

Out with the old, in with the new: More valid measures of switch cost and retrieval time in the task span procedure

GORDON D. LOGAN

Vanderbilt University, Nashville, Tennessee

Two experiments provided new measures of switch cost and retrieval time in the task span procedure. In Experiment 1, subjects were given lists of six task names to remember, followed by six targets on which to perform the tasks named in the list. The lists contained alternations and repetitions, and switch costs were estimated by comparing reaction time (RT) on alternation and repetition trials. The experiment also included memory span and single task conditions, so switch costs could be estimated by subtracting the sum of the RT in those conditions from the task span RT, as in the original report (Logan, 2004). The data suggested that the original measure of switch cost was invalid and that the new measure was preferable. In Experiment 2, subjects performed each task on the list twice. Retrieval was required on the first but not on the second trial in each pair. Retrieval time was estimated by comparing the RT on trials that required retrieval with trials that did not require retrieval. This measure was more valid than the RT in the memory span condition of Experiment 1, which was used in the original report.

Task-switching experiments are a popular way to investigate executive control (Monsell, 2003). In many procedures, reaction time (RT) and error rate are higher when subjects switch between tasks than when they repeat tasks. These *switch costs* are robust and easy to replicate, but their interpretation is controversial. Switch costs may reflect top-down control processes or bottom-up facilitation (Logan & Bundesen, 2003) or interference from previous trials (Allport, Styles, & Hsieh, 1994). To address this controversy, Logan (2004) developed the *task span procedure*, illustrated in Figure 1. Subjects receive a study list of task names to remember, followed by a test list of target stimuli. They perform each task on the list on the corresponding target. This procedure requires top-down control. Subjects must initiate retrieval of task names and initiate task switches when task names change. Logan reported two measures of switch costs that provide insight into these top-down control processes. The measures require strong assumptions. The purpose of this article is to test those assumptions and provide alternative measures of switch costs in the task span procedure.

Logan (2004) first measured switch costs by comparing RTs in the task span procedure with the sum of RTs in a memory span condition and a single task condition. Task span RTs include the time required to retrieve the next task name, switch tasks, and then perform the task on the current target. Memory span RT measures the time re-

quired for retrieval, and single task RT measures the time required to perform the task. The difference between the sum of these RTs and the task span RT measures switch cost, which was around 500 msec. This measure assumes that memory span and task span retrieval processes are the same and that single task and task span target processes are the same. Experiment 1 of the present article tested the validity of these assumptions.

Logan (2004) also measured switch costs by requiring subjects to perform each task 2, 3, or 4 times before proceeding to the next task. The first trial in each series required subjects to retrieve the next task name on the list, switch tasks, and perform the task on the current target. Subsequent trials required subjects only to perform the established task on the current target. The difference in RT between the first and subsequent trials, which was about 1,700 msec, included retrieval time and task-switching time, but failed to separate the two. The present Experiment 2 provides a method for separating retrieval time from task-switching time.

Logan's (2004) experiments did not allow comparisons of task repetitions and alternations within runs of trials. Lists were constructed by concatenating random orders of three task names; repetitions could occur only on Trials 4, 7, and 10 within a run and occurred on only one third of those trials. The present experiments allowed comparisons of repetitions and alternations within runs of trials. Subjects were given lists of six task names, three from each of two tasks. There are 20 ways to order three sets of two task names (see Table 1). Across lists, repetitions occurred on 40% of the trials, equally often in each serial position. This procedure provides a direct measure of switch costs—the difference in RT on repetition as opposed to alternation trials—that can be compared with Logan's measure.

This research was supported by National Science Foundation Grant BCS 0133202. I am grateful to Julie Delheimer for testing the subjects and analyzing the data. Correspondence concerning this article may be addressed to G. D. Logan, Department of Psychology, Vanderbilt University, Nashville, TN 37203 (e-mail: gordon.logan@vanderbilt.edu).

Table 1
Lists of Task Names Used in the Experiments

Original Sequence	Opposite Sequence	Sequential Effect	Number of Repetitions	Number of Alternations
HHHOOO	OOHHHH	RRARR	4	1
HOOOHH	OOHHHO	RARRA	3	2
HOOHHH	OOHHOO	ARRAR	3	2
HOOHOO	OOHOHH	RAAAR	2	3
HOOOHO	OOHHOH	RARAA	2	3
HOHHOO	OHOHHH	AARAR	2	3
HOHHOO	OHOHHH	ARARA	2	3
HOHOHH	OHOHHO	AAARA	1	4
HOHOHH	OHOHHO	ARAAA	1	4
HOHOHO	OHOHOH	AAAAA	0	5

Note—H, Hi—Lo task; O, Odd—Even task; R, repetition; A, alternation.

approximately 60 cm. Responses were collected from the numeric keypads.

Procedure. Each set of trials began with a warning display indicating the start of the study phase (*Study phase: 6 trials*). It was displayed until the subjects pressed a key. Then six task names were presented at a rate of 1 per sec (500 msec per name and then a 500-msec blank screen). Then a 1,000-msec warning display appeared, indicating the start of the test phase (*Test phase: 6 trials*). It was followed by a 100-msec blank screen, and then a digit that was exposed until the subject responded. Then the screen was blank for 100 msec, and the next digit appeared. This continued until six digits had been presented, whereupon the warning display for the next study phase appeared. There were 40 study–test blocks in each condition.

In the task span condition, subjects were told to remember the names on the study list and perform each task on the corresponding digit in the test phase. In the memory span condition, subjects were told to remember the names on the study list and then report them, in order, in response to digit prompts in the test phase. In the single task condition, subjects saw six repetitions of one of the task names in the study phase and were told to perform that task on each digit in the test phase. There were 20 different lists of task names in the task span and memory span conditions (see Table 1), and each was presented twice in random order in the two conditions. In the single task condition, subjects saw one task name for the first 20 sets of trials and the other task name for the last 20 sets. The six targets for each set of test trials were selected at random without replacement from eight possible targets and were presented in random order.

In the task span and single task conditions, all of the subjects pressed the 1 key for “high” and “odd” and the 3 key for “low” and “even.” In the memory span condition, all of the subjects pressed the 7 key for “Hi—Lo” and the 9 key for “Odd—Even.” Subjects were told to respond as quickly and accurately as possible. All subjects began with the single task conditions to familiarize them with the tasks and the sequence of study–test trials. Half of the subjects did the Hi—Lo task before the Odd—Even task, and half did the opposite. After the single task trials, half of the subjects performed the task span procedure before the memory span procedure, and half did the opposite.

Results and Discussion

The probability of performing correctly on all six test trials— $P(\text{Perfect})$ —was .913 for the single task condition, .851 for the memory span condition, and .776 for the task span condition. A one-way ANOVA revealed a significant main effect of condition [$F(2,30) = 7.74, MS_e = .009, p < .01$]. Nonorthogonal contrasts showed that $P(\text{Perfect})$ was significantly worse in the task span condition than in the memory span condition [$F(1,30) = 5.00, p < .05$] and the single task condition [$F(1,30) = 16.68, p < .01$];

the difference between the memory span and single task conditions was not significant [$F(1,30) = 3.42, p = .07$].

Mean RTs for repetitions and alternations were calculated for each condition for each subject, averaging over the second and subsequent trials in each test block. The accuracy of each response was determined independently. Only RTs from correct responses were included in the means. The means across subjects are plotted as a function of condition in Figure 2. Accuracy scores also appear in the figure. Accuracy was high, and there was no evidence of a speed–accuracy trade-off, so the analyses focus on RT.

Repetitions and alternations occurred in the task span and memory span conditions. Only repetitions occurred in the single task condition. In the task span condition, alternations were 378 msec slower than repetitions, similar to switch costs observed in many procedures (Monsell, 2003). In the memory span condition, alternations were only 38 msec slower than repetitions, which suggests that retrieving a different task name contributes little to switch costs in the task span procedure.

The significance of these differences was assessed by calculating contrasts from a one-way ANOVA on the mean RTs in the five conditions (single task, memory span repetition, memory span alternation, task span repetition, task span alternation). The overall effect of condition was significant [$F(4,60) = 142.68, MS_e = 31,524.76, p < .01$]. A contrast comparing repetitions with alternations across the task span and memory span conditions was significant [$F(1,60) = 498.44, p < .01$; the main effect of repetition], as was a contrast comparing the task span condition with the memory span condition [$F(1,60) = 21.96, p < .01$; the main effect of task span vs. memory span], and a contrast comparing the difference between repetitions and alternations in the task span and memory span conditions [$F(1,60) = 14.67, p < .01$; the interaction between repetition and task]. Fisher’s least significant difference (LSD) for $p < .05$ showed that the difference between repetitions and alternations was significant in the task span condition but not in the memory span condition. Nonorthogonal contrasts showed that the single task condition was faster than the task span condition [$F(1,60) = 212.78, p < .01$] but slower than the memory span condition [$F(1,60) = 13.27, p < .01$].

Logan (2004) estimated switch costs by subtracting the sum of single task and memory span RTs from task span RT. The present experiment tests the validity of that estimate. The sum of single task and memory span RTs for repetitions and alternations are plotted in Figure 2. The sum was 580 msec smaller than task span RT on alternation trials, similar to the 479 msec difference Logan observed. However, the sum was 240 msec smaller than task span RT on repetition trials, which is significant at $p < .05$ according to Fisher’s LSD test and inconsistent with the hypothesis that task span repetition RT equals the sum of single task and memory span RT.

Serial position effects provide a further test of the validity of Logan’s (2004) measure of switch cost. Mean RTs for repetition and alternation trials are plotted as a function

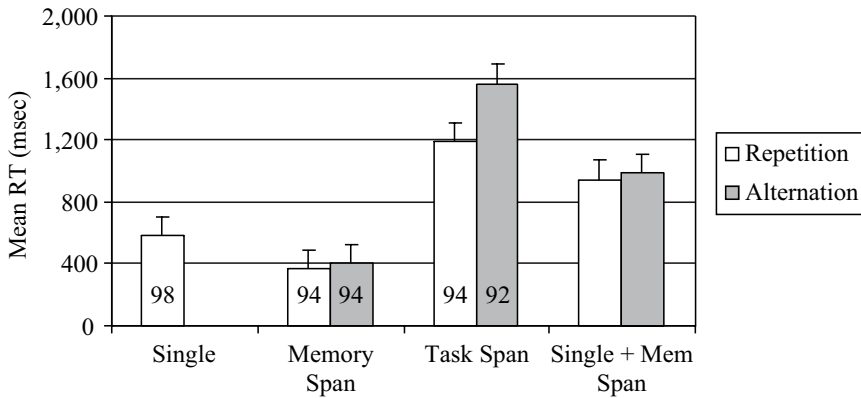


Figure 2. Mean reaction time (RT) for repetitions and alternations in single task, memory span, and task span conditions in Experiment 1. Both repetitions and alternations occur in the memory span and task span conditions. Only repetitions occur in the single task condition. The error bars represent the 95% confidence intervals for the means based on Fisher's least significant difference calculated from a one-way ANOVA on the five conditions (single task, memory span repetition, memory span alternation, task span repetition, and task span alternation). The numbers inside the bars are the percentage of correct responses in the corresponding condition. Also plotted is the sum of single task and memory span RT for repetitions and alternations.

of Serial Positions 2–6 for each condition in Figure 3. The first serial position is excluded from these analyses because RTs were very long (averaging 1,815, 828, and 842 msec for task span, memory span, and single task, respectively), and because repetitions and alternations cannot be defined. Figure 3 shows a concave-upward serial position effect in the task span condition and relatively flat serial position effects for single task and memory span conditions. No combination of single task and memory span RTs can produce the upward concavity observed in the task span RTs, which invalidates the hypothesis that task span RT is the sum of single task and memory span RTs.

The significance of these differences was tested with contrasts constructed from a 5 (condition: single task, memory span repetition, memory span alternation, task span repetition, task span alternation) \times 5 (serial position: 2, 3, 4, 5, and 6) ANOVA on the mean RTs, using the error term for condition \times serial position interaction ($MS_e = 28,201.93$). The quadratic trend in the serial position effect was strong and significant in the task span condition [$F(1,240) = 60.22, p < .01$], weak but significant in the memory span condition [$F(1,240) = 4.92, p < .05$], and nonsignificant in the single task condition [$F(1,240) = 0.11, p = .74$]. The quadratic trend in the task span condi-

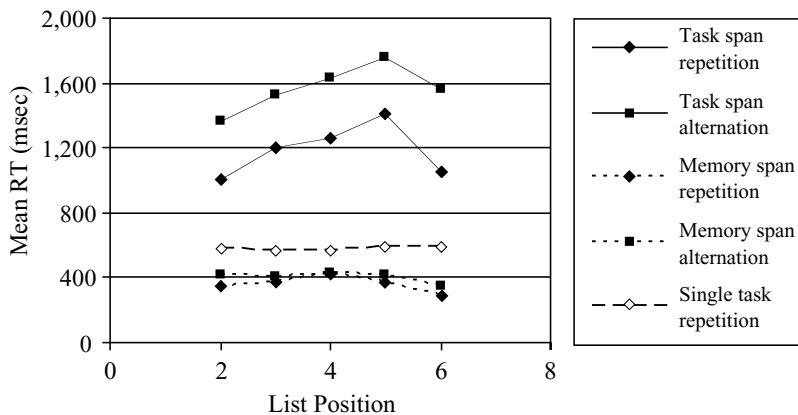


Figure 3. Mean reaction time (RT) for repetitions in single task, memory span, and task span conditions as a function of serial position in Experiment 1. Both repetitions and alternations occur in the memory span and task span conditions. Only repetitions occur in the single task condition. RTs for the first serial position are excluded, because they are exceptionally long and because they cannot be classified as repetitions or alternations.

tion was significantly stronger than the quadratic trend in the memory span condition [$F(1,240) = 15.35, p < .01$] and the single task condition [$F(1,240) = 22.62, p < .01$].

The data suggest that retrieval was different in the task span and memory span conditions. In the task span condition, subjects retrieved the task names one at a time, responding to targets between successive retrievals. In the memory span condition, subjects retrieved all of the task names at the start of the test phase, queuing them for report. Indeed, RT was very fast in the memory span condition, averaging 383 msec.

EXPERIMENT 2

Experiment 2 separated switch cost from name retrieval time. Subjects were given lists of six task names, followed by a series of 12 targets. They were told to perform each task on the list on 2 successive targets. If their study list was Hi-Lo, Odd-Even, Hi-Lo, Odd-Even, Hi-Lo, and Odd-Even, they performed the magnitude task on the first and second targets, the parity task on the third and fourth targets, and so on. The first trial in each pair of test trials requires retrieval of a task name. The trial is a *retrieved repetition* if the task name is the same as the previous one and a *retrieved alternation* if the task name is different. The second trial in each pair does not require retrieval because subjects repeat the task from the previous trial. These trials are *instructed repetitions*.

The contrast between instructed repetitions and retrieved repetitions provides a measure of retrieval time. Instructed repetitions do not require retrieval or task switching, whereas retrieved repetitions require retrieval but not task switching. Thus, differences in RT measure retrieval time. Retrieved alternations require retrieval and task switching, so the contrast between retrieved alterna-

tions and retrieved repetitions reflects the switching time, as in Experiment 1.

Method

Subjects. Sixteen subjects from the same pool served in a 1-h session for pay.

Apparatus and Stimuli. These were the same as in Experiment 1.

Procedure. The procedure was the same as in the task span condition in Experiment 1 except that there were 12 test trials instead of 6. There were 40 study-test blocks; each of the lists in Table 1 was presented twice, in random order.

Results and Discussion

Mean $P(\text{Perfect})$ was .734. This is approximately the same as $P(\text{Perfect})$ in the task span condition of Experiment 1, even though subjects performed the tasks twice as often and took twice as long to complete the test phase. This suggests that there is little trade-off between processing and storage in the task span procedure (Logan, 2004).

Mean RT was calculated for correct responses for retrieved repetitions, retrieved alternations, instructed repetitions following retrieved repetitions, and instructed repetitions following retrieved alternations. The first pair of trials in each series of test trials was excluded, because it could not be classified as a repetition or an alternation. The means across subjects and the accuracy scores appear in Figure 4. Accuracy was high, and there was no speed-accuracy trade-off, so the analyses focus on RT.

The RT data replicated the switch costs seen in Experiment 1. Retrieved alternations were 284 msec slower than retrieved repetitions. The RT data also provide a measure of retrieval time. Retrieved repetitions were 489 msec slower than the instructed repetitions. This difference is more than 100 msec longer than the memory span RTs in Experiment 1, which included time for response selec-

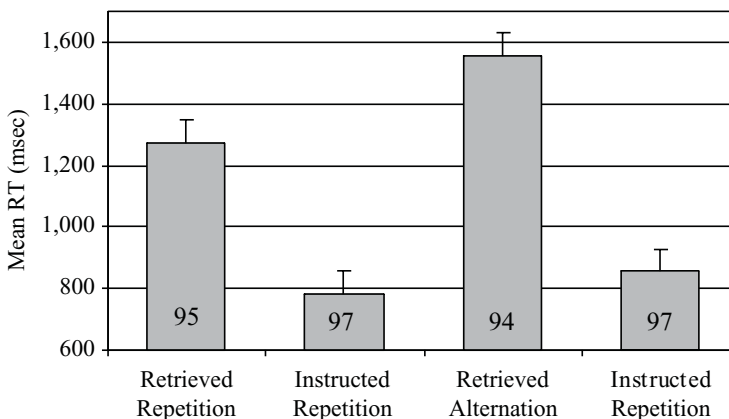


Figure 4. Mean reaction time (RT) for retrieved repetitions, instructed repetitions following retrieved repetitions, retrieved alternations, and instructed repetitions following retrieved alternations (from left to right) in Experiment 2. The error bars represent the 95% confidence intervals for the means, based on Fisher's least significant difference calculated from the interaction between repetition versus alternation and retrieve versus repeat. The numbers inside the bars are the percentage of correct responses in the corresponding condition.

tion and execution as well as retrieval. Thus, memory span RTs may underestimate retrieval time in the task span procedure. Instructed repetitions were 71 msec slower after retrieved alternations than after retrieved repetitions. This difference may reflect a carryover of interference from the retrieved alternation trial (Allport et al., 1994).

These conclusions were supported by a 2 (retrieval: retrieve vs. repeat) \times 2 (repetition: repetition vs. alternation) ANOVA on the mean RTs. There were significant main effects of retrieval [$F(1,15) = 125.21$, $MS_e = 45,300.41$, $p < .01$] and repetition [$F(1,15) = 35.82$, $MS_e = 14,106.97$, $p < .01$] and a significant interaction between them [$F(1,15) = 18.64$, $MS_e = 9,785.00$, $p < .01$]. A contrast comparing retrieved repetitions with retrieved alternations was highly significant [$F(1,15) = 65.94$, $p < .01$]; a contrast comparing instructed repetitions after retrieved repetitions and retrieved alternations approached significance [$F(1,15) = 4.12$, $p = .06$].

GENERAL DISCUSSION

Experiment 1 showed that the assumptions underlying the subtractive method Logan (2004) used to estimate switch costs in the task span procedure are not valid. Task span RT for repetition trials was larger than the sum of single task and memory span RTs, though the assumptions predict no difference. Also, task span RTs showed serial position effects that could not be produced by combining single task or memory span RTs. Experiment 1 provided an alternative measure of switch costs—the difference between task repetition and task alternation RT—that produced values typical of the literature (Monsell, 2003).

Experiment 2 adapted Experiment 1's method to separate switch costs from retrieval time. Switch costs were similar to those in Experiment 1, but the retrieval times were substantially longer than the memory span RTs in Experiment 1, providing further evidence that memory span RT underestimates retrieval time in the task span procedure.

Together, the two experiments provide new measures of switch cost and retrieval time to replace the old measures which they showed were invalid. These measures strengthen the interpretation of performance in the task span procedure, allowing it to provide better insight into top-down executive control processes. With these measures, the task span procedure may advance our understanding of executive control in many areas of psychology, including cognitive science, cognitive neuroscience, clinical science, individual differences, and lifespan development.

REFERENCES

- ALLPORT, D. A., STYLES, E. A., & HSIEH, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: MIT Press.
- LOGAN, G. D. (2004). Working memory, task switching, and executive control in the task span procedure. *Journal of Experimental Psychology: General*, **133**, 218-236.
- 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 & Performance*, **29**, 575-599.
- MONSELL, S. (2003). Task switching. *Trends in Cognitive Sciences*, **7**, 134-140.

(Manuscript received April 15, 2005;
revision accepted for publication June 13, 2005.)