

Is it harder to switch among a larger set of tasks?

Félice van 't Wout, Aureliu Lavric and Stephen Monsell

University of Exeter

This is the posted prepublication version of an article published in:

Journal of Experimental Psychology: Learning, Memory, and Cognition, 2015, Vol. 41, No. 2, 363–376 <http://dx.doi.org/10.1037/a0038268>

© 2014 American Psychological Association

This article may not exactly replicate the authoritative document published in the APA journal. It is not the copy of record.

Correspondence concerning this article should be addressed to Félice van 't Wout, now at

School of Experimental Psychology
University of Bristol
12a, Priory Road
Bristol BS8 1TU

E-mail: fmvantwout@gmail.com

Author Note

The work described in this paper was carried out as part of a PhD project by FvtW under the supervision of SM and AL, supported by a studentship from the Economic and Social Research Council (UK).

Abstract

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

Key words: Task switching, task cuing, task-set retrieval, procedural working memory.

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

As is well known (for reviews see: Kiesel et al., 2010; Monsell, 2003; Vandierendonck, Liefoghe & Verbruggen, 2010), there is a substantial task-switch cost: reaction times (RTs) and error rates are greater when the task changes, compared to when it repeats (e.g. Allport, Styles & Hsieh, 1994; Meiran, 1996; Rogers & Monsell, 1995). When the interval between the cue onset and the stimulus onset is increased to allow preparation for the upcoming task, the switch cost decreases (e.g. Meiran, 1996; Monsell & Mizon, 2006).

This reduction in switch cost (RISC effect) is thought by many to index active reconfiguration of task-set (e.g. Mayr & Kliegl, 2000; Rogers & Monsell, 1995; Rubinstein, Meyer & Evans, 2001; see Schneider & Logan, 2005 for an alternative account and Forrest, Monsell & McLaren, 2014, for counter arguments). Hypothesised sub-processes of task set reconfiguration include: reorienting attention to the relevant stimulus attribute (Longman, Lavric & Monsell, 2013, Longman, Lavric, Munteanu & Monsell, in press; Mayr, Kuhns & Rieter, 2012; Meiran, 2000), retrieving a task code (Altmann & Gray, 2008) or goal representation (Goschke, 2000) and retrieving the required task's S-R rules into procedural working memory (Mayr & Kliegl, 2000), and suppressing the previous (now irrelevant) task-set (Mayr & Keele, 2000; Schuch & Koch, 2003). In this paper, we ask whether the difficulty of making the required task-set the most active depends on the number of potential task-sets in play.

Procedural working memory, and the process of promoting a task-set into a most-active state, are theoretical constructs in need of clarification. The task-switching literature contains both "retrieval" and "activation" conceptions of the promotion process. According to the former, sets of S-R rules (e.g. Mayr & Kliegl, 2000; Rubinstein et al., 2001) or incomplete sets (Lien, Ruthruff, Remington & Johnston, 2005) are discretely retrieved or "loaded" into a PWM store. According to the second, illustrated by connectionist accounts of task-set control (e.g. Brown, Reynolds & Braver, 2007; Gilbert & Shallice, 2002) task-sets (represented in these models by task-nodes) have continuous degrees of activation which are determined by a combination of top-down control input, bottom-up activation from associated contexts and stimuli, and by their previous level of activation. In these accounts, there is no distinct PWM, only a most active task-set. We aim here not to distinguish these "retrieval" and "activation" conceptions of the task selection process and PWM, but to

establish the effects of a particular factor – the number of tasks in contention – on task-set retrieval/activation, as a constraint upon either type of account.

One explicit conception of PWM that incorporates elements of both “retrieval” and “activation” accounts is found in Oberauer’s (2009) general account of working memory. This model distinguishes, first, between declarative working memory and procedural working memory (PWM). The former holds representations (e.g. of objects or words) in an active state for operating upon, the latter makes processing possible by holding potentially relevant task-sets – the sets of operations or S-R rules. In Oberauer’s model, PWM consists of three components: the activated part of long-term memory (LTM), a component he calls the “bridge” (holding the S-R rules constituting the current task-set), and a “response focus” (essentially the response code most strongly activated by applying the current task-set rules to the current perceptual or mental contents of declarative WM). Procedures or task-sets that one has recently practiced are represented in the activated part of LTM. In a task-switching experiment, the several task-sets in play will be in this active state, competing for control over the bridge. The bridge represents the S-R rules currently governing performance. For effective performance, and specifically to avoid response ambiguity, it is assumed that the bridge’s content is restricted to a single task-set. (But S-R rules in the active part of LTM can still prime responses even though they are not in the bridge – hence the response congruence effects observed in numerous paradigms, including task-switching.) Hence task-switching requires updating the contents of the bridge with another task-set retrieved from the activated part of LTM. This retrieval process is assumed to be at least one source of task-switch costs.

If the retrieval of a task-set into PWM contributes to the cost of a task-switch, then if we make retrieval more difficult, this should increase the cost. To test this idea, Mayr and Kliegl (2000) manipulated the difficulty of retrieval. In their Experiment 1, participants were required to classify a set of centrally presented words, switching predictably either between a

size and a living/nonliving judgement (“low retrieval demand”), or between classification of the words’ colour and screen location during a practice session (“high retrieval demand”).

The switch cost was indeed found to be larger in the latter case. Furthermore, when the task cue explicitly displayed the S-R mappings (Experiment 3), this effect of retrieval demand was eliminated, suggesting that it is specifically the time required to retrieve the S-R rules that was responsible for the difference in switch cost.

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

One rationale for the intuition that the difficulty of task-set retrieval should increase with the number of tasks in play is that, in several other cases of human memory retrieval, retrieval does become more difficult the more candidates for retrieval there are in contention. Examples include the decrease in free recall performance with list length (Murdock, 1962), the decrease in cued recall with the number of paired associates learned (Aue, Criss & Fischetti, 2012, Experiment 5; Criss & Shiffrin, 2004), the “fan” effect (Anderson, 1974; Lewis & Anderson, 1976), and Hick's law (Hick, 1952; see Schneider & Anderson, 2011, for a recent treatment in terms of memory retrieval). Of course these examples differ in various ways from retrieval of a task set. The fan effect – the increase in the time to retrieve a fact associated with a concept as a function of the number of facts learned about that concept – involves competition among memory elements associated with a single cue. Hick's law, the logarithmic increase in choice reaction time with the number of S-R mappings may, in Oberauer's (2009) terms, be seen as a consequence of competition between the S-R rules

represented *within* the PWM "bridge", whereas we are interested here in competition among *sets* of rules for *access to* the bridge. There is no a priori requirement that the set-size effects observed in these other forms of memory retrieval should apply to task-set activation. Indeed, Oberauer (2009) explicitly asserts that while the Hick's law set-size effect – RT increasing with number of S-R rules for a task-set held in the bridge – is expected because of competition for the response 'focus', there should be no effect of the number of activated task sets in LTM on retrieval time into the bridge.

Da Silva Souza, Oberauer, Gade and Druey (2012) recently acquired some evidence in support of this latter assumption. Participants were cued to perform one of two or one of three digit classification tasks (odd/even, high/low, inner/outer¹), and the switch cost was no larger with 3 than with 2 tasks. But, several other task-switching studies have also manipulated the number of tasks in play, with mixed results. Rubin and Meiran (2005, Experiment 2) used two perceptual tasks (colour and shape classification) and two spatial tasks (upper/lower and left/right classification); the two sets of tasks used separate responses. When participants switched between 2 or 3 tasks in a task-cuing paradigm (with CSI of 100 or 1000 ms), the switch cost did not increase with the number of tasks. However, the comparison between the 2-task and 3-task conditions seems less than perfect because in the 3-task condition one of the tasks required separate responses. In another experiment (Kessler & Meiran, 2010), participants switched between a colour and a shape task on 75% of all trials. On the remaining 25% of trials participants performed either just one or any of three other perceptual tasks (judging the size, thickness or fill of a clover shape), so that participants were essentially switching between 3 or 5 tasks. Because only the colour and shape task trials were included in the analysis, task frequency was matched between the two conditions. No effect of number of tasks on the switch cost was found. But again, the

¹ If the digits 2 to 9 are used as stimuli, then the "outer" digits are 2, 3, 8 and 9, and the others "inner"

additional tasks always required a set of different stimuli and responses, complicating the interpretation of this null result. Kray, Li and Lindenberger (2002) required participants to switch between 2, 3 or 4 word classification tasks in a task-cuing paradigm (CSI was 1000 ms); the switch cost was substantially larger with more tasks. However, this effect was attributable almost entirely to faster RTs on repeat trials in the 4-task condition, which does not sit easily with any retrieval account. Buchler, Hoyer and Cerella (2008), had participants switch unpredictably between 2, 3 or 4 mathematical tasks, with no opportunity for advance preparation. Although the frequency with which each task was performed across set sizes was controlled, task recency was not (see below). And the outcome was equivocal: the switch cost was nearly twice as large in the 4-task condition as in the 2-task condition, but this difference was not statistically significant. Emerson and Miyake (2003; Experiment 4) required participants to switch between 2 or 3 mathematical tasks, using a list paradigm. They hypothesized that switching between 2 tasks might be harder than switching between 3 tasks, based on the "backwards-inhibition" or "n-2 task repetition" effect (e.g. Arbuthnott & Frank, 2000; Mayr & Keele, 2000) – slower RTs for a task that was more recently switched away from (i.e. a longer RT on the last trial of an ABA than of a CBA sequence of tasks, where A, B and C are tasks). Emerson and Miyake found no effect of number of tasks on the RTs but the error switch cost was larger in the 2-task condition. They speculated that a counteracting 'fan effect' may have caused the lack of interaction in the RTs. Finally, Kleinsorge and Apitzsch (2012) found that switch costs were larger with 4 tasks than with 2 (particularly at short CSI) in an unpredictable task-cuing condition but not in a predictable switching condition.

Taken together, the results are evidently inconsistent, with some experiments showing a larger switch cost with more tasks (Buchler et al., 2008; Kleinsorge & Apitzsch, 2012; Kray et al., 2002), others the opposite (Emerson & Miyake, 2003) and yet others no effect of

number of tasks on the switch cost (Da Silva Souza et al., 2012; Kessler & Meiran, 2010; Rubin & Meiran, 2005). Moreover, the number of tasks manipulation in these studies is often confounded with changes in their recency and/or frequency. Simply increasing the number of tasks performed in a fixed number of trials necessarily decreases both the frequency and the average recency with which each task has been performed. Over a typical one-session experiment, performance on a given task improves with practice, so an effect of frequency would be expected. And the ease with which a task-set is retrieved or activated seems very likely to reflect the recency with which that task was last performed.

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

In addition to examining the effect of the number of tasks on the switch cost, we also assessed another potential index of competition among the task-sets currently in play: the effect of response congruence. With just two tasks mapped to the same response set, it is typically found that congruent stimuli (mapped to the same response by both tasks) are

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

Experiment 1

On each trial, following a task cue, participants classified one of a set of 32 visual objects (see Figure 1 for examples). The set was designed so that each object afforded a binary classification on 5 orthogonal dimensions. In one session, the participant could be cued to perform any of the 5 corresponding classification tasks. In the other session, participants performed (and knew that they would perform) only 3 of the 5 tasks. Different sets of stimuli (32 “animals” or 32 “trees”) were used in the two sessions, so that the number of tasks (3 or 5) could be manipulated within subjects, without performance in the second session being affected by exposure in the previous session to a sub- or super-set of the same classifications. For two of the tasks in each session, which we refer to as the *probe* tasks, recency and

frequency were matched across sessions, to allow an estimate of the effect of number of tasks in play uncontaminated by differences in the recency and frequency with which the current or previous task had been performed. To test our expectation that any effect of the number of tasks in play would be on task-set preparation rather than the execution of a prepared task-set, we compared blocks of trials in which the CSI was short (100 ms) to blocks in which the CSI was long (1300 ms), and encouraged preparation. To control for any carryover of task-set activation or competition from the previous trial (“task-set inertia”), the response-stimulus interval (RSI) was the same with both CSIs, and we limited the critical analysis to two-trial sequences on which both tasks were probe tasks.

Method

Participants Twenty four participants aged between 18 and 22 ($M = 19.5$), 4 male and 20 female, took part in two sessions on consecutive days, and were paid between £12.80 and £16.00, depending on the speed and accuracy of their performance (see below).

--- *Insert Table 1 about here* ---

--- *Insert Figure 1 about here* ---

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

For each set, either of two spoken words (duration = 350 ms) could be used to cue each of the five classifications. The cue always changed from one trial to the next, to avoid confounding the effects of task-repetition and cue-repetition (Logan & Bundesen, 2003;

Monsell & Mizon, 2006). The two cue words used for each attribute are shown in Table 1, as are the assignments of the two values of that attribute to left and right response keys.

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

Trials began with a 500 ms blank screen, followed by presentation of a fixation dot. In the long CSI blocks the spoken task cue began 100 ms after the fixation dot, and the stimulus was displayed 1300 ms after the cue's onset, replacing the fixation dot. In the short CSI blocks, the cue word onset was 1300 ms after the onset of the fixation dot, and the stimulus was presented just 100 ms after the onset of the cue word. (Hence the response-stimulus interval was the same in long and short CSI blocks.) The stimulus was displayed until a response was detected. Responses were made using the left or right arrow key of a PC keyboard, pressed by the index finger of the left or right hand. If the wrong key was pressed, "Error!" was displayed on the screen for an extra 1000 ms. Otherwise a new trial began immediately.

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

each error). When participants improved on this score in a subsequent block of the same CSI, a bonus (£0.20) was awarded.

Of the five tasks of each set, two were chosen to serve as *probe* tasks matched for recency and frequency across the 3-task and 5-task conditions: the pattern and head classification tasks (for the animal stimuli) and the number and trunk classification tasks (for the tree stimuli). The two tasks of each pair were approximately matched for difficulty based on pilot data. Which of the other 3 tasks served as the non-probe task in the 3-task condition was varied over participants so that each was used equally often. To match the recency and frequency of the probe tasks, new trial sequences were generated for each participant, as follows. Every participant (P_1) was yoked to another participant (P_2). First, a 3-task sequence containing a random sequence of equal numbers of trials with tasks A and B (probe tasks) and C (non-probe task) was generated for participant P_1 . From this a 5-task sequence was created for participant P_2 by randomly replacing all instances of task C with equal numbers of instances of task C, D and E. As a result, probe task trials occupied exactly the same positions in the trial sequence for the 3-task and the 5-task participant with this stimulus set. Consequently, for AA and BB task repetitions and AB and BA task switches, the comparison between the 3-task and 5-task condition was unconfounded by effects of the frequency or recency of the task used on either the current or previous trial. (Control of the previous trial is important because recency/frequency might affect not just the difficulty of retrieving the next task-set but also the difficulty of suppressing the previous task-set.) An unavoidable corollary of this manipulation is that the proportion of switch trials was slightly higher in the 5-task condition (on average 74.2%) than in the 3-condition (66.7%). However, only probe task trials followed by probe task trials (AA, BB, AB and BA transitions, whose frequency did

not depend on number of tasks) were analysed for the critical frequency- and recency-matched comparison.²

The session began with a block of 32 practice trials per task for each of the three or five tasks to be performed in that session. Although these were single-task blocks, the two word cues for that task were also presented, alternating from trial to trial, with a CSI of 1300 ms. Half the participants practiced the tasks in one order, and the other half in the reverse order. These 160 (or 96 in the 3-task condition) practice trials were followed by 2 additional practice blocks (one containing 20 long CSI trials, the other containing 20 short CSI trials) which required participants to switch between tasks in the same way as in the experimental blocks. The practice phase lasted approximately 15 minutes, and the experimental phase about 45 minutes.

Results and Discussion

Correct responses with very short (<200 ms) or very long (>3000) reaction times (RTs) were excluded from analysis (0.02% of the correct responses), as were trials following an error. In each of the analyses described below, task assignment was entered as a between-subjects variable with two levels (5 trees and 3 animals versus 5 animals and 3 trees), to remove any variance attributable to differences in performance between the tasks per se from the error term. Because we were not interested in the effects of task per se, interactions with task assignment are not reported.

Effect of number of tasks for the probe tasks Mean correct RTs and error rates for two-trial sequences involving only the probe tasks (i.e. controlled for recency and frequency of the present and previous task) are shown in Figure 2 (left panel) as a function of number of tasks and CSI. A 2 (3 or 5 tasks) x 2 (long or short CSI) x 2 (switch or repeat) repeated measures

² This manipulation also makes the relative task frequencies unequal within the 5-task condition while they are equal in the 3-task condition.

ANOVA was conducted on the mean correct RTs and error rates, averaging over the two tasks.

Figure 2 (left panel) shows that there was no detectable effect of number of tasks on RT, $F < 1$. In fact, overall mean RT was identical in the 5 and 3-task conditions (531 ms), a difference of 0 ± 8 ms³. There were significant main effects of a task switch, $F(1,22) = 53.76$, $p < 0.001$, and CSI, $F(1,22) = 573.41$, $p < 0.001$. These interacted so that the switch cost reduced from 51 ms at the short CSI to 22 ms at the long CSI, a significant reduction in switch cost (or “RISC”) effect of 29 ms or 56%, $F(1,22) = 38.76$, $p < 0.001$. The switch cost and its reduction with preparation were very similar in the 5-task condition (from 50 ms to 24 ms; a RISC of 26 ± 8 ms) and in the 3-task condition (from 52 ms to 20 ms; a RISC of 32 ± 7 ms), a difference of 6 ± 12 ms ($F < 1$ for the 2- and 3-way interactions). Error rates showed a significant switch cost of 1.2%, $F(1,22) = 13.89$, $p < 0.001$; the CSI effect of 0.7% was not significant, $F(1,22) = 2.89$, $p = 0.103$. The error switch cost reduced from 2.1% in the short CSI to 0.2% in the long CSI, $F(1,22) = 7.74$, $p = 0.011$. The error rate ANOVA showed no significant interactions involving the number of tasks, $F_s < 1.13$, but there was a near-significant main effect: participants made $1 \pm 0.5\%$ fewer errors in the 5-task (3.5%) than in the 3-task condition (2.5%), $F(1,21) = 3.88$, $p = 0.062$.

Hence, when recency and frequency are controlled there is no evidence that retrieving the appropriate task-set is harder when there are more tasks in play. (If anything, there were slightly fewer errors with 5 tasks, but this did not interact with switching or preparation.) However we cannot affirm the null hypothesis on the basis of the conventional statistical tests reported above. We therefore used Bayesian statistics to compare the degree to which the data support the null hypothesis (H_0 – no effect of number of tasks) relative to an effect of number of tasks (H_1). Because previous studies have found mixed results with regards to the

³ $\pm \dots$ indicates the standard error of the mean difference.

effect of the number of tasks on task switching performance, the prior probabilities assumed for H_1 and H_0 were 0.5 (“uninformative priors”). To estimate the likelihood of the null hypothesis being correct given the data we applied Masson’s (2011) method, which uses the the sum of squares of the effect of interest and the error term associated with that effect to calculate the Bayes factor and posterior probabilities for H_1 and H_0 . The method assumes a prior distribution of possible effect size parameter values that is a standard normal distribution centered on the value of the effect size observed in the data and extending over the distribution of observed data (Raftery, 1999), also referred as ‘unit information prior’ (Kass & Wasserman, 1995). As can be seen from Table 2, for each of our three contrasts of interest (main effect of number of tasks; its interaction with switch/repeat; the three-way interaction between number of tasks, switch/repeat and CSI) we found “positive” evidence in support of the null-hypothesis, and not even “weak” evidence in support of H_1 ⁴: these results are consistent with the conclusion that it is not harder to retrieve a task-set when one must select from among a larger set of active tasks, when task recency and frequency are controlled.

--- *Insert Figure 2 about here* ---

--- *Insert Table 2 about here* ---

Effect of number of tasks for all trials The above analysis was restricted to trial sequences controlled for task recency and frequency. To explore the importance of this control, we repeated the analysis, now including all trial sequences (see Figure 2, right panel), so that the analysis was no longer restricted to the 4 in 9 trials on which a probe task trial followed a probe task trial. Now the outcome looks quite different: for task-switch trials at the short CSI,

⁴ We used Raftery’s (1995) categorisation of degrees of evidence: “weak” if the posterior probabilities are between 0.50 and 0.75, “positive” for posterior probabilities of 0.75-0.95, “strong” for posterior probabilities of 0.95-0.99 and “very strong” for posterior probabilities above 0.99.

RT was longer for 5 tasks (729 ms) than for 3 tasks (685 ms), a difference of 44 ± 14 ms. The relevant interactions were significant: switch costs were larger in the 5-task (65 ms) than in the 3-task (43 ms) condition (a difference of 22 ± 6 ms), $F(1,22) = 19.27$, $p < 0.001$; the effect of CSI was greater in the 5-task (247 ms) than in the 3-task condition (222 ms, a difference of 25 ± 10 ms), $F(1,22) = 10.09$, $p = 0.004$; and the switch cost was reduced to a greater extent in the 5-task condition (from 103 ms to 27 ms; a RISC of 76 ± 6 ms) than in the 3-task condition (from 61 to 25 ms, a RISC of 36 ± 8 ms), $F(1,22) = 23.64$, $p < 0.001$ (a difference of 40 ± 9 ms).

Consistent with the RT analysis, the reduction in error rate with preparation was slightly larger in the 5-task (3%) than in the 3-task condition (2.5%; a difference of $0.5 \pm 1.1\%$), though not significantly so, $F < 1$. In the error rates the main effect of number of tasks was in the opposite direction to RTs – slightly more errors with 3 tasks (3.8%) than with 5 (3.2%; a difference of $0.6 \pm 0.3\%$) – though this effect was not significant, $F(1,22) = 4.07$, $p = 0.056$.

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

Effects of task recency To examine the effects of recency more directly, we sorted the data for each number of tasks condition into three categories according to the lag since the last trial on which the same task had been cued: 2-3 trials back, 4-5 trials back, and 6-7 trials back, separately for short and long CSI trials, pooling over task (to ensure that, for longer lags, there were no empty cells). Figure 3 shows this analysis restricted to probe-probe task trial

sequences (left panel), and to an analysis pooling over all tasks (right panel). Only effects of and interactions with the linear component of the effect of recency are reported below.

If the time consumed by task-set retrieval is influenced by the task's recency of usage, RTs should increase with lag specifically on unprepared (short CSI) switch trials. The probe task trials, for which task frequency is unconfounded with number of tasks, should provide a "pure" estimate of the effect of recency. The data in Figure 3 (left panel) show a small recency effect in the short CSI condition (slope 4 ± 3 ms per intervening trial), and none in the long CSI condition (slope -4 ± 2 ms); the slopes differed significantly, $F(1,22) = 5.56$, $p = 0.028$, though in separate ANOVAs the main effect of lag was not quite significant in either the long CSI, $F(1,22) = 3.20$, $p = 0.087$, or the short CSI, $F(1,22) = 2.83$, $p = 0.106$. No interaction involving number of tasks was significant, $F < 1$. For the % error data, there were no significant effects or interactions involving task recency ($F < 1$, except for recency x number, $F(1,22) = 4.06$, $p = 0.056$).

--- Insert Figure 3 about here ---

The pattern of results for the uncontrolled tasks in Figure 3 (right panel) is similar: the main effect of task recency was significant, $F(1,22) = 5.73$, $p = 0.026$ but, as for the probe task trials, this recency effect originated entirely from short CSI blocks (slope 9 ± 3 ms per intervening trial), and was slightly negative in long CSI blocks (slope -2 ± 1 ms), $F(1,22) = 10.97$, $p = 0.003$. However, unlike the controlled data, RTs for unprepared trials appear longer, and the recency slope somewhat greater, for the 5-task condition (6 ± 1 ms) than for the 3-task condition (1 ± 2 ms), $F(1,22) = 8.04$, $p = 0.01$. As the controlled data indicated no effect of number of tasks per se, we can attribute this additional difference only to the lower average frequency with which the uncontrolled tasks are performed in the 5-task condition. This difference in recency slope between the 5- and 3-task conditions was larger in the short CSI (slopes 12 ms and 5 ms, respectively; a difference of 7 ± 3 ms, $F(1,22) = 5.15$, $p = 0.033$)

than in the long CSI (slopes -1 ms and -3 ms, respectively; a difference of 2 ± 2 ms, $F(1,22) = 1.04$), though the 3-way interaction was not significant, $F(1,22) = 1.41$, $p = 0.247$. The error rate analysis yielded similar trends, i.e. larger recency slopes for the 5-task ($0.9 \pm 0.6\%$) than for the 3-task ($-0.2 \pm 0.2\%$) condition, $F(1,22) = 2.95$, $p = 0.1$; and a larger recency slope in short CSI ($0.9 \pm 0.6\%$) than long CSI blocks ($-0.2 \pm 0.2\%$), $F(1,22) = 3.63$, $p = 0.07$. For the linear trend of recency and its other interactions, $F < 1$.

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

Effects of response congruence In the 5-task condition, there were 5 possible levels of response congruence (defined by the number of irrelevant attributes mapped to the same/opposite response), ranging from completely congruent (all four irrelevant attributes mapped to the correct response) to completely incongruent (all four irrelevant attributes mapped to the opposite response). In the 3-task condition, there were only 3 levels of congruence (2, 1 or 0 irrelevant attributes mapped to the same response). Is there more detectable conflict when four irrelevant attributes are mapped to the wrong response than when only one or two are? To examine the congruence effect with 3 and 5 tasks independent of the recency and frequency of the tasks performed on the current and immediately preceding trial, the analysis was restricted to probe-probe sequences.

Figure 4 shows the effect of degree of congruence for each combination of switch/repeat and number of tasks, with the abscissa representing the proportion of attributes that are incongruent (thus aligning the extreme values for the 3- and 5-task conditions). Allowing for noise, the data suggest a graded effect of degrees of response congruence, reasonably well described by regression lines fitted to these functions. An ANOVA on the 3-

tasks RTs showed a significant effect of degree of congruence, $F(2,44) = 3.72$, $p = 0.032^5$, with a significant linear trend: $F(1,22) = 6.27$, $p = 0.02$, and no quadratic trend, $F < 1$. An ANOVA on the 5-tasks RTs showed a significant effect of degree of congruence, $F(4,88) = 3.14$, $p = 0.03$, with a significant linear trend: $F(1,22) = 10.14$, $p = 0.004$, and no other significant trend components, $F < 1$. An ANOVA on the 3-task error rates found no significant main effect of congruence, $F(2,44) = 1.23$, $p = 0.303$ (nor linear trend). An ANOVA on the 5-task error rates yielded no significant main effect of congruence, $F(4,88) = 2.13$, $p = 0.083$, but the linear trend was significant, $F(1,22) = 5.57$, $p = 0.028$.

--- Insert Figure 4 about here ---

To compare these effects across the 3- and 5- task conditions, we express the congruence effect as the “congruence slope”: the difference between the values estimated from the regression line for 0% congruent and 100% congruent estimated from the regression line. (I.e. we scale the x-axis so that the interval between perfect congruence and perfect incongruence is the same for 3 and 5 tasks). For RTs the slope was slightly larger for 5 tasks (-22 ± 7 ms) than for 3 tasks (-13 ± 5 ms), but this difference of 9 ± 9 ms was not significant, $F(1,22) = 1.13$. Unusually, the congruence effect was slightly larger on repeat (-24 ± 6 ms) than on switch (-11 ± 6 ms) trials, although the slope difference of 13 ± 10 ms was not significant, $F(1,22) = 2$, $p = 0.171$). The interaction between effects of number and switch approached significance, $F(1,22) = 3.6$, $p = 0.071$, reflecting a larger congruence slope on repeat trials for the 5-task (-41 ± 12 ms) than for the 3-task condition (-8 ± 9 ms), but not on switch trials, for which the 5-task slope was in fact smaller (-3 ± 11 ms) than the 3-task slope (-18 ± 6 ms).

For error rates there was no significant difference between the 5-task slope (-2.0%) and the 3-task slope (-1.1% ; a difference of $0.9 \pm 1.3\%$), $F < 1$. In contrast to the RT analysis,

⁵ Huyn-Feldt corrected

congruence slopes were slightly larger on switch (-2.6%) than on repeat trials (-0.5%; a difference of $2.1 \pm 1.4\%$), although this difference was also not significant, $F(1,22) = 2.03$ $p = 0.169$. Finally, there was no significant interaction between number and switch, $F < 1$, and the error rate pattern did not support the near-significant interaction in the RTs (i.e. congruence slopes for switch and repeat trials were -3.1% and -1.0%, respectively, in the 5-task condition; and -2.0% and -0.1%, respectively, in the 3-task condition).

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

N-2 Repetition effect As described in the Introduction, Emerson & Miyake (2003, Experiment 4) attributed their finding that switch costs were larger with fewer tasks, to the “n-2 repetition effect” – longer RTs for a task that was more recently switched away from. To look for an n-2 effect in our data, we conducted a one-way ANOVA (restricted to switch trials from the 3-task session) comparing trials on which participants performed a task they had more (ABA) or less (CBA) recently switched away from. There was no evidence of an n-2 repetition effect. Participants were actually faster on ABA (560 ms) than ABC (566 ms; a difference of 6 ± 5 ms) sequences, $F(1,23) = 1.11$ (ns), but made slightly more errors on ABA (5.3%) than ABC sequences (4.3%; a difference of $1 \pm 0.6\%$), $F(1,23) = 3.31$ (ns).

Summary Participants were required to switch among 3 or 5 tasks. For two of those tasks (probe tasks), recency and frequency were matched between the 3- and 5-task conditions. When the analysis was restricted to probe task trials preceded by probe task trials, there was no evidence that having a greater number of tasks in play made performance more difficult, or increased the difficulty either of switching or of preparing to switch to a new task. (In fact the overall error rate was slightly higher overall for 3 tasks, but we do not dwell on this

difference as it was not found in Experiment 2). The importance of the potential confound of number of tasks with recency or frequency was demonstrated by an analysis including trial sequences in which number of tasks was confounded with recency and frequency. Now RTs were significantly longer for unprepared switch trials in the 5-task than in the 3-task condition, giving the appearance of a greater difficulty of task-set retrieval with more tasks in play.

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

Experiment 2

Do these conclusions generalise to other pairs of tasks? One property of the tasks used in Experiment 1 is that they involved attention to different perceptual attributes of an object, and those attributes (head, tail, etc.) were separated by spatial location. The response congruence effects were quite small, suggesting that participants were quite good at filtering out the irrelevant attributes, even on task-switch trials. Hence it could be argued that this is a

situation where competition was not so much in the retrieval of the relevant task-set rules, or in attending to the relevant dimension, as in the potential targets for spatial attention (cf. Longman et al., 2012, 2013) — and we know that Hick's law does not apply to eye movements to a target (e.g. Kveraga, Boucher & Hughes, 2002). Experiment 2 therefore used a similar design, but with a word as the stimulus and an orthogonal set of five lexical classification tasks, for each of which attention was required to a different semantic/phonological property of the word, rather than a spatially distinct attribute of an object.

There were two other minor changes. In Experiment 1, the two tasks we selected as recency- and frequency-controlled probe tasks were the same for all participants, but in Experiment 2 each pair of tasks served equally often as probe tasks. The other change was that only one set of tasks was used in Experiment 2 for the 3- and 5-tasks conditions (because we found it hard to generate more than one set of five orthogonal task pairs applicable to words). Participants performed five tasks in one session, and a subset of three in the other session, with order counterbalanced. To minimise any effect in the second session of earlier exposure to (a sub-set of) the tasks, the sessions were about two weeks apart.

Method

Participants Twenty participants aged between 18 and 41 ($M = 22.3$), 5 male and 15 female, took part in two sessions separated by at least 11 ($M = 14$) days. Participants were paid between £12.80 and £16.00 (using the same performance-related incentive as in Experiment 1).

Design The stimuli were a set of 32 object names, each classifiable along five dimensions, four semantic properties and one phonological. The five possible classifications, and the associated pair of task cues⁶ used for each were:

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

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

Ten participants completed the 3-task condition in the first session and the 5-task condition two weeks later. The other ten had the 5-task condition first, and it was made clear in the second session that only three of the five tasks would ever be performed in the second session. In addition, in both sessions, before the practice blocks, participants were shown the list of 32 words and the tasks applicable for that session, to familiarise them with the words and their categorisations.

Otherwise, Experiment 2 was identical to Experiment 1 with respect to numbers of practice and experimental trials, the method of constructing trial sequences, and the alternation of long (1300 ms) and short (100 ms) CSI blocks.

--- *Insert Table 3 about here* ---

⁶ Using one of the two potential response category labels as the cue results in faster responses when the cue matches the response category (e.g. Monsell & Mizon, 2006; Schneider & Logan, 2005) but this effect is orthogonal to those of interest here, and does not appear to modulate the effectiveness of preparation.

Results

The data were analysed in the same way as for Experiment 1. Application of the same criteria as in Experiment 1 resulted in the exclusion of 0.4% of the correct responses from the computation of mean correct RTs.

--- Insert Figure 5 about here ---

Effect of number of tasks for the probe tasks Mean RT and error rates for the controlled task sequences, shown in the left panel of Figure 5, largely replicate the result obtained in Experiment 1 (cf. left panel of Figure 2). There was little sign of any overall effect of number of tasks in play and, more important, no evidence for RTs on unprepared switch trials being selectively prolonged by an increase in the number of tasks, as would be expected if promoting a task-set to most-active state (after stimulus onset) was more difficult with more task-sets active. The mean RT difference between the 5-task (852 ms) and the 3-task condition (843 ms), a difference of 9 ± 27 ms, was not significant, $F < 1$. The overall reduction in switch cost with opportunity for preparation (from 66 to 32 ms) was a significant 34 ms or 52%, $F(1,19) = 7.99$, $p = 0.011$. This RISC effect did not differ significantly between the 5-task (27 ± 17 ms) and 3-task conditions (42 ± 19 ms), $F < 1$ for the 3-way interaction. Indeed the difference was in the opposite direction to what would be expected if task-set retrieval were slower with more tasks to select from.

There were significant main effects of CSI, $F(1, 19) = 97.86$, $p < 0.001$, and switching, $F(1,19) = 20.12$, $p < 0.001$. The effect of a switch did not differ significantly between the 5-task (56 ± 14 ms) and 3-task conditions (42 ± 11 ms), $F < 1$. There was no significant interaction between the effects of CSI and number of tasks, $F < 1$.

In the analysis of error rates, only the switch cost ($1.9 \pm 0.5\%$) was significant, $F(1,19) = 13.31$, $p = 0.002$. There was a small reduction in switch cost with preparation ($1.1 \pm$

1.5%), but it was not significant, $F(1,19) = 1.57$, $p = 0.225$. For the effect of, and all interactions with, number of tasks, $F < 1$.

As for Experiment 1, we applied Bayesian statistics to the three effects of interest (main effect of number of tasks; number of tasks x switch interaction; number of tasks x switch x CSI interaction) to assess the support for the null hypothesis. As can be seen from Table 2, we obtained “positive” evidence in support of the null hypothesis for the main effect of number of tasks and the three-way interaction (number of tasks x CSI x switch). For the two-way interaction (number of tasks x switch), the posterior probability for H_0 was 0.74 (borderline between “weak” and “positive” support). For all three effects, the evidence in support of H_1 was not even “weak” (posterior probabilities ≤ 0.25).

These results, like those of Experiment 1, strongly suggest that the number of tasks in play does not affect task switching performance and preparation. We note that a) the analysis was powerful enough to detect other effects (switch costs, preparation effects, etc.); (b) the analyses including all trials detected substantial effects of number of tasks in spite of being “diluted” by the controlled trials; c) the Bayesian analysis provided “positive” evidence in support of the null hypothesis. However, better estimates can be obtained by combining the results of the two experiments. First we repeated the ANOVA described above (with the within-subjects factors number of tasks, CSI and switch/repeat), but with an added between-subjects variable of “experiment”. Again, we found no evidence that number of tasks affected performance in any way: for the relevant effects (number of tasks main effect; the two-way interaction with switch/repeat; the three-way interaction with switch/repeat and CSI), $F < 1$. Second, we then repeated the Bayesian analysis on this combined data set. As can be seen in Table 2, for each of the three effects of interest, the evidence in support of the null hypothesis is “positive”, and generally, the probability of the null relative to the

probability of an effect is better than one or both of the individual experiments. This more powerful combined analysis thus provides further support for the critical null-hypotheses.

Effect of number of tasks for all trials As for Experiment 1, we repeated the above analysis, but now including all trials. As may be seen from Figure 5 (right panel), the pattern of results closely resembled that from the equivalent analysis for Experiment 1 (cf. right panel of Figure 2). Mean RTs were now longer in the 5-task (870 ms) compared to the 3-task condition (839 ms), although this difference of 31 ± 27 ms was not significant, $F(1,19) = 1.38$, $p = 0.254$. The analysis also revealed a significant switch cost of 69 ± 10 ms, $F(1,19) = 46.31$, $p < 0.001$; a CSI effect of 179 ± 19 ms, $F(1,19) = 90.13$, $p < 0.001$, and a reduction in switch cost of 40 ± 8 ms or 45%, $F(1,19) = 22.47$, $p < 0.001$. As was the case in Experiment 1, the largest RT difference between the 5-task and the 3-task condition was on short CSI switch trials. This resulted in a larger switch cost in the 5-task (89 ± 14 ms) compared to the 3-task condition (49 ± 9 ms), $F(1,19) = 12.91$, $p = 0.002$, and a marginally larger CSI effect, $F(1,19) = 3.566$, $p = 0.074$. The RISC effect was also larger in the 5-task condition (49 ± 14 ms) compared to the 3-task condition (30 ± 12 ms), though the 3-way interaction was not significant, $F(1,19) = 0.81$, $p = 0.379$. Comparing the 5- and 3-task conditions for each combination of long and short CSI and task switch and repeat trials, only the 75 ± 33 ms difference for short CSI switch trials was significant, $F(1,19) = 5.26$, $p = 0.033$.

The error data analysis revealed only a significant switch cost of $2.8 \pm 0.4\%$, $F(1,19) = 42.33$, $p < 0.001$, and a marginally significant RISC of $1.1 \pm 0.6\%$, $F(1,19) = 3.88$, $p = 0.064$ (for main effect and all interactions with number of tasks, $F < 2.67$).

Effects of task recency When this analysis was restricted to probe task sequences (left panel of Figure 6), the linear effect of recency (slope of 9 ms) approached significance, $F(1,19) = 3.70$, $p = 0.070$. Although the recency slope was larger in the short CSI (11 ± 8 ms) than in the long CSI (7 ± 6 ms), the interaction was not significant, $F < 1$. The difference between the

5-task (8 ± 6 ms) and 3-task (10 ± 8 ms) recency slopes was also not significant, $F < 1$, and neither was the 3-way interaction, $F < 1$. (When the recency slopes for short and long CSI trials with 3 and 5 tasks were analysed individually, none of the linear trends were significant, all $F_s < 2.45$, ns.) In the error analysis, neither the linear trend of recency nor any of its interactions were significant, all $F_s < 1$.

When all switch trials were included in the analysis (right panel of Figure 6), the linear trend of recency (slope 5 ms) was significant, $F(1,19) = 5.00$, $p = 0.038$. As in Experiment 1, the recency slope was larger in the short CSI (9 ± 3 ms) than in the long CSI (2 ± 4 ms); and larger with 5 tasks (7 ± 4 ms) than with 3 (4 ± 4 ms), though neither interaction was significant, $F(1,19) = 2.78$, $p = 0.112$ and $F < 1$, respectively. The 3-way interaction was also not significant $F(1,19) = 2.79$, $p = 0.112$. However, separate 2 (5 or 3) \times 3 (lag) ANOVAs for each CSI showed that the recency slope was (almost) significantly larger with 5 tasks (14 ± 5 ms) than with 3 tasks (4 ± 4 ms) in the short CSI, $F(1,19) = 4.32$, $p = 0.051$, but not in the long CSI, $F < 1$. In the error data, the linear trend of recency and all interactions with this factor were not significant, $F < 1$.

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

--- *Insert Figure 6 about here* ---

Effects of response congruence It can be seen in Figure 7 that the effects of response congruence were much stronger than in Experiment 1. The 3-task RT ANOVA showed a

significant main effect of congruence (slope -91 ± 35 ms), $F(2,38) = 21.30$, $p < 0.001^7$, captured by the significant linear trend, $F(1,19) = 33.31$, $p < 0.001$ (quadratic trend $F < 1$). The 5-task RT ANOVA revealed a congruence slope of similar size (-103 ± 47 ms), $F(4,76) = 9.91$, $p < 0.001$, with a linear trend component, $F(1,19) = 23.78$, $p < 0.001$, and no other significant trend components ($F < 2.30$, ns). The 3-task error ANOVA showed a significant main effect of congruence (slope $-10\% \pm 3.1\%$), $F(2,38) = 35.27$, $p < 0.001$, with a linear trend, $F(1,19) = 51.88$, $p < 0.001$, and no quadratic trend, $F = 3.21$ (ns). A much smaller congruence slope was found in the 5-task error data (slope $-3.7 \pm 2.7\%$), though the main effect was still significant, $F(4,76) = 2.89$, $p = 0.036$, as was the linear trend, $F(1,19) = 9.83$, $p = 0.005$, (all other trend components, $F < 1$).

To compare congruence slopes between the 3-task and the 5-task conditions, 2 (3 or 5) x 2 (switch or repeat) ANOVAs were run on the RT and error congruence slopes. The RT congruence slope did not differ significantly between the 3-task (-91 ± 35 ms) and 5-task condition (-103 ± 47 ms), $F < 1$, or between switch (-85 ± 33 ms) and repeat trials (-109 ± 59 ms), $F < 1$. The interaction between effects of number of tasks and a task switch was also not significant, $F = 1.23$ (ns). The error congruence slope was much larger in the 3-task ($-10 \pm 3.1\%$) than in the 5-task condition ($-3.7 \pm 2.7\%$), $F(1,19) = 28.21$, $p < 0.001$, and it was of similar size for switch ($-7.8 \pm 4.0\%$) and repeat trials ($-5.9 \pm 2.7\%$), $F < 1$. The interaction was also not significant, $F < 1$.

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

N – 2 repetition effect As in Experiment 1, there was no evidence for an n-2 task repetition effect. Participants were actually faster on (854 ms) ABA than on CBA sequences (871 ms; a

⁷ Huyn-Feldt corrected

difference of 17 ± 11 ms), $F(1,23) = 2.60$ (ns), but they made slightly more errors on ABA (8.3%) than CBA sequences (7.3%; a difference of $1.0 \pm 0.7\%$), $F(1,23) = 1.98$ (ns).

--- *Insert Figure 7 about here* ---

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

In both experiments the probe task analysis was restricted to trials for which the preceding trial was also a probe task, in case the switch cost was affected by the recency/frequency of the preceding as well as the current task. To examine the contribution of the previous task we examined the difference in switch costs with 3 and 5 tasks for the following three trial types: probe-probe sequences (4/9 of all trials), sequences of probe task following any task (2/3 of all trials) and sequences of any task following any task (see Table 4). Switch costs were substantially and significantly larger with 5 tasks only for the last of these; there was little difference between the probe-probe and any-probe sequences. Apparently it was the frequency/recency of the *current* task that mattered, not that of the *preceding* task.

--- *Insert Table 4 about here* ---

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

General Discussion

The aim of these experiments was to help elucidate the process of promoting a task-set to the most-active state, which in many accounts is conceptualised as retrieving it (or its S-R rules) into working memory (Mayr & Kliegl, 2000) or, more specifically, a protected current-task buffer in procedural working memory (Oberauer's, 2009, "bridge"). In particular, we asked whether this process is influenced by the number of alternative task-sets from which the required task-set must be selected. Previous experiments have yielded inconclusive results, possibly because of lack of control of the recency and frequency with which a task was

performed, or because the "extra" tasks were not afforded by the same stimuli. In the present experiments, in separate sessions participants switched among three and five tasks afforded by the same stimuli and using the same responses. For two of the tasks ("probe tasks"), recency and frequency were matched between sessions. A natural expectation, based on other cases of memory retrieval, would be that retrieval time would increase as a function of the number of task-sets in play. This should be reflected in a larger switch cost when switching among five tasks, particularly at the short cue-stimulus interval, which allows no opportunity for advance preparation.

In both experiments, when the analysis was restricted to sequences of the two tasks matched for recency and frequency, number of tasks was found to have no effect on overall RT, on the switch cost, or on the reduction in switch cost with preparation. There was thus no sign that having to select from a larger set of tasks per se caused extra difficulty on unprepared switch trials. However, when we pooled over all trials, and hence included data from the tasks not matched for recency and frequency, a different result was obtained: now participants were slower with five tasks, particularly when there was no opportunity for advance preparation. Although the pattern of results in this analysis was slightly different between Experiment 1 (participants were exclusively slower on short CSI switch trials, as indicated by the significant 3-way interaction) and Experiment 2 (participants were substantially slower on short CSI switch trials, but also somewhat slower on short CSI repeat trials, hence no significant 3-way interaction), the contrast to the analysis restricted to probe tasks is marked. Thus, had we not controlled for recency and frequency, we would have been tempted to conclude that the time required to upload a task-set into PWM increases as a function of the number of alternative task-sets. Indeed, the same confound may have led others to conclude that it is harder to switch among a larger set of tasks (Buchler et al., 2008; Kray et al., 2002).

So what caused participants to be slower with in the 5-task condition in the uncontrolled data? Overall, the probability of a task switch was slightly higher with five tasks, but this is unlikely to have caused the result, as a higher proportion of switch trials normally results in smaller switch costs (Monsell & Mizon, 2006; Mayr, Kuhns & Rieter, 2012). To explore the possibility that differences in task recency caused the apparent effect of number of tasks, task switch trials were analysed as a function of the task's recency. For tasks matched for frequency (probe tasks) RTs increased slightly as a function of lag when there was no opportunity for preparation, suggesting that task recency does have a small effect on the difficulty of retrieval. For data including uncontrolled tasks, there was an additional effect for unprepared trials that we can attribute only to the effect of task frequency. Hence as well as demonstrating that there is no effect of number of tasks in play when recency and frequency are controlled, the data are consistent with the idea that task-set activation on a switch trial is easier when the task has been more recently and frequently performed. In addition, there was very little difference between the switch costs for probe task trials preceded by probe task trials, and probe task trials preceded by all task trials, suggesting that it was the recency/frequency of the current, not the preceding, task that mattered.

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

and less interference on incongruent trials. It remains that interference did not *increase* with the number of tasks, consistent with our finding that switch costs and its reduction with preparation is not larger with five tasks than with three.

The demonstration that promoting a task-set into the most-active state is no harder when there are more tasks in play is consistent with Oberauer's (2009) theory of procedural working memory, or at least with his speculation that, because the activated part of LTM which holds the currently irrelevant task-sets is capacity-unlimited, promoting one to the bridge should not take longer the greater the number of tasks in play. Oberauer made this prediction on the basis of his assumption that declarative and procedural WM operate by analogous principles. However, recent data from his laboratory on declarative working memory (Da Silva Souza et al., 2012) appear to violate this principle! In their experiment participants were cued to perform one of up to 3 mathematical tasks (big/small, inner/outer, high/low) on an item selected from up to 3 lists (presented on the top, centre and bottom of the screen) of 3 digits each. The number of lists and the number of tasks were varied independently between blocks. The position of the task cue on the screen indicated the relevant list and item within that list. Task switch costs and list switch costs were underadditive, supporting Oberauer's assertion that procedural and declarative representations are selected in parallel. However, contrary to expectations, whereas task-switch costs were not larger with three tasks than with two, list-switch costs were largest with three lists at a short CSI, suggesting a difference between declarative and procedural WM processes. Da Silva Souza et al. (2012) explain this difference by suggesting that whereas the capacity-limited declarative buffer may sometimes hold more than one list (hence the increase in list switch costs with the number of lists), the capacity-limited PWM buffer only ever holds one task at a time (hence no effect of the number tasks on the switch cost). But as Da Silva Souza et al. did not match the recency and frequency with which each list (or item

within each list) was retrieved, it remains possible that their number of lists effect for declarative WM results from a confound of the recency and frequency with which each list was accessed.

Our results also have more general implications for theories of task switching. In the Introduction we outlined accounts of task switching that have conceptualised the task-set selection process in terms of “activation” or “retrieval”. Neither kind of account makes explicit assumptions about what makes this task-selection process more or less difficult. Indeed, it is worth noting that although many theories of task switching assume that the process of “retrieving” a task-set contributes to the switch cost (e.g. Rubinstein et al.’s (2001) two-stage model, Mayr & Kliegl’s (2000) LTM hypothesis, Rogers & Monsell’s (1995) task-set reconfiguration theory), none of these theories describe the factors that constrain this retrieval process in any detail. Only Mayr & Kliegl explicitly sought to manipulate retrieval demand, but did so by comparing retrieval of qualitatively different kinds of task set. Hence, the factors influencing the process of retrieval/activation remain largely unknown. Here we have provided one important constraint on future theories: the process is (counterintuitively) not influenced by the number of tasks in play.

As a final observation, we note that there have been objections (e.g. Logan and Bundesen, 2003; Schneider and Logan, 2005) to our assumption that performance in the task-cuing paradigm involves the application on each trial of two levels of action control: selection of a task-set followed by selection of a response using the set of S-R rules specified by that task-set. Forrest, Monsell and McLaren (2014) defend this assumption in detail. We note here only that our finding of no effect of number of tasks in play on task-set retrieval provides one new reason for the distinction between the two levels, for selection of responses *within* a choice-reaction task-set *does* show effects of set-size (i.e. number of S-R rules), as reflected in Hick's law. One might, of course, wonder whether Hick's law too could arise

from a confound between recency/frequency and set size. This seems unlikely given the results of classic and more recent manipulations of probability and repetition in choice RT (Hyman, 1953; Kornblum, 1968; Schneider & Anderson, 2011). Moreover, van 't Wout (in preparation) has demonstrated that the Hick's law effect of number of S-R rules on choice RT within one task survives the same control of recency and frequency that we applied in the present experiments. Hence insensitivity to the number of alternatives does appear to be unique to task-set activation, and this fact needs to be accommodated by models of task-set control.

References

- Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press.
- Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review, 115*, 602-639.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*, 451-474.
- Arbuthnott, K., & Frank, J. (2000). Executive control in set switching: Residual switch costs and task-set inhibition. *Canadian Journal of Experimental Psychology, 54*, 33-41.
- Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associate information in memory: evidence from cued recall. *Journal of Memory and Language, 66*, 109-122.
- Brown, J. W., Reynolds, J. R., & Braver, T. S. (2007). A computational model of fractionated conflict-control mechanisms in task-switching. *Cognitive Psychology, 55*, 37-85.
- Buchler, N. G., Hoyer, W. J., & Cerella, J. (2008). Rules and more rules: The effect of multiple tasks, extensive training, and aging on task-switching performance. *Memory & Cognition, 36* (4), 735-748.
- Criss, A. H., & Shiffrin, R. M. (2004). Pairs do not suffer interference from other types of pairs or single items in associate recognition. *Memory & Cognition, 32*, 1284-1297.
- Da Silva Souza, A., Oberauer, K., Gade, M., & Druey, M. D. (2012). Processing of representations in declarative and procedural working memory. *The Quarterly Journal of Experimental Psychology, 65* (5), 1006-1033.

- Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*, 257-303.
- Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, J., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General*, *137* (1), 131-148.
- Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*, 148-168.
- Forrest, C.L.D., Monsell S., & McLaren I.P.L. (in preparation). Is performance in task-cuing experiments mediated by task-set selection or associative compound retrieval?
- Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*, 297-337.
- Goschke, T. (2000). Intentional reconfiguration and involuntary persistence in task set switching. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 331-355). Cambridge, MA: MIT Press.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11-26.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *53*, 188-196.
- Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, *90*, 928-934.
- Kessler, Y., & Meiran, N. (2010). The reaction-time task-rule congruency effect is not affected by working memory load: Further support for the activated long-term memory hypothesis. *Psychological Research*, *74*, 388-399.

- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin*, *136* (5), 849-874.
- Kleinsorge, T., & Apitzsch, N. (2012). Task preparation based on precues versus memory: Precues lead to superior performance with two tasks but not with four tasks. *Journal of Cognitive Psychology*, *24* (2), 140-156.
- Kornblum, S. (1968). Serial-choice reaction time: Inadequacies of the information hypothesis. *Science*, *159*, 432-434.
- Kray, J., Li, K. Z., H., & Lindenberger, U. (2002). Age-related changes in task switching components: The role of task uncertainty. *Brain and Cognition*, *49*, 363-381.
- Kveraga, K., Boucher, L., M., & Hughes, H. C. (2002). Saccades operate in violation of Hick's law. *Experimental Brain Research*, *146*, 307-314.
- Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology*, *8*, 311-335.
- Lien, M. C., Ruthruff, E., Remington, R. W., & Johnston, J. C. (2005). On the limits of advance preparation for a task switch: Do people prepare all the task some of the time or some of the task all the time? *Journal of Experimental Psychology: Human Perception and Performance*, *31* (2), 299-315.
- Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance*, *29* (3), 575-599.
- Longman C. S., Lavric, A., & Monsell, S. (2013). More attention to attention? An eye-tracking investigation of selection of perceptual attributes during a task switch. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

- Longman, C., Lavric, A., Muteanu, C., & Monsell, S. (in press). Attentional inertia and delayed orienting of spatial attention in task-switching. *Journal of Experimental Psychology: Human Perception and Performance*.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679-690.
- Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, 129, 4-26.
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 1124-1140.
- Mayr, U., Kuhns, D., & Rieter, M. (2012, July 30). Eye movements reveal dynamics of task control. *Journal of Experimental Psychology General*. Advance online publication. doi: 10.1037/a0029353
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22 (6), 1423-1442.
- Meiran, N. (2000). Modeling cognitive control in task-switching. *Psychological Research*, 63, 234-249.
- Monsell, S. (1984). Components of working memory underlying verbal skills: A 'distributed capacities' view—A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 327–350). Hillsdale, NJ: Erlbaum.
- Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved Mysteries of the Mind* (pp. 93-148). Hove, E. Sussex: Erlbaum (UK) Taylor and Francis.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7 (3), 134-140.
- Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance*, 32 (3), 493-516.

- Murdock, B. B. Jr. (1962). The serial position effect in free recall. *Journal of Experimental Psychology*, 64, 482-488.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In Davidson, R. J., Schwartz, G. E., and Shapiro, D., (eds), *Consciousness and Self-Regulation: Advances in Research and Theory*. Plenum Press.
- Oberauer, K. (2009). Design for a working memory. *Journal of Learning and Motivation*, 51, 45-100.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111-196). Cambridge: Blackwell.
- Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian information criterion for model selection.". *Sociological Methods & Research*, 27, 411–427.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124 (2), 207-231.
- Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task switching paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1477-1491.
- Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27 (4), 763-797.
- Schneider, D. W., & Anderson, J. R. (2011). A memory-based model of Hick's law. *Cognitive Psychology*, 62, 193-222.
- Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: A short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General*, 134, 343-367.

- Schuch, S., & Koch, I. (2003). The role of response selection for inhibition of task sets in task shifting. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 92-105.
- Shaffer, L. H. (1965). Choice reaction with variable S-R mapping. *Journal of Experimental Psychology*, 70, 284-288.
- Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 89-103.
- Van 't Wout, F. (in preparation). The contribution of stimulus recency and frequency to Hick's law.
- Van 't Wout, F., Monsell, S., & Lavric, A. (in preparation). The effect of task-set complexity on preparation for a task switch: Evidence for a limited capacity component of procedural working memory.
- Vandierendonck, A., Liefoghe, B., & Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychological Bulletin*, 136 (4), 601-626.

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

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

	$p(H_0 D)$	$p(H_1 D)$
Experiment 1		
Number of tasks	0.83	0.17
Number of tasks x switch	0.83	0.17
Number of tasks x switch x CSI	0.81	0.19
Experiment 2		
Number of tasks	0.81	0.19
Number of tasks x switch	0.74	0.26
Number of tasks x switch x CSI	0.79	0.21
Experiments 1 & 2 combined		
Number of tasks	0.87	0.13
Number of tasks x switch	0.81	0.19
Number of tasks x switch x CSI	0.83	0.17

Table 2 Posterior probabilities for H_0 and H_1 given the data in Experiment 1 and 2.

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

Table 3 The 32 word stimuli used in Experiment 2, and their respective classifications along the five dimensions.

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

Table 4 Differences between the 5-task and 3-task switch costs in ms (% correct) for the following trial sequences: probe task following probe task (PP), probe task following any task (AP) and any task following any task (AA). * = $p < 0.05$.

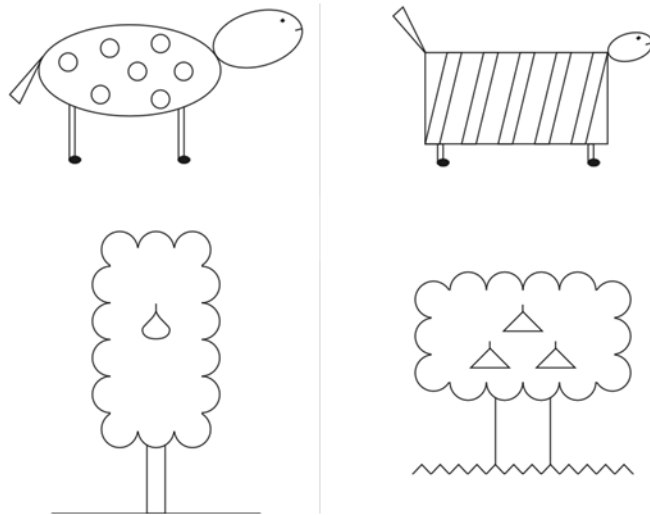


Figure 1 Examples of the two sets of stimuli used in Experiment 1. The “animal” and “tree” stimuli on the left required a left response in all tasks and differ on all five dimensions from the stimuli on the right, which therefore required a right response in all tasks.

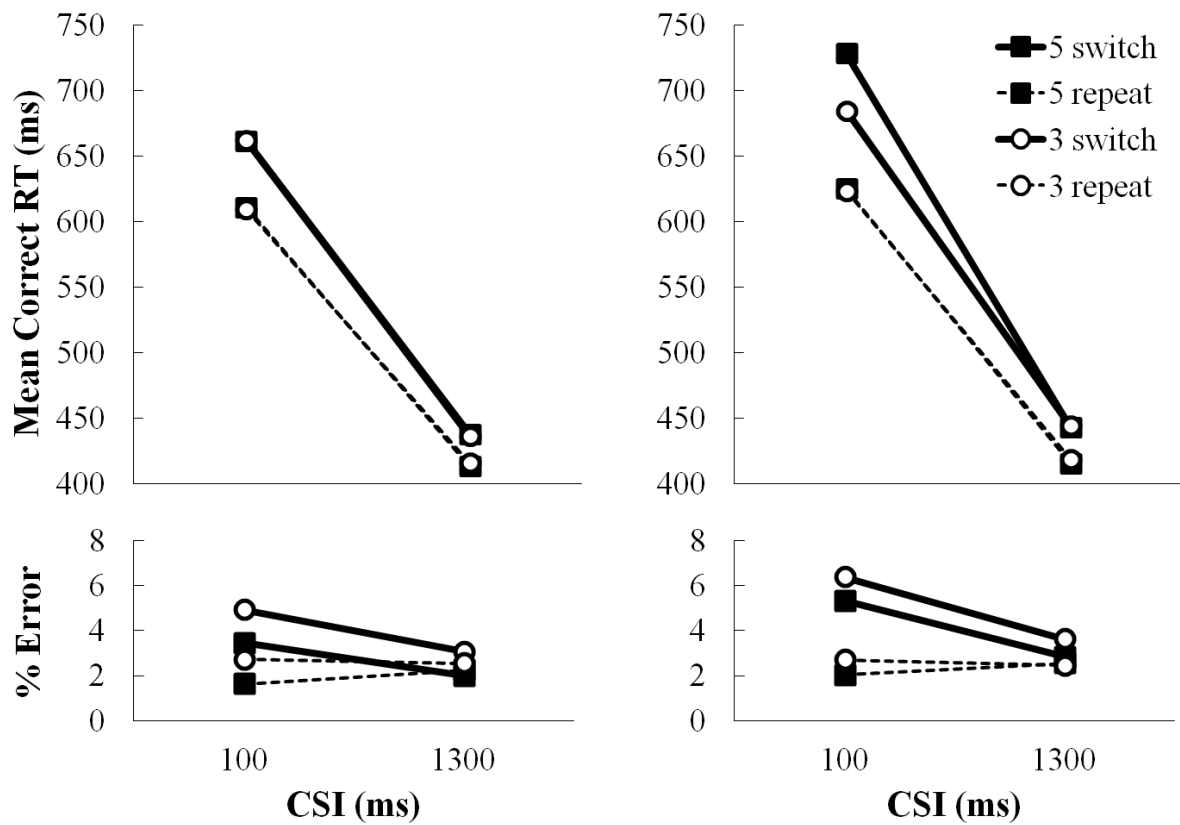


Figure 2 Experiment 1: Mean correct RT (top) and % error (bottom) data with three and five tasks, on switch and repeat trials, as a function of CSI, plotted separately for probe task trials with frequency and recency matched (left) and all trials (right).

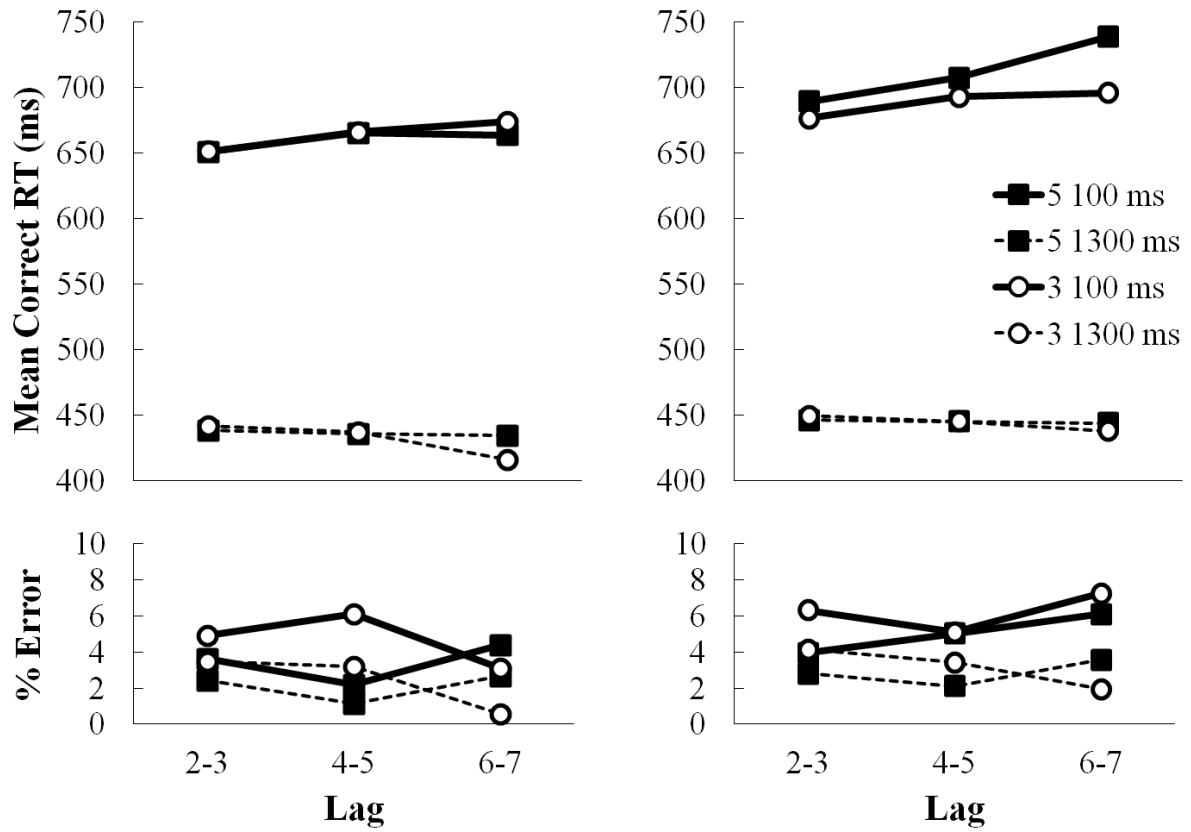


Figure 3 Experiment 1: Mean correct RTs (top) and % error (bottom) data with three and five tasks, on long and short CSI switch trials, as a function of the lag (how many trials back the same task was last performed), plotted separately for probe task trials (left) and all task trials (right).

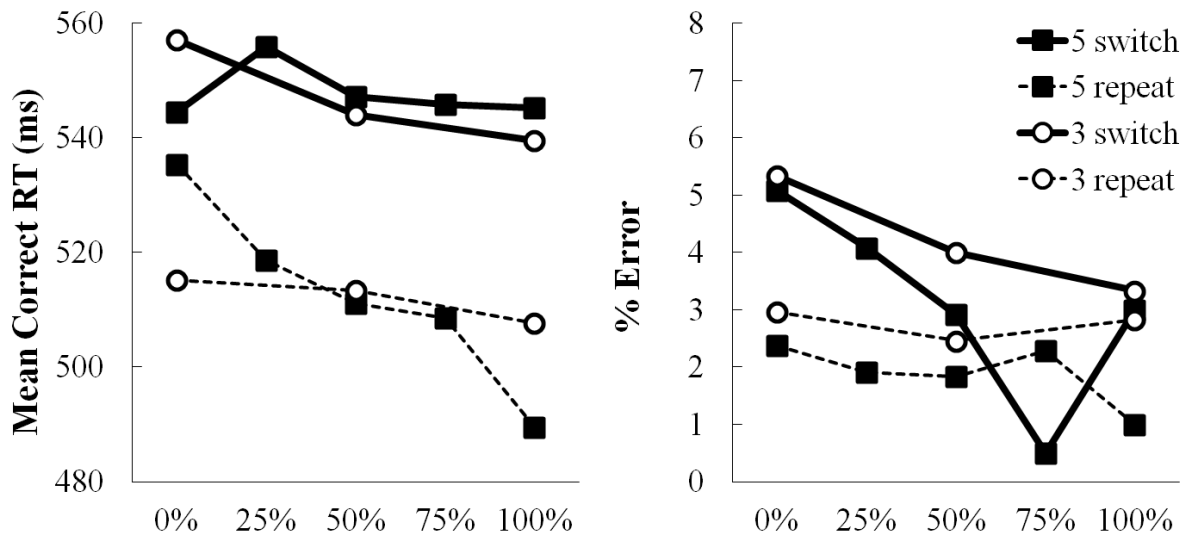


Figure 4 Experiment 1: Mean correct RTs (top) and % error (bottom) data with three and five tasks, on switch and repeat trials, as a function of response congruence: the proportion of irrelevant dimensions mapped to the correct response.

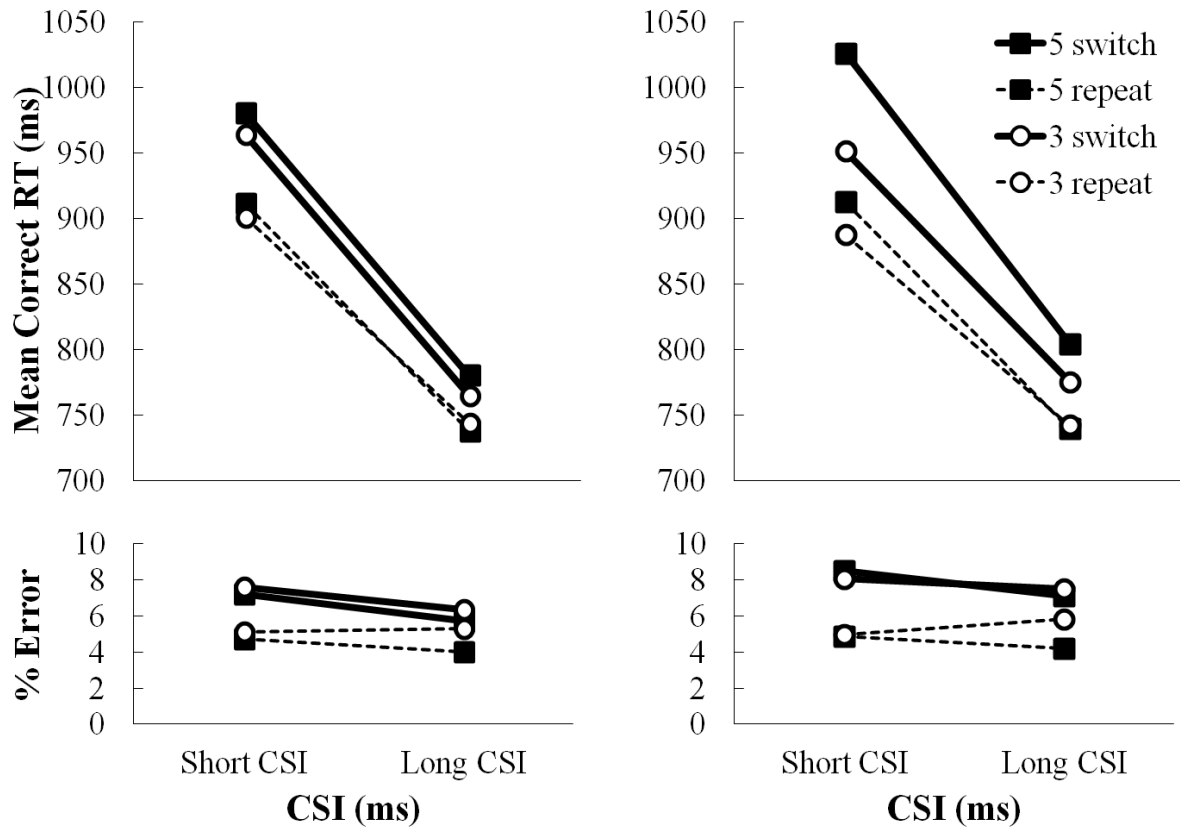


Figure 5 Experiment 2: Mean correct RT (top) and % error (bottom) data with 3 and 5 tasks, on switch and repeat trials, as a function of CSI, plotted separately for probe task trials with frequency and recency matched (left) and all trials (right).

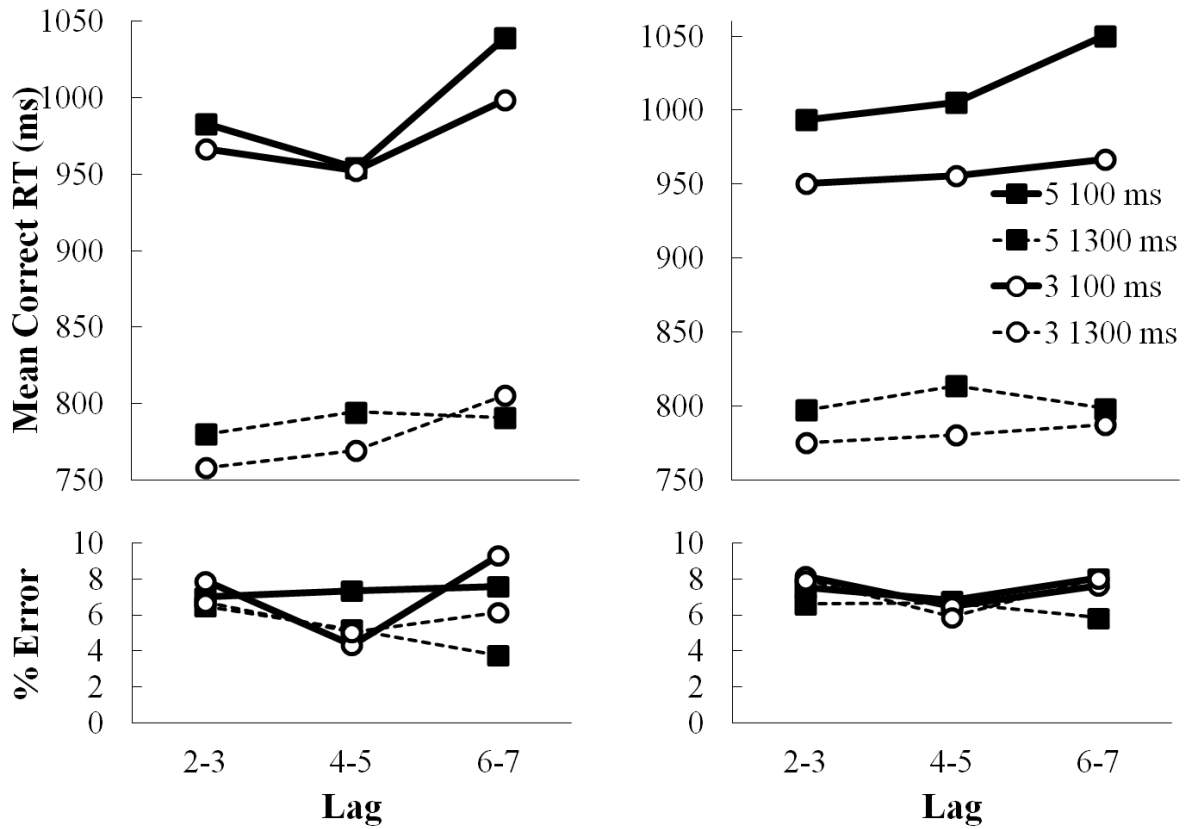


Figure 6 Experiment 2: Mean correct RTs (top) and % error (bottom) data with three and five tasks, on long and short CSI switch trials, as a function of the lag (how many trials back the same task was last performed), plotted separately for probe task trials (left) and all task trials (right).

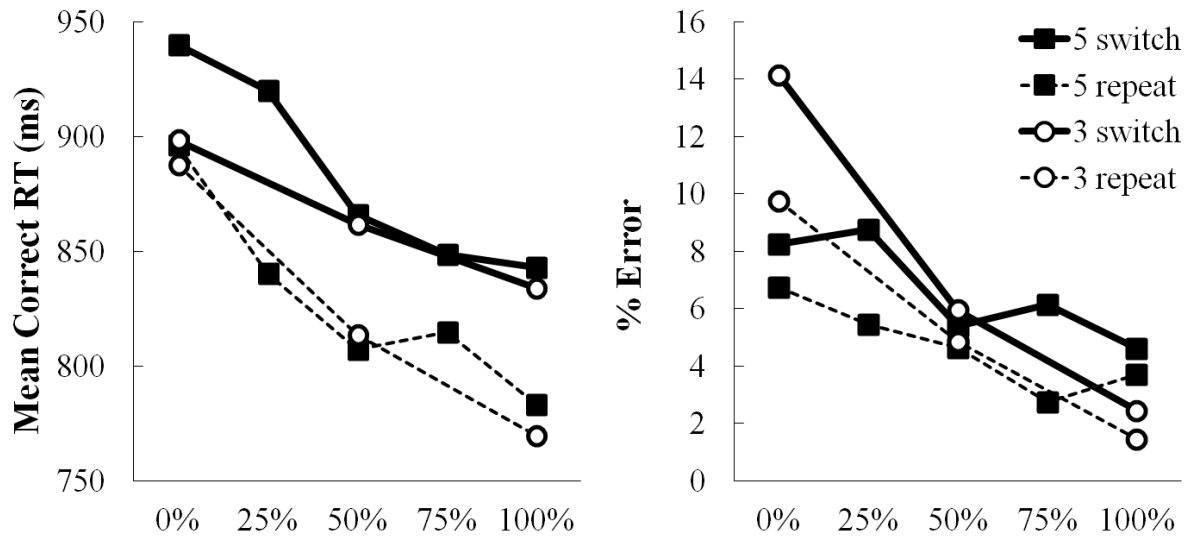


Figure 7 Experiment 2: Mean correct RTs (top) and % error (bottom) data with three and five tasks, on switch and repeat trials, as a function of response congruence: the proportion of irrelevant dimensions mapped to the correct response.