

Working Memory, Task Switching, and Executive Control in the Task Span Procedure

Gordon D. Logan
Vanderbilt University

Four experiments explored the *task span procedure*: Subjects received lists of 1–10 task names to remember and then lists of 1–10 stimuli on which to perform the tasks. Task span is the number of tasks performed in order perfectly. Experiment 1 compared the task span with the traditional memory span in 6 practiced subjects and found little difference. Experiment 2 compared the task span and the memory span in 64 unpracticed subjects and also found little difference. Experiment 3 compared practice with consistent and varied lists to address retrieval from long-term memory. Experiment 4 manipulated the number of task switches and found that it had little effect on task spans. The results suggest there is no trade-off between storage and task switching, which supports some theories of executive control and challenges others.

Executive control is the instrument of volition. It is the process by which the mind programs itself and exercises control over the execution of its programs. Executive control is an important topic in many areas of psychology, including cognitive science, cognitive neuroscience, life span development, psychopathology, and individual differences. Executive control appears in various guises, so it is studied in various ways. Two of the most popular domains in which it is studied are *working memory* and *task switching*. This article concerns the relation between working memory and task switching and what it reveals about the nature of executive control. It introduces a new experimental paradigm called the *task span procedure* that was designed to reveal the interaction between working memory and task switching: Subjects are given a list of tasks to perform and then a series of stimuli to perform them on. The task span procedure requires working memory to store the list of task names and keep track of progress through the list. It requires task switching because successive tasks on the list are generally different. It also requires subordinate processes, controlled by the executive, to perform the tasks on the list. The task span procedure addresses several issues that distinguish theories of working memory, including the nature of capacity limitations, the nature of information loss, and the role of long-term memory in working-memory spans. The task span procedure also addresses issues in the task-switching literature, including what is involved in changing from one task to another.

The Task Span Procedure

In principle, the task span procedure can involve any number of tasks of any kind. In everyday life, it could involve a series of chores you hope to accomplish on a weekend or a series of meetings in your office at work. In this article, the task span procedure involves three tasks that can be performed on numbers that are presented as digits (e.g., 3) or words (e.g., *nine*). The three tasks are *magnitude judgments* (i.e., is the number greater than or less than 5?), *parity judgments* (i.e., is the number odd or even?), and *form judgments* (i.e., is the number a digit or a word?). The task names specify the alternative responses and the left-to-right mapping of those responses onto response keys. Thus, *Hi–Low* is the name for the magnitude task, *Odd–Even* is the name for the parity task, and *Digit–Word* is the name for the form task. Subjects are given lists of 1–10 task names. The lists are constrained so that each task is usually different from the one before it and the one after it. Consequently, subjects must switch tasks before performing each judgment. The names on the list are presented one at a time and then the target stimuli are presented one at a time. There is one target for each name on the list. The target stimuli are the digits 1, 2, 3, 4, 6, 7, 8, and 9 and the words *one*, *two*, *three*, *four*, *six*, *seven*, *eight*, and *nine*. As is typical in studies of task switching, these stimuli are ambiguous, in that all three tasks can be performed on each of the stimuli.

The task span procedure provides two major dependent variables: accuracy and reaction time (RT). In order to perform correctly, subjects must (a) retrieve the task name that is appropriate for the current stimulus, (b) engage the task set appropriate to the task name, and (c) perform the appropriate task on the target stimulus. Tasks can be performed correctly only if each of these steps is executed correctly. The task span is the number of tasks that can be executed perfectly in order. The time to perform a task—RT—is the sum of the times required to execute each of these steps. Thus, RT provides insight into the retrieval processes involved in recalling task names, the executive processes involved in task switching, and the subordinate processes involved in performing the tasks. These measures bear on theories of working memory and theories of task switching.

Additional materials are on the Web at <http://dx.doi.org/10.1037/0096-3445.133.2.218.supp>.

This research was supported by Grants BCS 0133202 and BCS 0218507 from the National Science Foundation. I am grateful to Julie Delheimer for testing the subjects and helping with the data analysis and to Courtney Shipper for helping with the data analysis. I am grateful to Kate Arrington, Michael Kane, Nachshon Meiran, and Bob Logie for comments on the article.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, Vanderbilt University, Nashville, TN 37203. E-mail: gordon.logan@vanderbilt.edu

Working Memory and Executive Control

Working memory is the executive's desktop. It is involved in virtually every voluntary task a person performs. It is a short-term memory that stores goals, important inputs, and partial results and keeps them accessible. It keeps track of progress in multistep processes and evaluates the effectiveness of performance. In task-switching situations, it keeps track of the different task sets and the current task the person is performing. As is evident in these examples, the defining characteristic of working memory for many researchers is simultaneous processing and storage—keeping one thing in mind while doing another.

There are many theories of working memory, ranging from informal frameworks that summarize large data sets to formal computational models that predict the data quantitatively. Many of the dominant positions were summarized recently in a book edited by Miyake and Shah (1999). The theories differ in their approach to three major issues that can be addressed with the task span procedure: the nature of capacity limitations, the nature of the loss of information from working memory, and the role of long-term memory in performing working memory tasks.

Capacity Limitations in Working Memory

A fundamental property of working memory is that its capacity is limited. The number of items that can be recalled perfectly in order is limited to three to seven, depending on the nature of the items and the manner in which the limit is measured (Cowan, 2001; Miller, 1956). All theories of working memory assume its capacity is limited. They differ primarily in the assumptions they invoke to explain the limit. Some theories assume that a limited capacity for activation constrains the number of items that can be kept sufficiently active to be recalled (Anderson, Reder, & Lebiere, 1996; Daneman & Carpenter, 1980; Just & Carpenter, 1992; Logan, 1978, 1979; see Kahneman, 1973; Moray, 1967; Posner & Boies, 1971). Other theories assume that activation decays over time and only a limited number of items can be kept active enough to be recalled (Anderson et al., 1996; Baddeley, 1986, 1996; Baddeley & Hitch, 1974; Cowan, 1995, 1999; Hitch, Towse, & Hutton, 2001). Still other theories assume that similar items interfere with each other, and that limits the number of items that can be kept sufficiently distinct from each other to be recalled accurately (Baddeley & Logie, 1999; Cowan, 1999; Hasher & Zacks, 1988).

Early theories of working memory predicted a trade-off between processing and storage, assuming that a single source of capacity limited executive processing and that executive processing was responsible for at least some of the storage in span tasks (e.g., Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Logan, 1978, 1979). Many experiments sought evidence for the predicted trade-off, but few found it, so most modern theories of working memory do not predict strong trade-offs between processing and storage (Engle, Kane, & Tuholski, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kieras, Meyer, Mueller, & Seymour, 1999; Schneider & Detweiler, 1987). In Baddeley's recent theorizing, for example, the executive has no capacity for storage; all storage is accomplished by subordinate systems, such as the articulatory loop and the visuospatial sketchpad (Baddeley, 1996, 2000; Baddeley & Logie, 1999). However, a few experiments have shown trade-offs between concurrent memory loads and executive processes in-

involved in preparing for upcoming tasks (Logan, 1978, 1979), generating unique responses (Baddeley, Eldridge, Lewis, & Thompson, 1984), and manipulating stored items (Anderson et al., 1996). Two formal theories of working memory, based on the production system architecture, assume a limited capacity for activation that is shared between processing and storage: the adaptive control of thought—rational (ACT-R) theory of Anderson and colleagues (Anderson et al., 1996; Lovett, Reder, & Lebiere, 1999) and the CAPS theory of Just and Carpenter (1992). These theories predict a trade-off between processing and storage.

The task span procedure permits two tests of the trade-off between processing and storage. First, performance in the task span procedure, which requires storage, retrieval, task switching, and task execution, can be compared with performance in the standard memory span procedure, which requires only storage and retrieval. If the extra processes in the task span procedure trade off with storage and retrieval, then task spans should be smaller than memory spans. Experiments 1 and 2 tested these predictions, comparing a *perform condition* in which subjects performed the task span procedure with a *recall condition* in which subjects simply recalled task names. Second, task spans can be compared in conditions that vary in the amount of executive processing they require. If there is a trade-off between processing and storage, task spans should be smaller when more executive processing is required. Experiment 4 tested this prediction by comparing task spans in conditions in which subjects had to switch task sets several times during the retrieval period with conditions in which subjects had to switch task sets only a few times.

Loss of Information From Working Memory

A fundamental property of working memory is that it is transient. It holds information for a short period of time, while it is required for the task at hand, and then the information is lost. There are two main theoretical accounts of information loss in working memory—interference and decay—that date to studies of short-term memory in the 1960s. The interference account proposes that information is lost because of competition from other information (Baddeley & Logie, 1999; Cowan, 1999; Schneider & Detweiler, 1987). A strong interpretation of the interference account claims that time in storage does not predict loss. Information is held in short-term memory until it is displaced by other information (Waugh & Norman, 1965). The decay account proposes that information dissipates over time even if there is no competing information in memory. A strong interpretation of the decay account claims that time in storage is the only factor that predicts loss (Doshier & Ma, 1998). The process of decay begins when attention is switched away from the item and continues unabated as time passes (Anderson et al., 1996; Anderson & Matessa, 1997; Baddeley & Logie, 1999; Cowan, 1999; Kieras et al., 1999; Lovett et al., 1999).

The evidence for interference versus decay is mixed (for a recent review, see Nairne, 2002). Some researchers have shown poorer performance on memory span tasks with similar items, supporting the interference account (e.g., Conrad, 1964), but more recent studies have shown that manipulation of the time in storage has strong effects on performance (Baddeley, Thomson, & Buchanan, 1975; Cowan et al., 1992; Cowan, Wood, Nugent, & Treisman, 1997; Doshier & Ma, 1998). The task span procedure allows strong tests of the interference and decay hypotheses. The

interference hypothesis may be tested by plotting the probability of retrieving an item against list length for the task span procedure and the memory span procedure. If all that matters is the number of items that intervene between encoding and retrieval, then retrieval probability should be the same for the two procedures at equivalent list lengths, even though the extra processing required in the task span procedure should stretch retrieval over a longer period of time (Waugh & Norman, 1965). The decay hypothesis may be tested by plotting the probability of retrieving an item against the time at which it was retrieved, which is reflected in RT. The plot of retrieval accuracy against time from the task span procedure can be compared with a similar plot from the memory span procedure. If all that matters is the amount of time between encoding and retrieval, then retrieval probability should be the same for the two procedures at equivalent retrieval times (Doshier & Ma, 1998). Experiments 1, 2, and 4 tested these predictions.

Role of Long-Term Memory in Working Memory

A fundamental property of working memory is that it is distinguishable from long-term memory. Many researchers consider working memory to be the active portion of long-term memory—those long-term memory traces that exceed some threshold of activation (Anderson et al., 1996; Cowan, 1995, 1999, 2001; Engle et al., 1999; Lovett et al., 1999; see Norman, 1968). Other researchers consider working memory and long-term memory to be separate but interacting (Baddeley & Logie, 1999; Kieras et al., 1999; Schneider & Detweiler, 1987). A related issue that is not discussed as extensively is the role of long-term memory in span performance in tasks that are used to measure working-memory capacity. It is well established that long-term memory contributes to performance on the traditional memory span task (Cowan, 2001; Watkins, 1974). Thus, it is likely that long-term memory also contributes to performance on working memory span tasks.

The most extreme argument for the role of long-term memory in memory span tasks is Ericsson and Kintsch's (1995) theory of long-term working memory. They argued that expertise in a domain allows subjects to use skill-related retrieval structures to access information in long-term memory as easily as they would access currently active information in working memory. These retrieval structures allow experts to increase their digit span from Miller's (1956) 7 ± 2 to more than 80 (Ericsson, Chase, & Faloon, 1980). Ericsson and Kintsch (1995) described several other examples from the expertise literature and argued that college students have enough expertise in reading to support long-term working memory for information in text. It is possible that nonexpert subjects use more modest but nevertheless effective long-term memory retrieval structures to support their performance in memory span tasks. Cowan's (1995) idea of *virtual short-term memory* articulates this view: Subjects may use information readily available in long-term memory to supplement what they hold in short-term memory.

The timing of retrieval provides particularly clear evidence of the role of long-term memory in memory span tasks. If RT for retrieving each item is plotted against its position in the list, the data often show a "scalloped" pattern that suggests subjects retrieve the list from long-term memory in chunks (Anderson & Matessa, 1997; Kahana & Jacobs, 2000). RT is slow for the first item in a chunk because the chunk must be retrieved from long-term memory and "unpacked." RT is fast for subsequent items in

the chunk because they have been unpacked and remain active in short-term (or working) memory. To my knowledge, no studies of working memory spans have examined the timing of recall in this manner (but see Cowan et al., 2003). Many researchers are interested in individual differences and collapse the data from each subject into a single span measure to correlate with other ability measures (e.g., Engle et al., 1999).

The task span procedure provides an opportunity to examine the contribution of long-term memory to span performance. Subjects respond after retrieving each task name, and the RT of each response can be plotted against list position. A scalloped pattern in this plot would indicate retrieval of chunks from long-term memory. A flat pattern would suggest that all items in the span are simultaneously active in working memory, as most theories of working memory seem to assume. Alternatively, a flat pattern could indicate that each item is retrieved directly from long-term memory. Experiments 1–3 examine these predictions.

Task Switching and Executive Control

The ability to switch flexibly between tasks is the pinnacle of human cognition and the hallmark of executive control. It allows people to adapt to changing demands in the environment and it allows them to approach the same situation from different perspectives. But it comes with a cost: Task switching takes time and produces interference, as is evident in a variety of procedures that compare performance when tasks change with performance when tasks repeat (Allport, Styles, & Hsieh, 1994; Arrington & Logan, in press; Gopher, Armony, & Greenspan, 2000; Jersild, 1927; Logan & Bundesen, 2003; Meiran, 1996; Rogers & Monsell, 1995). The interpretation of switch costs is controversial. Many researchers interpret switch costs as reflecting the time required to reconfigure the cognitive system. However, they disagree on what is involved in reconfiguration.

Some researchers assume that reconfiguration only involves processes that operate on working memory: Changing the contents of working memory is sufficient to change performance (i.e., changing goals and changing stimulus–response mapping rules; Mayr & Kliegl, 2000; Rubinstein, Meyer, & Evans, 2001; Sohn & Anderson, 2001). Sohn and Anderson (2001) proposed a production-system theory of task switching within Anderson's ACT-R model. A production system is a set of elementary condition–action rules called productions. A production is "fired"—the action is executed—when its condition is satisfied. In ACT-R, the condition of a production consists of a goal and a stimulus input from the environment, both of which are represented in working memory. Goals enable task performance. If a goal of a production is not set, then the condition of the production will not be satisfied and the production will not fire. Instantiating goals in working memory is sufficient to enable the productions that comprise a task set, and changing goals in working memory is sufficient to change task set. Changing goals disables the productions relevant to the old task set and enables the productions relevant to the new task set.

Other researchers argue that reconfiguration involves processes outside of working memory as well as processes that operate on working memory. Changing goals is not sufficient. The subordinate processes must also be reprogrammed (Logan & Gordon, 2001; Meiran, 2000). Logan and Gordon (2001) proposed a theory of task switching in which an executive process controls the

stimulus selection processes in Bundesen's (1990) theory of visual attention (TVA) and the response selection processes in Nosofsky and Palmeri's (1997) exemplar-based random walk model (EBRW). In Logan and Gordon's (2001) theory, it is not sufficient to instantiate goals in working memory. The goals must be translated into "parameters" that control the operation of TVA, and the parameters must be communicated to TVA and EBRW to enable them to perform the new task.

Still others argue that switch costs reflect long-term memory processes that are outside of working memory. Switch costs reflect interference from past task sets or past associations to the stimuli under different task sets (Allport et al., 1994; Allport & Wylie, 2000). This approach says more about how past task sets are disabled than about how new task sets are enabled. A strong interpretation of this approach is that there is no active reconfiguration. However, a computational instantiation of this approach by Gilbert and Shallice (2002) requires an act of control by an executive process to enable performance whether or not the task set switches. The executive has to turn on task-demand nodes in a connectionist network in order to select the appropriate task. This is equivalent to Sohn and Anderson's (2001) assumption that instantiating a task goal is sufficient to establish and change a task set (also see Rubinstein et al., 2001).

The task span procedure provides a way to determine whether reconfiguration involves more than changing goals in working memory by comparing task span performance with appropriate control conditions. The test phase of the task span procedure requires three steps: (a) retrieving the next task on the list, (b) reconfiguring the cognitive system to perform the retrieved task, and (c) performing the retrieved task on the target stimulus. RT is the sum of the times required to accomplish each of these steps. The duration of the first step can be estimated by measuring RT in the test phase of a standard memory span procedure, in which subjects recall task names. The duration of the third step can be estimated by measuring RT in single-task control conditions, in which subjects execute the same task on each target stimulus. If the cognitive system must be reconfigured to establish a new task set, then RTs in the task span procedure will exceed the sum of the RTs in the memory span procedure and single-task controls by a substantial amount (i.e., by the time required for the second step). If changing goals or mapping rules in working memory is sufficient to enable performance on the retrieved task, then the first and third steps are all that is required to perform appropriately. RT in the task span procedure should equal the sum of the RTs in the memory span procedure and the single-task control conditions. Experiment 1 tested this prediction.

Experiment 1

The first experiment involved three conditions: a *perform* condition that implemented the task span procedure—subjects were given lists of task names and then lists of stimuli to perform the tasks on—a *recall* condition that implemented the memory span procedure—subjects were given lists of task names and then cues to recall the names—and *single-task* control conditions—subjects performed the same task on each target. All three conditions involved the same study–test format in order to equate timing and control for fatigue and expectancy effects. A list of 1–10 task names was presented in the study phase and then a list of 1–10 target stimuli was presented in the test phase.

Experiment 1 had four purposes. The first was to assess the trade-off between processing and storage by comparing task spans in the perform condition with memory spans in the recall condition. The second was to compare time-based decay and item-based interference accounts of information loss from working memory by plotting memory performance against list position and retrieval time. The third was to assess the role of long-term memory retrieval in working memory performance by examining RTs for each list position in the test phase to see if they show a scalloped pattern with longer RTs at chunk boundaries. The fourth was to assess the role of reconfiguration of subordinate processes in task switching by comparing RTs in the perform condition with the sum of RTs in the recall condition and the single-task condition.

Method

Subjects

The subjects were 6 graduate and undergraduate students recruited from the general university population and paid for participating in nine 1-hr sessions.

Apparatus and Stimuli

The stimuli were displayed on Gateway 2000 Crystalscan 1024 NI monitors controlled by Gateway 2000 486 computers. The task names were *Hi–Low*, *Odd–Even*, and *Digit–Word*, and the target stimuli were the digits 1, 2, 3, 4, 6, 7, 8, and 9 and the words *one*, *two*, *three*, *four*, *six*, *seven*, *eight*, and *nine*. The task names and targets were 5 mm high. *Hi–Low* was 20 mm wide, *Odd–Even* was 25 mm wide, and *Digit–Word* was 30 mm wide. Digit targets were 3 mm wide. Word targets were 10 mm (*one*, *two*, *six*), 12.5 mm (*four*, *nine*), and 15 mm (*three*, *seven*, *eight*) wide. Viewing distance was not controlled but was approximately 60 cm. At this distance, 1 cm is approximately 1 degree of visual angle.

Each study list was preceded by a display that said "STUDY: List Length *N*," where *N* ranged from 1 to 10. This display was exposed until the subject pressed the space bar. Then the screen went blank for 1,000 ms and the sequence of study stimuli began. Each test list was preceded by a display that said "TEST: List Length *N*," where *N* ranged from 1 to 10 and had the same value as *N* in the study list. This display was exposed for 1,000 ms. Then the screen went blank for 1,000 ms and the sequence of test items began. When presented in the study list, each task name was exposed for 500 ms and followed by a 500-ms blank screen before the next task name was presented. When presented in the target list, each target stimulus on the test list was exposed until the subject responded. Then the screen went blank for 1,000 ms and the next target stimulus was exposed.

Subjects responded by pressing keys on the numeric keypad. In the task span condition and single-task control conditions, subjects pressed 7 for "High," 9 for "Low," 4 for "Odd," 6 for "Even," 1 for "Digit" and 3 for "Word." The mapping of response categories onto keys was consistent with the ordering of the words in the task names. In the memory span procedure, subjects pressed one of the two keys in the row that the task was mapped onto without considering how the task applied to the target stimulus. Subjects were free to press either key. Thus, subjects pressed either 7 or 9 for "Hi–Low," either 4 or 6 for "Odd–Even," and either 1 or 3 for "Digit–Word."

Procedure

Each subject served in nine sessions: three in the perform condition, three in the recall condition, and three in the single-task conditions. Half of the subjects performed the three single-task sessions first, then the three recall sessions, and then the three perform sessions. The other half performed the three perform sessions first, then the three recall sessions, and

then the three single-task sessions. Each subject did one task for each of the single-task sessions. All subjects did the magnitude task (Hi–Low) in the first single-task session, the parity task (Odd–Even) in the second single-task session, and the form task (Digit–Word) in the third single-task session.

Each session involved 100 study and test lists. There were 10 replications of each list length from 1 to 10. Within each replication, list lengths were presented in ascending order and all list lengths were presented before the next replication began. Thus, the length of the first list was 1, the second list was 2, and so on up to list length 10 for the 10th list. The cycle was repeated for lists 11 through 20, 21 through 30, and so on.

Each list involved a study phase and a test phase. The study phase began with a display that announced the phase and the list length (e.g., STUDY: List Length 1). Subjects pressed the space bar to begin the study list. Once they pressed it, the computer determined the timing of the study list. One task name was presented every second. The test phase began immediately after the study phase with a display that announced the phase and the list length (e.g., TEST: List Length 1). This display was exposed for 1,000 ms. Then after the screen went blank for 1,000 ms, the test list was presented. The target stimuli in the test phase were displayed until subjects responded, after which the screen went blank for a 1,000-ms response-to-stimulus interval (RSI). Thus, timing depended partly on the subject and partly on the computer. Subjects were allowed to take breaks before beginning the study lists.

The order of task names on the study lists was random with the constraint that all three task names were presented before any task name repeated. Each study list was constructed by creating a list of 12 task names that consisted of four consecutive sets of the 3 task names, with the 3 names in each set in a separate random order. A study list of length N was created by taking the first N items from this list. The order of target stimuli was random as well. Each test list was constructed by randomly ordering the 16 target stimuli and taking the first N of them to create a test list of length N . Separate random orders were created for each session for each subject.

Subjects were given instructions daily during the 9 days of the experiment. They were given an overview of all of the conditions in the experiment in the first session. They were given detailed instructions about each condition in the first session they experienced it (i.e., on the first, fourth, and seventh sessions). They were given less detailed instructions in the second and third sessions in each condition, but care was taken in each session to make sure they understood what they were supposed to do.

Results and Discussion

Mean RTs and accuracy scores (the percentage correct) were calculated for each list position in each list length for the perform, recall, and single-task conditions for each subject, averaging over sessions. The mean RTs and mean accuracy scores across subjects for this experiment and the subsequent ones can be found in Appendixes A–D in the online version of this article, which is part of the PsycARTICLES database and is available on the Web at <http://dx.doi.org/10.1037/0096-3445.133.2.218.supp>. These means provide the basis for calculating task spans and memory spans and for analyzing list-position effects in the RT data.

Task Spans and Memory Spans

To estimate memory span, the probability of recalling all of the task names on a list in order perfectly, $P(\text{Perfect})$, was calculated for each list length in the recall condition. The means across subjects are plotted as a function of list length in Figure 1. Task span was estimated in two ways: with strict and lenient scoring. With strict scoring, subjects had to perform each task correctly. The probability of performing each task on the list in order

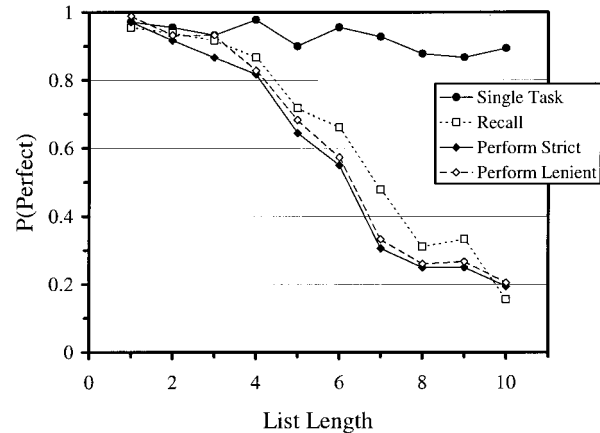


Figure 1. Probability of perfect performance as a function of list length for single-task, recall, and perform conditions in Experiment 1. Perform Strict = the perform condition scored strictly; Perform Lenient = the perform condition scored leniently; $P(\text{Perfect})$ = the probability of recalling all of the task names on a list in order perfectly.

perfectly was calculated for each list length. With lenient scoring, performance was considered correct if subjects performed the task correctly or if they made a *within-task error*, that is, if they pressed the wrong key for the right task (e.g., pressing the “Even” key for the stimulus 3 if the task was Odd–Even). Lenient scoring is equivalent to the scoring procedure used to estimate memory span. The probability of responding with the appropriate task perfectly in order was calculated for each list length in the perform condition. The means across subjects for strict and lenient scoring are also plotted as a function of list length in Figure 1. The probability of performing each of the tasks on a single-task list perfectly was also calculated and plotted in Figure 1. The numerical value of each point in Figure 1 can be calculated by converting the percentage of correct responses in Appendix A (which is available on the Web at <http://dx.doi.org/10.1037/0096-3445.133.2.218.supp>) to probabilities and multiplying the probabilities across list position for each list length.

The nature of capacity limitations in working memory. Memory span is defined in several ways (Cowan, 2001). In this article, I define memory span as the maximum list length that can be recalled in order perfectly 50% of the time (Miller, 1956). I calculated memory spans and task spans by finding the point at which the lines connecting the points in Figure 1 intersect $P(\text{Perfect}) = .5$. By this criterion, the memory span was 6.9 items in the recall condition and the task span was 6.2 items with strict scoring and 6.3 items with lenient scoring in the perform condition. The requirement to perform the tasks as well as recall their names reduced the memory span by 0.7 items—about 10%—suggesting a weak trade-off between processing and storage.

In the recall condition, the probability of perfect performance declined slowly from list length 1 to list length 4 and then declined more rapidly after that. The average value across list lengths was 0.63. In the perform condition, the pattern was similar, except that the probability of perfect recall was slightly smaller at each list length except 1 and 10. With strict scoring, the average across list lengths was 0.58, so the average difference between the functions was 0.05. With lenient scoring, the average across list lengths was

0.60. These data suggest there was little or no trade-off between processing and storage.

The probabilities of perfect performance were analyzed in a 3 (condition: perform, recall, or single task) \times 10 (list length) analysis of variance (ANOVA), using strict scoring for the perform condition. The ANOVA yielded a significant main effect of condition, $F(2, 10) = 13.12$, $p < .01$, $MSE = 0.161$, $\eta^2 = .724^1$; a significant main effect of list length, $F(9, 45) = 39.04$, $p < .01$, $MSE = 0.021$, $\eta^2 = .886$; and a significant interaction between condition and list length, $F(18, 90) = 13.94$, $p < .01$, $MSE = 0.011$, $\eta^2 = .736$. The main effect of condition and the interaction between list length and condition were due to the difference between the single-task condition on the one hand and the recall and perform conditions on the other. Single-task performance was near perfect, regardless of list length ($M = 0.93$). The data from the recall and perform conditions were analyzed in a separate 2 (condition) \times 10 (list length) ANOVA. List length was the only significant effect in that ANOVA, $F(9, 45) = 33.71$, $p < .01$, $MSE = 0.033$, $\eta^2 = .871$. Neither the main effect of condition, $F(1, 5) < 1.0$, $MSE = 0.081$, $\eta^2 = .074$, nor the interaction between condition and list length, $F(9, 45) = 1.68$, $p = .12$, $MSE = 0.006$, $\eta^2 = .251$, were significant.

The loss of information from working memory. Item-dependent interference theories make the strong prediction that performance should depend on the number of items that intervene between encoding and retrieval independent of the time that has elapsed (Waugh & Norman, 1965). The plots in Figure 1 are consistent with this prediction. Time-dependent decay theories make the strong prediction that performance should depend on the time at which the items were retrieved. Equivalent retrieval times for different conditions should produce equivalent retrieval accuracies (Doshier & Ma, 1998). To test this prediction, the data from the perform and recall conditions were plotted as a function of retrieval time in Figure 2. Retrieval time was calculated by adding RT to RSI. In this plot, the probability of perfect recall is much better in the perform condition than in the recall condition at equivalent retrieval times. Points along the functions can be compared using Fisher's least significant difference (LSD) test, computed from the error term for the interaction between list length

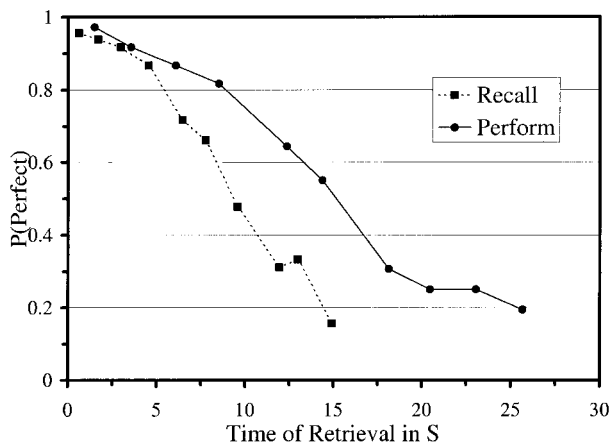


Figure 2. Probability of perfect performance as a function of time of retrieval for recall and perform conditions in Experiment 1. $P(\text{Perfect}) =$ the probability of recalling all of the task names on a list in order perfectly.

and condition in the ANOVA that compared the perform and recall conditions ($MSE = 0.007$, $df = 45$). The LSD value for $p < .05$ was 0.099. By this criterion, performance differs between conditions for retrieval times greater than 5 s. Thus, the data are inconsistent with time-dependent decay. They are more consistent with item-dependent interference.

Task Performance

The mean RTs for the perform, recall, and single-task conditions are presented as a function of list length and list position in Figure 3. RTs for list position 1 increased sharply with list length in the perform and recall conditions, as if subjects retrieved or rehearsed the whole list before proceeding. List length had small effects at other list positions and virtually no effect in the single-task condition. Consequently, the analysis focuses on list position effects, averaged across list length, which are presented in Table 1 for the recall, perform, and single-task conditions. In Table 1, the mean for list position 1 is the average of list lengths 1–10, the mean for list position 2 is the average of list lengths 2–10, and so on. Table 1 also contains the mean accuracy for each list position.²

The RT data were analyzed in a 3 (condition: perform, recall, or single task) \times 10 (list position) ANOVA. There were significant main effects of condition, $F(2, 10) = 50.74$, $p < .01$, $MSE = 396,426.94$, $\eta^2 = .910$, and list position, $F(9, 45) = 3.72$, $p < .01$, $MSE = 145,779.04$, $\eta^2 = .426$. The interaction between condition and list position was not significant, $F(18, 90) = 1.55$, $p = .091$, $MSE = 59,127.64$, $\eta^2 = .237$.

The role of long-term memory retrieval in working memory. The data from the recall and perform conditions show evidence of scalloping that is characteristic of retrieval of chunks from long-term memory (Anderson & Matessa, 1997; Kahana & Jacobs, 2000). RT was slow for the first item in the chunk and fast for the remaining items. The data suggest that subjects retrieved tasks in chunks of three. This is consistent with the structure of the memory lists, which were created by concatenating random orders of the three tasks. It is likely that subjects learned this constraint and exploited it to improve their ability to recall the tasks (see Bower & Winzenz, 1969; Chase & Simon, 1973; Ericsson & Kintsch, 1995; Hitch, Burgess, Towse, & Culpin, 1996; Stadler, 1995).

¹ The η^2 values reported here and elsewhere are partial η^2 .

² The RTs in Figure 3 and Table 1 are based on trials in which subjects responded correctly on the current trial regardless of errors they made on other trials in the list, reflecting standard practice in the RT literature. They average performance over trials in which the whole list was and was not recalled perfectly. Consequently, they may not reflect performance when subjects did recall the list perfectly. To assess this possibility, I calculated mean RTs for each list position and list length only for trials on which the whole list was recalled correctly. The pattern of the means with perfect performance was essentially the same as the pattern in Table 1. To compare them precisely, I recalculated the means in Table 1 using only those list lengths for which subjects performed at least one list perfectly. The mean RT across list position was 842 ms for perfect-performance scoring and 854 ms for the corresponding list lengths in standard scoring. The correlation between the means across conditions and list positions was .98. Thus, it is unlikely that the standard scoring method biased the measurements. However, it provides more observations for each condition for each subject, so the RT analyses in this experiment and subsequent ones used standard scoring.

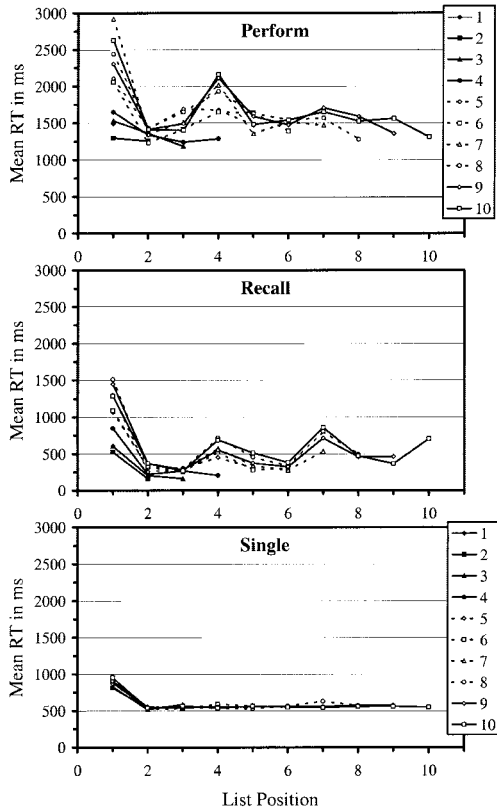


Figure 3. Mean reaction time (RT) as a function of list position (collapsed across list length) for perform, recall, and single-task conditions in Experiment 1.

The data in Figure 3 and Table 1 suggest an interaction between list position and condition because the recall and perform conditions showed a scalloped pattern but the single-task condition did not. The interaction between condition and list position was not significant, but the error term from the interaction ($MSE =$

59,127.64) was used to assess scalloping with planned contrasts. The contrast weights were 3, -2, -2, 3, -2, -2, 3, -2, -2, and 3 for list positions 1–10, respectively. The contrast was significant in the recall condition, $F(1, 90) = 37.45, p < .01$, and the perform condition, $F(1, 90) = 13.79, p < .01$, but not in the single-task condition, $F(1, 90) = 1.83, p = .18$. The pattern of scalloping was the same for the recall and perform conditions. A contrast evaluating the interaction between scalloping and recall versus perform was not significant, $F(1, 90) = 2.89, p = .09$. This suggests chunk size was the same—three items—for the perform condition and the recall condition despite the greater demands on processing in the perform condition.

The scalloping in Figure 3 and Table 1 may reflect a progressive reduction in uncertainty rather than chunk retrieval. Because of the way the lists were constructed, the first item in each set of three is one of three possibilities (i.e., Hi-Low, Odd-Even, or Digit-Word), the second item is one of two possibilities (e.g., if the first item was Hi-Low, then the second must be either Odd-Even or Digit-Word), and the third item is completely determined, given the first two items (e.g., if the first two items were Hi-Low and Odd-Even, the third must be Digit-Word). This predicts slow RT for the first item in each set of three, intermediate RT for the second, and fast RT for the third, which is similar to the observed pattern.

Progressive reduction in uncertainty can be distinguished from chunking by comparing the second and third items in each set of three. Progressive reduction in uncertainty predicts that the second will be slower than the third, whereas chunking predicts no difference between the second and the third. To test these predictions, I calculated contrasts that compared the second and third item in each set of three using weights 0, 1, -1, 0, 1, -1, 0, 1, -1, and 0 for list positions 1–10. The contrast was not significant in the recall condition, $F(1, 90) < 1.0$, or in the perform condition, $F(1, 90) < 1.0$, which is inconsistent with progressive reduction in uncertainty and consistent with chunking. Experiments 2 and 3 provide converging evidence on this issue.

Reconfiguration and task switching. RTs were much longer in the perform condition than in either the recall or the single-task

Table 1
Mean Reaction Times (RTs; in Milliseconds) and Percentages of Correct Responses as a Function of List Position in the Recall (Memory Span), Perform (Task Span), and Single-Task Conditions of Experiment 1

Measure	List position									
	1	2	3	4	5	6	7	8	9	10
Memory span condition										
RT	1,030	283	264	532	377	324	729	476	415	704
% correct	90	83	83	75	68	68	58	52	51	52
Task span condition										
RT	2,042	1,357	1,448	1,834	1,527	1,492	1,598	1,463	1,459	1,309
% correct	89	83	83	72	63	60	48	45	43	48
Single-task condition										
RT	901	540	550	550	554	557	572	570	569	550
% correct	99	98	99	99	98	98	98	98	99	98

conditions. This difference was expected because the perform condition requires the retrieval processes that are required in the recall condition and the task execution processes that are required in the single-task conditions. The important question is whether the perform condition requires more than these retrieval and execution processes, that is, whether there is more to task switching than changing goals and mapping rules in working memory. To address this question, mean RTs from the recall condition were added to mean RTs from the single-task condition and plotted together with the mean RTs from the perform condition in Figure 4. The sum of the recall condition and single-task condition underestimates the perform condition by 479 ms. Differences appeared at each list position except the first. The differences were 257, 390, 361, 551, 490, 621, 458, 519, and 484 ms for list lengths 2–10, respectively (see Table 1).

The main effect of condition was significant in the ANOVA on the RTs (see above). The difference between the three conditions was assessed with Fisher's LSD test, using the error term from the main effect of condition ($MSE = 396,426.94$, $df = 10$). The critical value for $p < .05$ was 256 ms. By this criterion, mean RT in the perform condition ($M = 1,612$ ms) was significantly longer than mean RT in the recall condition ($M = 517$ ms) and the single-task condition ($M = 616$ ms). Mean RT in the recall condition was not significantly different from mean RT in the single-task control condition. By the same LSD criterion, the 479-ms difference between the mean RT in the perform condition and the sum of the mean RTs in the recall and single-task conditions was significant.

Recall RTs were very fast, averaging 456 ms for list positions 2–10. This suggests that some of the retrieval process may have occurred during the 1,000-ms blank interval following the previous response. This possibility calls into question the use of recall RTs to estimate the task-name retrieval component of the task span procedure. However, subjects may also have retrieved task names during the RSI in the perform condition. Several investigators have argued that retrieval of task goals and mapping rules can happen in the preparatory interval before the target stimulus is presented (Mayr & Kliegl, 2000; Meiran, 1996; Rogers & Monsell, 1995; Sohn & Anderson, 2001). Experiment 2 manipulated RSI, including a 100-ms RSI that was intended to force retrieval to occur during the RT interval.

The difference between perform RTs and the sum of recall and single-task RTs may be due to the memory load in the perform

condition slowing subordinate processes (Anderson et al., 1996; Just & Carpenter, 1992). However, memory load is unlikely to account for much of the difference. Concurrent memory loads of seven or eight digits slow RT in comparable tasks by 50–100 ms (Logan, 1978, 1979), and the effective memory load was likely smaller in the perform condition. Many of the lists were smaller than seven or eight items and subjects tended to recall the lists in chunks of three, which suggests a memory load of three items.

The difference between perform RTs and the sum of recall and single-task RTs may have occurred because the perform condition required recall of mapping rules as well as task names to enable proper execution of the task, whereas the recall condition required only recall of task names. Mayr and Kliegl (2000) found that retrieving mapping rules took more time than simply retrieving task names did. Logan and Bundesen (2003) conducted an experiment with two of the three tasks used in the present experiment (Hi-Low and Odd-Even), in which subjects were either cued with the mapping rules (i.e., Hi-Low or Odd-Even) or with the names of the tasks (i.e., magnitude or parity). In their data, RT was only 49 ms longer with name cues than with mapping cues. This difference is not large enough to account for the 479-ms difference between perform RTs and the sum of recall and single-task control RTs in the present experiment.

Conclusions

The experiment addressed four issues in the literature on working memory and task switching. First, the equivalence of task spans and memory spans suggests there was no trade-off between processing and storage (also see Duff & Logie, 2001), which is consistent with some theories of working memory (Baddeley & Logie, 1999; Cowan, 1999; Engle et al., 1999; Kieras et al., 1999; Schneider & Detweiler, 1987) and inconsistent with others (Anderson et al., 1996; Just & Carpenter, 1992; Lovett et al., 1999). Second, the probability of perfect recall varied widely between the recall and perform conditions at equivalent retrieval times. This is inconsistent with time-dependent decay (Anderson et al., 1996; Anderson & Matessa, 1997; Baddeley, 1986, 1996; Hitch et al., 2001; Kieras et al., 1999) and consistent with item-dependent interference as the factor that determines loss of information from working memory. Third, RTs showed a scalloped pattern across list position that suggested retrieval of chunks from long-term memory. This is consistent with Ericsson and Kintsch's (1995) theory of long-term working memory and Cowan's (1995) concept of virtual short-term memory. It is inconsistent with the idea, implicit in many theories, that all of the items reported in span tasks are simultaneously active in working memory. Fourth, RT in the perform condition was longer than the sum of RTs in the recall and single-task conditions. This suggests that the perform condition required an extra process beyond those required in the recall and single-task conditions. This is inconsistent with theories of task switching that assume that changing goals in working memory is sufficient to enable a new task set (Mayr & Kliegl, 2000; Rubinstein et al., 2001; Sohn & Anderson, 2001) and consistent with theories that assume that subordinate processes must be reconfigured as well (Logan & Gordon, 2001; Meiran, 2000).

Experiment 2

The second experiment had three purposes. The first was to examine the trade-off between processing and storage with more

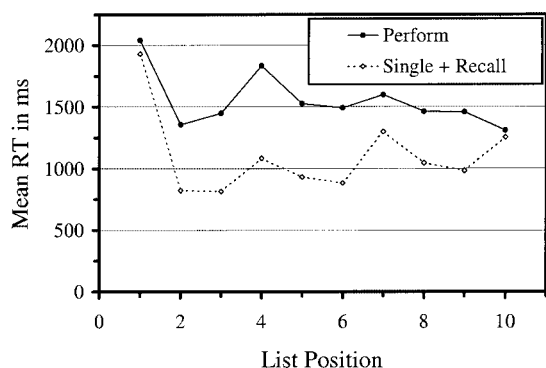


Figure 4. Mean reaction time (RT) in the perform condition and the sum of mean RTs in the single-task and recall conditions as a function of list length in Experiment 1.

statistical power. It replicated the comparison between the memory span and the task span conditions in a larger sample of subjects who were tested for a single session. The second was to examine time-dependent decay and item-dependent interference by manipulating the time between trials in the test phase. Half of the subjects were tested with a 100-ms RSI and half were tested with a 1,000-ms RSI. Item-dependent interference predicts no effect of RSI on memory span. Time-dependent decay predicts larger memory spans with shorter intervals.

The third purpose was to provide converging evidence for the role of long-term memory retrieval in task span and memory span performance. List length was manipulated coarsely, in steps of two or three, to provide subjects with a different structure to exploit in their chunking strategies (Bower & Winzenz, 1969; Chase & Simon, 1973; Ericsson & Kintsch, 1995; Hitch et al., 1996; Stadler, 1995). Half of the subjects experienced lists of 2, 4, 6, and 8 items and half experienced lists of 2, 3, 6, and 9 items. Subjects experiencing lists of 2, 4, 6, and 8 items should chunk their lists in sets of two or four. Consequently, they should be slow at list positions 1, 5, and 7 and fast at the other positions. Subjects experiencing lists of 2, 3, 6, and 9 items should chunk their lists in sets of three and, consequently, be slow at list positions 1, 4, and 7 and fast at the other list positions.

Method

Subjects

Altogether, 64 subjects from an introductory psychology class participated for course credit. Half of the subjects (32) had RSIs of 100 ms in the test phase and the other half (32) had RSIs of 1,000 ms. Within each of these conditions, half of the subjects (16) had list lengths of 2, 4, 6, and 8 and half (16) had list lengths of 2, 3, 6, and 9. None of the subjects had served in Experiment 1.

Apparatus and Stimuli

The apparatus and stimuli were the same as in Experiment 1, except for the RSI in the test phase and list lengths in the study and test phases. RSI was 100 ms for half of the subjects and 1,000 ms for the other half. List lengths were 2, 4, 6, and 8 for half of the subjects and 2, 3, 6, and 9 for the other half.

Procedure

The procedure was essentially the same as in the recall and perform conditions in Experiment 1. The major differences were that subjects performed 48 study–test sequences (240 trials) in each condition in a single session and there were 4 list lengths instead of 10. Half of the subjects within each group of 16 (i.e., within each combination of lists and RSI) had the recall condition before the perform condition and half had the conditions in the opposite order. Subjects were given practice with a single study–test sequence of list length 4 before the recall condition and practice with another single study–test sequence of list length 4 before the perform condition.

The lists were constructed in the same manner as in Experiment 1, by concatenating random orders of the three task names and selecting the first N for a list of length N . List length was manipulated in the same manner as in Experiment 1. Subjects had one study–test sequence with list length 2, then one with list length 3 (or 4), then one with list length 6, and then one with list length 9 (or 8) before repeating any list length. Then they experienced another set of the four list lengths in ascending order and then another, until they had 12 replications of the sequence.

Results and Discussion

Task Spans and Memory Spans

Mean RTs and percentages of correct responses were calculated for each list position in each list length for the perform and recall conditions for each subject. The mean RTs across subjects and mean accuracy scores across subjects are presented in Appendix B (which is available on the Web at <http://dx.doi.org/10.1037/0096-3445.133.2.218.supp>) as a function of list type (2468 vs. 2369) and RSI. The probability of responding perfectly is plotted as a function of list length for each list type and RSI in Figure 5.

The nature of capacity limitations in working memory. The memory spans for each combination of condition (recall and perform), list type (2468 vs. 2369), and RSI (100 or 1,000) were estimated from the points at which the lines in Figure 5 intersect $P(\text{Perfect}) = .5$. They are presented in Table 2. There was little difference between the memory span and the task span. Averaged over list type and RSI, the mean memory span was 6.9 in the recall condition and the mean task span was 6.4 in the perform condition—a difference of 0.5, which is about 7%. This suggests there was little or no trade off between processing and storage, replicat-

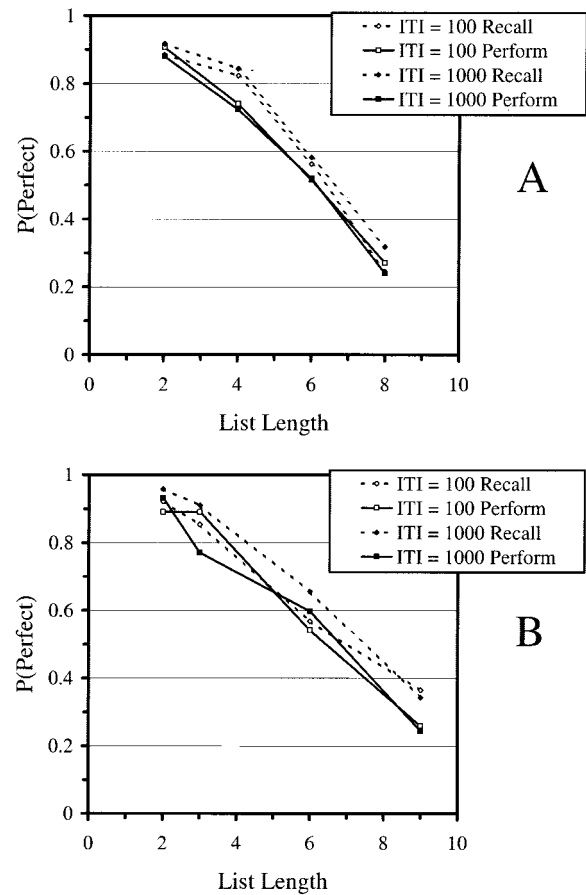


Figure 5. Probability of perfect performance as a function of list length for recall and perform conditions for intertrial intervals of 100 and 1,000 ms in Experiment 2. A: List length condition 2468. B: List length condition 2369. ITI = intertrial interval; $P(\text{Perfect})$ = the probability of recalling all of the task names on a list in order perfectly.

Table 2
Memory Spans and Task Spans as a Function of Response to Stimulus Interval (100 or 1,000 ms) and List Type (2468 vs. 2369) in Experiment 2

Span	2468		2369	
	100	1,000	100	1,000
Memory span	6.39	6.77	7.00	7.50
Task span	6.13	6.15	6.44	6.84

ing the results of Experiment 1 with a larger sample of subjects and a smaller amount of practice.

The probabilities of perfect performance in Figure 5 were subjected to 2 (condition: recall vs. perform) \times 2 (RSI) \times 4 (list length) ANOVAs. There was one ANOVA for the 2468 condition and another for the 2369 condition. The ANOVA for the 2468 condition revealed a significant main effect of list length, $F(3, 90) = 209.05, p < .01, MSE = 0.02, \eta^2 = .875$. No other effects or interactions were significant, including the main effect of condition, $F(1, 30) = 2.76, p = .11, MSE = 4.96, \eta^2 = .084$, and the main effect of RSI, $F(1, 30) < 1.0, MSE = 0.12, \eta^2 = .002$. The ANOVA for the 2369 condition showed a similar pattern. The only significant effect was the main effect of list length, $F(3, 90) = 189.34, p < .01, MSE = 0.05, \eta^2 = .863$. The main effect of condition approached significance, $F(1, 30) = 3.87, p = .058, MSE = 0.20, \eta^2 = .114$, but the main effect of RSI was far from significant, $F(1, 30) < 1.0, MSE = 0.12, \eta^2 = .005$.

The loss of information from working memory. RSI had little effect on the probabilities of perfect performance plotted as a function of list length in Figure 5, consistent with item-based decay (Waugh & Norman, 1965). To test time-based decay, the probabilities of perfect performance were plotted as a function of retrieval time in Figure 6. The time taken in retrieval did not predict the probability of perfect performance very well. At equivalent retrieval times, the probability of perfect performance was much higher in the perform conditions than in the recall conditions. Points along the functions can be compared using Fisher's LSD test computed from the error term for the interaction between list length, condition, and RSI ($MSE = 0.017$ for the 2468 condition and 0.027 for the 2369 condition). The LSD value for $p < .05$ was 0.06 for the 2468 condition and 0.08 for the 2369 condition. The data are more consistent with item-based interference (see Figure 5) than with time-based decay (see Figure 6).

Individual Differences in Memory Spans and Task Spans

I examined the relation between individual differences in memory span and task span by correlating performance in the recall and perform conditions. A composite span measure was calculated for each span for each subject by multiplying the probability of perfect performance at each list length by the list length and then summing over list lengths. For example, a subject with probabilities of perfect performance of 0.9, 0.8, 0.6, and 0.4 for lists of length 2, 4, 6, and 8, respectively, would have a composite span measure of 11.8. The correlation between the composite memory span measure and the composite task span measure was $.37, F(1, 30) = 4.71, p < .05$, for the 32 subjects with list type 2468 and $.39, F(1, 30) = 5.25, p < .05$, for the 32 subjects with list type 2369.

Collapsing across these groups, the correlation over all 64 subjects was $.37, F(1, 62) = 10.07, p < .01$. These correlations are similar in magnitude to correlations between other measures of working memory span and the traditional memory span (e.g., Daneman & Merikle, 1996).³

Task Performance: The Role of Long-Term Memory Retrieval in Working Memory

As in Experiment 1, mean RT increased with list length, and the effect was greatest on the first list position, as if subjects rehearsed the whole list before beginning to recall or perform. Mean RTs were collapsed across list length and plotted as a function of list position in Figure 7 in order to assess chunking. The means in Figure 7 show scalloped patterns consistent with chunk retrieval, like the pattern seen in Experiment 1. In this experiment, however, the pattern of scalloping depended on the way list length was manipulated. Subjects in the 2468 condition showed elevated RTs in list positions 1, 5, and 7. Subjects in the 2369 condition showed elevated RTs in list positions 1, 4, and 7. As in Experiment 1, the pattern of chunking was the same in the recall condition and the perform condition.⁴

Mean RT was longer for the 100-ms RSI than for the 1,000-ms RSI. This suggests that some of the memory retrieval or task switching or both were done during the RSI. In the 2468 condition, the difference in RT between RSIs was greater in the task span condition (523 ms) than in the memory span condition (343 ms), which suggests that both memory retrieval and task switching occurred in the longer RSI. In the 2369 condition, the difference in RT between RSIs was about the same in the task span condition (206 ms) and the memory span condition (268 ms), which suggests that only memory retrieval went on during the longer RSI.

Statistical support for these conclusions was sought in ANOVAs. The data from the 2468 condition were subjected to a 2 (condition: recall vs. perform) \times 2 (RSI) \times 8 (list position) ANOVA. It revealed significant main effects of condition, $F(1, 30) = 166.14, p < .01, MSE = 1,956,296.07, \eta^2 = .847$; RSI, $F(1, 30) = 8.17, p < .01, MSE = 2,662,181.80, \eta^2 = .214$; and list position, $F(7, 210) = 14.71, p < .01, \eta^2 = .329, MSE = 332,815.62$. The interaction between condition and list position

³ I calculated split-half reliabilities for the task span and memory span measures described in the preceding paragraph. Each subject experienced 10 blocks of trials in the perform condition, in which there was one set of study-test trials for each list length in each block. I calculated the task span separately for odd-numbered blocks and even-numbered blocks for each subject. The split-half reliability for odd and even blocks (adjusted by the Spearman-Brown prophecy formula) was $.86$ for subjects with list type 2468 and $.82$ for subjects with list type 2369. Collapsing over list types, the split-half reliability was $.85$. I calculated split-half reliability in the same way for the memory span measures in the recall conditions. It was $.87$ for subjects with list type 2468 and $.80$ for subjects with list type 2369. Collapsing over list types, the correlation was $.82$.

⁴ The 2369 perform condition shows the pattern of scalloping predicted by progressive reduction of uncertainty (see Figure 7B). RT is slow for the first item in each set of three, intermediate for the second, and fast for the third. It seems unlikely that subjects learned the constraints on list construction and exploited them in this condition because they had very little experience with the lists. Subjects in the 2468 condition, who had the same amount of experience with the lists, and subjects in Experiment 1, who had much more experience with the lists, did not show this pattern.

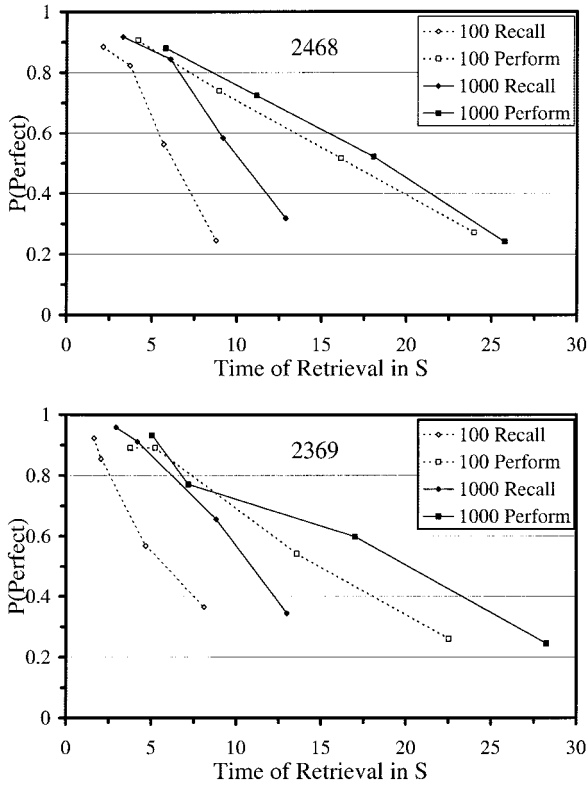


Figure 6. Probability of perfect performance as a function of time of retrieval for recall and perform conditions for intertrial intervals of 100 and 1,000 ms in Experiment 2. A: List length condition 2468. B: List length condition 2369. P(Perfect) = the probability of recalling all of the task names on a list in order perfectly.

was significant as well, $F(7, 210) = 2.97, p < .01, MSE = 243,846.49, \eta^2 = .090$. The pattern of scalloping was assessed with contrasts using the error term from the interaction between task type, RSI, and list position ($MSE = 243,846.49$). To test for elevation in the first, fifth, and seventh positions, the contrast weights were 5, -3, -3, -3, 5, -3, 5, and -3 for list positions 1-8, respectively. The contrast was significant for both RSIs in the recall condition, $F_s(1, 210) = 34.12$ and 11.31 for the 100-ms and 1,000-ms RSIs, respectively, both $ps < .01$, and for both RSIs in the perform condition, $F_s(1, 210) = 158.45$ and 75.06 for the 100-ms and 1,000-ms RSIs, respectively, both $ps < .01$. Scalloping was stronger in the perform conditions than in the recall conditions, $F(1, 210) = 11.38$ and 7.03 for the 100-ms and 1,000-ms RSIs, respectively, both $ps < .01$.

The data from the 2369 condition were subjected to a 2 (condition) \times 2 (RSI) \times 9 (list position) ANOVA. The pattern of results was essentially the same as in the 2468 condition: The main effects of condition, $F(1, 30) = 343.70, p < .01, MSE = 139,066.84, \eta^2 = .920$, and list position, $F(8, 240) = 24.64, p < .01, MSE = 193,923.22, \eta^2 = .451$, were significant, and the main effect of RSI approached significance, $F(1, 30) = 3.76, p = .063, MSE = 2,171,192.35, \eta^2 = .111$. The interaction between condition and list position was significant, $F(8, 240) = 8.29, p < .01, MSE = 137,135.79, \eta^2 = .217$. The pattern of scalloping was assessed with contrasts using the error term from the interaction between condition, RSI, and list position ($MSE = 137,135.79$),

A

B

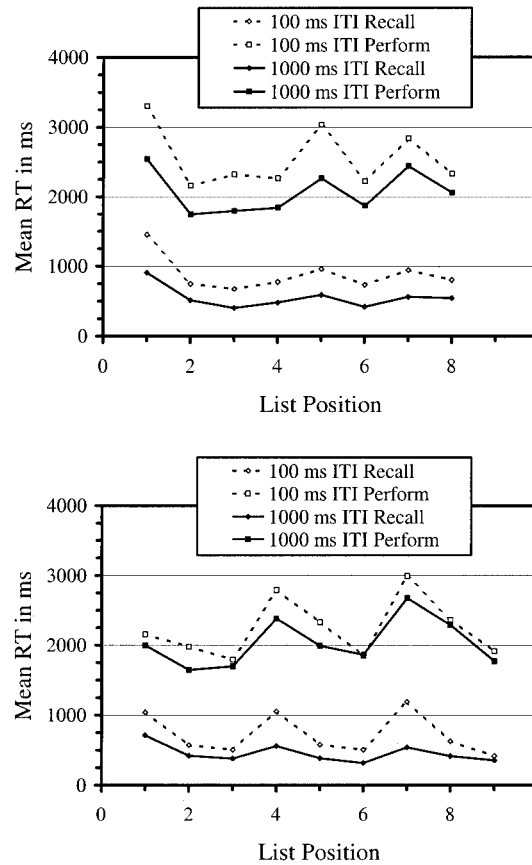
using contrast weights of 2, -1, -1, 2, -1, -1, 2, -1, and -1 for list positions 1-9, respectively, to test for elevation in the first, fourth, and seventh positions. The contrast was significant for both RSIs in the recall condition, $F_s(1, 240) = 185.78$ and 37.48 for the 100-ms and 1,000-ms RSIs, respectively, both $ps < .01$, and for both RSIs in the perform condition, $F_s(1, 240) = 400.40$ and 277.82 for the 100-ms and 1,000-ms RSIs, respectively, both $ps < .01$. Scalloping was stronger in the perform condition than in the recall condition for the 1,000-ms RSI, $F(1, 240) = 7.41, p < .01$, but not for the 100-ms RSI, $F(1, 240) < 1.0$.

Conclusions

Experiment 2 provided converging evidence for the conclusions drawn from Experiment 1: Task spans were essentially the same as the memory spans, suggesting that processing did not trade off with storage. RSI had little effect on task spans or memory spans, suggesting that item-based interference is a better explanation of loss from working memory than is time-based decay. The RTs in the test phase showed a pattern of scalloping suggestive of chunking that followed the pattern of list lengths.

Experiment 3

The third experiment sought converging evidence that the scalloped pattern in the RT data reflects chunking. Subjects performed



A

B

Figure 7. Mean reaction time (RT) as a function of list position (collapsed across list length) for recall and perform conditions for intertrial intervals (ITIs) of 100 and 1,000 ms in Experiment 2. A: List length condition 2468. B: List length condition 2369.

the task span procedure with consistent and varied lists. With consistent lists, they saw the same list of tasks for 10 consecutive study–test trials before proceeding on to the next list. Under these conditions, they should learn the sequence of tasks on each list and be able to retrieve the next task directly from long-term memory, without having to retrieve chunks into working memory. RT should be slow for the first task on the list but it should be fast for each subsequent task with no evidence of the scalloped pattern that is characteristic of chunking. With varied lists, subjects saw a different list on each study–test trial so they could not benefit from learning previous lists. Under these conditions, they should recall the list in chunks and show the scalloped pattern of RTs that is characteristic of chunking.

The lists were constructed in the same way for the consistent- and varied-list conditions, by concatenating random orders of the three tasks and choosing the first N of them for list length N . Thus, the two conditions had the same reduction in uncertainty across the three tasks in each set. If the scalloped pattern is due to progressive reduction in uncertainty, it should appear with both consistent and varied lists. If the scalloped pattern is due to chunk retrieval, it should appear only with varied lists.

Subjects were given practice with each task in two single-task control sessions to familiarize them with the tasks and to provide baseline data against which to assess performance in the consistent- and varied-list conditions.

Method

Subjects

The subjects were 6 graduate and undergraduate students from the general university community who were paid for participating in ten 1-hr sessions. None had served in Experiment 1 or 2.

Apparatus and Stimuli

These were the same as in Experiment 1. In the test phase, the RSI was 1,000 ms.

Procedure

The procedure was the same as in Experiment 1. Subjects performed two single-task sessions, one as practice before beginning the consistent- and varied-list sessions and one after they completed those sessions. They performed 110 trials with each task in each session (i.e., two cycles through the 55 trials required to give them lists of length 1–10). After one session of single-task practice, half of the subjects performed the four sessions with consistent lists before the four sessions with varied lists. The other half of the subjects performed the four sessions with varied lists before the four sessions with consistent lists. In the varied-list conditions, a new list was prepared on each study trial by randomly ordering the three tasks in the same way as in Experiment 1. In the consistent-list condition, one list of each length was prepared at the beginning of each session, and subjects experienced each of those lists 10 times in a session. In both varied- and consistent-list conditions, subjects saw all 10 replications of one list length before progressing on to the next list length.

Results and Discussion

Task Spans

The probability of performing each task on a list perfectly was calculated for each list length in the single-task, consistent-list, and

varied-list conditions. The means across subjects are plotted in Figure 8. Performance in the single-task conditions was about the same as in Experiment 1. Subjects were 80% correct or better for each list length. Performance in the consistent-list condition was indistinguishable from performance in the single-task condition. Apparently, subjects learned the lists well over the 10 repetitions. Performance in the varied-list condition was worse than performance in the consistent-list and single-task conditions. It decreased substantially as list length increased. Nevertheless, it was substantially better than performance with equivalent amounts of practice in Experiment 1 (see Figure 1). Apparently, experiencing the same list length 10 times in succession produces better performance than experiencing a different list length on each trial.

These conclusions were supported by a 3 (list type: single task, consistent, varied) \times 10 (list length) ANOVA on the probabilities of perfect performance. There was a significant main effect of list type, $F(2, 10) = 608.04$, $p < .05$, $MSE = 0.015$, $\eta^2 = .992$; a significant main effect of list length, $F(9, 45) = 55.52$, $p < .01$, $MSE = 0.002$, $\eta^2 = .917$; and a significant interaction between list type and list length, $F(18, 90) = 12.24$, $p < .01$, $MSE = 0.003$, $\eta^2 = .710$.

Task Performance

Mean RTs were collapsed over list length and plotted as a function of list position in Figure 9. RT for the single-task condition was slow in the first list position but uniformly fast for subsequent list positions. RT for the varied-list condition showed a scalloped pattern. List positions 1, 4, 7, and 10 were slower than the others, suggesting that subjects recalled the list in chunks of three. RT for the consistent-list condition was faster overall than RT for the varied-list condition, suggesting a benefit from learning. Consistent-list RTs were slower than single-task controls, possibly because the consistent-list condition required task switching while the single-task condition did not. The most important result was the pattern of performance across list position. It resembled the single-task condition more than the varied-list condition. RT was slow for the first list position and then relatively uniform for the remaining positions, suggesting there was no chunking with consistent lists.

These conclusions received support in a 3 (list type: single task, consistent, varied) \times 10 (list position) ANOVA on the mean RTs.

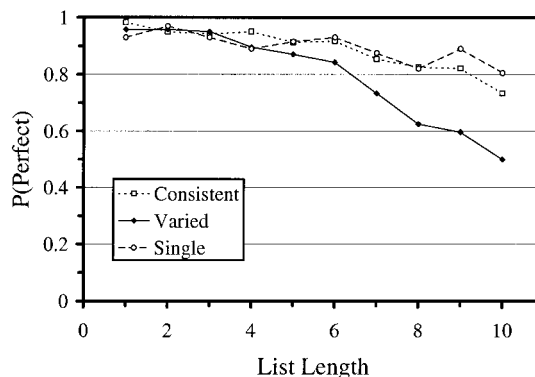


Figure 8. Probability of perfect performance in the consistent-list, varied-list, and single-task conditions in Experiment 3. $P(\text{Perfect})$ = the probability of recalling all of the task names on a list in order perfectly.

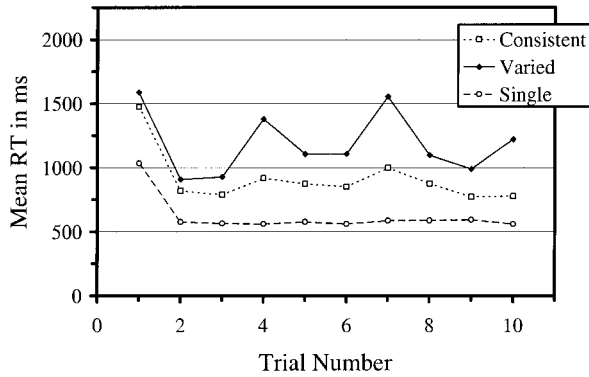


Figure 9. Mean reaction time (RT) as a function of list position (collapsed across list length) for consistent-list, varied-list, and single-task conditions in Experiment 3.

There was a significant main effect of list type, $F(2, 10) = 33.32$, $p < .05$, $MSE = 145,195.09$, $\eta^2 = .870$; a significant main effect of list position, $F(9, 45) = 5.88$, $p < .01$, $MSE = 102,664.79$, $\eta^2 = .540$; and a significant interaction between them, $F(18, 90) = 2.97$, $p < .01$, $MSE = 24,709.21$, $\eta^2 = .373$. The presence of the scalloped pattern was assessed with planned contrasts using the error term from the interaction. One condition compared the first position with the second and third in each set of three, using 3, -2, -2, 3, -2, -2, 3, -2, -2, and 3 as weights. It was significant in the single-task condition, $F(1, 90) = 6.90$, $p < .05$; the consistent-list condition, $F(1, 90) = 26.45$, $p < .01$; and the varied-list condition, $F(1, 90) = 99.61$, $p < .01$. Other contrasts compared the strength of the scalloped pattern between conditions, using the weights 3, -2, -2, 3, -2, -2, 3, -2, -2, and 3 for one condition and -3, 2, 2, -3, 2, 2, -3, 2, 2, and -3 for the other. The scalloped pattern in the varied-list condition was significantly stronger than the patterns in the consistent-list condition, $F(1, 90) = 11.70$, $p < .01$, and the single-task condition, $F(1, 90) = 27.04$, $p < .01$. The scalloped pattern in the consistent-list condition was not different from the pattern in the single-task condition, $F(1, 90) = 3.17$, $p = .079$. A final set of contrasts compared the second and third positions in each set of three, using the weights 0, 1, -1, 0, 1, -1, 0, 1, -1, and 0 to distinguish between chunking and progressive uncertainty reduction. None of the contrasts were significant in any condition, all $F_s(1, 90) < 1.0$, supporting the chunking account.

Conclusions

This experiment was designed to provide converging evidence for long-term memory retrieval in the task span procedure. The scalloped pattern in the function relating RT to list position appeared only in the varied-list condition, which supports the chunk retrieval hypothesis.

Experiment 4

The fourth experiment was designed to seek converging evidence for the null trade-off between task switching and storage. The task span procedure was modified to manipulate the number of task switches subjects had to perform. As before, subjects were given lists of two, four, or six tasks. The novel requirement was

that they had to perform each task several times. One group of subjects—in the *alternating-task* condition—performed the tasks in sequence in the order in which they appeared on the list two, three, or four times. Thus, subjects with a list of two tasks (Hi-Low, Odd-Even) that were performed three times would have six test trials in which they performed the following sequence of tasks: Hi-Low, Odd-Even, Hi-Low, Odd-Even, Hi-Low, Odd-Even. This alternating-tasks condition requires task switching after every target stimulus. The other group of subjects—in the *repeating-task* condition—performed each task on the list two, three, or four times before switching to the next task on the list. Thus, subjects with a list of two tasks (Hi-Low, Digit-Word) that were performed three times would have six test trials in which they performed the following sequence of tasks: Hi-Low, Hi-Low, Hi-Low, Digit-Word, Digit-Word, Digit-Word. This repeating-task condition requires task switching after two, three, or four task repetitions.

The alternating-task condition and the repeating-task condition require different amounts of task switching but involve the same number of task performances. Thus, differences between them are due to task switching rather than task performance. If the executive processing required for task switching trades off with storage in working memory, then task spans should be smaller in the alternating-task condition than in the repeating-task condition.

Method

Subjects

The subjects were 96 students from an introductory psychology course who served to fulfill course requirements. There were six groups of 16 subjects, formed by the factorial combination of number of repetitions (2, 3, or 4) and repeating tasks versus alternating tasks. None of the subjects had served in Experiment 1, 2, or 3.

Apparatus and Stimuli

The apparatus and stimuli were the same as in the previous experiments. The lists were constructed as before, by concatenating random orders of the three task names and selecting the first N of them. The timing parameters were the same as those in Experiment 1. The warning display for the study lists indicated the number of items on the study list (2, 4, or 6). The warning display for the test lists indicated the total number of test trials, which was the product of list length and the number of repetitions (i.e., 4, 8, or 12 for two repetitions; 6, 12, or 18 for three repetitions; 8, 16, or 24 for four repetitions).

Procedure

Each subject experienced a total of 576 test trials. Subjects who performed two repetitions of each task on each list had 24 study-test blocks. Subjects who performed three repetitions of each task on each list had 16 study-test blocks. Subjects who performed four repetitions of each task on each list had 12 study-test blocks. The study and test lists were constructed as in the previous experiments.

Results and Discussion

Task Spans

Mean RTs and percentages of correct responses were calculated for each list position in each trial number (List Position \times Number of Repetitions) for each subject. The probability of responding

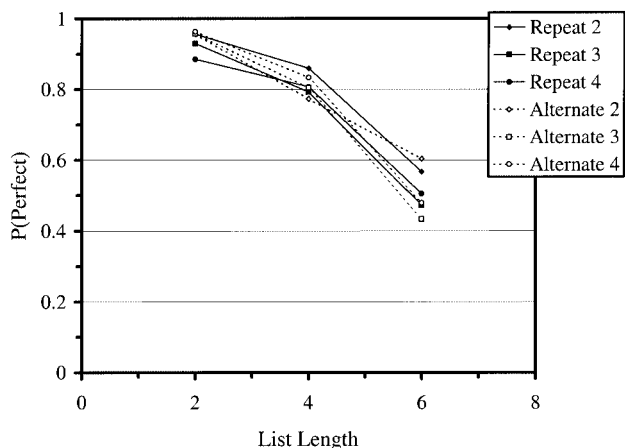


Figure 10. Probability of perfect performance as a function of list length for repeating-task (repeat) and alternating-task (alternate) conditions of Experiment 3. $P(\text{Perfect})$ = the probability of recalling all of the task names on a list in order perfectly.

perfectly is plotted as a function of list length for repeating and alternating tasks and number of repetitions in Figure 10.

The nature of capacity limitations in working memory. The probability of perfect performance decreased as list length increased, but it was not affected systematically by repeating versus alternating tasks or by the number of repetitions. Memory spans could not be estimated because the lowest probability of responding perfectly was greater than 0.5 in three of the six conditions. This implies task spans of 6 or larger, which is consistent with the results of Experiments 1 and 2. The lack of an effect of repeating versus alternating tasks suggests that task switching does not trade off with storage. If it did, accuracy should have been lower with alternating tasks, which required more task switching. The lack of an effect of number of repetitions, particularly in the alternating-task conditions, also suggests that task switching did not trade off with storage. If it did, accuracy should have been lower the greater the number of repetitions and, thus, the greater the number of task switches. The present data converge with the data from Experiments 1 and 2 in suggesting little or no trade-off between processing and storage.

The probabilities of perfect performance were analyzed in a 3 (list length) \times 2 (alternating tasks vs. repeating tasks) \times 3 (number of repetitions) ANOVA. The only significant main effect was the main effect of list length, $F(2, 180) = 333.10, p < .01, MSE = 0.02, \eta^2 = .787$. The main effects of alternating versus repeating tasks, $F(1, 90) < 1.0, MSE = 0.05, \eta^2 = .000$, and number of repetitions, $F(1, 90) = 1.54, p = .219, MSE = 0.05, \eta^2 = .033$, were not significant. The only significant interaction was the interaction between list length and number of repetitions, $F(4, 180) = 3.02, p < .05, MSE = 0.01, \eta^2 = .063$. It reflects a slight tendency for worse performance at the longer list length the greater the number of repetitions, which may be due to an increase in retention interval the greater the number of repetitions.

The loss of information from working memory. The close correspondence between probabilities of perfect performance for equivalent list lengths in Figure 10 is consistent with item-based interference. Figure 11 plots the probability of perfect performance as a function of retrieval time to test time-based decay. There were

large differences between conditions for equivalent retrieval times, which is inconsistent with time-based decay. The differences between conditions in Figures 10 and 11 can be assessed with Fisher's LSD test using the interaction between list length, number of repetitions, and alternating versus repeating tasks ($MSE = 0.01, df = 180$). The LSD value for $p < .05$ is 0.07.

Task Performance

Mean RT, collapsed across list length, is plotted as a function of list position for repeating- and alternating-task conditions in Figure 12. The data from the alternating-task conditions show little effect of the number of repetitions. By contrast, the data from the repeating-task conditions were strongly affected by number of repetitions. RT for the first trial in a run of repetitions was long, and RT for subsequent trials was short. The mean difference between the first trial of a run (a task alternation) and subsequent repeated trials (task repetitions) was 1,291 ms, 2,103 ms, and 1,757 ms for two, three, and four repetitions, respectively. This suggests that the mean time to retrieve a task goal, translate it to control settings, and communicate them to subordinate processes was 1,717 ms.

The RT data were analyzed separately for each number of repetitions. The two-repetition data were analyzed in a 2 (alternating vs. repeating tasks) \times 12 (list length) ANOVA. It produced a significant main effect of alternating versus repeating tasks, $F(1, 30) = 13.39, p < .01, MSE = 2,195,885.48, \eta^2 = .309$; a significant main effect of list position, $F(11, 330) = 38.40, p < .01, MSE = 161,213.84, \eta^2 = .561$; and a significant interaction between them, $F(11, 330) = 17.93, p < .01, MSE = 161,213.84, \eta^2 = .374$. The interaction was due to the effect of task switching that was present in the repeating-task condition but not in the alternating-task condition. This hypothesis was tested with a contrast analysis, using the error term from the interaction and contrast weights of 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, and -1 for list positions 1-12, respectively. The contrast was significant for both the repeating-task condition, $F(1, 330) = 496.60, p < .01$, and the alternating-task condition, $F(1, 330) = 18.43, p < .01$, but sub-

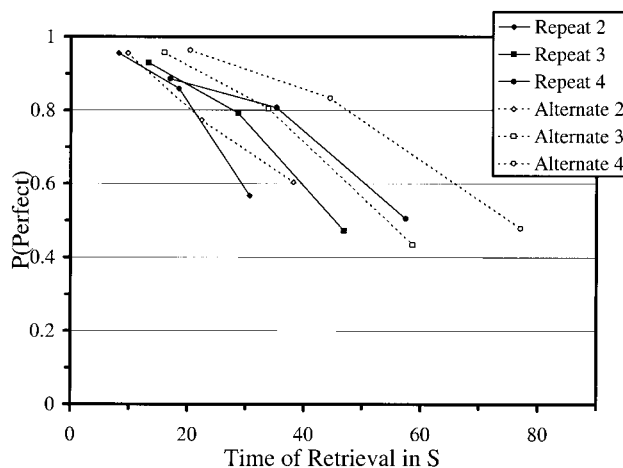


Figure 11. Probability of perfect performance as a function of time of retrieval for repeating-task (repeat) and alternating-task (alternate) conditions of Experiment 3. $P(\text{Perfect})$ = the probability of recalling all of the task names on a list in order perfectly.

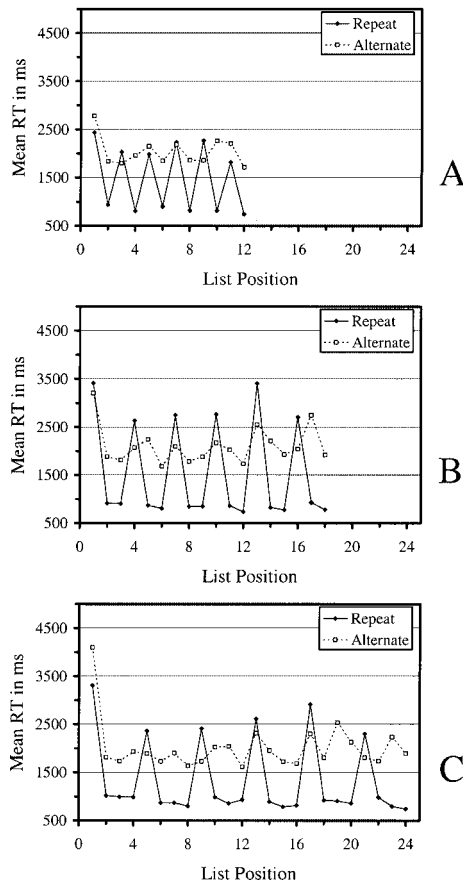


Figure 12. Mean reaction time (RT) as a function of list position (collapsed across list length) for repeating-task (repeat) and alternating-task (alternate) conditions of Experiment 3. A: Two repetitions. B: Three repetitions. C: Four repetitions.

stantially stronger in the repeating-task condition. A contrast testing the difference between these contrasts was highly significant, $F(1, 330) = 161.85, p < .01$.

The three-repetition data were analyzed in a 2 (alternating vs. repeating tasks) \times 18 (list length) ANOVA. It produced a significant main effects of alternating versus repeating tasks, $F(1, 30) = 11.41, p < .01, MSE = 4,066,959.58, \eta^2 = .276$, and list position, $F(17, 510) = 59.23, p < .01, MSE = 228,789.57, \eta^2 = .664$, and a significant interaction between them, $F(17, 510) = 26.83, p < .01, MSE = 228,789.57, \eta^2 = .472$. The interaction was subjected to a contrast analysis, using the error term from the interaction and contrast weights of 2, -1, -1, 2, -1, -1, 2, -1, -1, 2, -1, -1, 2, -1, and -1 for list positions 1–18, respectively. The contrast was significant for the repeating-task condition, $F(1, 510) = 1,237.18, p < .01$, and the alternating-task condition, $F(1, 510) = 38.26, p < .01$. A contrast testing the difference between these contrasts was highly significant, $F(1, 510) = 420.16, p < .01$.

The four-repetition data were analyzed in a 2 (alternating vs. repeating tasks) \times 24 (list length) ANOVA. It produced significant main effects of alternating versus repeating tasks, $F(1, 30) = 27.23, p < .01, MSE = 3,263,006.36, \eta^2 = .476$, and list position, $F(23, 690) = 49.96, p < .01, MSE = 216,209.17, \eta^2 = .625$, and

a significant interaction between them, $F(23, 690) = 16.35, p < .01, MSE = 216,209.17, \eta^2 = .353$. The interaction was assessed with a contrast analysis, using the error term from the interaction and contrast weights of 3, -1, -1, -1, 3, -1, -1, -1, 3, -1, -1, -1, 3, -1, -1, -1, 3, -1, -1, -1, 3, -1, -1, and -1 for list positions 1–24, respectively. The contrast was significant for the repeating-task condition, $F(1, 690) = 1,028.02, p < .01$, and the alternating-task condition, $F(1, 690) = 70.12, p < .01$. A contrast testing the difference between these contrasts was highly significant, $F(1, 690) = 280.59, p < .01$.

Individual Differences in Task Spans and Task Switching

To investigate the relation between individual differences in task span and task switching, I correlated measures of the task span with measures of task-switching time for subjects in the repeating-task condition. Task span measures were calculated as in Experiment 2, by multiplying the probability of perfect performance at a given list length by that list length and summing the products over list length to produce a single score for each subject. Task-switching times were calculated by subtracting mean RT for the second trial (the first repetition) in a run from mean RT for the first trial (an alternation) in a run. There were six differences for each subject (corresponding to the six peaks in the repeating-task functions in Figure 12). The average of those six differences was calculated to produce a single task-switching score for each subject. These scores were then correlated with the task span measures. The correlation was small and not significant, $r = -.16, F(1, 46) = 1.18, p = .28$, suggesting no relation between individual differences in task span and individual differences in task-switching time.⁵

Conclusions

Task spans were unaffected by the amount of task switching required. Task spans were the same for subjects who repeated tasks several times as for subjects who alternated between tasks on every trial. Moreover, task spans were the same regardless of the number of times the tasks had to be performed, which also determined the number of task switches that were required. This null effect of task switching on task span suggests that storage in working memory does not trade off with the executive processing required for task switching. This conclusion converges with the conclusion drawn from comparisons between task spans (which required task switching) and memory spans (which did not) in Experiments 1 and 2.

General Discussion

Four experiments examined the relation between the number of tasks a person could remember and execute perfectly—the task span—and the ability to switch from one task to another. There

⁵ I calculated the split-half reliability for the task span measure by calculating the span for odd and even blocks for each subject in the repeating-list condition. The split-half reliability (adjusted by the Spearman-Brown prophecy formula) was .84. I calculated the split-half reliability for the task-switching time measure by calculating mean RT for repetitions and alternations on odd and even blocks and subtracting the mean for repetitions from the mean for alternations for each subject. The split-half reliability for odd and even blocks was .94.

were four major conclusions: First, the experiments suggested there was no trade-off between processing and storage. In Experiments 1 and 2, task spans were about the same as memory spans despite the fact that the task span procedure required substantial amounts of task switching. In Experiment 4, task spans were the same when subjects switched every trial as when they switched every second, third, or fourth trial. Second, the data were more consistent with item-based interference than with time-based decay as the mechanism for loss of information from working memory. In Experiments 1, 2, and 4, there was little difference between conditions when retrieval probabilities were plotted against the number of items on the list, but there were dramatic differences between conditions when retrieval probabilities were plotted against retrieval time. Third, Experiments 1, 2, and 3 showed a scalloped pattern in retrieval times that suggested subjects retrieved the lists in chunks from long-term memory instead of keeping the entire list active in working memory. Fourth, Experiment 1 compared RTs in the task span procedure with the sum of RTs in the memory span and single-task conditions and found a substantial difference, suggesting that task switching requires more than changing goals or mapping rules in working memory. These conclusions bear on several issues in the literature on working memory and executive control.

Capacity Limitations in Working Memory

The evidence against strong trade-offs between processing and storage is consistent with a growing body of evidence for multiple-process or multiple-resource theories of working memory (Baddeley, 1996, 2000; Baddeley & Logie, 1999; Engle et al., 1999; Kane et al., 2001; Kieras et al., 1999; Schneider & Detweiler, 1987). The evidence is inconsistent with theories that propose that processing and storage are limited by the same capacity for activation (Anderson et al., 1996; Just & Carpenter, 1992; Lovett et al., 1999). This is an important conclusion, but its implications are not very direct. There are many possible architectures that involve multiple executive processes. A lot of theoretical work must be done to derive alternative hypotheses and discover ways to distinguish them. In my view, formal computational models are a necessary step in this development (see Logan & Gordon, 2001; also see Anderson et al., 1996; Kieras et al., 1999; Schneider & Detweiler, 1987). An adequate theory will not simply arise from the data, no matter how much we collect.

The Loss of Information From Working Memory

The present results consistently confirmed the strong prediction of item-based interference theories of information loss, that the number of items intervening between encoding and retrieval was the main determinant of performance (Waugh & Norman, 1965). The results consistently disconfirmed the strong prediction of time-based decay theories of information loss, that the amount of time between encoding and retrieval was the main determinant of performance (Doshier & Ma, 1998). The results are generally consistent with theories of working memory that attribute memory loss to interference (Baddeley & Logie, 1999; Cowan, 1999; Schneider & Detweiler, 1987) and are inconsistent with theories that attribute memory loss to decay (Anderson et al., 1996; Anderson & Matessa, 1997; Baddeley & Logie, 1999; Cowan, 1999; Kieras et al., 1999; Lovett et al., 1999). The contrast between

interference and decay is a complex issue that remains to be resolved (Nairne, 2002). The present research supports interference accounts but does not rule out all possible decay accounts. Again, I believe that formal computational models are a necessary part of the resolution (also see Nairne, 2002).

The Role of Long-Term Memory in the Task Span Procedure

In some respects, it is surprising that subjects did as well as they did on the task span procedure. We are accustomed to thinking of working memory capacity as a strong limit on performance, which implies that subjects should do well up to the limit of working-memory capacity and then fail consistently afterward. The accuracy data in Table 1 show they did much better than we might expect. Subjects responded correctly 48% of the time on the 10th item of the 10-item lists. There were six possible responses, so choosing one response at random would be correct one sixth or 16.7% of the time. There were three tasks, so choosing one task at random and performing it correctly would result in correct performance on one third or 33% of the trials. Moreover, subjects performed perfectly up to the 10th response only 25% of the time (i.e., $P[\text{Perfect}] = .25$ for list length of 9), so they were likely to have made at least one error before the 10th item. Apparently, they were able to recover from errors and salvage their performance on subsequent items (also see Anderson & Matessa, 1997).

The chunking apparent in the scalloping in the RT data provides some insight into the processes underlying this ability to perform with above-chance accuracy beyond the capacity of working memory and to recover from errors. Recalling chunks from long-term memory allows subjects to perform accurately beyond the limits of working memory because their performance is not based entirely on information in working memory (cf. Ericsson & Kintsch, 1995). Recalling chunks in series allows for recovery from error. If subjects remember that chunk $N + 1$ follows chunk N , they can make errors during chunk N and still recall chunk $N + 1$ correctly.

This analysis suggests an important role for long-term memory in executive control. Executive control need not depend only on what is present in working memory. It can also depend on what can be retrieved from long-term memory (Ericsson & Kintsch, 1995). Long-term memory retrieval is likely to be a deliberate, active process in which the executive decides it needs specific information at a specific point in time and engages a specific retrieval process to get it. Long-term memory retrieval may be one of the endogenous acts of control in the executive's repertoire that it engages to manage performance. Allport and colleagues emphasized the role that passive retrieval from long-term memory may play in task-switching experiments in discussing task set inertia (Allport et al., 1994) and stimulus-specific priming (Allport & Wylie, 2000; Wylie & Allport, 2000). The present analysis suggests that active retrieval may play an important role as well (also see Mayr & Kliegl, 2000).

By contrast, information that is present in working memory need not require a deliberate, active process to make it available (but see Nairne, 2002). The information is simply there to be used should the executive or the subordinates need it. Some executive processing may be necessary to keep information active in working memory (e.g., rehearsal), but that processing need not be locked so closely to the time at which the information is needed. Consequently, there may not be strong trade-offs between working

memory and executive processes like task switching. There may be stronger trade-offs between executive processes and deliberate retrieval from long-term memory.

Executive Control and the Task Span Procedure

Three procedures have been used to capture the endogenous act of control in task switching. Each procedure has problems that challenge its ability to capture executive processing (Logan, 2003). The alternating-task procedure compares performance when subjects alternate between tasks with performance when they repeat the same task. RT is longer with alternating tasks than with repeated tasks, and this difference is interpreted as a measure of task switching (Jersild, 1927; Rubinstein et al., 2001). Rogers and Monsell (1995) pointed out that the alternating and repeating conditions differ in memory load and in the requirement to keep track of which task to perform, and this may contribute to the difference in RT independent of task switching. They proposed the *alternating runs* procedure to deal with these problems. It requires subjects to alternate between runs of successive tasks (e.g., AABBAABB) so that alternations (AB, BA) and repetitions (AA, BB) are observed in the same sequence of trials. This equates memory load and the requirement to keep track of the current task. However, the alternating runs procedure does not control the time at which task switching begins. Subjects may prepare some processes in advance and wait for the next stimulus before switching other processes (Rogers & Monsell, 1995). The *explicit task-cuing* procedure was intended to allow control over the onset of task-switching processes (Meiran, 1996). A cue presented before each target tells subjects which task to perform. Trials are sorted into task alternations and task repetitions post hoc and switching time is inferred from the difference between them. Logan and Bundesen (2003, in press) found evidence that the explicit task-cuing procedure may not involve task switching. Together, the cue and the target uniquely specify the response, so subjects may use a "compound stimulus strategy" in which they choose the response associated with the combination of the cue and the target.

The task span procedure may overcome some of these difficulties. Memory load and the requirement to keep track of the current task can be controlled precisely (e.g., Experiment 4). Only target stimuli are presented in the test trials. Subjects must recall which task to perform on their own. Timing can be controlled to measure the duration of preparatory processes (e.g., Experiment 2). Task span performance can be compared with control conditions to isolate and measure the duration of various control processes (e.g., Experiment 1).

The Task Span and Other Memory Spans

This article compared the task span with the traditional memory span. Working memory is often studied with a third kind of span task, variously called the reading span (Daneman & Carpenter, 1980), the counting span (Case, Kurland, & Goldberg, 1982), and the operations span (Engle et al., 1992). In the reading span, subjects are required to read a series of sentences and remember the last word in each sentence. The reading span is the number of final words they can remember in order perfectly. In the counting span, subjects enumerate a series of displays and remember their numerosities. The counting span is the number of numerosities they can remember in order perfectly. In the operations span,

subjects see a series of arithmetic operations, each followed by an unrelated word. They must perform the operations and remember the words. The operations span is the number of words they can remember in order perfectly. In each case, the task requires subjects to store information while processing other information, which is the core idea in the concept of working memory.

Reading, counting, and operation spans are often used to study individual differences. They are correlated with the traditional memory span, but they correlate even more strongly with measures of cognitive abilities, like the Scholastic Aptitude Test (Daneman & Merikle, 1996). They are correlated with some measures of executive abilities but not with others (Miyake et al., 2000). Experiments 2 and 4 showed that the task span is correlated with the traditional memory span but not with task-switching time.

It would be interesting to compare the task span with the reading, counting, and operation spans. The task span and the reading, counting, and operation spans differ from the traditional memory span in that they all require subjects to perform another task in addition to the memory task. Reading, counting, and operation span procedures require a dual task during the study phase. The task span procedure requires a dual task during the test phase. Studies of long-term memory show different effects of dual tasks at study and at test. Dual tasks at study interfere with memory performance, but the dual tasks are performed well. Dual tasks at test do not interfere as much with memory performance, but the dual tasks are performed poorly (Baddeley et al., 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). Similar results may be found with the span procedures.

It would be interesting as well to investigate individual differences in the task span and relate them to individual differences in the reading, counting, and operation spans as well as to individual differences in cognitive abilities. The task span requires subjects to process and store information at the same time, so it may show the same pattern of correlation with measures of abilities as the reading, counting, and operation spans. However, the task span requires concurrent processing during retrieval, whereas the reading, counting, and operation spans require concurrent processing during encoding, so the patterns of individual differences may be different.

Conclusions

The results of the four experiments suggest that several executive processes underlie performance in the task span procedure. The processes that underlie storage seem to be separate from the processes that underlie task performance (Experiments 1, 2, and 4), and task switching appears to involve processes outside of working memory (Experiment 1). The data are more consistent with theories that assume multiple executive processes than with theories that assume a single executive. The task span procedure promises to yield new insights into working memory, task switching, and the interaction between them in complex performance.

References

- Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentions set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421–452). Cambridge, MA: MIT Press.
- Allport, A., & Wylie, G. (2000). Selective attention and task set: Activation and inhibition of competing tasks. In S. Monsell & J. Driver (Eds.),

- Attention and performance XVIII* (pp. 35–70). Cambridge, MA: MIT Press.
- Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.
- Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221–256.
- Arrington, C. M., & Logan, G. D. (in press). The cost of a voluntary task switch. *Psychological Science*.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49A*, 5–28.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Cognitive Sciences*, *4*, 417–423.
- Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*, 641–657.
- Baddeley, A., Eldridge, M., Lewis, V., & Thompson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*, 518–540.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge, England: Cambridge University Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.
- Bower, G. H., & Winzenz, H. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, *80*(2, Pt. 2).
- Bundesden, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.
- Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–84.
- Cowan, N. (1995). *Attention and memory*. Oxford, England: Oxford University Press.
- Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, England: Cambridge University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*, 1–17.
- Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*, 113–132.
- Cowan, N., Wood, N. L., Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science*, *8*, 290–295.
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.
- Doshier, B. A., & Ma, J. J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 316–335.
- Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *54A*, 31–48.
- Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 972–992.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, England: Cambridge University Press.
- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980, June 6). Acquisition of a memory skill. *Science*, *208*, 1181–1182.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Gilbert, S., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*, 193–251.
- Gopher, D., Armony, L., & Greenspan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General*, *129*, 308–339.
- Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.
- Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49A*, 116–139.
- Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*, 184–198.
- Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology* (Whole No. 89).
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
- Kahana, M. J., & Jacobs, J. (2000). Interresponse times in serial recall: Effects of intraserial repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1188–1197.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47–70.
- Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights

- into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 183–223). Cambridge, England: Cambridge University Press.
- Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, *107*, 32–63.
- Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 189–207.
- Logan, G. D. (2003). Executive control of thought and action: In search of the wild homunculus. *Current Directions in Psychological Science*, *12*, 45–48.
- 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*, 575–599.
- Logan, G. D., & Bundesen, C. (in press). Very clever homunculus: Compound stimulus strategies for the explicit task cuing procedure. *Psychonomic Bulletin & Review*.
- Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393–434.
- Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). Cambridge, England: Cambridge University Press.
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1124–1140.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1423–1442.
- Meiran, N. (2000). Reconfiguration of stimulus task sets and response task sets during task switching. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII* (pp. 377–399). Cambridge, MA: MIT Press.
- Miller, G. A. (1956). The magical number 7 plus or minus 2: Some limits on our capacity to process information. *Psychological Review*, *63*, 81–97.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, England: Cambridge University Press.
- Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84–92.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*, 522–536.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*, 391–408.
- Rogers, R. D., & Monsell, S. (1995). The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.
- 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*, 763–797.
- Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and memory* (Vol. 21, pp. 54–119). New York: Academic Press.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.
- Sohn, M.-H., & Anderson, J. R. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology: General*, *130*, 764–778.
- Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 674–685.
- Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin*, *81*, 695–711.
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.
- Wyllie, G. R., & Allport, D. A. (2000). Task switching and the measurement of “switch costs.” *Psychological Research*, *63*, 212–233.

Received May 7, 2003

Revision received January 28, 2004

Accepted January 28, 2004 ■