in fact, there was no reliable switch cost associated with colour-naming at all). If it is presumed that switch costs reflect a control operation that is needed to switch to a new task, this result is very difficult to understand. The Stroop colour-naming task is generally taken as the very paradigm of a "controlled" task, while word-reading is generally considered to be "automatic". On these grounds, one might suppose that it would take more executive control to switch to Stroop colour-naming than to word-reading, and that the switch costs associated with Stroop colour-naming would be larger than those associated with word-reading. The results of Experiment 5 of Allport et al. (1994) were inconsistent with this interpretation of switch costs. Their results led them to propose the Task-set Inertia (TSI) hypothesis. According to this hypothesis, switch costs derive from the persisting activation of task processes established during the performance of one or two trials of the previous task. Thus, Allport et al. proposed that, when participants are performing Stroop colour-naming, they have to strongly inhibit the word-reading process, and this inhibition then persists inertially. When they subsequently have to perform the word-reading task, this persisting inhibition delays the generation of the word-reading response. Furthermore, because word-reading is initially inhibited (and colour-naming still activated), information relevant to the colour-naming task interferes with word-reading, resulting in "reverse Stroop" interference effects. The TSI hypothesis does not predict similarly large switch costs when participants switch back to colour-naming, because participants need not normally inhibit colour-naming in order to read the word. Therefore, there is less (or even no) persisting inhibition of colour-naming that is "left over" from the word-reading trials.

The TSI hypothesis, therefore, strongly predicts that switch costs should change as a function of the task that participants are switching from, not as a function of the

task that they are switching to. Experiment 1 was designed to test this prediction. To anticipate, the results of Experiment 1 turn out to be generally consistent with the TSI hypothesis, indicating large proactive interference (PI) effects from a previous, competing task, affecting both Switch and Repeat trials following a switch of tasks. However, Experiments 2 and 3 provide further results which cannot be accounted for by the TSI hypothesis, as formulated by Allport et al. (1994). Thus, a new hypothesis is proposed based on stimulus-cued retrieval (c.f., Allport & Wylie, in press-b). According to this hypothesis, performance costs following a switch from a competing task reflect proactive task interference (as before); however, the PI arises as a form of long-term negative priming, resulting from prior, learned, associative S-R bindings between particular stimulus attributes and responses. We shall refer to this as the "retrieval hypothesis".

Experiment 1

Rogers and Monsell (1995) argued in favour of a stage-like control process reflected in RT switch costs, on the basis of their Experiment 6. In this experiment, they extended the alternating runs paradigm from four quadrants to eight cells. They presented a circle on the computer screen, and divided it into eight equal sections – like the slices of a pizza. Successive stimuli were presented in adjacent pizza-slices, in clockwise order. They instructed participants to do one task when the stimuli appeared, say, in the upper half of the screen, and the other task when stimuli appeared in the lower half. Thus, participants did four trials of one task before switching to four trials of the other task. Rogers and Monsell presented this experiment as a test of the TSI hypothesis. If RTs decreased gradually over the three successive Repeat trials following a switch, they argued, the TSI hypothesis would be supported; if RTs decreased abruptly after the first stimulus, a stage-like process would be upheld, and the TSI hypothesis would need revision. They found the latter, and concluded that switch costs were not due to TSI, but to the conjunction of endogenous and exogenous control operations, resulting in task-set reconfiguration.

While Rogers and Monsell (1995) demonstrated that participants exhibit longer RTs on the first trial of a run of a given task, and considerably shorter RTs thereafter, none of their experiments unambiguously addressed whether switch costs result from participants switching to a new task or from the current one. Because they required participants to switch between the same two, mutually exclusive tasks, with the same bivalent stimuli, they confounded switching from one competing task with switching to another competing task. Thus, their results appear compatible with a stage-like hypothesis, in which participants must complete some switching process before they can switch to a new task; but they are also compatible with a hypothesis like TSI which

proposes that residual switch costs merely represent proactive interference from task-sets that participants have recently switched from. Experiment 1 was therefore designed to investigate the extent to which residual switch costs reflect the demands of switching *to* the forthcoming task, or of switching *from* the previous task, using Rogers and Monsell's alternating runs paradigm.

This experiment used Stroop stimuli, like Experiment 5 of Allport et al. (1994). Furthermore, we used the same tasks, word-reading (*Word*) and colour-naming (*Colour*), as were used in that experiment. However, to be able to compare Switch and Repeat trial RTs from within a single block, we used the alternating runs paradigm of Rogers and Monsell (1995).

The experiment proceeded in three stages, and all participants completed the three stages in the same order. In Stage 1, participants switched between reading neutral colour words and naming the colour of rows of xs. Because all the stimuli were "neutral" in this stage, we shall refer to it as the *All-Neutral* condition. In the second stage (*Colour-Neutral*), *Word* involved participants reading incongruent Stroop stimuli while *Colour* remained as before: participants named the colour of neutral strings of xs. In the last stage (*All-Stroop*), *Word* involved reading incongruent Stroop stimuli, and *Colour* involved naming the colour of incongruent Stroop stimuli.

The experiment was principally designed to investigate *Word*. It will be remembered that, in their Experiment 5, Allport et al. (1994) found large switch costs associated with *Word* as well as large reverse Stroop interference effects. However, it was not clear from their study whether these effects resulted from a difference in the *Word* stimuli participants were switching to (which were neutral in one case and Stroop in the other) or from a difference in the *Colour* stimuli participants were switching from (which were likewise, neutral in one case and Stroop in the other). The design of Experiment 1 allows us to investigate this directly. In this experiment, the All-Neutral and All-Stroop conditions can be considered to be baseline conditions – the critical manipulation occurs in the Colour-Neutral condition. If participants incur large switch costs, and interference effects, when the *Word* stimuli they are switching to are Stroop stimuli, we would expect to find a large RT difference between RTs to *Word* in Colour-Neutral relative to All-Neutral. This is simply because the *Word* stimuli are neutral in the All-Neutral block and Stroop in the Colour-Neutral block. However, if the large switch costs, and interference effects result from persisting processes that were established during the performance of the previous task (the task participants are switching from), we would expect *Word* latencies in Colour-Neutral to be no different from those exhibited in All-Neutral. This is simply because the stimuli used for the *Colour* task are the same in these two conditions (neutral xs). Rather, we would expect to find a large RT difference between RTs to *Word* in All-Stroop relative to

Colour-Neutral because the stimuli used for *Colour* are neutral in the Colour-Neutral block and Stroop in the All-Stroop block.

Method

Participants. This study included 10 participants. All were recruited from the Oxford University Subject Panel, and were paid for their participation. Their ages ranged from 29 to 51 years with a mean age of 41.1 years. Of these 10 participants, 7 were female. All had normal, or corrected to normal acuity, and normal colour vision.

Materials and displays

Stroop stimuli. The Stroop stimuli consisted of six colour words: "red", "green", "blue", "yellow", "pink", and "brown". All stimuli were presented in 30 point, Times font, lower-case bold type face. These words were presented in one of six colours: *red*, *green*, *blue*, *yellow*, *pink*, and *brown*. In all cases, the word and the colour (in which the word was presented) were incongruent – for instance, the word "red" could be presented in any of the six colours except *red*. The colours' RGB (red, green, blue) breakdown were as follows: *red* = {65535, 0, 0}; *green* = {0, 65535, 0}; *blue* = {0, 0, 65535}; *yellow* = {65535, 65535, 0}; *pink* = {65535, 36980, 41383}; *brown* = {38928, 18252, 6757}.

Neutral stimuli. The neutral stimuli were intended to allow the execution of only one of the two tasks (*Colour* or *Word*). The neutral *Colour* stimuli consisted of a string of the character "x" that could be three to six characters long – to control for the length of the word strings. Thus, "xxx" was presented a number of times equal to the presentation of the word "red"; "xxxx" was presented a number of times equal to the sum of the presentation of the words "blue" and "pink"; and so on. These xs were presented in the same six colours as the words. The neutral stimuli for the *Word* task consisted of the six words presented in *black*.

All stimuli were presented using a Power Macintosh 7200/60 computer equipped with a custom-made input/output board (I/O board). This I/O board was fitted with a microsecond clock that was started when a stimulus was presented, and was stopped by the participant's vocal response. The participants responded into a microphone that was wired to a custom-made voice-key. When the signal from the microphone rose above a threshold level, the voice-key produced a signal that stopped the microsecond clock in the I/O board. The clock held the amount of time it had been running in memory until the experiment's software retrieved it, and this time represented the RT. The computer was equipped with a 15-in. multi-scan, colour monitor set at 640×480 resolution.

The display consisted of a large, black cross (16 cm \times 16 cm) that was presented against a grey background. This cross divided the screen into four quadrants. The horizontal bar of the cross was thicker than the vertical bar (horizontal bar = 3 mm; vertical bar = 1 mm). For half of the participants, all blocks started with the presentation of a stimulus in the upper, left quadrant. This was counter-balanced so that, for the other half of participants, the first stimulus was presented in the lower, right quadrant. Stimuli then proceeded, from quadrant to quadrant on successive trials, in a clockwise direction. On any trial, the stimulus was positioned in the quadrant such that the centre of the stimulus was 3.4 cm from both the vertical and horizontal bars of the cross.

Subjects were presented with different combinations of stimuli depending on the block of the experiment. In the first experimental block (All-Neutral), participants were presented with neutral *Colour* stimuli in, for instance, the upper two quadrants, and with neutral *Word* stimuli in the lower two quadrants. In the second block (Colour-Neutral), participants were again presented with neutral *Colour* stimuli in the upper two quadrants, and with

incongruent Stroop stimuli in the lower two quadrants. Finally, in the third block (All-Stroop), participants were presented with incongruent Stroop stimuli in all quadrants. In all three blocks, the task was to name the colour of the stimuli in the top two quadrants, and to read the word in the bottom two quadrants. This was counterbalanced across participants such that half were presented with *Colour* stimuli in the lower two quadrants and *Word* stimuli in the upper quadrants.

In all three blocks, we ensured that the response on any given trial could not be primed by the preceding stimulus or response. In All-Neutral, this simply meant ensuring that no stimulus or response was repeated on successive trials (e.g., the word "red" was not presented for reading on two successive reading trials, nor was the word "red" presented for reading followed by a string of xs that was coloured *red* or vice versa). In the other two blocks, however, we also ensured that in no case did either the *Word* concept or the *Colour* concept recur on successive trials. That is, if the word "red" – presented in *green* – was presented on trial n , we ensured that the stimulus presented on trial $n+1$ was neither red (in its word form or colour) nor green (in its word form or colour). This regimen precluded all item-priming (positive and negative – as well as repetition of distractors) between immediately successive stimuli.

Responses were recorded as correct or incorrect by the experimenter on-line, and a trial did not commence until the previous response was coded. The resulting inter-trial interval (ITI) was variable, and approximately 1000 ms, or longer, in duration. Subjects were encouraged to prepare themselves as far as possible for the task on the next trial during this interval. This ITI (1000 ms) represents a preparation interval at which the benefits of anticipatory task preparation have been found to be approximately asymptotic in a number of studies (e.g., De Jong et al., submitted; and Meiran et al., in press; Rogers & Monsell, 1995). At this interval, therefore, observable switch costs should be "residual" switch costs only (Meiran, 1996). Stimuli were presented for 3000 ms or until participants responded and responses of less than 150 ms were treated as errors.

Design and procedure

The experiment proceeded in three, discrete stages.

Stage 1 – All-Neutral. Subjects were seated before the computer monitor at a comfortable viewing distance (approximately 60 cm). They were told that this was a reaction time experiment, and that they should respond vocally (speaking into the microphone) as quickly as possible without sacrificing accuracy. The stimuli and tasks were then explained (coloured xs for *Colour* and words for *Word*). Subjects then completed 30 practice trials of All-Neutral. Following this, they proceeded to the experimental All-Neutral block (120 trials). Only the data from the experimental trials were included in the analyses.

Stage 2 – Colour-Neutral. Subjects were apprised of the nature of the stimuli in the Colour-Neutral block (i.e., coloured xs and incongruent Stroop words), and were told that the tasks remained the same. They were reminded of the need to respond quickly without sacrificing accuracy, and then completed 30 practice trials of Colour-Neutral followed by 120 trials of the experimental block. Only the data from the experimental trials were included in the analyses.

Stage 3 – All-Stroop. Subjects were told that, in this block, all stimuli would be coloured colour-words (i.e., Stroop stimuli), but that the tasks remained identical to those of the previous blocks. Once again, the need to respond quickly without sacrificing accuracy was impressed upon them. They then completed 30 practice trials of All-Stroop before proceeding on to complete 120 experimental trials. Only the data from the 120 experimental trials were included in the analyses.

This experiment had a 2 (*Colour* vs. *Word*) \times 2 (Switch vs. Repeat) \times 3 (All-Neutral, Colour-Neutral, All-Stroop) design. All participants completed the three blocks in the order outlined above: Stage 1, Stage 2, then Stage 3. We designed this order confound into the experiment because we suspected that participants' latencies to neutral stimuli would be different after they had experienced responding to incongruent Stroop stimuli. We therefore measured

RTs to All-Neutral stimuli first and gradually introduced incongruent Stroop stimuli in the successive stages of the experiment.

Results

In all three of the experiments presented here, the median RT of each condition was calculated for each participant. All of the RT analyses are based on these median scores.

All-Neutral

A repeated-measures ANOVA was used to analyse the data from the All-Neutral block. It had two factors: Task (*Colour* vs. *Word*), and Switch (Switch vs. Repeat). There was a significant main effect of Task, and there was a significant main effect of Switch, but there was no interaction between the two (Fig. 1). The main effect of Task [$F(1, 10) = 158.57, p < 0.0001$] resulted from participants reading words faster than they named colours. The main effect of Switch [$F(1, 10) = 16.26, p < 0.01$] resulted from participants responding more quickly on Repeat trials than they did on Switch trials. Post-hoc comparisons revealed that participants exhibited significantly longer latencies on Switch trials than on Repeat trials (switch costs) in both tasks: *Colour* ($p < 0.045$), and *Word* ($p < 0.035$). The mean percentage of errors participants made in All-Neutral was 4.4%. An analysis of these errors revealed no reliable differences between the four conditions ($p > 0.05$ in all cases).

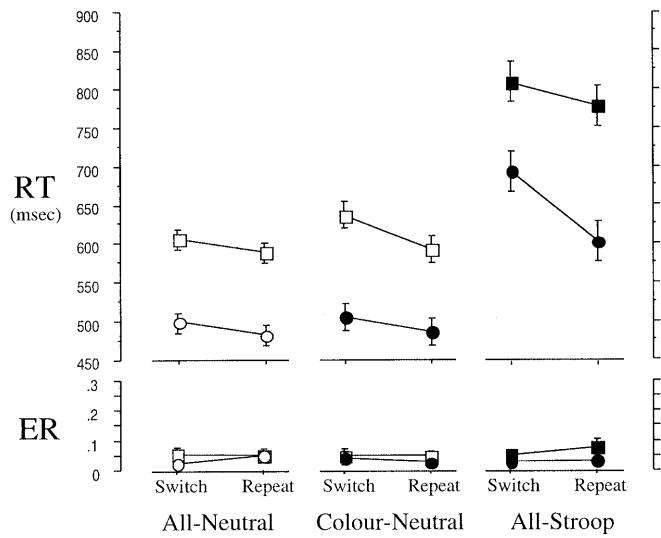


Fig. 1 RT and ER for Experiment 1. Error bars show 95% within-subject confidence intervals (Loftus & Masson, 1994). Open symbols denote neutral stimuli; filled symbols denote incongruent Stroop stimuli. Square symbols denote data from colour-naming, round symbols denote data from word-reading (RT Reaction times, ER error rates)

Colour-Neutral

The same analysis was carried out on the data from the Colour-Neutral block. The results are very similar to All-Neutral (Fig. 1). Again, there was a significant main effect of Task [$F(1, 10) = 200.72, p < 0.0001$] such that *Word* RTs were faster than *Colour* RTs. Furthermore, there was a significant main effect of Switch [$F(1, 10) = 51.85, p < 0.0001$] such that participants responded faster on Repeat trials than they did on Switch trials. Again, there was no interaction between these two factors; however, in this block post-hoc comparisons revealed that participants exhibited reliable switch costs only when performing the *Colour* task (*Colour*: $p < 0.01$; *Word*: $p < 0.2$). In this block, the mean percentage of errors was 4.2%. An analysis of these errors again revealed no reliable differences between the four conditions ($p > 0.05$ in all cases).

All-Stroop

The same analysis was carried out on the data from the All-Stroop block. We found a significant main effect of both factors – Task [$F(1, 9) = 69.83, p < 0.0001$] and Switch [$F(1, 9) = 65.13, p < 0.0001$] – and a significant interaction between the two [$F(1, 9) = 9.17, p < 0.015$]. As can be seen in Fig. 1, participants read the words more quickly than they named the colours; participants responded faster on Repeat trials than they did on Switch trials; but having to switch from *Colour* to *Word* resulted in especially slow RTs. Post-hoc comparisons revealed that the switch costs on *Word* were highly reliable ($p < 0.0001$), and the switch costs on *Colour* were marginally reliable ($p < 0.065$). The mean percentage of errors in this block was 4.8%. An analysis of errors revealed no significant differences between the four conditions ($p > 0.05$ in all cases).

It was possible that the large difference between Switch and Repeat trials in the *Word* task was modulated by practice. For instance, it could be that this effect was large early in the All-Stroop block, and became smaller as the block progressed. To test this, we split the data into three epochs, corresponding to the trials from the first third of the block, from the second third, and from the final third of the block. (This resulted in ten observations per participant per epoch.) We computed the median of the Switch and Repeat trials from each epoch, and analysed these data with a two-factor, repeated-measures ANOVA. The two factors were Switch (Switch vs. Repeat) and Epoch (1, 2, or 3). The only reliable effect was that of Switch [$F(1, 9) = 28.33, p < 0.001$]. Epoch was not reliable, nor did it interact with Switch. Thus, the effects on *Word* RTs appeared to be stable across the All-Stroop block.

Between block comparisons

To address the main purpose of the experiment, we made four comparisons: (1) RTs to the neutral *Word*

stimuli in the All-Neutral block were compared to RTs to the Stroop *Word* stimuli in the Colour-Neutral block; (2) RTs to the Stroop *Word* stimuli in the Colour-Neutral and All-Stroop blocks were compared with one another; (3) RTs to the neutral *Colour* stimuli (the xs) in the All-Neutral and Colour-Neutral blocks were compared; and (4) RTs to the neutral *Colour* stimuli in Colour-Neutral were compared with RTs to Stroop *Colour* stimuli in All-Stroop.

1. Word-reading: All-Neutral vs. Colour-Neutral. In this analysis, we compared participants' reading RTs across All-Neutral and Colour-Neutral with a repeated-measures ANOVA; the analysis had two factors: Block (All-Neutral vs. Colour-Neutral) and Switch (Switch vs. Repeat). In the All-Neutral block, participants were reading neutral words, and in the Colour-Neutral block, they were reading incongruent Stroop words. If the large switch costs and reverse Stroop effects associated with *Word* in the experiments reported by Allport et al. (1994) are simply due to the nature of the *Word* stimuli, we would expect to find a reliable interaction between Block and Switch.

The only reliable effect was the main effect of Switch [$F(1, 10) = 19.53, p < 0.001$]. This resulted from participants responding with longer latencies on Switch trials than on Repeat trials. Importantly, there was no interaction between Block and Switch ($F < 1$).

The same analysis was carried out on the errors made in these two blocks. No reliable differences were found.

2. Word-reading: Colour-Neutral vs. All-Stroop. In this analysis, we compared word-reading RTs in Colour-Neutral and All-Stroop with a repeated-measures ANOVA. This ANOVA had two factors: Block (Colour-Neutral vs. All-Stroop) and Switch (Switch vs. Repeat). In both blocks, participants were responding to identical stimuli in an identical way (i.e., they were always reading incongruent Stroop stimuli). The only difference between these two blocks was that, in Colour-Neutral, the *Colour* task involved participants naming the colour of neutral *Colour* stimuli, while in the All-Stroop block, the *Colour* task involved participants naming the colour of incongruent Stroop stimuli.

There was a main effect of Block [$F(1, 9) = 37.73, p < 0.001$], and of Switch [$F(1, 9) = 26.05, p < 0.001$], and there was a reliable interaction between the two [$F(1, 9) = 30.36, p < 0.001$]. As Fig. 1 shows, in Colour-Neutral, participants read the Stroop words with approximately the same latencies as they had read the neutral *Word* stimuli in All-Neutral. However, in All-Stroop (when the other task involved naming the colour of Stroop stimuli), there was a clear difference between participants' latencies on Switch and Repeat trials (post-hoc comparisons revealed this difference to be significant at $p < 0.0001$). Furthermore, Fig. 1 shows that participants' reading latencies were longer (overall) when the other task involved Stroop stimuli than when it involved neutral stimuli (reflected in the main effect of Block).

An analysis of the errors made in these two blocks found no reliable difference between conditions.

3. Colour-naming: All-Neutral vs. Colour-Neutral. Although this experiment was not designed to investigate *Colour* with the same rigour as *Word*, it was nevertheless of interest to see if a change in the *Word* stimuli would result in a change in the latencies associated with *Colour*. We therefore analysed participants' responses to the neutral *Colour* stimuli (xs) in All-Neutral and Colour-Neutral with a repeated-measures ANOVA. This ANOVA had two factors: Block (All-Neutral vs. Colour-Neutral) and Switch (Switch vs. Repeat). In both blocks, participants were responding to identical stimuli in an identical way (i.e., they were always naming the colour of strings of coloured xs). The only difference between these two blocks was that, in All-Neutral, the *Word* task involved participants reading neutral *Word* stimuli, while in the Colour-Neutral block, the *Word* task involved participants reading incongruent Stroop stimuli.

There was no main effect of Block ($p > 0.05$), but there was a main effect of Switch [$F(1, 10) = 14.23, p < 0.004$], and the interaction between Block and Switch was reliable [$F(1, 10) = 5.97, p < 0.04$]. As can be seen in Fig. 1, participants exhibited significantly longer latencies on *Colour* Switch trials in Colour-Neutral than they did in All-Neutral (post-hoc comparisons revealed this difference to be significant at $p < 0.003$). Subjects' RTs to Repeat trials were not reliably different between the two blocks.

4. Colour-naming: Colour-Neutral vs. All-Stroop. This analysis was carried out as another complement to the analyses of the *Word* data. The stimuli for *Colour* were different across these two blocks (they were neutral in Colour-Neutral, and incongruent Stroop in All-Stroop), and the *Colour* task was much more "difficult" in the All-Stroop block. If switch costs represent an index of the extent to which control processes are active in setting up the forthcoming task, we should find that switching interacts with Block in this comparison. However, the stimuli used for *Word* were the same in both conditions. Therefore, if switch costs simply represent a difference in PI on Switch and Repeat trials, we would expect no interaction between switching and Block.

The data from *Colour* in Colour-Neutral and All-Stroop were, therefore, analysed with a two-factor, repeated-measures ANOVA. The factors were Block (Colour-Neutral vs. All-Stroop) and Switch (Switch vs. Repeat). The main effect of Block [$F(1, 9) = 170.7, p < 0.0001$] and the main effect of Switch [$F(1, 9) = 28.08, p < 0.0005$] were both highly reliable. However, these two factors did not interact (Fig. 1).

Discussion

The primary aim of this experiment was to investigate the effects of these manipulations upon the performance of the *Word* task. Focusing on this task allows us to relate our findings to the already very large literature on Stroop tasks (MacLeod, 1991), and also to the similarly extensive literature on visual word recognition and oral

naming (Coltheart, 1987). We shall, therefore, begin by discussing the results from the *Word* task.

Word-reading

In the All-Neutral block, there was a small, but reliable switch cost associated with *Word*. This was found despite the fact that the stimuli used for *Word* and *Colour* unambiguously cued the relevant task in this block. On the strength of the evidence supplied by Jersild (1927) and Spector and Biederman (1976), we might have expected to find no switch cost; however, there are several possible reasons for this discrepancy. The stimuli used for both *Colour* and *Word* consisted of alphabetical characters that were presented in some colour. (The "neutral" *Word* stimuli, though presented in black – which was not a member of the response set for either colour-naming or word-reading – were nevertheless presented in a nameable colour.) This small amount of overlap between the stimuli for the two tasks might have been enough to engender some interference between the two tasks. Alternatively, the switch cost observed in All-Neutral may result from the overlap in the responses to the two tasks (c.f., Meiran, in press).

That a reliable difference was found between the Switch and Repeat trials of *Word* in the All-Neutral block is less important than the finding that this difference did not become larger when participants were asked to read incongruent Stroop stimuli in the Colour-Neutral block. This suggests that the control processes involved in a switch to reading were not affected by a change in the *Word* stimuli themselves (provided the *Colour* stimuli remain unchanged).

This lack of a difference between RTs when participants were asked to read neutral and Stroop stimuli is consistent with a large body of literature on the Stroop effect (MacLeod, 1991), but its importance in this paradigm should not be underestimated. The change in stimuli between the All-Neutral and Colour-Neutral conditions represents a change in the overlap between the stimuli used for the two tasks. In the All-Neutral condition, there was minimal overlap between the stimuli used for the two tasks; in the Colour-Neutral condition, the stimuli used for the *Word* task clearly also afforded the *Colour* task, and might therefore have been expected to "cue" the *Colour* task as a competing response. However, if any competitor-cueing took place, it was not evident in the RTs associated with the *Word* task.

This stable pattern of RTs in the *Word* task was radically changed when participants proceeded to the All-Stroop block. In this block, participants responded with much longer latencies on both Switch and Repeat trials than they had in the Colour-Neutral condition – despite the fact that the *Word* stimuli they were responding to (incongruent Stroop stimuli) were the same in the All-Stroop and Colour-Neutral blocks. This dramatic change in *Word* RT was brought about by a

change in the stimuli used for – and, therefore, the task requirements of – the *Colour* task. The fact that performance changes as a function of the task-demands of the *other* task strongly suggests that some form of PI is affecting performance on *Word* in the All-Stroop block. These results are, therefore, entirely consistent with the TSI hypothesis. Indeed, the pattern of results in the All-Neutral and All-Stroop blocks replicates Experiment 5 of Allport et al. (1994). According to Allport et al.'s interpretation, when participants were performing the *Colour* task in the All-Stroop block, they had to strongly suppress the *Word* task-set. This suppression, according to Allport et al. then persisted: the *Word* task was “negatively primed”. When the participants then had to perform the *Word* task on subsequent trials, this suppression evidenced itself as longer latencies on both the Switch and Repeat trials of the *Word* task. An alternative explanation, which Experiments 2 and 3 will support, was referred to as the retrieval hypothesis. According to the retrieval hypothesis, participants establish (learn) associative S-R links, or “bindings”, during performance of *Colour*, that are re-evoked, or retrieved, when they return to *Word* and encounter the same stimuli. Specifically, during the *Colour* task, participants form a binding between the colour of the stimulus and the relevant naming response, and between the written word and a null response (that is, a “don't respond” response). When they are subsequently asked to read the same stimuli (Stroop words) aloud, these latter bindings are re-evoked and must be overcome before a reading response can be made.

These results suggest that switch costs change largely as a function of the requirements of the other task. That is, switch costs change as a function of the task participants are switching from, not the task they are switching to. This, in turn, has profound implications for the study of task switching.

Colour-naming

Although this experiment was not designed to investigate *Colour* as formally as it was designed to investigate *Word*, the data from the *Colour* trials corroborates the explanation given above. In the All-Neutral block, a small but reliable difference was found between the Switch and the Repeat trials. This is similar to the difference found on *Word*, and is probably subject to a similar explanation. However, as with *Word*, the differences in RT within a block are less important than the differences between blocks.

In both the All-Neutral and Colour-Neutral blocks, the stimuli for the *Colour* task were coloured xs; the only difference between these conditions was the stimuli used for the *Word* task. Despite the fact that the *Colour* stimuli were the same in both blocks, the RTs associated with *Colour* differed in the All-Neutral and Colour-Neutral blocks: participants responded with longer latencies on the Switch trials of the Colour-Neutral

block than in the analogous condition of the All-Neutral block. This suggests that the processes used to perform the *Word* task, when the stimuli for *Word* are coloured, resulted in PI on *Colour*. It could be that participants partially inhibited the processes associated with colour-naming while responding to the Stroop *Word* stimuli in the Colour-Neutral block. Equally, it could be that they formed associative S-R bindings between the colours of the words and a “null” response. Regardless of which interpretation is correct, it is clear that the increased latencies associated with the Switch trials of the Colour-Neutral block cannot be due to a difference in the *Colour* stimuli themselves, because these stimuli are identical in both blocks, and therefore must be due to the change in the stimuli used for the *Word* task.

Furthermore, when participants proceeded from the Colour-Neutral block to the All-Stroop block, and performed the *Colour* task with incongruent Stroop stimuli, the difference between Switch and Repeat trials remained the same (i.e., there was no interaction between Block and Switch in the analysis of Colour-Neutral and All-Stroop *Colour* RTs). This too is consistent with the assertion that switch costs change as a function of the requirements of the “other”, competing task. In this case, the stimuli used for the *Colour* task changed from neutral to Stroop, but the stimuli used for the *Word* task remained the same, and so did the switch costs. Of course, the latencies associated with *Colour* in the All-Stroop block were far longer than in the Colour-Neutral block – reflecting the well-documented finding that participants have great difficulty performing *Colour* with Stroop stimuli – but the difference between RTs on Switch and Repeat trials was the same in the Colour-Neutral and All-Stroop blocks.

Taken together, these results are difficult to reconcile with any hypothesis that attributes switch costs to the control demands of the task to which the switch is made, without taking account of the preceding, competing task. This conclusion is reinforced by the results of both *Word* and *Colour* in this experiment.

Before proceeding, however, we should consider an alternative possible account of the data pattern (in particular, the switch costs) in Experiment 1, as follows. This account rests on the observation that in the All-Neutral and Colour-Neutral conditions, the task-stimuli distinctively cued the relevant tasks, but they did not do so in All-Stroop¹. Thus, it could be that during the first two blocks, participants learned to ignore the location cue, and then had to learn (or re-learn) to use this cue in the final (All-Stroop) block. (Recall that all participants proceeded through the three blocks in the same fixed order.) Thus, this hypothesis proposes that the effects we found in the All-Stroop block are due to a difference in the number of cues participants had available: In the All-Neutral and Colour-Neutral blocks, the relevant task was cued by both the task-stimuli and the location

¹We would like to thank D. Gopher for pointing out this possibility to us.

in which the stimuli were presented; in the All-Stroop block, the relevant task was cued *only* by the location in which the stimuli were presented.

There are several reasons to doubt this hypothesis. If the observed effects were simply the result of differences in task cueing, then we would expect there to be no difference in RTs between identical tasks in All-Neutral and Colour-Neutral. Contrary to this prediction, however, we found a reliable difference between the RTs on *Colour* Switch trials in the All-Neutral and Colour-Neutral blocks, despite the fact that participants had two cues to the relevant (identical) task in both blocks. Another reason to doubt this hypothesis is supplied by comparisons between the Colour-Neutral and All-Stroop blocks. If the observed differences in the switch costs for *Word* were due simply to a difference in the number of available task-cues, we should surely expect to find a similar effect on switch costs for *Colour*. While (unsurprisingly) participants' *Colour* latencies were longer in response to Stroop stimuli (in All-Stroop) than to Neutral stimuli (in Colour-Neutral), there was no difference in the switch cost between these two conditions. Accordingly, the difference in the *Word* switch cost between the Colour-Neutral and All-Stroop blocks is unlikely to be due to this cause. Finally, if the switch cost associated with *Word* in the All-Stroop block depended on the need to learn to use the location cue, we might expect that the switch cost (and *Word* RTs generally) would decrease as the All-Stroop block proceeded. When the *Word* data from this block were divided into three epochs and reanalysed with Epoch as a factor, there was no main effect of Epoch, nor did it enter into any interaction. In fact, the difference between Switch and Repeat trials became marginally larger as the All-Stroop block proceeded. (The switch costs of the three epochs were as follows: 100 ms, 96.6 ms, 128.9 ms.) Therefore, this hypothesis does not appear to be consistent with these data.

Rather, the results of Experiment 1 appear entirely consistent with the TSI hypothesis. When the stimuli used for one task afford two currently relevant tasks, the other, competing task must be suppressed – thereby allowing the relevant task to attain temporary dominance and to be performed. Clearly, performance of the *Colour* task, in response to Stroop stimuli, requires much greater suppression of the word-reading process than the complementary case (word-reading, and suppression of colour-naming). According to the TSI hypothesis, such active suppression is liable to persist, involuntarily, even after a switch of tasks, affecting both Switch and Repeat trials. The pattern of results in Experiment 1 conforms to these predicted effects.

However, as noted in the Introduction, the TSI account, in terms of persisting activation-suppression of competing processing pathways, encounters severe difficulties in explaining the results of Experiments 2 and 3. Moreover, it is not the only possible account of the results just described. The pattern of results in Experiment 1 is equally consistent with a revised hypothesis, in terms

of stimulus-cued retrieval of prior associative learning (Allport & Wylie, in press-b). According to this revised hypothesis, participants form, or learn, associative S-R bindings between stimulus attributes and particular responses. Thus, on the *Colour* trials of the All-Stroop block, participants learned to associate the colours of the stimuli with a colour-naming response, and to associate the orthography of the stimuli to a null response. When they were later required to return to *Word*, in the All-Stroop condition, the Stroop stimuli cued the associative bindings formed during the earlier *Colour* trials because the stimuli used for *Word* were now of the same type as those previously used for *Colour*. The proactive task interference that we have observed, according to this revised hypothesis, is thus a form of retrieval-based long-term negative priming (c.f., Lowe, 1998; Neill, 1997). We shall return to this hypothesis later, in relation to Experiments 2 and 3.

The interpretation of switch costs

It would be convenient if the difference between RTs on Switch and Repeat trials could be taken to represent the time needed to “reconfigure” the system to perform the currently relevant task. In the studies of Jersild (1927), and those carried out by Spector and Biederman (1976), switch costs (or “shift costs”) were the difference between “pure” and alternating, or mixed, lists (in their Experiment II, Spector and Biederman recorded discrete RTs rather than recording list completion times; however, they still blocked pure and mixed conditions). Following Rogers and Monsell (1995), the term switch cost has come to refer to the difference between a Switch trial and the immediately subsequent trial (the Repeat trial). This shift away from comparing pure and mixed lists, and towards comparing RTs on trials within a single block is based on Rogers and Monsell's observation that the former comparison confounds several factors. While this observation is important, the results from this experiment show that the comparison between Switch and Repeat trials in the alternating runs paradigm introduces confounds of its own.

When one compares Switch and Repeat trials in the alternating runs paradigm, one is assuming that the Repeat trials offer an appropriately stable baseline. This assumption, in turn, is based upon the assumption that participants effect complete task-set reconfiguration during the Switch trial, and that the Repeat trial is, therefore, unaffected by the switch of task. Rogers and Monsell (1995) made an effort to test this assertion in their Experiment 6, and found that RTs did not diminish over the course of three successive Repeat trials. This result appears to support the idea that Repeat trials offer a stable, fully reconfigured baseline against which to measure Switch trial RTs. Indeed, Rogers and Monsell asserted that this provides strong evidence that participants have achieved complete task-set reconfiguration by the end of the Switch trial.

The results from the present experiment contradict this assertion. The RTs associated with *Word* Repeat trials increased dramatically when the stimuli associated with the *Colour* task were altered to afford both tasks. Thus, when the two tasks were divergent, and both tasks were performed on overlapping stimulus sets, Repeat trials suffered considerable interference from the other task. This indicates that complete reconfiguration is not attained during Switch trials, and that Repeat trials can not be assumed to represent a stable baseline that can be used in any simple way to gauge the cost of switching task-set.

In fact, this experiment calls into question the validity of the notion of switch costs as a measure of the time taken by participants to execute a shift of task-set. Rather, there appear to be several factors influencing RT on Switch trials – among which is an interference effect that is “left over” from the previous task, and that continues to be evident on the Repeat trials.

Experiment 2

Perhaps the most striking feature of Experiment 1 was what could be considered a “reverse Stroop” effect seen on *Word* in the All-Stroop block, by comparison with *Word* responses to neutral stimuli, in the All-Neutral block. This effect was largest on Switch trials (approximately 200 ms), but was approximately 120 ms even on Repeat trials. Experiment 2 explores these PI effects on the *Word* task further.

Suppose that, after having performed the All-Stroop condition for some time, participants were returned (without any pause) to the Colour-Neutral condition, continuing to perform alternating runs of *Colour* and *Word*. What effects should we expect to see on the performance of *Word* in the new Colour-Neutral condition? If participants are able to achieve complete task-set re-configuration after one trial of a task (i.e., during the course of a Switch trial), we should expect *Word* RTs to return abruptly, after one trial, to the levels found in Colour-Neutral in Experiment 1. In contrast, if PI from a previous task-condition continues to affect performance over many trials, as predicted by the TSI hypothesis of Allport et al. (1994) [and as predicted also by the retrieval hypothesis (c.f., Allport & Wylie, in press-b)], RTs in the *Word* task should gradually diminish, over a number of trials – eventually returning to levels comparable to those seen in the Colour-Neutral condition of Experiment 1.

Suppose, further, that after some number of alternating runs in the Colour-Neutral condition, participants were *again* returned to the All-Stroop condition (again without any pause) and once more continued with alternating runs of *Colour* and *Word*. Should we expect to see *Word* RTs increase abruptly to the levels found in Experiment 1, in the All-Stroop condition? Or might these performance costs build up gradually, as a cumulative effect of PI from the Stroop *Colour* task, over a number of cycles in the All-Stroop condition?

To examine these questions, we designed an experiment in which participants successively performed the All-Stroop and Colour-Neutral conditions of Experiment 1. They performed each condition for six cycles² (where a “cycle” is four successive trials, two trials of *Colour* and two of *Word*) before being shifted to the other stimulus condition, without interruption.

Method

Participants. Four men and five women participated in this experiment. Participants ranged in age from 23 to 52 with a mean age of 37.89 years. All were recruited from the Oxford University Subject Panel, and were paid for their participation. Participants all had normal or corrected to normal acuity, and normal colour vision. None of these participants had taken part in Experiment 1.

Stimuli and apparatus. The apparatus and the general display was the same as that used for Experiment 1. The task stimuli, however, were somewhat different. The neutral *Colour* stimuli used for this experiment were filled rectangles of the relevant colour. The rectangles were the same height as the word stimuli, and were equal in width to a five-character word (the approximate mean word length of the words used). The stimulus set was also slightly altered. The words (as well as the colours) used were: RED, GREEN, BLUE, PURPLE, PINK, and ORANGE. All *Word* stimuli were presented in bold, upper case to decrease the difference in the amount of colour present in the Stroop stimuli and the neutral, *Colour* stimuli. Furthermore, a black outline was added around the letters of the words to better define their boundaries. In all other respects, the word stimuli were similar to those in Experiment 1 (i.e., they were presented in bold, 30 point, Times font). The RGB breakdown of the colours used was as follows: red = {65535, 0, 0}; green = {0, 65535, 0}; blue = {0, 0, 65535}; purple = {54020, 1700, 65535}; pink = {65535, 36980, 41383}; orange = {65535, 25738, 652}.

Procedure. All participants were seated at a comfortable viewing distance from the computer screen (approximately 60 cm), and were instructed about the nature of Stroop stimuli. The alternating runs paradigm was then explained to them, and they were told that when the stimuli appeared in one half of the grid, they were to do one task (i.e., name the colour), and that when it appeared in the other, they were to do the other task (i.e., read the word). To control for eye movements (c.f., Rogers & Monsell, 1995), we counterbalanced participants across four conditions: two participants were asked to name the colour (the *Colour* task) when the stimuli appeared in the upper half of the grid, and to read the word (the *Word* task) when the stimuli appeared in the lower half of the grid; three participants were asked to do the reverse; two participants were asked to do *Colour* when the stimuli appeared in the left half of the grid, and to do *Word* when the stimuli appeared in the right half of the grid; and the remaining two participants were asked to do the reverse. All participants completed three practice blocks of 30 trials – entirely comprised of incongruent Stroop stimuli – to familiarise themselves with the requirements of the tasks. The data from these practice trials were excluded from the analyses.

²Prior to running this experiment, we had no way of knowing how many cycles would be necessary to allow RTs to return to levels comparable to those seen in the Colour-Neutral condition of Experiment 1. Six cycles was arbitrarily chosen as a value that should allow participants to go at least some of the way towards achieving pre-All-Stroop RTs.

After completing the practice blocks, participants completed another short block (6 cycles, or 24 trials, long – referred to as a “mini-block”) of precisely the same tasks: *Word* in one half of the grid, *Colour* in the other; all stimuli were incongruent Stroop stimuli (i.e., the All-Stroop condition). They then proceeded (without a pause) to complete 24 trials in which they responded to neutral *Colour* stimuli and incongruent Stroop *Word* stimuli (i.e., the Colour-Neutral condition). In the Colour-Neutral mini-block, participants continued to perform *Word* in one half of the grid and *Colour* in the other – in the same configuration as for the All-Stroop mini-block. Thereafter, they regularly alternated between All-Stroop and Colour-Neutral mini-blocks until they had completed 5 mini-blocks. (A set of 5 mini-blocks is referred to as a mega-block.) Between mega-blocks, participants were allowed to rest.

Because we were concerned that an abrupt, un cued change in stimulus type might confuse participants, we included a tone at the beginning of every cycle (except the very first cycle of a mega-block). This tone acted as a cue that told the participants when they could expect the stimuli to change from All-Stroop to Colour-Neutral (or from Colour-Neutral to All-Stroop). When the stimuli changed from All-Stroop to Colour-Neutral, the cue was a low, 220-Hz tone (150 ms long); when the stimuli reverted to All-Stroop from Colour-Neutral, the cue was a high, 800-Hz tone (two bursts – each 45 ms long, with a 60-ms delay between them). To avoid confounding the start of the new block with the occurrence of a tone, a cue was included at the beginning of each cycle (except the very first cycle of a mega-block): while participants were responding to All-Stroop stimuli, the cue was always the high, 800-Hz tone; while they were responding to Colour-Neutral stimuli, the cue was always the low, 220-Hz tone. The tone always occurred immediately prior to the *Colour* Switch trial.

An individual trial consisted of the presentation of a stimulus for 3000 ms or until the participant responded. As in Experiment 1, participants responded verbally into a microphone that was attached to a voice-key. This was followed by a 500-ms RSI before the next stimulus was presented in the next location (clockwise) in the 2×2 grid. Responses that were less than 150 ms were treated as errors. Responses were coded (as correct or incorrect) on-line by the experimenter. The experiment was also recorded on audio tape to minimise experimenter error.

Since we were interested in the effect that any given mini-block could have on the subsequent mini-block, the first mini-block of each mega-block was excluded from the analysis (since it was preceded only by rest). This excluded mini-block was always an All-Stroop mini-block. Subjects completed six mega-blocks in the

experiment, yielding 12 observations at each position in each of the six cycles of a mega-block. We increased the number of observations in each cell to 24 by collapsing the data across cycles 1 & 2, 3 & 4, and 5 & 6.

Results

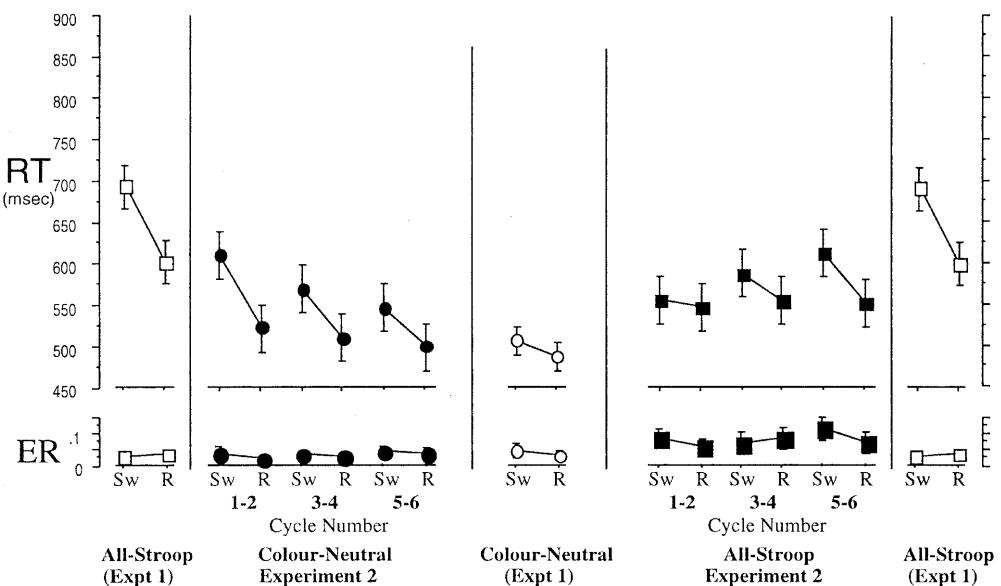
Word-reading

Colour-Neutral. As in Experiment 1, the results of primary interest are the RTs to the *Word* trials in the Colour-Neutral mini-block: these are shown in Fig. 2 (filled circles) on the left; as are the RTs to the comparable *Word* trials from the Colour-Neutral condition in Experiment 1 (open circles). The RTs from the All-Stroop mini-block are also shown (filled squares) on the right of the figure. An inspection of Fig. 2 reveals that RTs for both Switch and Repeat trials of the Colour-Neutral block gradually diminish over the course of successive cycles. However, they never attain the speed with which participants responded in the Colour-Neutral condition of Experiment 1.

This observation was confirmed by subjecting the data to a two-factor, repeated-measures ANOVA. The two factors were Cycle (Cycle 1-2, 3-4, and 5-6) and Switch (Switch vs. Repeat). Only the two main effects of Cycle [$F(2, 16) = 4.54, p < 0.035$] and of Switch [$F(1, 8) = 5.58, p < 0.05$] were reliable. The effect of Cycle resulted from participants responding with shorter RTs as the cycles progressed. The effect of Switch resulted from participants responding with longer RTs on Switch trials than on Repeat trials.

Inspection of Fig. 2 suggests that RTs on both Switch and Repeat trials declined over the course of the Colour-Neutral mini-block. Planned comparisons confirmed that RTs on Switch trials declined from Cycles 1-2 to Cycles 5-6 ($p < 0.01$). The smaller decline in RT on the

Fig. 2 RT and ER for *Word* from Experiment 2 (filled symbols). Error bars show 95% within-subject confidence intervals. Data from Experiment 1 (open symbols) are included for reference. Square symbols denote data from the All-Stroop mini-block; round symbols denote data from the Colour-Neutral mini-block



Repeat trials (from Cycles 1–2 to Cycles 5–6) was not reliable within the Colour-Neutral mini-block. However, when RTs were compared across mini-blocks, there was no reliable difference between the latencies displayed on the last Repeat trials of the All-Stroop mini-block and the first Repeat trials of the Colour-Neutral mini-block. In contrast, there was a reliable difference between the latencies on the last Repeat trials of the Colour-Neutral mini-block and the first Repeat trials of the All-Stroop mini-block ($p < 0.025$).

To further test whether, after six cycles of the Colour-Neutral mini-block, participants had attained a “fully prepared” state, a between-subjects, between-experiments analysis was run on the data from Cycles 5–6 of Experiment 2 and the data from the Colour-Neutral block of Experiment 1. This analysis had one within-subjects factor (Switch: Switch vs. Repeat), and one between-subjects factor (Experiment: 1 vs. 2). The effect of Switch was reliable [$F(1, 17) = 8.75, p < 0.01$], as was the interaction between Switch and Experiment [$F(1, 17) = 4.41, p < 0.05$]. The effect of Switch results from participants responding with shorter latencies on Repeat trials than on Switch trials. The interaction results from participants in Experiment 2 responding with longer latencies on Switch trials of Cycles 5–6 than participants in Experiment 1 (the RTs on Repeat trials were roughly comparable across the two experiments; Fig. 2).

Colour-Neutral errors. The same repeated-measures analysis was applied to the error data. There were no reliable effects, nor were there any reliable interactions.

All-Stroop. The data from the All-Stroop mini-block were analysed with the same repeated-measures analysis. The main effect of Cycles was reliable [$F(2, 16) = 5.12, p < 0.025$]. This resulted from participants responding with longer latencies as the cycles proceeded. The main effect of Switch was also reliable [$F(1, 8) = 6.9, p < 0.035$]. This resulted from participants responding

with longer latencies on Switch than on Repeat trials. The interaction between Cycles and Switch was not reliable.

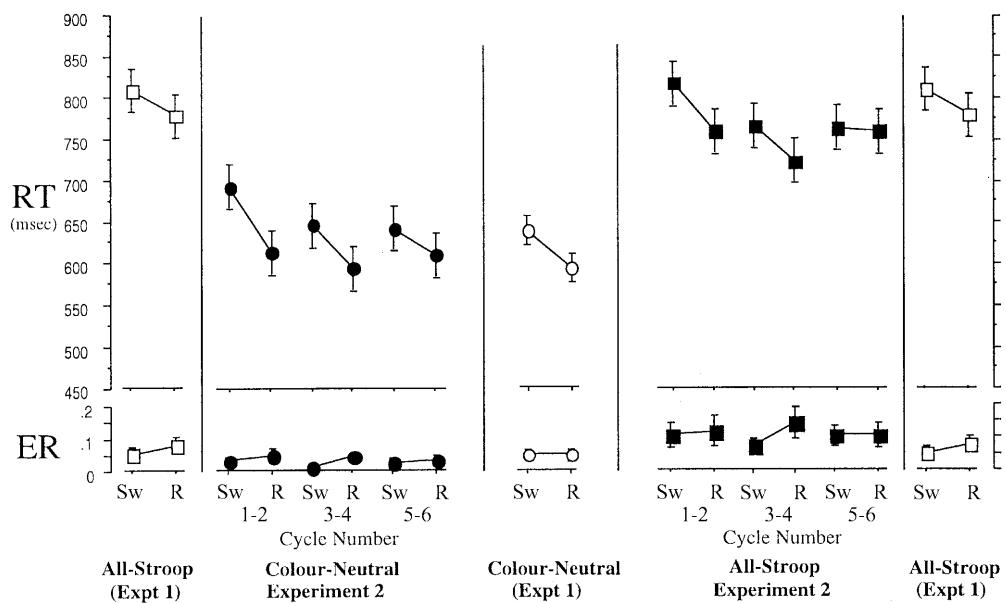
All-Stroop errors. The same repeated-measures analysis was applied to the error data. There were no reliable effects, nor were there any reliable interactions.

Colour-naming

Colour-Neutral. Figure 3 depicts the RTs to the *Colour* trials of the Colour-Neutral mini-block (filled circles) on the left, with the RTs from the comparable conditions of Experiment 1 included for comparison (open circles). Data from the *Colour* trials of the All-Stroop mini-block are also shown (filled squares: Experiment 2; open squares: Experiment 1) on the right of the figure. Inspection reveals that, though RTs were much slower – overall – during the All-Stroop mini-block compared to the Colour-Neutral mini-block, RTs on Repeat trials in the Colour-Neutral condition remained broadly stable as a function of Cycle. In contrast, RTs on Switch trials of the Colour-Neutral mini-block appeared to shorten as the cycles progressed.

These observations were confirmed by subjecting the data to a two-factor, repeated-measures ANOVA with the same factors as those used in the analysis of *Word* (Cycle and Switch). The main effect of Cycle approached significance [$F(2, 16) = 2.77, p < 0.095$]. The main effect of Switch was highly reliable [$F(1, 8) = 38.53, p < 0.001$], and confirms that participants responded with longer latencies on Switch trials than on Repeat trials. The interaction between Cycle and Switch was marginally reliable [$F(2, 16) = 3.27, p < 0.065$]. This reflects the fact that most of the change in RT on Switch and Repeat trials tended to occur between Cycles 1–2 and Cycles 3–4.

Fig. 3 RT and ER for the *Colour* task of Experiment 2 (filled symbols). Error bars show 95% within-subject confidence intervals. Data from Experiment 1 (open symbols) is included for comparison. Square symbols denote data from the All-Stroop mini-blocks; round symbols denote data from the Colour-Neutral mini-blocks



As already noted, Fig. 3 suggests that RTs on Switch trials in Colour-Neutral declined as the cycles progressed, but that RTs on Repeat trials did not. This was confirmed by planned comparisons. The difference between Switch trial RTs on Cycles 1–2 and Switch trial RTs on Cycles 5–6 was highly reliable ($p < 0.0005$); the difference between the Repeat trials on these cycles was not. Furthermore, the RTs on both Switch and Repeat trials on Cycles 5–6 were comparable to RTs in Experiment 1 in the Colour-Neutral condition (despite the difference in the type of Neutral stimuli used).

Colour-Neutral errors. The same repeated-measures ANOVA was applied to the error data. There were no reliable effects.

All-Stroop. The same repeated-measures ANOVA was applied to the *Colour* data from the All-Stroop mini-block. Both of the main effects were reliable [Cycle: $F(2, 16) = 3.46$, $p < 0.06$; Switch: $F(1, 8) = 6.54$, $p < 0.04$] as was the interaction between Cycle and Switch [$F(2, 16) = 3.83$, $p < 0.05$]. This resulted from participants responding with faster RTs on Switch trials as they progressed through the cycles, but not on Repeat trials (with the exception of the Repeat trial on Cycles 3–4, which is associated with an increased error rate; Fig. 3). Most of this acceleration occurred between Cycles 1–2 and 3–4.

All-Stroop errors. The same repeated-measures ANOVA was applied to the error data. There were no reliable effects.

Discussion

Word-reading: the Colour-Neutral mini-block

The primary purpose of this experiment was to test the prediction that PI from the Stroop *Colour* task should continue to affect *Word* performance over a number of trials, after the participants were shifted from the All-Stroop to the Colour-Neutral condition, diminishing gradually over successive Colour-Neutral cycles. The results appear to support this prediction. The interference on both the Switch and Repeat trials diminished as the cycles progressed. Furthermore, as the comparison between RTs in Cycles 5–6 of the Colour-Neutral mini-block and RTs from the Colour-Neutral condition of Experiment 1 shows, this reduction of interference appears not to be complete, even after six cycles of the Colour-Neutral condition (Fig. 2). This suggests that the PI effects from the *Colour* trials in the All-Stroop mini-block last for longer than six cycles – that is, longer than 24 trials. Thus, Experiment 2 provides a better test of the TSI hypothesis than Experiment 6 of Rogers and Monsell (1995). In that experiment, they attempted to test TSI by including more than one Repeat trial, and they found that RTs did not decrease over successive Repeat trials. This is an important result, and it shows that the state of the system may remain relatively constant over the course of three Repeat trials. However,

our results clearly show two things: (1) there is a residual interference effect from *Colour* that diminishes only gradually, over more than six cycles; and (2) there is also a residual interference effect on the Repeat trial (as in Experiment 1). Both of these points are entirely consistent with the TSI hypothesis.

A closer consideration of the results from the Colour-Neutral mini-block, however, shows that the TSI hypothesis – as stated in Allport et al. (1994) – needs a major revision if it is to account for these data. One of the main features of the TSI account is that the control state of a given task-set persists inertially from previous trials. This should predict that, throughout the Colour-Neutral mini-block, when participants no longer have to inhibit *Word* to perform *Colour*, the *Word* task-set should become increasingly less suppressed as the cycles progress. Inasmuch as RTs decreased during the Colour-Neutral block, this prediction appears to be borne out. However, the TSI hypothesis – as stated – should further predict that this RT decrease should be monotonic. That is, RTs on the Switch trials of Cycles 3–4 should be no slower than RTs on Repeat trials of Cycles 1–2, and RTs on Switch trials of Cycles 5–6 should be no slower than RTs on Repeat trials of Cycles 3–4. This is simply because no intervening events occurred (in the *Colour* task) which should change the control state of the *Word* task-set. Figure 2 shows that this prediction is not borne out. The increase in *Word* RT on Switch trials, relative to the preceding Repeat trials, suggests that either (1) there is some additional, stage-like process required on each Switch trial, or (2) the interference from previous Stroop colour-naming trials is re-elicted on Switch trials and must be overcome before a reading response can be released. If we suppose that there is some additional process required on Switch trials, we must further suppose that the duration, and perhaps even the necessity, of this process depends upon whether, and how recently, a competing task has been performed. Importantly, this putative process cannot be supposed to be an “endogenous” control operation, since its duration appears to depend critically upon how recently another, competing task has been performed: in other words, its duration appears to depend upon external events, and not upon endogenous processes. Therefore, even if we suppose that there is a stage-like (control) process, which contributes to these Switch trial RTs, its duration appears to be affected by PI from previous Stroop colour-naming, the amount of interference gradually diminishing over successive cycles as Stroop colour-naming recedes into the past. (Note, this experiment is not able to distinguish whether the reduction of PI is dependent on the amount of time since the last instance of Stroop colour-naming, or on the number of intervening Stroop word-reading trials.)

The hypothesis that the interference from previous Stroop colour-naming trials is re-elicted, or retrieved, on Switch trials, however, will be supported by Experiment 3, and is supported by a consideration of the All-Stroop mini-blocks. We therefore turn to the All-Stroop mini-blocks now.

Word-reading: explaining the All-Stroop mini-block

The TSI hypothesis (Allport et al., 1994) is not sufficient to account for the non-monotonic decline in interference on word-reading in the Colour-Neutral mini-blocks. The TSI hypothesis is not sufficient, either, to account for evidence that this interference effect is, at least in part, item specific. (Indeed, no previous models of task-switching performance can account for such evidence.) Allport and Wylie (in press-b, Experiment 5) tested the hypothesis that the large interference effects found on Stroop word-reading following Stroop colour-naming were due to a general suppression of the word-reading pathway, (i.e., the word reading “task-set”). We reasoned that, if this was the case, after an intervening block of Stroop colour-naming, participants’ Stroop word-reading latencies should be longer (compared to word-reading performance prior to the Stroop colour-naming block), regardless of what words were to be read. On the other hand, if the PI from colour-naming was – at least in part – item specific, then participants’ word-reading latencies would be especially long when they read words that had also been presented as distractors in the intervening *Colour* block. This turned out to be the case. Allport and Wylie (in press-b) found that, while participants’ Switch trial latencies were increased for all words read, after the intervening *Colour* block, their latencies were especially long in response to words that had been presented as distractors during the intervening *Colour* block, and hence might be subject to long-term, item-specific negative priming.

This result cannot be accounted for by the hypothesis that the interference on word-reading was entirely due to persisting activation-suppression of the colour-naming and word-reading processing pathways, respectively. Such a hypothesis should predict only a general slowing of word-reading, and would not predict item-specific effects. This is simply because a pure activation-suppression account – affecting the whole processing pathway – does not include any mechanism whereby previous experience with individual stimuli can affect performance. For performance to be affected differentially by individual, “primed” and “not primed” stimulus items in this way, the system must retain past instances – either as some form of “event” memory, or as item-specific S-R associations.

Recently, Lowe (1998) found long-term, item-specific negative priming effects. In his experiments, some stimulus items were negatively primed one, three, six or nine times, and these items were probed after a 5-min delay. Provided the context was similar between prime and probe, he found evidence of long-term negative priming (i.e., negative priming that outlasted the 5-min interval) that was modulated by the number of times the individual stimuli had been negatively primed (i.e., the long-term negative priming was item specific). Both of these findings are difficult to account for on the supposition that negative priming effects are entirely the result of

persisting activation-suppression states. Rather, they suggest the involvement of memory (and stimulus-cued retrieval) processes in at least some forms of negative priming.

Applying this interpretation to the results of Experiment 2, it is proposed that, during the *Colour* trials of the All-Stroop mini-block, participants cumulatively learned (or strengthened) S-R associations, or bindings, (a) between each of the specific colours (of the Stroop stimuli) and a colour-naming response, and (b) between the specific word forms and a null (“inhibitory”) response. The S-R associations between the colours and a colour-naming response were already relatively well learned; in contrast, the S-R associations between the word forms and a null response were presumably new for these participants. When they were subsequently required to read the words (in the All-Stroop condition), we suggest, interference resulted because the same stimuli triggered retrieval of these competing S-R bindings, established during performance of the preceding *Colour* trials. (We refer to this as the retrieval hypothesis.) When participants then moved on to the Colour-Neutral mini-block, and ceased to perform *Colour* with incongruent Stroop stimuli, the only bindings that would continue to be updated – in response to Stroop stimuli – were between Stroop stimuli and a (positive) reading response. Therefore, though the Stroop stimuli would continue to re-evoke the bindings established during the *Colour* trials of the earlier All-Stroop mini-block(s) (resulting in interference with *Word*), as the Colour-Neutral mini-block proceeded this interference would be increasingly counterweighted by the re-learning of (item-specific) positive S-R bindings between the word-stimuli and the appropriate reading response. Word-reading RTs therefore became shorter as the Colour-Neutral mini-block proceeded.

Returning to the All-Stroop mini-blocks, Fig. 2 shows that, during these mini-blocks, *Word* RTs on Switch trials gradually increased. This is also inconsistent with the TSI hypothesis. If participants suppressed the task-set for *Word* to perform *Colour* (with incongruent Stroop stimuli), as the TSI hypothesis claims, we should expect them to show large costs on *Word* after just one (or, perhaps, two) Stroop *Colour* trial(s). That is, we would expect to find RTs to *Word* to increase dramatically in Cycle 1, and to remain relatively constant throughout the remainder of the All-Stroop mini-block. We would not expect to find Switch trial RTs gradually increasing during the course of this mini-block.

However, the hypothesis that the interference effects on *Word* in the All-Stroop condition are due to stimulus-triggered retrieval of the S-R associations, incrementally learned during the Stroop *Colour* trials, allows us to explain this pattern of results. As previously discussed, results reported by Allport and Wylie (in press-b, Experiment 5) showed that the PI effects found on Stroop word-reading (reverse Stroop effects), following

Stroop colour-naming, were due, at least in part, to the re-presentation of individual stimulus items that participants had previously encountered in the Stroop *Colour* context. The pattern of results suggested that item-specific negative priming might underlie the reverse Stroop effects found in switching contexts (the retrieval hypothesis). This hypothesis clearly deserves more direct testing.

While Experiment 2 was not designed with this hypothesis in mind, the retrieval hypothesis suggested the following post-hoc analysis. As each All-Stroop mini-block proceeded, an increasing proportion of the stimuli presented on *Word* trials would have been items that had previously been responded to, within the same mini-block, in the *Colour* task. Thus, in Cycles 1–2 of any All-Stroop mini-block, relatively few stimulus items would have, so far, been negatively primed by *Colour* performance during that mini-block. In contrast, by Cycles 5–6, most of the word items would have been negatively primed by individual *Colour* trials in that mini-block. Negative priming from the Stroop *Colour* task would, therefore, be expected to affect an increasing proportion of *Word* trials as the Cycles progressed. Hence, the median RT would tend to increase over the course of the mini-block.

To test this interpretation, the RTs associated with word-stimuli that had been negatively primed in this way during the preceding course of each All-Stroop mini-block were extracted, and analysed separately. The two data-sets are shown in Fig. 4. For this analysis, we used a repeated-measures ANOVA with two factors: Cycle (Cycles 1–2, Cycles 3–4, Cycles 5–6) and Switch (Switch vs. Repeat). It will be recalled that, when all of the data were included, both factors were reliable.

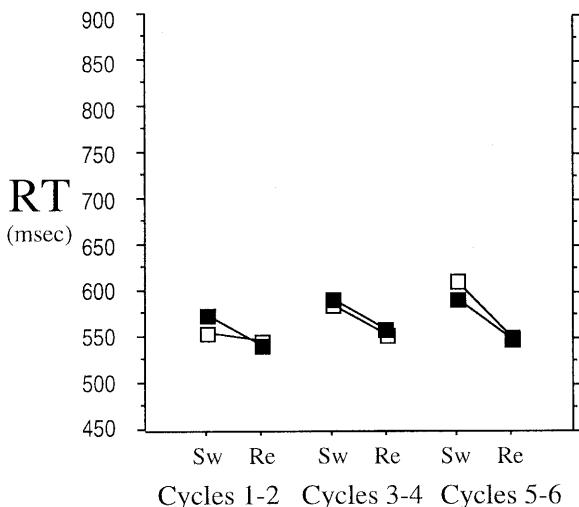


Fig. 4 RT for *Word* in the All-Stroop block of Experiment 2. The unfilled squares correspond to the whole data-set; the filled squares correspond to those trials that were negatively primed (Sw Switch, Re Repeat)

When the data from only the negatively primed items were analysed, only the effect of Switch was reliable [$F(1, 8) = 6.00, p < 0.05$]. The lack of an effect of Cycle or of an interaction between Switch and Cycle accord well with the predictions of the retrieval hypothesis. According to the retrieval hypothesis, during the Colour-Neutral mini-block Stroop stimuli were associated only with word-reading responses. When participants subsequently began the All-Stroop mini-block, these recent S-R associations (from the previous Colour-Neutral mini-block) biased responses to Stroop stimuli in favour of word-reading, and we found that participants produced relatively fast Switch RTs (Cycles 1–2). However, this is not the case for items (i.e., word forms) that participants have already had to actively suppress during the All-Stroop *Colour* trials of the current mini-block. These items were now associated with a null response, and consequently *Word* RTs to these items were longer. As the Cycles progressed, an increasing proportion of *Word* items would have been encountered in the *Colour* task context and, therefore, been associated recently with a null response code. Thus, although – overall – Switch trial median RTs increased gradually as the Cycles progressed, the RTs to items that had been negatively primed in this way remained more or less constant³.

The interpretation of switch costs

Word-reading. Experiment 2 extends our critique of switch costs in two ways. First, it extends our argument that Repeat trials do not offer a stable baseline of “task preparation” or “task readiness” against which to measure Switch trial RTs. In the course of the Colour-Neutral mini-blocks, RTs on Repeat trials gradually decreased. In the All-Stroop mini-blocks, response times on Repeat trials were longer throughout, relative to the last cycle of Colour-Neutral and relative to the All-Neutral condition of Experiment 1. Therefore, if one wishes to calculate switch costs by subtracting Repeat trial RTs from Switch trial RTs, it is difficult to know which Repeat trial RTs to use. Evidently, the Repeat trials in Colour-Neutral do not represent a “completely prepared state,” given the considerably faster RTs that participants demonstrated in Experiment 1 (in either the All-Neutral or the Colour-Neutral conditions). Thus, this experiment provides grounds for misgivings about inferences based on switch costs in the alternating runs paradigm.

³It might be argued that the lack of a significant effect of Cycle in the analysis of the negatively primed RTs is not surprising, given that we have decreased the number of observations per data point. Admittedly, this analysis is not a powerful one. An experiment specifically designed to address this issue is needed. However, while we are not able to make strong claims, it is nevertheless important that we find the predicted pattern of results when we analyse the appropriate subset of the data.

The second critique of switch costs offered by Experiment 2 is an extension of the finding that switch costs changed as a function of the requirements of the *other* task. Experiment 1 showed that switch costs changed as a function of the task that participants were switching from, rather than as a function of the task they were switching to; Experiment 2 adds to this finding by showing the extended time course of this effect. The stimuli used for the *Word* task in Experiment 2 were always incongruent Stroop stimuli. Hence, the changes in RT and switch costs cannot be attributed to changes in the intrinsic requirements of the *Word* task – these effects must be attributed to the task participants were switching from: the *Colour* task. Furthermore, the time course of these effects is quite long – indeed, we underestimated the number of cycles needed when we designed this experiment. We had supposed that six cycles (24 trials) of the Colour-Neutral condition, following the All-Stroop condition, would be enough for the *Word* RTs to approach levels comparable to those of the Colour-Neutral condition in Experiment 1. Clearly, we were mistaken in this. The after-effects of having performed the *Colour* task with incongruent Stroop stimuli apparently lasted longer than 24 trials.

Colour-naming. A remarkable feature of the data from the two *Colour* conditions was the similarity of the switch costs in the All-Stroop and the Colour-Neutral conditions (Colour-naming switch costs are shown in Fig. 5). Although (as expected) the *Colour* RTs to Stroop stimuli were all longer than those to neutral stimuli, the switch costs were very similar between the two conditions, and they changed in a similar way over successive cycles. This observation is quite important. It is tempting to assume that switch costs can be taken as an index of the amount of executive control necessary to switch to a new task. However, this supposition is not supported by the present data – nor by the data from

Experiment 1 (Fig. 5). Stroop colour-naming has been taken as a paradigm case of a task that requires considerable executive control; hence, according to this interpretation, switch costs should be larger when participants are switching to Stroop colour-naming than when they are switching to neutral colour-naming. Contrary to this assumption, there was no reliable difference between the switch costs associated with these two stimulus types.

Experiment 3

We have presented evidence here and elsewhere (Allport & Wylie, in press-a, in press-b; Allport et al., 1994; Meuter & Allport, in press), that the performance costs following a switch between divergent S-R tasks represent, in large measure, a form of PI from the preceding task. In Experiment 2, however, several features of the results appeared inconsistent with the earlier account of Allport et al. (1994) of such PI, in terms of persisting activation-suppression of task processing pathways (“task-set inertia”). These results, and others reported by Allport and Wylie (in press-b) led us to propose that it is learned S-R associations (i.e., associative weight changes) that persist, rather than patterns of activation and suppression.

Because the retrieval hypothesis proposes that S-R associations are *learned*, it implies that they may be incrementally modified on every trial. If so, the more often a given S-R mapping is executed with a given stimulus type, the stronger the S-R links will become between those stimuli and the responses associated with that task. Thus, if participants perform the *Word* task with Stroop stimuli twice as often as they perform the *Colour* task with the same Stroop stimuli, we would expect the Stroop stimuli to be more strongly associated with word-reading responses than colour-naming responses. This was the case in Experiment 2: participants read Stroop stimuli in both the All-Stroop mini-blocks and the Colour-Neutral mini-blocks, but they only named the colour of Stroop stimuli in the All-Stroop mini-blocks. If we reversed this ratio, such that participants performed the *Colour* task with Stroop stimuli twice as often as they performed the *Word* task with the same Stroop stimuli, our hypothesis would predict that the Stroop stimuli should become more strongly associated with colour-naming. If so, we would expect to find more PI on *Word*, in response to Stroop stimuli, than was found in Experiment 2.

We tested this prediction in this experiment using the same basic paradigm that was used in Experiment 2, but with the following difference. In Experiment 2, the *Colour*-task stimuli switched in successive mini-blocks between neutral and Stroop, while the stimuli for the *Word* task were Stroop stimuli throughout. In Experiment 3, the *Word*-task stimuli switched (in successive mini-blocks) between neutral and Stroop, while the stimuli for the *Colour* task remained Stroop stimuli

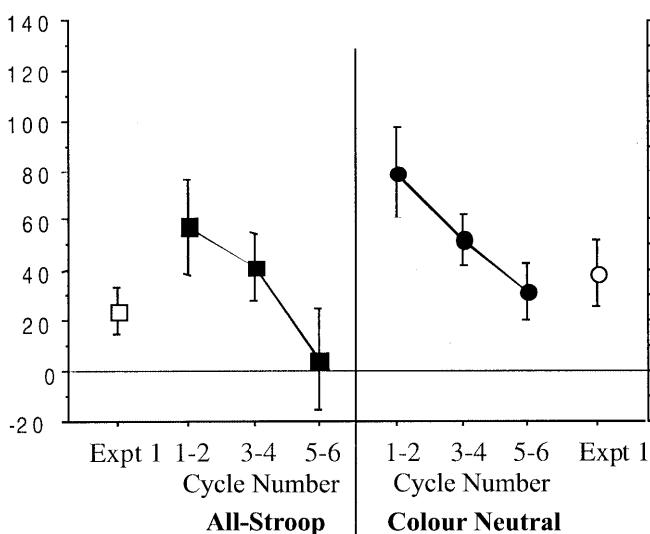


Fig. 5 Switch costs for *Colour*. Solid symbols represent data from Experiment 2; open symbols represent data from Experiment 1

throughout. Thus, in this experiment, for the *Colour* task participants always encountered incongruent Stroop stimuli; but, for the *Word* task, they encountered neutral and incongruent Stroop stimuli in alternate mini-blocks.

Method

Participants. Four men and four women participated in this experiment. Participants ranged in age from 28 to 50 with a mean age of 38.6 years. All were recruited from the Oxford University Subject Panel, and were paid for their participation. Participants all had normal, or corrected to normal acuity, and normal colour vision. None of these participants had taken part in Experiments 1 or 2.

Stimuli and apparatus. The apparatus and general display was exactly the same as that used for Experiment 2. The stimuli used for the All-Stroop mini-blocks were also the same (i.e., they were incongruent Stroop stimuli). However, the stimuli used for the other mini-blocks differed: in this experiment, the stimuli used for the *Word* task in these other mini-blocks were neutral stimuli. These mini-blocks will, therefore, be referred to as Word-Neutral mini-blocks. The neutral *Word* stimuli consisted only of the outlines of the letters. These stimuli thus appeared to be transparent, and to have no colour. The stimuli used for the *Colour* task in the Word-Neutral mini-blocks were always incongruent Stroop stimuli.

Procedure. The procedure used was similar to that used for Experiment 2. The only difference was that Word-Neutral mini-blocks were presented to these participants where Colour-Neutral mini-blocks had been presented to participants in Experiment 2.

Results

Word-reading

To investigate the effect of stimulus type on word-reading, the data from the two mini-blocks of *Word* (Word-Neutral and All-Stroop) were analysed together (Fig. 6). This repeated-measures ANOVA had three factors: Mini-block (Word-Neutral vs. All-Stroop), Cycles (Cycles 1–2, 3–4, 5–6), and Switch (Switch vs. Repeat). All three main effects were reliable: Mini-block [$F(1, 7) = 22.23, p < 0.01$], Cycle [$F(2, 14) = 4.33, p < 0.035$], Switch [$F(1, 7) = 28.13, p < 0.01$]. The effect of Mini-block resulted from participants responding with longer latencies when reading Stroop stimuli (the All-Stroop mini-block) than when reading neutral stimuli (the Word-Neutral mini-block). The effect of Cycle resulted from a tendency for participants to respond with longer latencies on the later cycles. The effect of Cycle was different in the two mini-blocks, resulting in a reliable interaction between Mini-block and Cycle [$F(2, 14) = 5.66, p < 0.025$]. This interaction was due to participants' RTs in the Word-Neutral mini-block being shorter in Cycles 1–2 than in Cycles 3–4 or

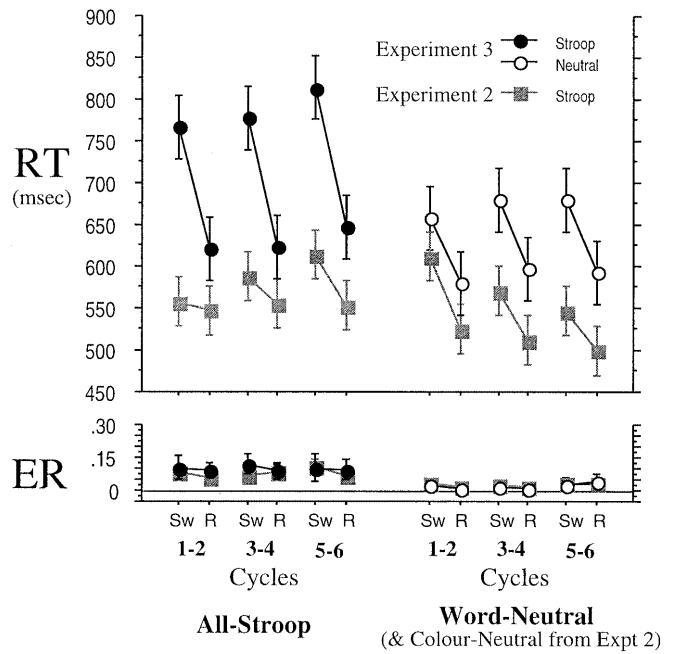


Fig. 6 RT and ER for word-reading from Experiment 3 (filled circles denote data from the All-Stroop mini-block; open circles denote data from the Word-Neutral block). Error bars show 95% within-subject confidence intervals. Word Data from Experiment 2 (Colour-Neutral) is included for comparison (grey, filled squares) (SW Switch) (R Repeat)

5–6 ($p < 0.01$ in both cases). In the All-Stroop mini-block, RTs in Cycles 1–2 were not reliably different from RTs in Cycles 3–4, but RTs in Cycles 5–6 were reliably longer than in the previous Cycles ($p < 0.001$ in both comparisons).

Finally, Mini-block interacted with Switch [$F(1, 7) = 13.01, p < 0.01$]. This resulted from a larger difference between Switch and Repeat trial RTs in the All-Stroop mini-block than in the Word-Neutral mini-block.

Errors for word. The same analysis was carried out on the error data. No effects were reliable.

Colour-naming

The same repeated-measures ANOVA was applied to the *Colour* data from the two mini-blocks (Fig. 7). Although RTs to *Colour* were not the focus of this experiment, this contrast represents the analogue of the contrast between *Word* RTs in the Colour-Neutral and All-Stroop mini-blocks of Experiment 2. It was, therefore, of interest to see whether *Colour* RTs would differ between the Word-Neutral and All-Stroop mini-blocks in this experiment, despite the fact that the only difference between these conditions consisted in the stimuli used for the *other* task (*Word*).

The only reliable main effect was that of Switch [$F(1, 7) = 13.96, p < 0.01$]. This resulted from participants responding with longer RTs on Switch trials than

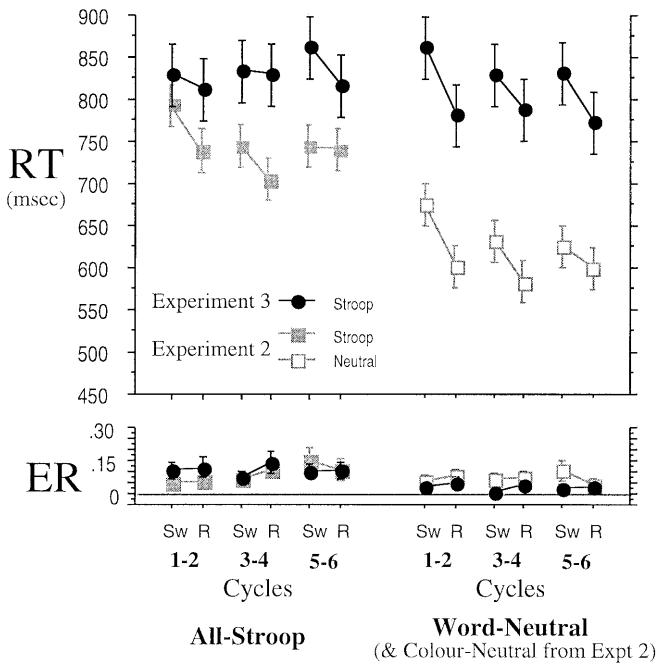


Fig. 7 RT and ER for *Colour* from Experiment 3 (filled circles). Error bars show 95% within-subject confidence intervals. Data from Experiment 2 is included for comparison (grey, filled squares denote data from the All-Stroop mini-block; open grey, squares denote data from the Colour-Neutral mini-block)

on Repeat trials. The interaction between Switch and Mini-block approached significance [$F(1, 7) = 3.74, p < 0.095$]. This marginally reliable effect resulted from participants responding, on Switch trials, with broadly similar latencies in the two mini-blocks, whereas Repeat trial RTs were consistently shorter in the Word-Neutral than in the All-Stroop mini-block.

Errors for colour. The same analysis was carried out on the error data. No effects were reliable.

Between experiment comparisons

Because the same paradigm was used in Experiments 2 and 3, we can compare participants' performance across the two experiments. These experiments were designed to investigate performance on *Word*; for brevity, therefore, we only report comparisons between the *Word* blocks (Fig. 6). The most straightforward comparison is between the All-Stroop mini-blocks in the two experiments. These data were analysed with a three factor, mixed, between- and within-subjects ANOVA. Experiment was the between-subjects factor (Experiment 2 vs. 3); the within-subjects factors were Cycles (Cycles 1–2, 3–4, 5–6) and Switch (Switch vs. Repeat). All three main effects were reliable: Experiment [$F(1, 15) = 7.43, p < 0.015$], Cycles [$F(2, 30) = 10.41, p < 0.001$], and Switch [$F(1, 15) = 33.71, p < 0.0001$]. The effect of Cycle resulted from participants responding with longer latencies as the cycles progressed. The effect of Experi-

ment resulted from participants in Experiment 2 responding with shorter latencies than participants in Experiment 3. The effect of Switch resulted from participants responding with longer RTs on Switch trials than on Repeat trials. Experiment also interacted with Switch [$F(1, 15) = 13.82, p < 0.01$]. This interaction resulted from participants in Experiment 3 exhibiting a larger difference between Switch and Repeat trials than participants in Experiment 2 (Fig. 6).

Despite the difference in the stimuli used in the Colour-Neutral and Word-Neutral mini-blocks (incongruent Stroop stimuli vs. neutral stimuli, respectively), it is nevertheless instructive to compare the *Word* RTs from these two mini-blocks (Fig. 6). The same mixed, between- and within-subjects ANOVA was used to analyse the data from these blocks as was used in the previous analysis for the two All-Stroop conditions. Of the main effects, only Switch reached significance [$F(1, 15) = 20.99, p < 0.001$]; however, Experiment was marginally reliable [$F(1, 15) = 3.44, p < 0.085$]. The effect of Switch resulted from participants responding with shorter latencies on Repeat trials than on Switch trials. The effect of Experiment resulted from participants reading incongruent Stroop words in Experiment 2 faster than they read neutral words in Experiment 3. The two-way interaction between Cycles and Experiment was also reliable [$F(2, 30) = 6.46, p < 0.01$]. This resulted from RTs decreasing across cycles in Experiment 2, and slightly increasing across cycles in Experiment 3.

Discussion

Experiment 3 was designed to investigate the prediction that larger PI effects on *Word* would be found, relative to Experiment 2, in the All-Stroop mini-blocks when the set of Stroop stimuli was associated with the *Colour* task more often than with the *Word* task. In Experiment 2, the ratio of Stroop colour-naming to Stroop word-reading was 1:2 in favour of word-reading. In Experiment 3, the ratio of Stroop colour-naming to Stroop word-reading was 2:1 in favour of colour-naming. We hypothesised that, if the PI effects in Experiments 1 and 2 were due to learned S-R associations, this change in ratio should increase the PI effects on *Word* in the All-Stroop mini-blocks of Experiment 3 relative to the same mini-blocks of Experiments 2. As Fig. 6 shows, the data clearly support this prediction.

Furthermore, performance on *Word* was strongly affected by the extent to which the *Word* stimuli overlapped with the stimuli used for *Colour*. We found a large amount of interference on *Word* in the All-Stroop mini-blocks, where the stimuli used for the two tasks were identical, and we found much less interference on *Word* in the Word-Neutral mini-blocks, where the stimuli used for *Word* were not coloured. This result suggests that the stimuli themselves cue the S-R associations that participants have learned during previous

trials. The more the stimuli for *Word* overlap with the stimuli previously associated with *Colour*, the more they will cue the *Colour* task, and the more interference will result.

The interpretation of switch costs

Word-reading. Because the All-Stroop mini-block of Experiment 3 was identical to the All-Stroop mini-block of Experiment 2, we compared All-Stroop word-reading latencies across experiments. This showed that *Word* latencies were longer in Experiment 3 than in Experiment 2, but also that the difference between Switch and Repeat trial RTs was larger in Experiment 3 than in Experiment 2 (Fig. 6). This difference in switch costs across the experiments is striking, and requires an explanation.

In the All-Stroop mini-blocks the task requirements for both *Colour* and *Word* were precisely the same in Experiment 2 and Experiment 3. Therefore, the medium-to short-term requirements of the All-Stroop mini-blocks in the two Experiments were the same. To account for this pattern of results, a hypothesis is needed that explicitly includes the long-term, cumulative effects of previous processing. Unlike the retrieval hypothesis, none of the hypotheses currently available in the task-switching literature includes a mechanism to account for such effects.

Colour-naming. In Experiment 3, the stimuli for the *Colour* task remained constant, and it was the stimuli for *Word* that changed across mini-blocks. Thus, the most straightforward comparison is between *Colour* RTs in the All-Stroop and Word-Neutral mini-blocks, in response to Stroop stimuli in both cases. This comparison revealed that there was a difference in RTs on Repeat trials in these two mini-blocks: participants responded with longer RTs on Repeat trials in the All-Stroop mini-block than in the Word-Neutral mini-block. Response times on the Switch trials, however, did not differ between mini-blocks. This has the curious result that switch costs were *smaller* during the All-Stroop mini-block than during the Word-Neutral mini-block. This finding is very difficult to reconcile with the view that these RT differences (so-called switch costs) provide a behavioural measure of the amount of control necessary to switch from one task to another. This conclusion becomes even stronger when the *Word* switch cost data are also considered.

In Experiment 1, for instance, large switch costs were associated with *Word*, only when bivalent (Stroop) stimuli were used for both the *Word* task and the competing *Colour* task. A model that attributes switch costs to the time it takes some stage-like process to switch, or “reconfigure” the system might account for this by proposing that, in this case, the reconfiguration takes longer because the task-stimuli do not uniquely cue the relevant task. Thus, the longer switch costs represent the extra time the reconfiguration takes

in the face of interference from another, potentially relevant task. However, such an analysis should surely predict that the *same* pattern would be evidenced in the *Colour* switch costs in the corresponding case (the Word-Neutral condition in Experiment 3). Larger switch costs should be found for *Colour* when the *Word* stimuli are the same as the *Colour* stimuli (the All-Stroop condition) compared to a condition in which they are different (the Word-Neutral condition). Contrary to this, we found smaller switch costs for *Colour* in the All-Stroop condition than in the Word-Neutral condition.

Concluding remarks

These experiments have shown that switch costs, on their own, cannot be used as an index of the extent to which control processes are active in the alternating runs paradigm of Rogers and Monsell (1995). It is possible that, in some circumstances, switch costs can provide useful insights into executive task control; however, in the experiments presented here, a consideration of the switch costs does not seem to add anything to the interpretation of the data. In making this assertion, we are not claiming that “control processes” do not exist – they manifestly do, as evidenced by the fact that participants successfully and consistently switch from one task to the other, making very few errors. Neither are we claiming that these control processes take zero time to execute, or that they are, in principle, unmeasurable. One measurement technique that has already proved useful, and should continue to yield enlightening results, is the manipulation of the preparation interval (e.g., Allport et al., 1994; De Jong, in press; Mayr and Keele, in press; Meiran, 1996; Meiran et al., in press; Rogers and Monsell, 1995; and De Jong et al., submitted).

However, the experiments presented here have provided evidence that is inconsistent with all previous task-switching models. The persisting interference on *Word* performance in the Colour-Neutral mini-block of Experiment 2 is inconsistent with any model that claims (or assumes) that task-set reconfiguration is completed in the course of a single Switch trial (e.g., Rogers and Monsell, 1995; Meiran, 1996; and De Jong et al., submitted). Although this persisting interference is predicted by the TSI hypothesis (Allport et al., 1994), the non-monotonic decline in interference is not consistent with that hypothesis. Moreover, neither stage-like models of switch costs nor the TSI hypothesis is able to account for the item-specific effects, found in the data from the All-Stroop mini-block of Experiment 2. In contrast, a new hypothesis, based on learned associations between stimulus representations and response representations, does very much better. This hypothesis is similar to learning and retrieval-based theories of negative priming (Allport, Tipper & Chmiel, 1985; Lowe, 1998; Neill, 1997; Neill, Valdes, Terry & Gorfein, 1992). On every trial, participants

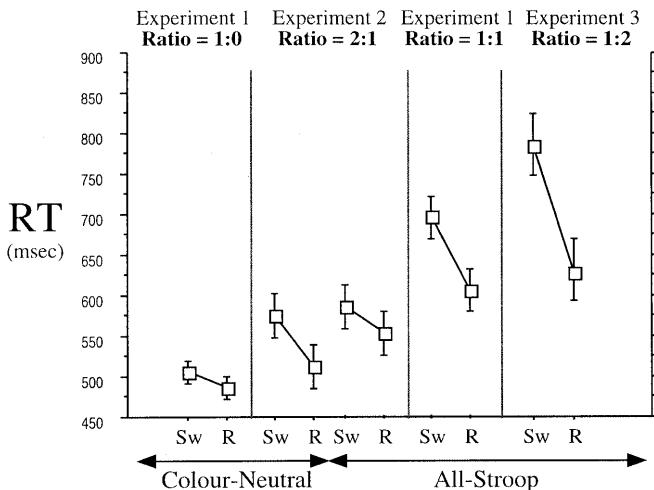


Fig. 8 RT for Stroop word-reading across the three experiments as a function of *Word* to *Colour*

modify the associations between stimulus and response representations [and, perhaps, also the representation(s) associated with the goal]; and, on every trial, participants are influenced by previously learned S-R associations.

One of the implications of this interpretation is that performance should be affected, monotonically, by the proportion of trials on which participants have learned to associate Stroop stimuli with *Colour* vs. *Word*. That is, the ratio of Stroop colour-naming to Stroop word-reading should affect the amount of PI found on *Word*. The comparison of Stroop word-reading across the various conditions of the three experiments speaks directly to this point. In the Colour-Neutral condition of Experiment 1, participants had only ever read Stroop word stimuli (i.e., the ratio of *Word* to *Colour* was 1:0). In Experiment 2, the ratio of *Word* to *Colour* responses (to Stroop stimuli) was 2:1. In the All-Stroop condition of Experiment 1, the ratio was 1:1 (in the medium term). In Experiment 3, Stroop stimuli were presented for colour-naming on all trials and for word-reading on only half of *Word* trials – making the ratio 1:2. If learned S-R associations are important in determining performance, we should find an increasing amount of interference on *Word* as the ratio increasingly favours the *Colour* task. As Fig. 8 shows, this is precisely what is found.

Despite the secondary differences in the methods used in the three experiments, a between-subjects ANOVA was applied to the Stroop word-reading data from each of the conditions in which participants read Stroop words in the three experiments. This was done to provide a starting point for future research into the effect of the ratios of competing tasks on task performance. This analysis had one between-subjects factor (Ratio) and one within-subjects factor (Switch). Both of the main effects were reliable: Switch [$F(1, 42) = 69.82, p < 0.0001$], Ratio [$F(4, 42) = 8.21, p < 0.0001$]. The

interaction between Switch and Ratio was also highly reliable [$F(4, 42) = 7.33, p < 0.0001$]. This analysis must be treated with some caution, as the stimuli used in Experiment 1 differed somewhat from those used in Experiments 2 and 3. Nevertheless, there appears to be a robust effect of ratio across the three experiments. This result is entirely consistent with a memory- and retrieval-based explanation of these effects (the retrieval hypothesis), but is difficult to reconcile with any model of task-switching performance that does not explicitly include memorial processes.

Acknowledgements Glenn Wylie was funded by a Studentship from the J. S. McDonnell Foundation and the Pew Charitable Trusts of St Louis, USA, through the Oxford McDonnell-Pew Centre for Cognitive Neuroscience, UK (McDonnell-Pew Cognitive Neuroscience Program Grant no. 93-44). We gratefully acknowledge this support.

References

- Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421–452). Cambridge, MA: MIT Press.
- Allport, D. A., Tipper, S. P., & Chmiel, N. R. J. (1985). Perceptual integration and post-categorical filtering. In M. I. Posner, & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 107–132). Hillsdale, NJ: Erlbaum.
- Allport, A., & Wylie, G. (in press-a). Task switching: positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, A. M. Treisman (Eds.), *Attention, space and action: studies in cognitive neuroscience*. Oxford, England: Oxford University Press.
- Allport, A., & Wylie, G. (in press-b). 'Task-switching', stimulus-response bindings, and negative priming. In S. Monsell, & J. Driver (Eds.); *Control of cognitive processes: attention and Performance XVIII*. Cambridge, MA: MIT Press.
- Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Coltheart, M. (Ed.). (1987). *The psychology of reading. Attention and performance XII*. Hove, England: Erlbaum.
- De Jong, R. (in press). An intention-activation account of residual switch cost. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes: attention and performance XVIII*. Cambridge, MA: MIT Press.
- Goschke, T. (in press). Decomposing the central executive: persistence, deactivation, and reconfiguration of voluntary task set. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes: attention and performance XVIII*. MIT Press, Cambridge, MA.
- Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, 89.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review*, 1, 476–490.
- Los, S. A. (1996). On the origin of mixing costs: exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, 94, 145–188.
- Lowe, D. (1998). Long-term positive and negative identity priming: evidence for episodic retrieval. *Memory and Cognition*, 26, 435–443.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, 109, 163–203.
- Mayr, U., & Keele, S. (in press). Changing internal constraints on action: the role of backward inhibition. *Journal of Experimental Psychology: General*.

- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1-20.
- Meiran, N., Choren, Z., & Sapir, A. (in press). Component processes in task shifting. *Cognitive Psychology*.
- Meuter, R. F. I., & Allport, D. A. (in press). Bilingual language switching in naming: asymmetrical costs of language selection. *Journal of Memory and Language*.
- Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1291-1305.
- Neill, W. T., Valdes, L. A., Terry, K. M., & Gorfein, D. S. (1992). Persistence of negative priming: II. Evidence for episodic trace retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 993-1000.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1-18). New York: Plenum.
- Rogers, R. D., & Monsell, S. (1995). The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207-231.
- Spector, A., & Biederman, I. (1976). Mental set and shift revisited. *American Journal of Psychology*, 89, 669-679.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.