Cooper, Richard P. and Wutke, Karolina and Davelaar, Eddy J. (2012) Differential contributions of set-shifting and monitoring to dual-task interference. Quarterly Journal of Experimental Psychology 65 (3), pp. 587-612. ISSN 1747-0218.

Differential contributions of set-shifting and monitoring to dual-task interference

Journal Article

http://eprints.bbk.ac.uk/4176

Version: Accepted (Refereed)

Citation:

Cooper, R.; Wutke, K.; Davelaar, E.. (2012)
Differential contributions of set-shifting and monitoring to dual-task interference –
Quarterly Journal of Experimental Psychology (In Press)

To appear in the *Quarterly Journal of Experimental Psychology*

Differential Contributions of Set-Shifting and Monitoring to Dual-Task Interference[*]

Richard P. Cooper, Karolina Wutke and Eddy J. Davelaar

Birkbeck, University of London

It is commonly argued that complex behaviour is regulated by a number of "executive functions" which work to co-ordinate the operation of disparate cognitive systems in the service of an overall goal. However, the identity, roles, and interactions of specific putative executive functions remain contentious, even within widely accepted tests of executive function. The authors present two experiments that use dual-task interference to provide further support for multiple distinct executive functions and to establish the differential contributions of those functions in two relatively complex executive tasks – Random Generation and the Wisconsin Card Sorting Test. Results are interpreted in terms of process models of the complex executive tasks.

Keywords: Executive functions, set-shifting, monitoring, dual-task interference, random number generation, Wisconsin Card Sorting Test.

Recent years have witnessed substantial interest, and progress, in understanding the mechanisms that support the control or regulation of complex cognitive tasks. There is now a substantial literature that posits that such control is effected by a number of separable functions, such as *set-shifting* and *response inhibition*. These so-called *executive functions* are held to be general control functions that operate across tasks. Thus, set-shifting is held to be invoked whenever it is necessary to switch from one task to another. This occurs in a range of laboratory paradigms where participants must periodically switch from one stimulus-response set to another (see Kiesel et al., 2010, Monsell, 2003, and Vandierendonck et al., 2010, for reviews), as well as in naturalistic settings such as responding to an interruption (e.g., answering a telephone call while writing a manuscript). Similarly, response inhibition is held to be invoked whenever it is necessary to inhibit a prepotent response. This is commonly studied in the colour-naming condition of the Stroop task (e.g., Friedman & Miyake, 2004; Miyake et al, 2000), but the same function is held to be involved when resisting either temptation or the production of an habitual response (e.g., when resisting the urge to answer a ringing telephone).

There is now converging evidence from behavioural, neuropsychological and neuroimaging studies for a number of specific control functions. Consider set-shifting. Behaviourally, as indicated above, this is generally considered to be evidenced in tasks where it is necessary to switch from one response set to another. In the laboratory it was first assessed by tasks where participants are required to complete one block of N trials of type A, one block of N trials of type B, and one block of 2N trials which alternate between type A and B (Jersild, 1927). In this design the key dependent measure is the difference in time between the mixed block and the sum of the A and B blocks. More contemporary behavioural

---

[*] This research was supported by a faculty research grant from Birkbeck, University of London, to Richard P. Cooper. We are grateful to Erik Altmann, Ines Jentzsch, André Vandierendonck and two anonymous reviewers for constructive comments on early drafts of this manuscript. Correspondence should be sent to Richard P. Cooper, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom. E-mail: R.Cooper@bbk.ac.uk

approaches to set-shifting use compound stimuli (e.g., number/letter pairs) where the part of the stimulus to which a response is required (number or letter) differs across trials and is indicated, for example, by the position of the stimuli on screen (Meiran, 1996; Rogers & Monsell, 1995). Responses are slower on switch trials than non-switch trials, even when switching occurs in a predictable fashion and ample time is allowed between trials. Broad neuropsychological support for the concept has been adduced from patients with frontal brain injury who show deficits in switching between response sets, as demonstrated by high rates of perseverative errors in tasks such as the Wisconsin Card Sorting Test described below (e.g., Milner, 1963; Nelson, 1976; Stuss et al., 2000), as well as group studies of frontal patients on standard laboratory switching tasks such as those described above (Shallice et al., 2008). Finally, neuroimaging studies of switching tasks suggest a common mechanism for task setting or maintenance localised within left dorsolateral frontal cortex (e.g., Brass et al., 2005; MacDonald et al., 2000). Similar behavioural, neuropsychological and neuroimaging evidence supports a domain-general process of response inhibition (e.g., Aron et al., 2003, 2004; Verbruggen & Logan, 2008). Beyond set-shifting and response inhibition, numerous other candidate executive functions have been proposed. These include memory related functions such as the active maintenance (e.g., Baddeley, 1996; Miller & Cohen, 2001; Smith & Jonides, 1997; Sakai, Rowe & Passingham, 2002), updating (e.g., Miyake et al., 2000; Schmeichel, 2007) and refreshing (Raye et al., 2007) the contents of short-term or working memory, as well as more general process control functions such as response selection (e.g., Bunge et al, 2002; Hegarty, Shah & Miyake, 2000; Szmalec, Vandierendonck & Kemps, 2005), selective attention (Smith & Jonides, 1999), monitoring/checking (Shallice et al., 2008), task management (Smith & Jonides, 1999), and dual-task coordination (e.g., Emerson et al., 1999; Logie et al., 2004).

While the range and number of executive functions that have been proposed suggests that significant progress has been made in understanding cognitive control, cataloguing and understanding the operation and interaction of executive control functions in specific tasks remains a significant challenge (Cooper, 2010). One of the more ambitious behavioural studies that has attempted to make progress on this issue is that of Miyake et al. (2000), who were primarily concerned with the functions of response inhibition, set-shifting and memory updating / monitoring (i.e., “monitoring and coding incoming information for relevance to the task at hand and then appropriately revising the items held in working memory”; Miyake et al., 2000, p. 57). The study used an individual differences methodology with over 130 participants, each of whom completed 14 different tasks. Nine of these tasks were argued to be “simple”, in that they were held primarily to tap just one of three specific executive functions. Thus three simple tasks were intended to tap response inhibition, three to tap set-shifting and three to tap memory updating / monitoring. Performance on these tasks was analysed using Confirmatory Factor Analysis (CFA) and found to be explicable in terms of three correlated but distinct factors. The remaining five tasks (the Wisconsin Card Sorting Test, the Tower of Hanoi, Random Number Generation, Operation Span and Dual Tasking) were more complex and considered potentially to involve multiple executive functions. Miyake et al. attempted to determine the involvement of their three factors in these complex tasks using the statistical technique of Structural Equation Modelling (SEM), with the three factors established from their earlier CFA corresponding to latent variables within a structural equation model. For example, performance on the Tower of Hanoi problem, which involves moving disks from one peg to another subject to various constraints in order to achieve a goal configuration (e.g., Simon, 1975), was argued to be related specifically to the response inhibition function.

Subsequent studies using similar sets of tasks have replicated key aspects of Miyake et al.'s (2000) results (Friedman et al, 2008; Friedman & Miyake, 2004). However, a serious limitation of the original study is its reliance on a two-stage process to infer the involvement of executive functions in complex tasks. As previously noted, CFA is first used to establish that the elementary executive tasks tap separable functions and then SEM is used to establish the involvement of those functions in the more complex executive tasks. One concern is that both of these methods involve an element of subjectivity in model selection (e.g., MacCallum & Austin, 2000). More critically, since the factors extracted from this first phase of analysis were used in a second phase to establish the extent to which performance on the complex executive tasks was a function of the individual factors, any statistical error in the first phase is likely to be compounded in the second phase. This is of particular concern because the factors established from the initial CFA were highly correlated.

To illustrate, consider the analyses conducted by Miyake et al. (2000) concerning the factors underlying the production of perseverative errors on the Wisconsin Card Sorting Test (Milner, 1963). The test requires participants to sort cards according to a changing criterion. Each card shows a number of coloured shapes, e.g., four red circles or two blue triangles. Participants sort the cards into piles that match "target" cards according to the number, type or colour of the shapes on the cards. After each sorting attempt they are told only if they were correct or incorrect. To perform well on the test, participants must infer the experimenter's sorting criterion, but the experimenter changes this criterion when the participant achieves a run of ten (or in some versions six) correct sorts. Perseverative errors arise when a participant continues to sort by an old, no longer valid, criterion, despite receiving negative feedback. Such errors may relate to poor set-shifting (i.e., failure to switch to a new sorting criterion), to inadequate response inhibition (i.e., failure to inhibit sorting to an old and now falsified rule), or even to failure to monitor for and integrate negative feedback. Miyake et al. considered five possible models relating their three factors (set-shifting, memory updating and response inhibition) to the number of perseverative errors produced by their participants. The model they endorsed consists of one path from the set-shifting factor, with a standardised path coefficient of 0.38 for the set-shifting factor. However, the fit of this model ($\chi^2(32) = 25.45$; IFI = 1.06)[1] is only slightly better than that of a model with paths from both set-shifting and response inhibition ($\chi^2(31) = 25.02$; IFI = 1.06), and the model with one path from response inhibition also produces a reasonable fit ($\chi^2(32) = 30.59$; IFI = 1.01), with a standardised path coefficient of 0.33 for the response inhibition factor. Critically, three different models produce qualitatively similar fits to the data. While goodness-of-fit statistics may be used to rank order these models, the models are based on factors inferred from the inter-correlations between performance on the nine simple tasks. Any statistical error in these factors could easily alter this rank ordering.

For example, each of Miyake et al.'s (2000) participant's score on the latent factor of response inhibition is inferred largely from their scores on the antisaccade, stop-signal and Stroop tasks. These are reported to have pairwise intercorrelations of $r = 0.19$ (antisaccade against stop-signal), $r = 0.20$ (antisaccade against Stroop interference) and $r = 0.18$ (stop-signal against Stroop interference). Yet in a near replication of this part of initial study, Friedman and Miyake (2004) report correlations of $r = 0.16$, $r = 0.23$ and $r = 0.15$ respectively. As one should expect from any empirical study, even with over 100 participants the reported values of r clearly include some statistical error. Consequently, the inferred

---

[1] IFI is Bollen's (1989) Incremental Fit Index. Miyake et al. (2000) suggest that a value of more than 0.95 indicates a good fit.

values of the latent variable corresponding to response inhibition (and, by extension, the latent variables corresponding to the other executive functions) for each subject can only be approximations. Thus, while Miyake et al.'s conclusion – that "shifting ability is a crucial component of perseverative errors in the WCST" (Miyake et al., p. 50) – is plausible, further investigation is warranted using complementary methodologies which circumvent the statistical limitations of the combined CFA / SEM approach and which therefore allow convergence onto a particular cognitive architecture.

The current pair of studies attempts to establish the role of elementary executive functions in complex tasks (including the WCST) using a different approach – one based on a dual-task methodology. Thus, the studies reported here document the effects of different secondary tasks (held to tap different executive functions, analogous to the elementary executive tasks of Miyake et al., 2000) on performance of different primary tasks (analogous to the complex executive tasks of Miyake et al., 2000). In all cases, the elementary executive tasks involve auditory input and vocal output, while the complex executive tasks involve visual input and manual output.

Historically, dual-task methodology has been used with great success, particularly in the study of attentional processes and response selection (see, Pashler, 1994, for a review) and in the fractionation of subprocesses involved in short-term memory (e.g., Baddeley & Hitch, 1974). The former case is exemplified by work on the Psychological Refractory Period (PRP), in which participants are required to perform concurrently two simple stimulus-response tasks where the stimuli are presented in quick succession. Typically, the response to the stimulus that is presented second is delayed, and this delay does not decrease appreciably when the interval between the two stimuli is decreased. This appears to imply some degree of serial processing in these simple tasks. Pashler (1994) argues that the data indicate that stimulus processing may occur in parallel, but that there is a bottleneck in the response selection stage such that only one response may be selected at a time, but this position is not universally accepted. Thus, Meyer and Kieras (1997) argue that PRP effects arise from conflicts at the motor level which can be minimised by strategic scheduling of task-related cognitive operations within a central processor, and provide a computational instantiation of this in terms of their Strategic Response Deferment model, which is itself embedded within the EPIC cognitive architecture.

In the latter case, the basic approach is to require participants to perform simultaneously two more complex tasks – a primary task and a secondary task – with non-overlapping input and output requirements. If performance on the primary task is impaired relative to a single-task condition or relative to a condition with another secondary task, then it is inferred that the original two tasks share some processing step or resource. Thus, dual-task interference is a key source of support for the various functional components – the phonological loop, the visuo-spatial sketch pad and the central executive – of the multi-component working-memory model of Baddeley and Hitch (1974).

Following in this latter tradition, a number of dual-tasking studies have sought to fractionate cognitive processes by impairing or degrading performance of Baddeley and Hitch's central executive. Two main lines of work may be identified. Some studies conceived of the central executive as a single non-decomposable system that may be affected in a graded or all-or-none way by a secondary task (e.g., Dunbar & Sussman, 1995). Using this approach, for example, Szmalec, Vandierendonck and Kemps (2005) argued that the response selection process held by Pashler (1994) to be an information processing bottleneck was a function of the central executive rather than of a visual or spatial working memory slave system (see also Szmalec & Vandierendonck, 2007). Other studies have considered dual-tasking to be one of the functions supported by the central executive (e.g., Baddeley, 1996; see also Baddeley et al., 1997; Miyake et al. 2000). On this latter approach, concurrent task interference would reflect the efficiency of an individual's dual-tasking ability, but could not be used to further decompose or fractionate the central executive.

This paper develops a third line of work. Baddeley et al. (1998) used a dual-task paradigm in which the primary task was random generation. This task requires that the participant generates from a given response set (e.g., digits) a series of responses that are as random as possible. It was shown that the degree of randomness was affected by the secondary task. When the secondary task was verbal fluency – i.e., to generate as many words as possible in a given period that start with a given letter – the degree of randomness in the primary task was lower than when the secondary task was either to count from a given digit or to recite the alphabet. Baddeley and colleagues interpret their results as reflecting the use of a common switching process by both random generation and category fluency – a position discussed further below. For our purposes, the Baddeley et al. (1998) study raises the possibility of using a dual-task methodology to decompose the central executive and thereby test Miyake et al.'s (2000) SEM analysis of the diversity of executive functioning.

The remainder of this paper reports two experimental studies in which two different primary tasks – random generation, held by Miyake et al. to draw on both memory updating / monitoring and response inhibition, and the Wisconsin Card Sorting Test, held by Miyake et al. to draw primarily on set-shifting – are each coupled with three secondary tasks. The secondary tasks – the digit switching task, the 2-back task and the go-no go task – are similarly held to each draw primarily upon different executive functions (set-shifting, memory updating / monitoring, and response inhibition, respectively). The differential interference patterns that result support a view of the central executive as comprising multiple distinct executive functions, with functions related to set-shifting and memory updating / monitoring playing different roles in the two primary tasks. However, the interference patterns differ from what would be expected given the CFA / SEM analyses of Miyake et al. (2000). The results are interpreted within the context of verbal process accounts of the primary tasks.

Experiment 1: Random Sequence Generation

Interest in random generation as a cognitive task stems from the fact that human participants are poor at producing "random" sequences. Their attempts typically exhibit standard biases. For example, if the task is to generate a sequence of random digits between 0 and 9 (inclusive), participants will typically produce repeat responses (e.g., "2" followed immediately by "2") at less than chance rates (e.g., Rapoport & Budescu, 1997; Towse, 1998). When items are generated at a fixed pace, increasing that pace typically results in greater redundancy (i.e., greater predictability of response on trial *n* given the responses on previous trials; Baddeley, 1966, but see Towse, 1998). Randomness is also influenced by factors such as the size of the response set, and whether that response set is externalised (as on a keyboard) or maintained internally (as in the generation of spoken responses) (Towse, 1998).

In reviewing the literature at the time, Miyake et al. (2000) note two ways in which executive functions might contribute to performance in random generation tasks. First, following Baddeley et al. (1998), they note the need to suppress stereotyped responses. This, they suggest plausibly, requires a response inhibition function. Second, following Jahanshahi et al. (1998), they note the likely involvement of a memory updating / monitoring process in keeping track of recent responses to check that they conform to the participant's concept of randomness. As Miyake and colleagues note, conceivably both functions are involved and conceivably the two functions have different effects on different measures of randomness. Consistent with this, Miyake et al. (2000) analysed their participants' responses on a random digit generation task according to several distinct measures of randomness as recommended by Towse and Neil (1998). One measure, equality of response usage, specifically assesses the relative frequency of the different possible responses – if a sequence is random then each response should be approximately equally frequent. Other measures, which we refer to collectively as measures of sequence stereotypy, assess the relative frequency of pairs or larger sequences of responses. Subsequent analyses by Miyake et al. using Structural Equation Modelling based on the individual difference factors established from participant performance on simple executive function tasks suggested that measures which could be interpreted as indicating biases towards or away from stereotyped sequences were associated with the executive function of response inhibition, while measures which could be interpreted as indicating biases related to (in)equality of response usage were associated with the executive function of memory updating.

However, as argued in the introduction, this SEM analysis is based on factors extracted from a preliminary Confirmatory Factor Analysis – factors which are subject to statistical error and which are themselves correlated. The dual-task paradigm offers the prospect of providing an alternative (and potentially complementary) methodology for establishing the contributions of putative executive functions to random generation behaviour because in contrast to individual differences methodology it allows the experimenter to manipulate the demands of the secondary tasks and measure the effects of such manipulations on measures of randomness. Experiment 1 therefore explores the effects of combining random generation with tasks that are assumed to tap different simple executive functions. In all cases the simple executive tasks involve auditorily presented stimuli and vocal responses. The random generation task involves visual presentation of the response set and manual selection from that set, as described below. The random generation task used here thus differs in modality from that used by Miyake et al. (2000) and Baddeley et al. (1998). The difference is motivated only by the requirement that input and output modalities of the primary and secondary tasks should not be shared. While it is clear that the response format has a substantive effect on the sequences generated in random generation tasks (e.g., Towse, 1998), a processing account of random generation (see below) suggests that the executive function requirements of the two task variants should be similar.

Two hypotheses follow from the structural equation models endorsed by Miyake et al. (2000). First, if response inhibition is required to suppress stereotyped response sequences, then such responses should be more frequent when random generation is coupled with a response inhibition task than with a task that primarily taps some other executive function. Second, if memory updating / monitoring is required to monitor response frequency, then measures of equality of response usage should be disturbed when random generation is coupled with a memory updating / monitoring task in comparison to when it is coupled with a task that primarily taps some other executive function.

Alternative hypotheses, however, may be derived from process-oriented models of random generation (e.g., the mathematical model of Rapoport & Budescu, 1997, and the verbal model of Baddeley et al., 1998). Within those models, random generation involves several stages: (a) generation of a candidate response, which is based on applying an existing response schema (e.g., increment the previous response by 2, or decrement it by 1); (b) checking that the candidate response is sufficiently random with respect to recent responses; (c) if so, producing the response; but (d) if not, inhibiting the candidate response, switching to a new response schema and returning to step (a). Based on this verbal process model, response inhibition only comes into play when the checking process – a process that requires monitoring of the record of recent behaviour – detects that the candidate response should be inhibited. If this checking process were to be impaired, e.g., by concurrent performance of a

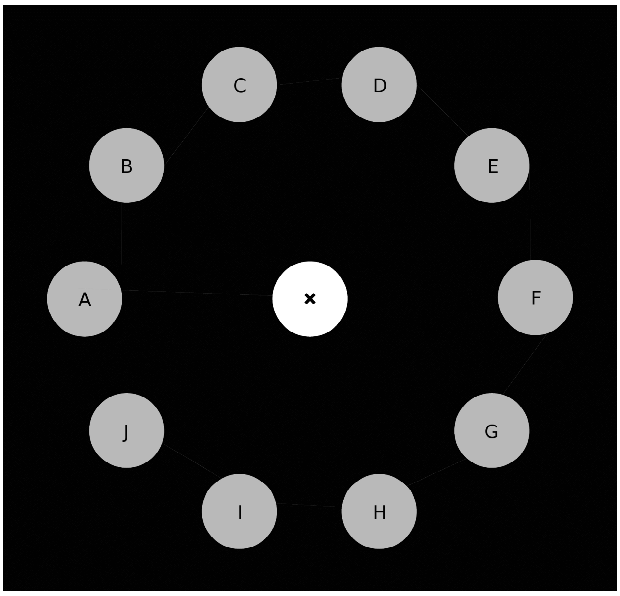

Figure 1: The Random Sequence Generation Task display (Experiment 1). On each trial, participants were required to use the mouse to select one letter at random. The mouse was automatically repositioned to the centre circle after each trial. Peripheral disks were normally green but flashed white when selected.

secondary task that also shares monitoring requirements, then one would expect less randomness in the generated sequence. While this may be reflected in a decrease in equality of response usage, because regulating equality of response usage requires some record of the relative frequency of responses, such a decrease cannot be guaranteed as highly regular strategies (e.g., counting) can still produce low scores on this measure. Moreover, given the process model any change in equality of response usage would not be diagnostic of response inhibition because that dependent measure is also a function of the veracity of monitoring and checking process. A clearer prediction relates to measures of sequence stereotypy, which on this account should be markedly influenced by concurrent performance of a task that demands frequent memory updating and monitoring.

Method

Participants. Thirty-six participants (15 male, 21 female; average age 26 years 11 months) took part in the experiment. Participants were recruited from the university's volunteer participant panel, which includes undergraduate and postgraduate students, as well as interested laypeople. All participants were paid £5 for their effort.

Design. Participants completed the primary random sequence generation task four times: first as a single task, and then with each of three auditory-vocal tasks: the digit-switching task, the 2-back task, and the go-no go task. The order in which the auditory-vocal tasks were administered was fully counterbalanced. The experiment therefore employed a within-subjects design where the independent variable, secondary task, had four levels: none, digit-switching, 2-back and go-no go. Responses on the random generation task were self-paced. The key dependent variables were therefore response rate and measures of randomness in the random sequence generation task (in each of the four conditions), and accuracy for each of the auditory-vocal tasks.

The Random Sequence Generation Task. In each block of the random sequence generation task, participants were required to generate a sequence of 100 letters (in the range

'A' to 'J') using a computer mouse to select each letter from a clock-face type display as shown in Figure 1. The mouse was initially positioned in a disk at the centre of the screen. To select a letter, participants had to move the mouse to one of the ten locations arranged in a circle around the central disk, and click. After selecting a letter the mouse automatically returned to the central position in preparation for the next selection. Participants were instructed to make their selections as "random as possible". Performance was self-paced but if no selection was made within five seconds a null response was recorded and the screen was reset. In reality, this only occurred during practice trials.

The Auditory-Vocal Tasks. In the digit-switching task, modelled on that of Monsell (2003), participants heard a series of digits (either 1, 2, 3, 4, 6, 7, 8 or 9) at a rate of one digit every 2.5 seconds. Participants were initially required to respond "high" if the digit was greater than 5 and "low" otherwise. After 4 trials of this form, a tone was presented indicating that the required responses had changed and that participants were to respond "odd" if the digit was 1, 3, 7 or 9 and "even" otherwise. Tones were presented after every 4 trials throughout the task to indicate that a switch between the two response sets was required. The dependent variable was accuracy, i.e., the number of correct trials divided by the total number of trials.

In the 2-back task, participants heard a series of digits (in the range 1 to 9, with each digit being equally likely) at a rate of one digit every 2.0 seconds. They were required to respond vocally with "yes" if the current digit was the same as that two trials before. The dependent measures were accuracy (the number of hits and correct rejections divided by the total number of trials) and sensitivity (d′, calculated according to standard principles of signal detection theory).

In the go-no go task, participants heard a series of either single beeps (duration 100 msec) or double beeps (duration 200 msec), with the interval between beep onsets varying randomly from 1.5 seconds to 2.5 seconds. They were required to generate a vocal response (the word "yes") as quickly as possible after each single beep, but to produce no response after a double beep. Single beeps occurred on 5 out of every 6 trials. As in the 2-back task, the dependent measures were accuracy and sensitivity.

Procedure. Each participant completed four blocks of the visual-manual random sequence generation task. Prior to the first block, participants were familiarised with the random generation task through eight practice trials. They then completed the first block (i.e., 100 trials) of the generation task. During the second, third and fourth blocks, participants were required to perform the random sequence generation task while simultaneously completing one of the three auditory-vocal tasks. The order of completion of the auditory-vocal tasks was counter-balanced across participants, with all participants completing each task. In all cases, practice on the auditory-vocal tasks was given prior to performance of the dual-task condition, and the auditory-vocal tasks were continued for as many trials as needed for completion of one block of 100 responses on the random generation task. Upon completion of the fourth and final block, participants were thanked, paid, and debriefed. The complete experimental session lasted approximately 20 minutes.

So as to avoid machine-related interference between concurrent tasks, one PC was used to administer the random generation task and a second was used to administer the auditory-vocal tasks. Participants sat at a comfortable distance in front of the monitor attached to the PC that administered the random generation task and interacted with that PC through a mouse controlled by their preferred hand. In blocks 2, 3 and 4 they wore noise-reducing headphones through which auditory stimuli were presented and directed their vocal responses to a microphone positioned in front of the monitor. The experimenter sat behind the participant and manually recorded all responses to each auditory-vocal task.

Dep. Var.	**Exp. Val.**	**Control**	**Digit-Switching**	**2-Back**	**Go-No Go**
RT (ms)	–	766 (174)	726 (229)	708* (173)	669* (166)
R	0.000	0.962 (0.567)	2.048* (1.912)	1.979* (1.620)	1.196 (1.017)
RNG	0.000	0.300 (0.068)	0.410* (0.151)	0.461* (0.166)	0.388* (0.158)
RR	0.100	0.014 (0.024)	0.004 (0.010)	0.002* (0.004)	0.005* (0.014)
AA	0.200	0.259 (0.143)	0.328 (0.252)	0.424* (0.279)	0.334* (0.260)
OA	0.100	0.131 (0.070)	0.136 (0.085)	0.097 (0.102)	0.130 (0.096)
TPI	0.598	0.440 (0.118)	0.363* (0.150)	0.267* (0.161)	0.349* (0.159)

Table 1: Actual and expected mean values of each dependent measure of the random generation task and for each condition (Experiment 1). Standard deviations are shown in brackets. * indicates significantly different from the control condition, at $p < 0.05$ (corrected). RT = Response Time; R = Redundancy score; RNG = Evans' (1978) Random Number Generation score; RR = Repeat Response score; AA = Adjacent Associates score; OA = Opposite Associates score; TPI = Turning Point Index.

Results

Descriptive Statistics and Preliminary Analyses. In analysing random generation performance it is common to pre-process response sequences to generate a set of indices of randomness for each sequence (see, e.g., Towse & Neil, 1998). These indices measure different aspects of randomness, and can be treated as the task's dependent measures. This approach was followed here. In addition to mean response time (RT), six indices were computed for each block of responses:[2] R, the redundancy score, which measures the relative use of each response option; RNG, the random number generation score of Evans (1978), which measures the relative use of each pair of responses (e.g., response F followed by response A, or response B followed by response C); RR, the proportion of responses that are repeats of the previous response; AA, the proportion of responses that are spatially adjacent to the previous response (e.g., responding F followed by G or E, or A followed by J or B); OA, the proportion of responses that are spatially opposite to the previous response (e.g., responding F followed by A or H followed by C); and TPI, the turning point index – a measure of the relative number of times responding switches from a clockwise to a counter-clockwise pattern and vice versa.

For each measure of randomness, there is an expected value corresponding to true randomness. R and RNG are entirely independent measures, with deviations from the expected values on R and RNG reflecting biases away from equality of response usage and equality of bigram usage respectively. RR, AA and OA each measure specific bigram biases. Values of RR above 0.1, AA above 0.2 and OA above 0.1 indicate a bias towards repeat, adjacent and opposite response pairs respectively. These three measures are not independent. Rather, they provide a more fine-grained decomposition of the RNG measure. Lastly, TPI is a

---

[2] See Towse and Neil (1998) for formulae for the less obvious measures (R, RNG and TPI). Note that in comparison to Towse and Neil (1998), TPI values reported here are divided by 100 so that, like most other measures of randomness used here, they lie within the 0 to 1 range.

trigram measure. Low values of TPI indicate a tendency to move around the clock-face in one direction (either clockwise or counter-clockwise).

Mean scores for each dependent measure in each condition are shown in Table 1. The table also shows the expected value of each measure for a sequence where successive responses are independent (i.e., random).

In order to explore the independence, or otherwise, of the six measures of randomness, correlation coefficients were calculated for each pair of such measures in the control condition (i.e., when random generation was performed in isolation). No association was found between R and RNG ($r = 0.001$), consistent with the claim that R and RNG measure distinct aspects of randomness. (Indeed, R is not significantly correlated with any other measure of randomness.) Some significant correlations were found between the various bi-gram measures, specifically between RNG and AA ($r = 0.433$, $p = 0.008$, two-tailed) and between AA and OA ($r = -0.628$, $p < 0.001$, two-tailed). These reflect the facts that a) AA and OA measure biases to the production of specific bigrams which, together, make up the RNG score and b) a bias towards one type of bi-gram (e.g. opposite associates) necessarily results in a bias away from other types of bi-grams (such as adjacent associates). The lack of any significant correlations involving the RR score conceivably reflects the limited range of this measure due to a floor effect, rather than the lack of association.

Participants performed at near ceiling levels on each of the auditory-vocal tasks. Mean response times were significantly faster in the go-no go task (624.6 msec) than the 2-back task (993.1 msec; $t(27) = 12.47$, $p < 0.001$), which in turn were significantly faster than in the digit-switching task (1173.3 msec; $t(28) = 5.00$, $p < 0.001$).[3] Mean accuracy was significantly less in the 2-back task (0.947) than in the digit-switching task (0.981; W(34) = 70.0, $p < 0.001$) and than in the go-no go task (0.990; W(34) = 70.0, $p < 0.001$). Signal processing analysis on those tasks for which it is appropriate showed good sensitivity (2-back: $d' = 1.301$; go-no go: $d' = 3.150$). Further analysis of the auditory-vocal task data is reported in the appendix.

Effects of Secondary Tasks. Our primary interest is in the differential effects of the three secondary tasks. Thus, data from the control condition were ignored in the initial analysis of secondary task effects and one-way within-subjects ANOVAs were conducted on the data from the three experimental conditions. As shown in Table 2, there were significant effects of task on all dependent measures except RR, which in any case was near floor.

In order to consider the possible differential effects of the auditory-vocal tasks on each dependent measure, the control condition was used to establish baselines for each dependent measure, allowing us to determine whether concurrent performance of the different auditory-visual secondary tasks resulted in more or less extreme deviations from the baseline. Figure 2 shows how the various dependent measures pattern for the three dual-task conditions when all dependent measures in the experimental conditions are converted to z-scores with respect to their distribution in the control condition. Informally, it appears that R is affected greatly by concurrent performance of either the digit-switching task or the 2-back task, but only slightly by concurrent performance of the go-no go task. In contrast, all other measures of randomness appear to be most affected by concurrent performance of the 2-back task, with the digit-switching and go-no go tasks having similar, lesser, effects.

The reliability of the above observations was tested with a number of post-hoc t-tests using a Bonferroni correction to guard against type 1 errors. Given that there are three

---

[3] Some RT data for the secondary tasks was lost due to lack of sensitivity of the voice-key. The analyses reported here are thus based on a subset of participants for which RT data was available. This accounts for the variation in reported degrees of freedom across comparisons.

	F ratio	Probability	Effect Size
RT	$F(1.610, 56.355) = 3.669$	$p = 0.041$	$\eta^2 = 0.095$
R	$F(2, 70) = 12.401$	$p < 0.001$	$\eta^2 = 0.262$
RNG	$F(2, 70) = 9.458$	$p < 0.001$	$\eta^2 = 0.213$
RR	$F(1.687, 59.040) = 1.764$	$p = 0.179$	$\eta^2 = 0.048$
AA	$F(1.467, 51.360) = 8.224$	$p = 0.001$	$\eta^2 = 0.190$
OA	$F(1.696, 59.366) = 13.302$	$p < 0.001$	$\eta^2 = 0.275$
TPI	$F(2, 70) = 11.933$	$p < 0.001$	$\eta^2 = 0.254$

Table 2: ANOVA results showing effects of task for each dependent measure from Experiment 1. (Note: Greenhouse-Geisser corrections to degrees of freedom have been reported where appropriate to correct for violations of sphericity.) RT = Response Time; R = Redundancy score; RNG = Evans' (1978) Random Number Generation score; RR = Repeat Response score; AA = Adjacent Associates score; OA = Opposite Associates score; TPI = Turning Point Index.

auditory-vocal tasks (and hence three pair-wise comparisons) and the post-hoc comparison requires bi-directional hypotheses, we adopt a critical value of $0.05 \div 6 = 0.008$. At this highly conservative level, participants generated letters significantly more rapidly when simultaneously performing the go-no go task than the digit-switching task ($t(35) = 2.95$, $p = 0.003$). Responses were also generated more rapidly in the go-no go task than in the 2-back task, though the difference was not significant at the adjusted level ($t(35) = 2.17$, $p = 0.019$). The sequences generated with the go-no go task were also characterised by significantly lower R scores (versus digit-switching: $t(35) = 4.40$, $p < 0.001$; versus 2-back: $t(35) = 4.34$, $p < 0.001$). In both cases (RT and R), the difference between the digit-switching task and the 2-back task did not approach significance (RT: $t(35) = 0.68$, *n.s.*; R: $t(35) = 0.36$, *n.s.*). To summarise, while responses in all dual-task conditions were generated more quickly than in the control condition (presumably due to a practice effect), responses were generated at a significantly slower rate in the digit-switching and 2-back conditions than in the go-no go condition. A similar effect occurs with R scores, with them being similarly (and significantly) affected when random generation was combined with the digit-switching and with 2-back tasks, but less affected when the secondary task was the go-no go task.

For the five other dependent measures (RNG, RR, AA, OA and TPI), Figure 2 shows that concurrent performance of the 2-back task had the greatest effect (in comparison to the control condition), with the digit-switching and go-no go tasks producing similar more moderate effects. Thus, RNG score was significantly greater in the 2-back condition than the digit-switching condition ($t(35) = 2.82$, $p = 0.004$) or the go-no go condition ($t(35) = 3.78$, $p < 0.001$), but did not differ reliably between the digit-switching and go-no go conditions ($t(35) = 1.57$, *n.s.*). This pattern was repeated for AA (2-back versus digit-switching: $t(35) = 3.39$, $p = 0.001$; 2-back versus go-no go: $t(35) = 2.83$, $p = 0.004$; digit-switching versus go-no go: $t(35) = 0.37$, *n.s.*), OA (2-back versus digit-switching: $t(35) = 4.87$, $p < 0.001$; 2-back versus go-no go: $t(35) = 3.46$; $p = 0.001$; digit-switching versus go-no go: $t(35) = 0.87$, *n.s.*), and TPI (2-back versus digit-switching: $t(35) = 4.85$, $p < 0.001$; 2-back versus go-no go: $t(35) = 3.56$; $p = 0.001$; digit-switching versus go-no go: $t(35) = 0.64$, *n.s.*).[4]

---

[4] RR showed the same pattern of effects, though for this dependent measure the effects were not statistically significant: 2-back versus digit-switching: $t(35) = 1.65$, $p = 0.053$; 2-back versus go-no go: $t(35) = 1.60$, $p = 0.059$; digit-switching versus go-no go: $t(35) = 0.54$, $p = 0.295$. The failure of these tests to reach statistical significance is, we suggest, the result of a floor effect in the RR measure.

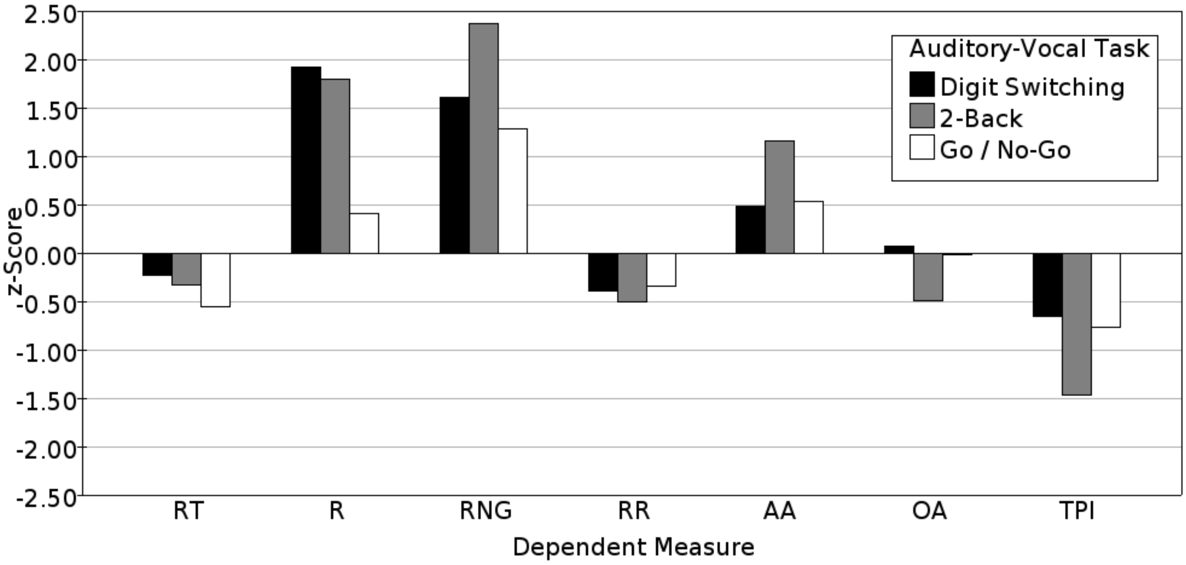

Figure 2: Effects of auditory-vocal task on dependent measures of the random generation task (Experiment 1). All scores are z-scores calculated with respect to the mean and standard deviation of the corresponding variable in the control condition. RT = Response Time; R = Redundancy score; RNG = Evans' (1978) Random Number Generation score; RR = Repeat Response score; AA = Adjacent Associates score; OA = Opposite Associates score; TPI = Turning Point Index. Note that RR, AA and OA are not independent. Rather, they offer a finer-grained analysis of RNG.

The effect of combining random generation with the auditory-vocal tasks may be summarised as follows: a) for all measures apart from response time, the go-no go task produced the least interference; b) the digit-switching and 2-back tasks had similar effects on R scores, resulting in significant increases in both scores (taking them further from their observed values in the control condition); however, c) on all measures of bi-gram associations (RNG, RR, AA and OA) and tri-gram associations (TPI), combining random generation with the 2-back task was found to consistently result in more interference (in the form of more extreme divergence from the control conditions) than combining it with either the digit-switching or the go-no go task. Moreover the digit-switching and go-no go tasks had statistically equivalent effects on these measures. Thus, combining random generation with the 2-back task yielded a higher RNG score, fewer repeat responses, more adjacent responses, fewer opposite associates, and a lower turning point index than when either the digit-switching or the go-no go task was combined with random generation. The digit-switching and go-no go tasks produced similar mean values for all five indices, which in all five cases were closer to the control means.

Discussion

The results of Experiment 1 suggest that, as hypothesised, the executive or control requirements of a complex task may be fractionated by combining that task with different secondary tasks that have different executive requirements, particular when the complex task produces multiple dependent measures that are potentially differentially dependent on different executive functions. Thus, if we assume that the digit-switching task primarily taps the executive function of set-shifting, the 2-back task primarily taps the executive function of memory maintenance, monitoring and updating, and the go-no go task primarily taps the executive function of response inhibition – assumptions to which we return both below and in the General Discussion – then the cognitive processes which are reflected in the R score (equality of response usage) appear to be dependent more on set-shifting and memory maintenance, monitoring and updating than on response inhibition, while the processes reflected in the measures of sequence stereotypy appear to be more dependent on memory maintenance, monitoring and updating than on either set-shifting or response inhibition.

This interpretation, in terms of the differential executive requirements of the auditory-vocal tasks, has clear implications for discriminating between the hypotheses based on the structural equation model of Miyake et al. (2000) and the more process-oriented accounts of Rapoport and Budescu (1997) and Baddeley et al. (1998) that served in part to motivate Experiment 1. Consider first the structural equation model endorsed by Miyake et al. Recall that on the basis of this model it was hypothesised that stereotyped response sequences would be more frequent when random generation was coupled with a response inhibition task (the go-no go task) than with a task tapping some other executive function. This was not found. Contrary to the model, no measures of response stereotypy (i.e., bi-gram or tri-gram association scores) were affected more by response inhibition task than by the other secondary tasks. This result cannot be taken as strong evidence against the Miyake et al. position on measures of prepotent associates, however, as it is possible that the response inhibition task was insufficiently challenging. In the go-no go task, inhibition of a response was required on only one in every six trials, accuracy was very high, and response times were generally fast.

More problematic for the position of Miyake et al. are the results related to equality of response usage. It was also hypothesised on the basis of Miyake et al.'s favoured structural equation model that this measure would be most disrupted when random generation was coupled with a memory updating / monitoring task. While equality of response usage was disrupted by such a secondary task, it was equally disrupted by the set-shifting secondary task. Paralleling the putative argument related to response inhibition, one might suggest that the set-shifting requirements of the digit-switching task are in some sense greater than the memory updating / monitoring requirements of the 2-back task, and thus the digit-switching task impacted more on equality of response usage than would have been the case if the two secondary tasks were balanced for their executive function difficulty. Note though that while response times in the digit-shifting task were significantly slower than in the 2-back task, accuracy was significantly higher. So an account in terms of differential difficulty of the executive requirements of the secondary tasks is not satisfactory. Thus, the predictions derived from the structural equation model found by Miyake et al. (2000) to produce the best fit to the data are not supported by the dual-task results.

The verbal process model fares better. On the basis of this account it was hypothesised that measures of randomness related to sequence stereotypy would be impaired when random generation was combined with a secondary memory updating / monitoring task. This is precisely what was found, with all five sequence stereotypy measures (RNG, RR, AA, OA and TPI) being most strongly affected (i.e., deviating more, either positively or negative) relative to control when the random generation task was coupled with the 2-back task. Thus, we interpret our results as supporting the claim that the 2-back task impairs memory monitoring / updating during the random generation task, leading to an impairment in the detection of patterns.

RNG is a composite measure that reflects RR, AA, OA and other possible bi-gram subsequences. As such, it should not be surprising that these measures pattern together. However the three submeasures provide some insight into the way in which generation changes when coupled with the 2-back task: The number of adjacent associates increases at the expense of opposite associates and repeat responses. On the verbal process model, this suggests that for this version of random generation the response schema *move to an adjacent position* in the absence of effective memory monitoring / updating, is more likely to be selected (or at least not rejected) than the schemas *move to the opposite position* or *repeat the previous response*.

Neither the structural nor the verbal process accounts provide an immediate explanation of the other main feature of the data, namely that R, equality of response usage, was equally and significantly adversely affected when random generation was coupled with either the digit-switching or the 2-back task, but barely affected when random generation was coupled with the go-no go task. We have not considered how response frequencies might be balanced during random generation, but it would seem that there are two possibilities: either that a process of monitoring detects frequent candidate responses before they are produced and such responses are subsequently inhibited, or that monitoring detects infrequent responses and such responses are deliberately selected. Both possibilities imply that R is likely to be adversely affected if monitoring is impeded, as we assume it is when simultaneously completing the 2-back task. Thus, both accounts are consistent with the effect of the 2-back condition on R. However, the fact that R is significantly adversely affected by digit-switching suggests that this deliberate selection involves both memory and set-shifting functions.[5]

A further point worthy of comment concerns the number of repeat responses (the RR score) across the various conditions. The strong tendency to produce fewer repeat responses than would be expected by chance is frequently reported in random generation tasks (e.g., Baddeley et al., 1998; Rapoport & Budescu, 1997; Towse & Valentine, 1997), and it was apparent in our control condition (mean RR = 0.014, when chance would dictate RR = 0.100). However, the tendency to avoid such responses was even stronger in all dual-task conditions. This supports an account in which responses are automatically inhibited after they are generated, and in which this inhibition needs to be actively overcome in order to produce a repeat response (Baddeley et al., 1998; Towse & Valentine, 1997), rather than one in which repeat responses are rejected as part of the monitoring process.[6]

There is an alternative explanation for the results of Experiment 1 that we have not considered. The interference pattern could result from a single factor (task difficulty or cognitive load) if the various dependent measures are differentially dependent on that factor. In particular, the interference pattern could result if a) the 2-back task is harder than the digit-switching task which is in turn harder than the go-no go task, and b) the R score is heavily dependent on central resources and so is affected substantially by tasks of even moderate difficulty (i.e., both the 2-back and digit-switching tasks), but c) the various bi-gram scores are only mildly dependent on central resources and so are only substantially affected when random generation is combined with the most difficult task – the 2-back task. Experiment 2 is designed in part to rule out this possibility by considering interference effects on a further

---

[5] An alternative possibility is that set-shifting is itself dependent on memory processes (Altmann & Gray, 2008). This may account for similar effects of the digit-switching and 2-back tasks on the R score. It would suggest, however, that all dependent measures should pattern similarly with respect to these secondary tasks. This is not what was observed.

[6] It is possible that lower than chance RR scores contribute to higher than chance AA scores. Adjacent associate responses may be due in part to inhibition of a repeat response and the resultant selection of a response that is spatially close but distinct from the repeat response. Such a response would be an AA response. We are grateful to Erik Altmann for raising this possibility. However, while such a mechanism coupled with increased inhibition of repeat responses may contribute to the increase in AA scores in the dual task conditions, it cannot be the only factor, as the increase in AA scores is far greater than the decrease in RR scores.

primary task that is frequently held to require the executive function of set-shifting, the Wisconsin Card Sorting Test (e.g., Miyake et al., 2000). A second purpose of Experiment 2 is to further demonstrate the utility of the dual-task methodology for fractionating executive functions by applying it with a different primary task.

Experiment 2: Wisconsin Card Sorting Test

The Wisconsin Card Sorting Test (WCST), as described in the introduction, requires participants to use feedback (positive or negative) to sort cards according to a criterion that may change during the test without notice. Traditionally the test was considered to assess frontal function, with frontal patients typically producing high rates of perseverative error responses following a change by the experimenter in the sorting criterion. But the test is multi-componential (Reverberi et al., 2005; Stuss et al., 2000). In order to avoid perseverative responses, participants must respond appropriately given negative feedback. They must deduce that the previous criterion no longer applies, they must inhibit any response based on that criterion, they must deduce a new plausible criterion, and they must switch to this new criterion.

Miyake et al. (2000) endorse the view that perseverative errors on the task result from a failure by the participant to switch the sorting criterion in response to negative feedback. This is supported by their structural equation modelling, though as we have argued this support is not strong. Dunbar and Sussman (1995) argue for a related account of the origins of perseverative errors. They used a dual-task paradigm in which WCST was paired with a range of tasks varying in working memory requirements and found that perseverative errors were more frequent when the secondary task involved the phonological loop (e.g., when participants were required to repeat a nonsense syllable) than when it did not (e.g., when it involved addition or tone detection). Dunbar and Sussman interpret their results as indicating that perseverative errors arise from a failure to maintain information in the phonological loop. It is unclear what information must be maintained in the phonological loop to avoid such errors, but presumably it involves verbalisation of the sorting criterion hypothesised by the participant. In fact, more recent work suggests that verbal mediation may facilitate task switching (e.g., Baddeley, Chincotta & Adlam, 2001; Kirkham, Cruess & Diamond, 2003; see Cragg & Nation, 2010, for a review). The accounts can therefore be seen to be related.

Two alternative accounts of WCST errors come from the neuropsychological literature. Stuss et al. (2000) had patients with frontal brain lesions perform the WCST under several conditions, manipulating the degree of instructional support given about possible rules and rule changes. Different subgroups of frontal patients were prone to different types of error in the different conditions. Thus, patients with dorsolateral or superior medial frontal lesions were especially prone to perseverative errors, but superior medial patients tended to make fewer set loss errors – errors in which a string of correct responses that had obtained positive feedback is followed without obvious reason by an erroneous response – than patients with left or right unilateral dorsolateral lesions, while patients with inferior medial frontal lesions were specifically prone to set loss errors and relatively immune to perseverative errors. The implication of these results is that different regions of prefrontal cortex support different functions required for successful completion of the WCST.

A second neuropsychological study, that of Reverberi et al. (2005), sheds further light on the possible deficits underlying patient errors. Reverberi et al. were specifically concerned with the rule-induction component of WCST and its possible role in the avoidance of perseverative errors. They used a related task – the Brixton spatial rule-attainment task (Burgess & Shallice, 1996) – which requires that participants extrapolate a temporally evolving spatially defined sequence based on a series of exemplars. As in the WCST, the rule underlying the sequence defined by the exemplars changed without warning. For some patients, poor performance on the test could be accounted for by poor working memory. This was not the case for left lateral patients, who appeared to be specifically impaired on inducing possible rules based on the known exemplars.

Together, these neuropsychological studies suggest that the processes involved in the generation and regulation of WCST performance may be fractionated, with separable processes responsible for the prevention of perseverative errors, the prevention of set loss errors, and the generation of candidate rules (i.e., sort by number, colour, or shape). This suggests a further set of hypotheses, namely that completion of the WCST with different secondary tasks that interfere with these subprocesses will bias participants towards different types of errors. More specifically, the view considered above that perseverative errors result from a failure to switch the sorting criterion implies that perseverative errors should increase when WCST is paired with a secondary task that loads on set-shifting (e.g., the digit-switching task) relative to secondary tasks of similar difficulty that do not (e.g., the 2-back task or the go-no go task).

Method

Participants. Forty-eight participants (19 male, 29 female; average age 26 years 9 months) took part in the experiment. Participants were recruited from the university's volunteer participant panel, which includes undergraduate and postgraduate students, as well as interested laypeople. All participants were paid £5 for their effort. No participants completed both Experiments 1 and 2.

Design. As in Experiment 1, participants completed the primary task – which in this case was a computer-administered 64-card version of the Wisconsin Card Sorting Test – four times: first as a single task, and then in three dual-task conditions with each of three auditory-vocal tasks. The experiment therefore employed a within-subjects design where the independent variable (secondary task) had four levels: none, digit-switching, 2-back and go-no go. The key dependent variables were the standard ones for the WCST, namely number of categories achieved and the number and type of errors made in each condition. Accuracy on the secondary tasks was also recorded. The auditory-vocal tasks were exactly as in Experiment 1. Thus, the only difference between Experiment 1 and Experiment 2 was in the choice of visual-manual task.

The Wisconsin Card Sorting Test. The WCST was administered according to the "64A" procedure of Stuss et al (2000). Thus, 64 cards were used in each administration, the correct sorting category changed when participants achieved a sequence of 10 consecutive correct responses, and participants were not informed of such changes. The initial correct sorting category (colour, number or shape) was varied over blocks.

Procedure. Each participant completed four blocks of the Wisconsin Card Sorting Test.[7] Prior to the first block, participants were familiarised with the sorting aspect of the task through four practice trials. They then completed the first block (i.e., 64 cards) of the WCST.

---

[7] The use of WCST in a within-subjects design requires some justification. The task's novelty, and in particular the unannounced sorting criterion changes, are commonly considered to be part of the difficulty of the task. However, the task has been used before in a within-subjects context (Stuss, 2000; see also Reverberi et al, 2005). Its use here is justified because a) all participants have a chance to learn the task rules during the control condition, b) order of tasks in the experimental conditions is counterbalanced, and c) we are not interested in the processes concerned with learning that the sorting criterion changes – merely in the behaviour following a change. Note also that even in the standard administration participants are only naïve to the first criterion change – participants then negotiate up to five more criterion changes during the task.

Dependent Variable	Control		Digit-Switching		2-Back		Go-No Go	
Categories	3.77	(1.10)	2.00*	(1.52)	2.67*	(1.46)	3.21*	(1.39)
Correct	49.21	(6.74)	40.13*	(9.95)	43.25*	(9.85)	46.54*	(7.72)
TFC	14.08	(7.22)	33.27*	(21.56)	23.79*	(19.49)	21.00*	(15.78)
CPE	10.02	(4.81)	16.96*	(8.15)	15.42*	(6.34)	11.83*	(6.03)
NPE	4.77	(2.45)	6.92*	(4.01)	5.33	(4.97)	5.63	(3.05)
PP	0.69	(0.09)	0.71	(0.11)	0.77*	(0.13)	0.68	(0.13)

Table 3: Mean (and standard deviations) of dependent measures derived from WCST performance in the four conditions of Experiment 2. TFC = Trials to First Category; CPE = Classical Perseverative Errors; NPE = Non-Perseverative Errors; PP = Perseveration Proportion. * indicates values significantly different ($p < 0.05$, corrected) from the control condition.

During the second, third and fourth blocks, participants were required to perform the WCST while simultaneously completing one of three auditory-vocal tasks. The order of completion of the auditory-vocal tasks was counter-balanced across participants, with all participants completing each task. In all cases, practice on the auditory-vocal tasks was given prior to performance of the dual-task condition, and the auditory-vocal tasks were continued for as many trials as needed for completion of one block of the WCST. Upon completion of the fourth and final block, participants were thanked, paid, and debriefed. The experimental session lasted approximately 30 minutes.

So as to avoid machine-related interference between concurrent tasks, one PC was used to administer the WCST and a second to administer the auditory-vocal tasks. Participants sat at a comfortable distance in front of the monitor attached to the PC that administered the WCST and interacted with that PC through a mouse operated by their preferred hand. In blocks 2, 3 and 4 they wore noise-reducing headphones through which auditory stimuli were presented and they directed their vocal responses to a nearby microphone. The experimenter sat behind the participant and recorded all responses to each auditory-vocal task.

Results

Descriptive Statistics and Preliminary Analyses. Participant performance on the Wisconsin Card Sorting Test was scored according to standard procedures (Heaton, 1981), yielding scores for total number of categories achieved, number of cards correctly sorted, number of trials to achieve the first category (TFC), number of classical perseverative errors (CPE) and number of non-perseverative errors (NPE). Since the number of errors varied across condition, an additional dependent measure was calculated – the proportion of errors that were perseverative (PP). This was computed as the ratio of the number of classical perseverative errors divided by the total number of errors. Means for all dependent measures in each condition are shown in Table 3.

Participants performed best in the control condition, achieving more correct sorts, more categories, taking fewer cards to obtain the first category, and making fewer errors of each type, despite them being naïve to the task. With the exception of the PP measure, all dependent measures were strongly inter-correlated (e.g., categories versus correct: $r = 0.899$, $p < 0.001$; categories versus TFC: $r = -0.665$, $p < 0.001$; categories versus CPE: $r = -0.965$, $p < 0.001$; categories versus NPE: $r = -0.857$, $p < 0.001$; but categories versus PP: $r = 0.207$, $p = 0.159$). This is to be expected. For a participant to achieve many categories s/he must sort most cards correctly, which in turn means that s/he must make relatively few errors (perseverative or non-perseverative). S/he must also achieve the first category relatively quickly.

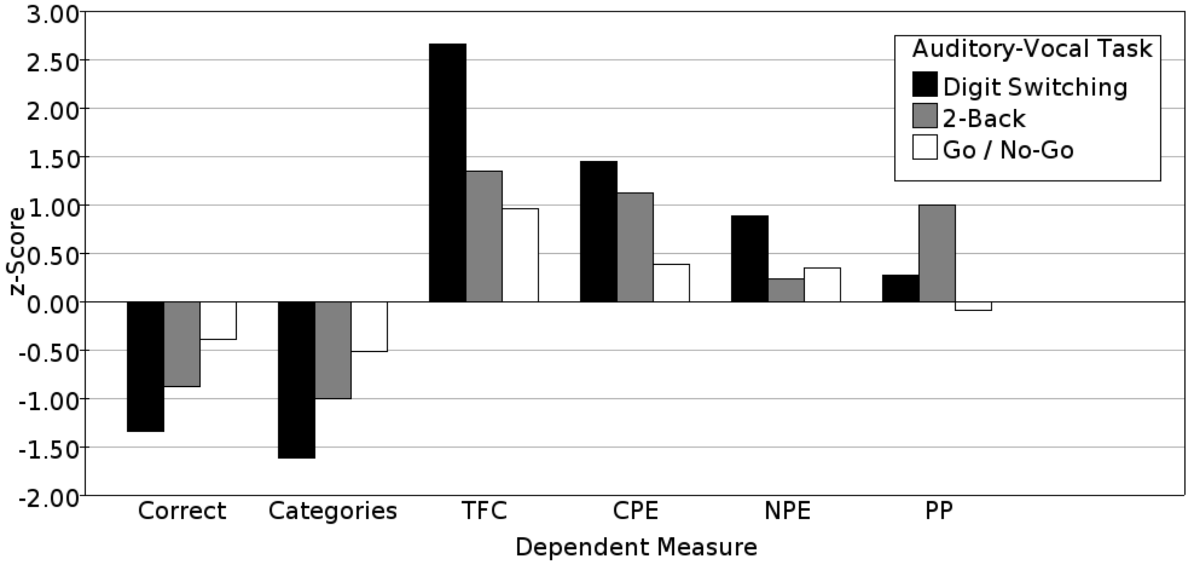

Figure 3: Deviations from mean control values for each dependent measure of the WCST across the three dual-task conditions of Experiment 2. As in Figure 2, all scores are z-scores calculated with respect to the mean and standard deviation of the corresponding variable in the control condition. TFC = Trials to First Category; CPE = Classical Perseverative Errors; NPE = Non-Perseverative Errors; PP = Perseveration Proportion.

As hypothesised, performance was poorest in the dual-task digit-switching condition, though the effect of this dual-task was not restricted to perseverative errors – it was accompanied by an increase in trials to first category (TFC) and an increase in non-perserverative errors (NPE). Because of this increase in non-perseverative errors, the *proportion* of errors that were perseverative (PP) was greatest not in the digit-switching condition but in the 2-back condition.

With respect to auditory-vocal task performance, and as in Experiment 1, mean response times were significantly faster in the go-no go task (762.3 msec) than in the 2-back task (1079.5msec; $t(41) = 16.86$, $p < 0.001$), which in turn were significantly faster than in the digit-switching task (1312.3msec; $t(41) = 14.58$, $p < 0.001$). The difference in mean accuracy on the digit-switching (0.897) and the 2-back (0.911) tasks was not significant ($t(47) = 1.08$), but accuracy on both tasks was significantly less than that on the go-no go task (0.935; digit-switching versus go-no go: W(48) = 336, $p = 0.008$, uncorrected; 2-back versus go-no go: W(48) = 367.5, $p = 0.012$, uncorrected). Signal processing analysis on those tasks for which it is appropriate showed good sensitivity (2-back: d' = 1.293; go-no go: d' = 2.896).

Effects of Secondary Task. In order to visualise the dual-task interference on the various dependent measures, all dual-task scores for all dependent measures were converted to *z*-scores based on the means and standard deviations derived from the control condition distributions. The result is shown graphically in Figure 3.

It is clear from the figure that the digit-switching task has the greatest effect on the first five dependent measures, while the 2-back task generally has an intermediate effect with the go-no go task having the least effect. However, this pattern is not reflected for PP (the proportion of errors that are perseverative). Thus, while the number of classical perseverative errors is greatest for the digit-switching condition, the number of non-perseverative errors is also high in this condition, and so if error rate across conditions is controlled, the proportion

Dependent Variable	**Statistic**	**Probability**	**Effect Size (partial η^2)**
Categories	F(2, 94) = 21.512	$p < 0.001$	$\eta^2 = 0.314$
Correct	F(2, 94) = 12.613	$p < 0.001$	$\eta^2 = 0.212$
TFC	F(1.740, 81.796) = 9.867	$p < 0.001$	$\eta^2 = 0.174$
CPE	F(1.719, 80.774) = 13.937	$p < 0.001$	$\eta^2 = 0.229$
NPE	F(2, 94) = 3.259	$p = 0.043$	$\eta^2 = 0.065$
PP	F(2, 94) = 7.741	$p = 0.001$	$\eta^2 = 0.141$

Table 4: Summary of ANOVA results for all WCST dependent variables from Experiment 2. TFC = Trials to First Category; CPE = Classical Perseverative Errors; NPE = Non-Perseverative Errors; PP = Perseveration Proportion. (Note: Greenhouse-Geisser corrections have been used where Mauchley's test suggested violation of sphericity.)

of errors that are perseverative is greater when WCST is combined with the 2-back task than when it is combined with either of the other auditory-vocal tasks.

As in Experiment 1, all participants performed the control condition (WCST alone) prior to the three experimental conditions. Following the logic of analysis for Experiment 1, data from the control condition were therefore ignored and one-way ANOVAs were conducted on the data from the three experimental conditions. As shown in Table 4, there were significant effects of task on all dependent measures.

Post-hoc *t*-tests were performed to assess the reliability of the differential effects. Assuming a Bonferroni corrected critical value of 0.008 as in Experiment 1, significantly fewer categories were achieved in the digit-switching condition than in the 2-back condition ($t(47) = 3.48$, $p = 0.001$) and in the 2-back condition than in the go-no go condition ($t(47) = 3.27$, $p = 0.001$). Similarly, significantly fewer cards were correctly sorted in the digit-switching condition than the 2-back condition ($t(47) = 2.51$, $p = 0.008$), and in the 2-back condition than the go-no go condition ($t(47) = 2.88$, $p = 0.003$). Significantly more trials were required to obtain the first category during the digit-switching condition than in the 2-back condition ($t(47) = 2.84$, $p = 0.003$) or the go-no go condition ($t(47) = 4.21$, $p < 0.001$), but the difference between the 2-back condition and the go-no go condition was not significant ($t(47) = 1.19$, *n.s.*).

With respect to errors, more perseverative errors were made in the digit-switching condition than in the 2-back condition ($t(47) = 1.52$, *n.s.*), and more in the 2-back condition than the go-no go condition ($t(47) = 4.52$, $p < 0.001$), but only in the latter case was the difference statistically significant. The pattern for non-perseverative errors was similar, with more non-perseverative errors in the digit-switching condition than the go-no go condition ($t(47) = 2.01$, *n.s.*) and marginally more in the go-no go condition than the 2-back condition ($t(47) = 0.41$, *n.s.*). Only the difference between the extremes (i.e., between the digit-switching and 2-back conditions) reached corrected statistical significance ($t(47) = 2.54$, $p = 0.007$).

Turning to the derived measure, the proportion of perseverative errors was significantly greater in the 2-back condition than either the digit-switching condition ($t(47) = 2.847$, $p = 0.003$) or the go-no go condition ($t(47) = 3.478$, $p = 0.001$), but the measure did not differ reliably between the digit-switching and go-no go conditions ($t(47) = 1.316$, *n.s.*). Further analysis of the auditory-vocal task data is reported in the appendix.

Discussion

At first glance it appears that, as hypothesised, performance on WCST was more affected by a secondary task held primarily to involve set-shifting (the digit-switching task) than secondary tasks held primarily to involve other executive functions (the 2-back task and the go-no go task). However, the decrement in performance was evidenced not only by an increase in classical perseverative errors in the critical condition, but also by an increase both in non-perseverative errors and in the number of trials required to attain the first rule, and by a decrease both in cards correctly sorted and in categories achieved. Thus, the effect of the digit-switching task on the WCST was not selective.

This non-selectivity could be taken to suggest that the digit-switching task is simply more resource-intensive than the two other auditory-vocal tasks. Response times were, after all, slower in this task than in either of the other auditory-vocal tasks. Much of Figure 3, for example, is easy to interpret if the digit-switching task is harder than the 2-back task and that in turn is harder than the go-no go task. However, any attempt to account for these results by arguing that the secondary tasks vary along a single dimension (e.g., task difficulty) fails to account for the effects of secondary task on the perseveration proportion measure, which was significantly inflated in the 2-back condition, or for the effect of the 2-back condition on measures of bi-gram associations in Experiment 1. These imply, if anything, that the 2-back task is the most difficult of the three. A more subtle conclusion is therefore required.

What processes must occur following an error (and so negative feedback) on the WCST in order to avoid a subsequent error? One must internalise the negative feedback, reject the current hypothesis (e.g., that the sorting criterion is *colour*), and generate an alternative hypothesis (e.g., that the sorting criterion could be *shape*). Perseverative errors are likely to arise if one were to ignore negative feedback (through failure to monitor incoming information). In contrast, if one were to reject the current hypothesis but select an alternative at random, then both perseverative and non-perseverative errors are likely to arise. Thus, we suggest that the interference pattern supports a componential account of WCST in which simultaneous performance of the 2-back task impairs monitoring – a process essential to success on the 2-back task – while simultaneous performance of the digit-switching task impairs some other process such as hypothesis generation.

Since multiple factors may result in perseverative errors this componential analysis and the relation between the current studies and patient studies must remain somewhat tentative. In particular, it is conceivable that different factors lie behind the increase in perseverative errors reported here when the WCST is combined with the 2-back task than in the dual-task studies of Dunbar and Sussman (1995) and in the patient studies of Stuss et al. (2000). Most critically, however, the data appear to rule out any account of the interference effects based on a single task dimension (such as difficulty). As in Experiment 1, the results support a multi-component account of executive functioning. Equally, while the interference pattern is consistent with a set-shifting factor playing a major role in WCST performance, it does not support the *single* factor model (based on the set-shifting factor) endorsed by Miyake et al. (2000) arising from their SEM analysis.

General Discussion

The theoretical orientation from which the experiments reported here were conducted is one in which behaviour is controlled by a set of separable control processes, with the different control processes being recruited differentially by the three secondary tasks. Results from Experiments 1 and 2 support this. Different secondary tasks lead to different patterns of interference on the two primary tasks. Critically, the interference patterns of the secondary tasks on primary task performance cannot be explained under the assumption that the

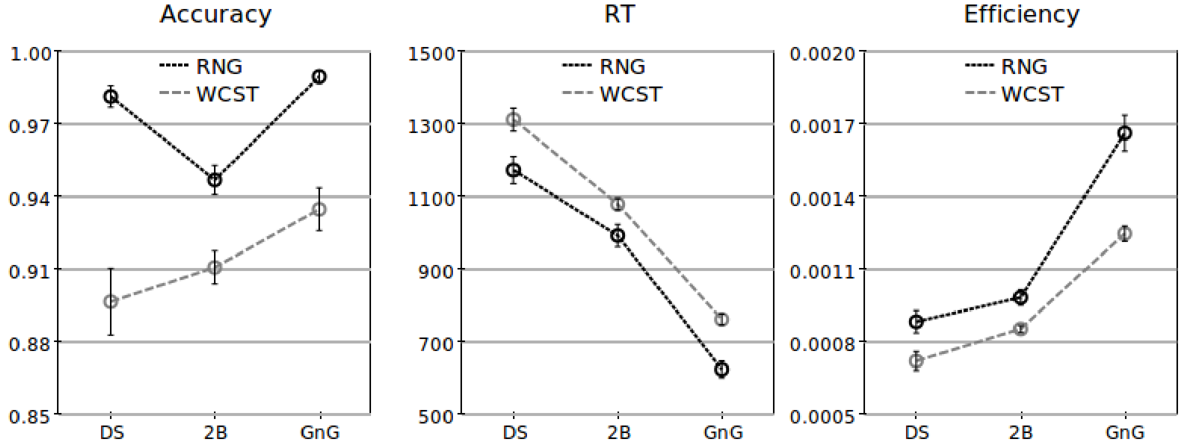

Figure 4: A comparison across Experiment 1 and Experiment 2 of the mean values of three dependent measures from the auditory-vocal tasks. Error bars represent one standard error deviation from the mean. Left panel: Accuracy (i.e., proportion of correct responses). Centre panel: Response time in milliseconds. Right panel: Efficiency, defined as accuracy divided by response time. DS = digit-switching task; 2B = 2-back task; GnG = go-no go task.

different secondary tasks vary on a single dimension (e.g., task difficulty). Moreover, the interference patterns argue against the involvement of the specific executive functions attributed by Miyake et al. (2000) to random generation and WCST performance on the basis of statistical analysis of individual differences data. In contrast, they support an account of both tasks in which both set-shifting and memory updating / monitoring play critical roles.

Comparison of Experiment 1 and Experiment 2

Experiment 1 and Experiment 2 differ only in the primary task used. Secondary task performance on the two experiments may therefore be directly compared. Figure 4 shows the effect of the different primary tasks (from Experiment 1 and Experiment 2) on secondary task accuracy, response time and efficiency (defined as accuracy divided by response time). All secondary tasks were performed more slowly and with more errors in Experiment 2 than in Experiment 1. Accuracy on the digit-switching task was particularly affected when that task was coupled with the WCST in comparison to when it was coupled with random generation, but as shown in the right-most panel, ranking secondary tasks by efficiency produces the same ordering in the two experiments.

One reason why the data shown in Figure 4 are important is that rate of responding differed across the secondary tasks. Thus, the inter-trial interval of the digit-switching task was 2.0 seconds, but for the 2-back task it was 2.5 seconds and for the go-no go task it varied between 1.5 and 2.5 seconds. These rates were set during pilot work so as to allow sufficient time for participants to respond while still making each task moderately demanding. However, given this, participants may have developed different time-sharing strategies or different relative prioritisations of the primary and secondary tasks in the different experimental conditions across the two experiments. Analysis of response times to secondary tasks (presented above but shown graphically in Figure 4, centre panel) shows no evidence of varying task prioritisation over Experiment 1 and Experiment 2. Thus, while responses to the secondary tasks were slower in Experiment 2 than in Experiment 1, the rank ordering of response times to the secondary tasks did not vary across the experiments. Moreover, the slowing in response times when the secondary tasks were paired with WCST in comparison to random generation was similar across the secondary tasks (approximately 100 msec). In other words, while performance of WCST slows down concurrent performance on the auditory-vocal tasks in comparison to random generation, its effect is similar on all three auditory-vocal tasks. To consider this further, suppose that participants prioritised (say) digit-switching over the primary task (random generation) in Experiment 1 but the primary task (WCST) over digit-switching in Experiment 2. This could explain the difference in accuracy on the digit-switching task in the two experiments (see Figure 4, left panel), but it would suggest that digit-switching should have a minimal effect on WCST performance. As shown in Figure 3, this was not the case.

Methodological Considerations

Methodologically, a key aspect of both experiments is that the primary tasks provide multiple dependent measures. This raises the possibility of dissociations between the dependent measures, and the dual-task methodology succeeds precisely because different secondary tasks differentially affect the various dependent measures. Thus, the approach developed here may in principle be applied with other primary tasks to decompose the involvement of different control subprocesses in the performance of those tasks, provided that the primary task yields multiple dependent measures that are not themselves all highly correlated.

While the dual-task approach to the fractionation of executive functions appears successful, three methodological concerns need to be countered. First, we have assumed that the auditory-vocal tasks are “process-pure” (Jacoby, 1991) with respect to their executive function involvement. Thus, we have assumed that the 2-back task primarily taps memory updating / monitoring and not response inhibition or set-shifting. Parallel assumptions apply to the digit-switching task and the go-no go task. These assumptions are likely to hold only as a first approximation. For example, it is highly plausible that digit-switching invokes not just mechanisms that support set-shifting but also mechanisms that support both response inhibition and memory updating / monitoring. Response inhibition may be required as on any trial there are two possible responses (e.g., the correct response to the stimulus “7” is either “odd” or “high”), one of which should not be produced. Memory updating / monitoring may be required because the task changes (from odd/even to high/low) every four trials, and the current task must therefore be maintained in working memory. Similarly, good performance on the 2-back task is likely to involve executive functions beyond memory updating / monitoring. Response inhibition, for example, may be required to inhibit positive responses to foils occurring 1-back and 3-back in the stimulus sequence. Worse, we have no clear account at a process level of precisely what is involved in the function described as memory updating / monitoring. We have simply assumed that this function is shared by the 2-back task and the working memory tasks used to tap this function by Miyake et al. (2000). The methodology would undoubtedly be improved if process-pure auditory-vocal tasks could be devised. However, it is unclear whether any executive task can be reasonable said to tap one and only one putative executive function. In our view, the strength of the effects observed in the current studies indicate that secondary tasks which *primarily* tap one putative executive function rather than another are sufficient to demonstrate varying forms of central executive involvement in the primary tasks.

A second methodological concern is that the secondary tasks may not be sufficiently matched, either on their difficulty or on their time course. Thus, in both experiments participants performed better (with greater accuracy and faster responses) on the go-no go task than on either of the other secondary tasks. Note however that in Experiment 1 while accuracy on the go-no task (0.990) was greater than accuracy on the digit-switching task

(0.981), the difference was not significant. Moreover, in Experiment 2 accuracy on the digit-switching task (0.897) was slightly worse than on the 2-back task (0.911). Thus, while the go-no-go task may be argued to be the easiest of the three secondary tasks, the accuracy scores imply that the digit-switching and 2-back tasks cannot be ordered on a single dimension of task difficulty, and moreover that, at least when coupled with the random generation task, the digit-switching and go-no go tasks are of similar difficulty.

Finally, a third potential methodological objection is that dual-tasking may impose its own burden on set-shifting. That is, participants might cope with the dual-tasking situation by alternating between the primary task and the secondary task, and this alternation will presumably put high demands on set-shifting. We have already discussed previous studies which suggest that dual-tasking ability is distinct from the executive functions tapped by the secondary tasks. A further argument against the dual-task/set-shifting position derives from an analysis of the decrement in primary task performance from the control conditions during each auditory-vocal task, and in particular the correlation between this decrement and auditory-vocal task performance. In Experiment 1, these correlations were uniformly strong and held for all dependent measures (i.e., not just those related to set-shifting). The correlations were less strong in Experiment 2 (possibly reflecting greater variability in participant performance), but again there was no clear relation between the correlations and dependent measures which did, or did not, relate to set-shifting. Thus, it appears that if dual-tasking does impose a burden on set-shifting, that burden does not exhaust set-shifting capacity.

Response Inhibition

The go-no go condition was included in both experiments in order to determine the extent to which performance of each primary task was dependent upon the putative executive function of response inhibition. But of the three secondary tasks, the go-no go task produced the least interference on virtually all dependent variables in both experiments. The relative lack of effects could be taken to suggest that neither primary task involves response inhibition to any significant degree. While this may be the case, it would be premature to draw negative conclusions for the lack of effects relating to response inhibition. At least three other alternatives (besides the null hypothesis) need to be considered. First, the lack of significant effects could be due to the relative ease of the go-no go task (as discussed above), or the fact that good performance on the task presumably draws only intermittently on response inhibition. Alternatively, the lack of effects could be due to the go-no go task taxing some element of executive functioning that differs from response inhibition in the sense used by Miyake et al.

In fact, response inhibition is likely to be more difficult to isolate within the dual-tasking framework than set-shifting or monitoring because it is held to be invoked only when an habitual response must be suppressed. In the experiments considered here, it was required on only 1 in 6 trials of the auditory-vocal task. It is not possible to design a secondary task that could draw continuously on the construct, for in such a task there could be no habitual response. An alternative possibility, if response inhibition is to be assessed within the dual-task approach, could be to yoke the primary and secondary tasks such that events in the primary and secondary tasks which require response inhibition occur simultaneously. Even this approach is fraught with difficulties, however, as response inhibition may plausibly be a "global" process such that inhibiting multiple near simultaneous responses is easier (i.e., less resource intensive, or dependent solely on a global "withhold all responses" process) than inhibiting just one of several near simultaneous responses (e.g., Coxon et al., 2009).

But does the go-no go task tap response inhibition? We have discussed response inhibition in terms of an ability to inhibit a prepotent response, and this would seem to be an appropriate characterisation of the processes required when performing the Stroop task, particularly given current process accounts of Stroop performance (e.g., Cohen & Huston, 1994). However there is an alternative possible characterisation, namely that response inhibition is specifically involved only when one is required to cancel production of a response *after* it has been selected. This view of response inhibition is consistent with the work of Pashler (e.g., Pashler, 1994) and Verbruggen, Vandierendonck and colleagues (e.g., Szmalec et al., 2005; Verbruggen, Liefooghe & Vandierendonck, 2006) on response selection as a distinct process which precedes response production. While performance on the go-no go task plausibly assesses the former sense of response inhibition, the later is more plausibly assessed by performance on the so-called stop-signal task, where participants are required to make a simple choice discrimination *unless* the stimulus is followed by a beep. In the standard administration of the task, the time between the stimulus and the beep is adjusted on a participant-by-participant basis to produce a stopping accuracy of, for example, 50%. Our use of the go-no go task as a response inhibition task is justified by the relation between the go-no go and the Stroop tasks.

Conclusion

We have demonstrated in a random generation task that different measures of randomness are differentially affected by secondary tasks that primarily tap set-shifting and monitoring. Similarly, different measures of WCST performance are differentially affected by such secondary tasks. Taken together, the interference profiles are not consistent with a unitary resource-based model of cognitive processing in which tasks vary along a single dimension (task difficulty or cognitive load). Rather, they argue for a decomposition of central processing in which different aspects of the primary tasks draw differentially on processes supporting set-shifting and monitoring. In the language of Pashler (1994), the results suggest that central processing is not fully parallel, but is itself subject to bottlenecks related to shifting task set and memory updating / monitoring.

We have not sought to address a more fundamental issue. Can the executive functions of set-shifting and monitoring be decomposed, or are they in some sense primitive or atomic? It is our suspicion that set-shifting at least is non-atomic (see Altmann & Gray, 2008). To demonstrate this, however, more subtle techniques will be required.

References

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review*, *115*(3), 602–639.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience*, *6*(2), 115-116.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170-177.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Journal of Experimental Psychology, 18,* 334-336.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 5-28.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641-657.

Baddeley, A. D., Della Sala, S., Papagno, C., & Spinnler, H. (1997). Dual-task performance in dysexecutive and nondysexecutive patients with a frontal lesion, *Neuropsychology*, *11*(2), 187-194.

Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *51*(4), 819-852.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47-89). New York: Academic.

Bollen, K.A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research*, *17*(3), 303-316.

Brass, M., Derrfuss, J., Forstmann, B., & Cramon, D. Y. V. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences*, *9*(7), 314-316.

Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. E. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *NeuroImage*, *17*(3), 1562-1571.

Burgess, P.W. & Shallice, T. (1996). Bizarre responses, rule detection and frontal lobe lesions. *Cortex*, *32*(2), 241-259.

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

Cooper, R. P. (2010). Cognitive control: Componential or emergent? *Topics in Cognitive Science*, *2*(4), 598-613.

Coxon, J. P., Stinear, C. M., & Byblow, W. D. (2009). Stop and go: The neural basis of selective movement prevention. *Journal of Cognitive Neuroscience*, *21*(6), 1193-1203.

Cragg, L. & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science*, *2*(4), 631-642.

Dunbar, K., & Sussman, D. (1995). Toward a cognitive account of frontal lobe function Simulating frontal lobe deficits in normal subjects. *Annals of the New York Academy of Sciences*, *769*, 289-304.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N. & Emslie, H. (2000). A neural basis for General Intelligence. *Science*, *289*(5478), 457-460.

Emerson, M. J., Miyake, A., & Rettinger, D. A. (1999). Individual differences in integrating and coordinating multiple sources of information, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1300-1321.

Evans, F. J. (1978). Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society*, *12*, 35-38.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101-135.

Heaton, R. K. (1981). A manual for the Wisconsin card sorting test. Odessa FL: Psychological Assessment Resources.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, *28*, 376–385.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

Jahanshahi, M., Profice, P., Brown, R. G., Ridding, M. C., Dirnberger, G., & Rothwell, J. C. (1998). The effects of transcranial magnetic stimulation over the dorsolateral prefrontal cortex on suppression of habitual counting during random number generation. *Brain, 121*(8), 1533-1544.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology* (Whole No. 89).

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

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behaviour on a dimension-switching task. *Developmental Science*, *6*(5), 449-467.

Logie, R. H., Cocchini, G., Delia Sala, S., & Baddeley, A. D. (2004). Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's Disease. *Neuropsychology*, *18*(3), 504-513.

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modelling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

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

Miller, E.K. & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423-1442.

Meyer, D.E. & Kieras, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3–65.

Milner, B. (1963). Effects of different brain lesions on card-sorting. *Archives of Neurology*, *9*, 90-100.

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

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313-324.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220-244.

Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behaviour. *Psychological Review*, *104*(3), 603-617.

Raye, C., Johnson, M., Mitchell, K., Greene, E., & Johnson, M. (2007). Refreshing: A minimal executive function. *Cortex*, *43*(1), 135-145.

Reverberi, C., D’Agostini, S., Skrap, M., & Shallice, T. (2005). Generation and recognition of abstract rules in different frontal lobe subgroups. *Neuropsychologia*, *43*(13), 1924-1937.

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

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*(5), 479-484.

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, *136*(2), 241-255.

Shallice, T., Stuss, D.T., Picton, T.W., Alexander, M.P. & Gillingham, S. (2008). Mapping task switching in frontal cortex through neuropsychological group studies. *Frontiers in Neuroscience*, *2*(1), 79-85.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268-288.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5-42.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657-1661.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*. *15*(2), 201-292.

Stuss, D.T., Levine, B., Alexander, M.P., Hong, J., Palumbo, C., Hamer, L., Murphy, K.J. & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*, 388–402.

Szmalec, A. & Vandierendonck, A. (2007). Estimating the executive demands of a one-back choice reaction time task by means of the selective interference paradigm. *Quarterly Journal of Experimental Psychology, 60*, 1116-1139.

Szmalec, A., Vandierendonck, A., & Kemps, E. (2005). Response selection involves executive control: Evidence from the selective interference paradigm. *Memory & Cognition*, *33*, 531–541.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology*, *89*(1), 77-101.

Towse, J.N. & Neil, D. (1998). Analyzing human random generation behaviour: A review of methods used and a computer program for describing performance. *Behaviour Research, Instruments and Computers*, *30*(4), 583–591.

Towse, J. N., & Valentine, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology*, *9*(4), 381-400.

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

Verbruggen F., Liefooghe, B. & Vandierendonck, A. (2006). Selective stopping in task switching: The role of response selection and response execution. *Experimental Psychology*, *53*(1), 48-57.

Verbruggen, F., & Logan, G. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418-424.

Appendix: Analysis of Auditory-Vocal Task Behaviour

One important issue regarding the auditory-vocal task not addressed in the main body of this paper concerns whether there is evidence of a dual-task trade-off (i.e., whether good performance on the visual-manual task correlates with poor performance on the auditory-vocal tasks), or, alternatively, whether the dual-task data support a "general intelligence" account of individual differences (Spearman, 1904; Duncan et al., 2000), where participants who score highly on the auditory-vocal tasks are also less affected by the dual-task condition. To address this issue correlational analyses were conducted between accuracy on the auditory-vocal tasks and each dependent measure from each primary task.

A second issue concerns whether performance on the various auditory-vocal tasks correlate. The concept of general intelligence underlying a unitary central executive would suggest that there should be moderate-to-strong correlations between the various auditory-vocal tasks, while weak correlations between the auditory-vocal tasks (which we hold tap distinct executive functions) would support a “multiple processes” view. A second analysis of the auditory-vocal data was performed to address this issue.

Random Generation

Analysis of the correlations between secondary task performance and primary task dependent measures from Experiment 1 revealed that for the digit-switching and go-no go conditions no correlations reached significance (at the $p = 0.05$ level). Note that this analysis is limited by the high accuracy scores – greater than 98% – in the digit-switching and go-no go tasks. In the 2-back condition, however, where accuracy was slightly lower, the correlation between accuracy and RNG score was highly significant ($r = -0.566$, $p < 0.001$, two-tailed). This effect was also apparent for AA ($r = -0.366$, $p = 0.028$, two-tailed) and TPI ($r = 0.491$, $p = 0.002$, two-tailed). Participants who performed well on the 2-back task tended to also have low AA scores and high TPI scores, corresponding to greater randomness in their sequences. More generally, the direction of correlations argues against a dual-task trade-off. On the contrary, when participants performed the 2-back task with the random generation task, those who performed well on one task tended to also perform well on the other task.

Turning to the second analysis, assuming one-tailed tests and without correcting for multiple comparisons, there was a mild correlation between accuracy on the digit-switching and 2-back tasks ($r = 0.307$, $p = 0.034$), but not on the digit-switching and go-no go tasks ($r = -0.013$) or on the 2-back and go-no go tasks ($r = 0.112$). While one might interpret the lack of correlations here as arguing against a general factor underlying performance across the auditory-vocal tasks, this would be premature. Critically, as in the above analysis, high accuracy on the digit-switching and go-no go tasks limit the power of this analysis.

Wisconsin Card Sorting Test

Accuracy scores on the secondary tasks in Experiment 2 were lower than in Experiment 1, meaning that the ceiling effects are less likely to be an issue and correspondingly that the correlational analyses are more powerful. Nevertheless, paralleling the results of Experiment 1, analysis of the auditory-vocal data from Experiment 2 found that accuracy and sensitivity (where relevant) on the auditory-vocal task generally did not correlate significantly with any measures of WCST performance. For example, the correlations between cards correctly sorted and auditory-vocal accuracy were: $r = 0.179$ during digit-switching ($df = 46$, $p = 0.206$, two-tailed); $r = 0.267$ during 2-back ($df = 46$, $p = 0.067$, two-tailed); and $r = -0.046$ during go-no go ($df = 46$, $p = 0.757$, two-tailed). The only exceptions to this arose during the 2-back condition, where positive correlations were found between sensitivity and number of cards correctly sorted ($r = 0.286$, $p = 0.049$, two-tailed) and between accuracy and number of categories achieved ($r = 0.305$, $p = 0.035$, two-tailed), and a negative correlation was found between accuracy and number of classical perseverative errors ($r = -0.316$, $p = 0.029$, two-tailed). None of these correlations survive Bonferroni corrections for multiple tests.

Pair-wise correlations between accuracy measures across the auditory-vocal tasks were also calculated. In all cases and in contrast with Experiment 1, these were positive and significant (digit-switching and 2-back: $r = 0.345$, $p = 0.008$; digit-switching and go-no go: $r = 0.455$, $p = 0.001$; 2-back and go-no go: $r = 0.333$, $p = 0.010$, one-tailed probabilities in all cases as the relevant hypotheses in all cases were unidirectional). Thus, this analysis provides additional support for Miyake et al.’s (2000) claim that, despite individual differences in the efficiency of set-shifting, memory updating / monitoring and response inhibition, the factors are also mildly correlated across individuals.