The Role of Working Memory Capacity in Retrieval

Virginia M. Rosen
University of South Carolina

Randall W. Engle
Georgia Institute of Technology

The verbal fluency task requires generation of category exemplars and appears to be an example of what M. Moscovitch (1995) calls a strategic test of memory retrieval. Four experiments explored the role of individual differences in working memory (WM) capacity on verbal fluency under various secondary load conditions. High WM participants consistently recalled more exemplars. However, load conditions caused a decline in recall only for high WM participants. Low WM participants showed no effect of secondary workload on exemplar generation. WM group differences and load effects were observed even in the 1st min of retrieval, which suggests that differences were not due to differences in knowledge. A model of retrieval is supported that relies on cue-based-automatic activation, monitoring of output for errors, controlled suppression of previously recalled items, and controlled strategic search.

There is an old idea, first proposed by Baldwin (1894), that attentional resources are limited in some basic ways and that this limitation is reflected in the memory-span task. Baldwin and others argued that memory-span tasks reflect a fundamental aspect of intellectual abilities, both for the developing human and across individuals at given stages of development. The simple digit-span task, for example, consists of presenting a person with a list of digits and having the person recall the digits in the same order in which they were presented. The span score is the maximum number of digits that can be recalled in perfect order. The digit- and word-span tasks are both frequently found in standardized tests of intelligence even today. Although these two simple memory-span tasks distinguish certain subgroups such as children with learning disabilities, neither can be shown to consistently and reliably predict such mainstays of higher level cognition as reading or listening comprehension among the population with normal learning abilities (Perfetti & Lesgold, 1977).

Baddeley and Hitch (1974) proposed a system they re-

ferred to as *working memory* that combined aspects of the processing of information with the storage of intermediate products of that processing. The system consisted of two different data-representation elements: one for speech-based information called the *articulatory loop* and another for visual and spatial information called the *visuo-spatial sketchpad*. The third element of Baddeley and Hitch's model is a limited-capacity attention mechanism called the *central executive* that appears to be conceptually similar to what Norman and Shallice (1986) called the *supervisory attentional system* and related to what Posner and Snyder (1975) and Schneider and Shiffrin (1977) referred to as *controlled attention*. The central executive may also be related to the *anterior attentional system* proposed by Posner and Peterson (1990). And, studies of individual differences in working memory capacity (Turner & Engle, 1989) more than likely reflect individual variation in the limited-capacity attention mechanism.

Daneman and Carpenter (1980) developed the first valid measure of working memory capacity. Their reading-span task is really a dual task that requires the participant to read or listen to a series of sentences and, separately, to keep track of the last word of each sentence so that they can be recalled later. The span score is the maximum number of words that can be recalled perfectly. However, a wide variety of other complex dual-task measures have now been used to measure working memory capacity (Kyllonen & Christal, 1990; Salthouse, Mitchell, Skovronek, & Babcock, 1989; and Turner & Engle, 1989), and we have argued elsewhere (Cantor, Engle, & Hamilton, 1991) that these measures reflect a common mechanism.

Further, that mechanism is apparently of fundamental importance to higher level cognition because measures of working memory capacity reliably predict performance in a wide variety of real-world cognitive tasks. Significant relationships with measures of working memory capacity have been reported for reading comprehension (Daneman & Car-

penter, 1980, 1983); language comprehension (King & Just, 1991; MacDonald, Just, & Carpenter, 1992); learning to spell (Ormrod & Cochran, 1988); following directions (Engle, Carullo, & Collins, 1991); vocabulary learning (Daneman & Green, 1986); note taking (Kiewra and Benton, 1988); writing (S. L. Benton, Kraft, Glover, & Plake, 1984); reasoning (Kyllonen & Christal, 1990); and complex learning (Kyllonen & Stephens, 1990; Shute, 1991).

Notice that all of the tasks in the above list reflect some level of acquisition or learning. Our primary research question for the present set of studies is directed at whether working memory capacity is important to retrieval. It is our opinion that working memory capacity plays an important role in both acquisition and some types of retrieval. Moscovitch (1995) has distinguished between two different types of memory tests. *Associative memory tests* are those in which the cue leads to rather automatic retrieval of the target memory. An example of this would be when the face of a friend leads quickly and with no effort to the retrieval of that person's name. Moscovitch (1995) refers to the other type of memory test as "strategic" because the cue serves as the "starting point of a memory search that has elements in common with problem solving" (p. 1345). Moscovitch argues further that although the associative process is automatic and is mediated by the hippocampal system, the strategic process is effortful, under voluntary control, and mediated by the prefrontal cortex of the frontal lobes. The connection between controlled, effortful retrieval and the frontal lobes fits with developmental findings. Improvements in cognitive performance at the beginning of the life cycle and declines in performance at the end of the life cycle are often attributed to the development of the frontal lobes in children and their deterioration in the elderly (cf. Gerstadt, Hong, & Diamond, 1994; West, 1996). Thus, research on the role of individual differences in working memory capacity and controlled retrieval likely has implications for the role of development on retrieval as well.

One approach to examining the importance of controlled attention on retrieval is to have participants learn and retrieve under a mental load. For example, Baddeley, Lewis, Eldridge, and Thomson (1984) had individuals perform a variety of learning and memory tasks under conditions of load or no load. The memory tasks included paired-associates learning, free recall, sentence verification, and verbal fluency. The two types of concurrent load tasks were card sorting and maintaining a sequence of digits in memory. Baddeley et al. found that the concurrent load hurt acquisition in all of the tasks. Although retrieval was slowed somewhat under load, the probability of correct retrieval was unaffected by concurrent load. The exception was a verbal fluency task in which participants generated all the exemplars of the "animal" category they could think of over a 2-min period. In this task, the concurrent memory load caused a reduction of over 30% in the number of animal names retrieved.

A more recent article by Craik, Govoni, Naveh-Benjamin, and Anderson (1996) also showed that successful retrieval, at least in some tasks, was minimally affected by a concurrent divided-attention task. A divided-attention task was performed during encoding or during retrieval in single-trial free recall, paired associates, and single word-recognition tasks. These authors also manipulated the emphasis given to the primary and secondary tasks. Across tasks, divided attention at encoding hurt later recall, but performance on the divided attention task was unimpaired. Divided attention at retrieval had a slight affect on the success of retrieval, but performance on the divided attention task was slowed considerably. This finding suggested that controlled attention was necessary for encoding but that retrieval was relatively automatic and seized attention from the secondary task.

The evidence seems clear that although some forms of retrieval require controlled, effortful attention (e.g., strategic search), other forms occur relatively automatically. Our logic was that individual differences on measures of working memory or central executive capacity should be reflected in measures of controlled, attention-demanding retrieval but should not be important to measures of retrieval that are based on automatic activation. The category-generation task used by Baddeley et al. (1984) should be a task that demands strategic retrieval and lends itself to the study of the influences of individual differences in working memory on such tasks.

Our primary research question was whether working memory is important to strategic retrieval. However, we also assumed that our findings allow us to make some inferences relevant to the relationship between frontal lobe functions and retrieval. There is growing speculation that central executive functions are localized in some portion of the frontal lobe, probably the dorsolateral prefrontal cortex. As we said earlier, the central executive appears to correspond to the supervisory attentional system that Norman and Shallice (1986) attributed to the frontal lobes. Baddeley (1986) and Baddeley and Wilson (1988) have referred to the term *dysexecutive syndrome* when talking about the pattern of working memory and attentional deficits found in patients with frontal lobe damage, which suggests that a link exists between the syndrome and reduced functioning of the central executive. The attentional deficits of these patients with frontal lobe damage included difficulty with controlling, and dividing, attention.

Engle and Oransky (in press) extended the link between frontal lobes and central executive to the normal population when they observed that the ends of the dimension we call working memory capacity yields similar patterns as is found between patients with frontal lobe damage and individuals with undamaged frontal lobes. Therefore, the fact that frontal lobe damage often leads to deficits on the verbal fluency task (A. L. Benton, 1968; Milner, 1964) would lead us to expect differences on this task among individuals of differing working memory capacity.

The verbal fluency task has a venerable past. Thurstone (1938) included verbal fluency as part of the Primary Mental Abilities (PMA) Test. In a category-fluency task, an individual is told to generate category instances over a period of time, typically several minutes. Individuals typically emit "clusters," or groups of semantically related words with pauses between the clusters. For example, if the category were animals, an individual might say "wolf, deer,

bear," followed by a pause and then "dolphin, shark, whale." These clusters can be identified both by a judges' assessment of the semantic relatedness of the words within a cluster and by the time intervals between the spoken words. Words within the same cluster tend to be spoken in bursts with short intervals between the words. The intervals between the last word in one cluster and the first word in the next cluster tend to be longer. Further, over the retrieval period, the intercluster duration time tends to increase while the size of the clusters and the intervals between the words within a cluster tends to remain relatively stable (Gruenewald & Lockhead, 1980). Some of the studies described later explore these various aspects of the retrieval protocol.

We report the results of four experiments that address the role of individual differences in working memory and the further effects of a concurrent task on verbal fluency. The first study tested the prediction that individuals who score high on a measure of working memory capacity generate more members of a category than individuals who receive a low working memory-span score. The next three studies explored the effects of retrieval under various mental work loads for high- and low-span participants.

General Method

There were some procedures common to all of the studies reported here.

Participants

All participants were students at the University of South Carolina earning credits for their psychology classes. All had participated earlier in a task designed to measure their working memory capacity. The American Psychological Association's ethical guidelines for the treatment of human participants were followed in all four experiments.

Operation-Span Task

Participants in all four experiments were prescreened on a variation of the Turner and Engle (1989) operation-span task. The operation-span task consisted of three randomly ordered presentations of five sets of math operations with an accompanying word (e.g., IS $(2 \times 3) + 2 = 5$? DOG). Each set contained 2–6 math operations that appeared one at a time in the center of the computer screen. Participants were required to read each operation aloud, to verify aloud whether the answer was correct, and then to say the word that followed the operation. Each set was followed by a recall cue to write down the words that followed each math operation in the set, in their exact order, on a response sheet. The span score was then calculated by summing the total number of words recalled, across correctly recalled sets. The participants were then placed into one of two groups based on whether their performance fell into the lower (low-span) or upper (high-span) quartile of performance across all participants.

Category Task Procedure

Participants in all four experiments were video- or audio-taped while generating exemplars from the animals category. The experimenter was not present in the room during the generation task in the first and third experiments but was present in the second and fourth experiment, wherein participants concurrently tracked digits during category generation. The participants were told to start generating names as soon as they were given the category name. Participants in Experiments 1, 2, and 3 were told to keep repetitions to a minimum, whereas participants in Experiment 4 were not discouraged from making repetitions when generating animal names. Participants in all of the experiments were told to keep trying to retrieve new names throughout the duration of the experiment.

Scoring

In the first three experiments, the video- and audiotapes were transcribed, and interword retrieval times were measured through the use of a computerized timing program. The timing program was written with Micro Experimental Laboratory software (Schneider, 1988). The timing procedure involved having a scorer listen to the taped output of each participant while simultaneously looking at the written transcript of the output. The scorer pressed the space bar on a computer keyboard each time they heard a word, and the computer program stored the times between keypresses. A hyperbolic-based algorithm then registered cluster boundaries in the output of each participant on the basis of slope differences between these interword retrieval times (see Gruenewald & Lockhead, 1980, for details about the slope-difference algorithm). The algorithm required a slope and asymptote parameter for each participant. We obtained these parameters from fitting the mean number of animal names retrieved by each participant at 30-s intervals to a hyperbolic function.

We found that for some participants, the algorithm either defined each word in the cumulative output as a cluster boundary or defined the first 10–20 words of output as a single cluster. We used the following procedure to adjust the slope parameter in these cases. We examined the 1st min of retrieval for uniformity of interword intervals. If the interword intervals were uniform, then we did not adjust the slope on that individual. However, if we found a departure from uniformity in the interword intervals, then we increased or decreased the slope until we saw a cluster of two or more words appear at the point of departure in the output. For example, a point of departure might have consisted of two words that were separated by less than 1 s and then followed by another word 4+ s later. In other words, a point of departure in uniformity existed where there was a change in the slope of the cumulative output. We adjusted the slope and asymptote parameters for 4 participants in Experiment 1, 4 participants in Experiment 2, and 5 participants in Experiment 3. These participants were predominantly high-span participants. Although a hyperbolic function best rep-

resents the cumulative output function of free recall, it is not unusual to find that the output of some individuals is not perfectly fit by a hyperbolic function (Wixted & Rohrer, 1994). A discussion of output functions, and individual differences in approach to asymptote, is beyond the scope of this article.

In the first and second experiment only, we made a comparison between the number of cluster boundaries determined by the algorithm and the number of cluster boundaries determined by three judges. The transcripts for all of the participants were studied by three judges, who determined the placement of cluster boundaries based on semantic relatedness. Each judge rated each participant's transcript three times, placing a check mark beside each word that they were certain marked a semantic cluster boundary. Cluster boundaries for each judge were then defined as only those words that had three check marks placed beside them. An interrater reliability analysis (Shrout & Fleiss, 1979) was then conducted on the judges' ratings. The judging procedure and algorithm were based on Gruenewald and Lockhead (1980). Finally, we looked at the relationship between only those cluster boundaries as defined by the three judges and the cluster boundaries that were determined by the algorithm. This was done to explore the relationship between the algorithm's time-based cluster boundaries and the judges' semantic-based cluster boundaries. The algorithm-based cluster information was used in all of the analyses reported here.

Analyses

Five dependent measures were extracted from each participant's output: mean number of words retrieved; mean number of clusters retrieved; mean time between clusters; mean size of a cluster; and mean time within a cluster.

Experiment 1

The question addressed in the first experiment was whether individuals who score high on a measure of working memory capacity would generate more animal names than individuals who score low on the same measure. All of the participants in Experiment 1 generated animal names for 15 min.

Method

Participants

Thirty undergraduate students from the University of South Carolina participated in exchange for course credit in their psychology classes. The participants were assigned to low- and high-span groups based on their operation-span score. There were 15 participants in each span group. The mean operation-span score was 8 for the low-span and 24 for the high-span group.

Procedure

Participants were videotaped while generating animal names for 15 min. The category task procedure was previously described in the general methods section.

Results

Cluster Boundary Judging

We first examined the consistency of the semantic-based judgments of cluster boundaries among our three judges. An interrater reliability analysis produced a coefficient of .66 as an index of the level of cluster-boundary correspondence among our three judges. We then compared the semantic-based boundaries shared by all three judges with the algorithm-based cluster boundaries. A Pearson product-moment correlation (r) showed a significant relationship between the two cluster-boundary judging methods, $r(450) = .75$, $p < .0001$. We found that this relationship was maintained across all but the first 2 min of retrieval.

We believe that the lack of a relationship in the first 2 min was due to greater sampling among related subcategories in those first 2 min. For example, a participant might retrieve three names from a small dog category followed by three names from a large dog category. If the interval between the two groups of dog names was larger than the intervals between the names within each group, then the algorithm would have registered two clusters each containing three dog names. In contrast, the judges would have registered the two groups as one large cluster of six dog names because of their semantic relatedness. In fact, the average cluster size determined by the judges and the algorithm in the 1st min was 4 and 1.67 names, respectively. The difference in cluster size in the 1st min, as determined by the two methods, proved to be significant, $t(28) = 11.08$, $p < .01$. In spite of the lack of a correlation for the two judging methods in the first 2 min of retrieval, we felt confident about using the algorithm-based cluster information for the analyses.

Retrieval of Names

Table 1 contains the significance (F values) and mean square errors ($MSEs$) for the analysis of variance (ANOVA), with span as a between-subjects variable. Figure 1A shows that the high-span participants retrieved more animal names when compared to the number of animal names retrieved by the low-span participants, $F(1, 28) = 23.08$, $p < .0001$ (mean of 130 vs. 85 names). The span difference in the number of animal names retrieved was visible in the 1st min of retrieval, $F(1, 28) = 14.0, p < .001$ (mean of 27 vs. 20 names, high and low span, respectively). Figure 1B shows that the high-span participants were faster than the low-span participants to retrieve clusters of animal names, $F(1, 28) = 13.38$, $p < .001$ (mean of 8.95 vs. 11.70 s between clusters), which led to the retrieval of more clusters for the high- than low-span participants, $F(1, 28) = 24.49$, $p < .0001$ (mean of 99 vs. 70 clusters), as shown in Figure 1C. However, once a cluster was retrieved, there was

Table 1

Experiment 1: Analysis of Variance With Span as a Between-Subjects Variable and Cell Means of Measures

Measures	Source		Span (N = 15)	
	Span (A) (df = 1)	MSE (df = 28)	Low	High
No. of names retrieved	23.08****	658.09	85	130
No. of clusters retrieved	24.49****	252.82	70	99
Time between clusters (s)	13.38***	4.26	11.70	8.95
Size of cluster	4.35*	.02	1.19	1.31
Time within a cluster (s)	.59	.03	.90	.94

$*p < .05.$ $***p < .001.$ $****p < .0001.$

no difference in the times between words within a cluster for the high- and low-span participants, $F(1, 28) = .59, p < 1.0$ (.94 vs. .90 ms). The cluster sizes were slightly larger for the high- than low-span participants, $F(1, 28) = 4.35, p = .05$ (mean of 1.31 vs. 1.19 names). Although the algorithm-based cluster sizes were small, they were compatible with the algorithm-based cluster sizes found by Gruenewald and Lockhead (1980) and the semantic-based clusters found by Troyer, Moscovitch, and Winocur (1997).

Repetitions

Repetitions amounted to less than 5% of the data, but more repetitions were made by the low- than high-span participants (mean of 5.14 vs. 3.93 repetitions). There were too few repetitions to submit to a statistical analysis.

Discussion

The question addressed in the first experiment was whether individuals who score high on a measure of working memory capacity would generate more animal names than individuals who score low on the same measure. In the first experiment, high-span participants retrieved more animal names and clusters, were faster to retrieve the clusters, and had larger sized clusters when compared to the performance of the low-span participants on the same measures. These findings suggested that a relationship exists between central executive and frontal lobe functioning as measured by working memory span and fluency, respectively. However, the findings of the first experiment cannot answer any questions concerning the role of working memory capacity in retrieval. The second experiment was conducted in order to examine the role of working memory capacity in retrieval.

Experiment 2

The question addressed in the second experiment was how an attention-demanding concurrent task would affect the retrieval of high- and low-span participants. A concurrent task should have a detrimental effect on generation for those participants who use working memory capacity to generate animal names. One half of the participants in each span group generated animal names without a concurrent task. The other half of the participants in each span group generated animal names while concurrently tracking digits on a computer screen.

Method

Participants

Seventy undergraduate students from the University of South Carolina participated in exchange for course credit in their psychology classes. Participants were assigned to low- and high-span groups based on their operation-span score. There were 17 participants in each of the two low-span groups and in the high-span group that did not track digits. There were 19 participants in the high-span group that tracked digits. The mean operation-span score was 7 for the low-span and 24 for the high-span groups.

Procedure

In Experiment 2, participants generated animal names for 10 min. Half of the participants in each span group generated animal names for 10 min without a concurrent task. The other participants in each group generated animal names while tracking digits. The digit-tracking task was based on a task used by Jacoby, Woloshyn, and Kelley (1989) and was programmed with the Micro Experimental Laboratory software (Schneider, 1988). The digits 1–9 appeared one at a time in the four corners of a computer screen. The presentation rate for each digit was one per second, and the digit sequence occurred clockwise around the perimeter of the screen. Participants were told to press the space bar on the computer keyboard whenever they saw a third odd digit in a row appear in the sequence. The program registered responses and provided feedback depending on the response. Participants heard a high-pitched tone for correct responses and a low-pitched tone for incorrect responses and missed critical digits. There were 20 possible correct responses within the 10 min. All participants received two 2-min practice sessions of digit tracking. The first 2-min practice session involved tracking digits without retrieving exemplars from a category. The second 2-min practice session involved tracking digits while retrieving from the category "automobiles."

The experimenter was present during category generation for those participants who concurrently tracked digits, in order to monitor performance on the digit-tracking task. The experimenter sat a minimum distance of 3 ft behind the participants in order not to distract them. The only time that a participant was prompted was if they missed more than 3 of the 20 critical digits in a row. If the experimenter noticed that the participant missed 3 critical digits in a row, then the participant was reminded that they had to keep tracking the digits while they were saying animal names. Participants who missed more than 10 of the 20 critical digits during the 10-min session were eliminated from the analyses. Two high-span participants were eliminated, leaving 17 participants in that condition, as well as 17 participants in each of the remaining conditions, for a total of 68 participants in this study.
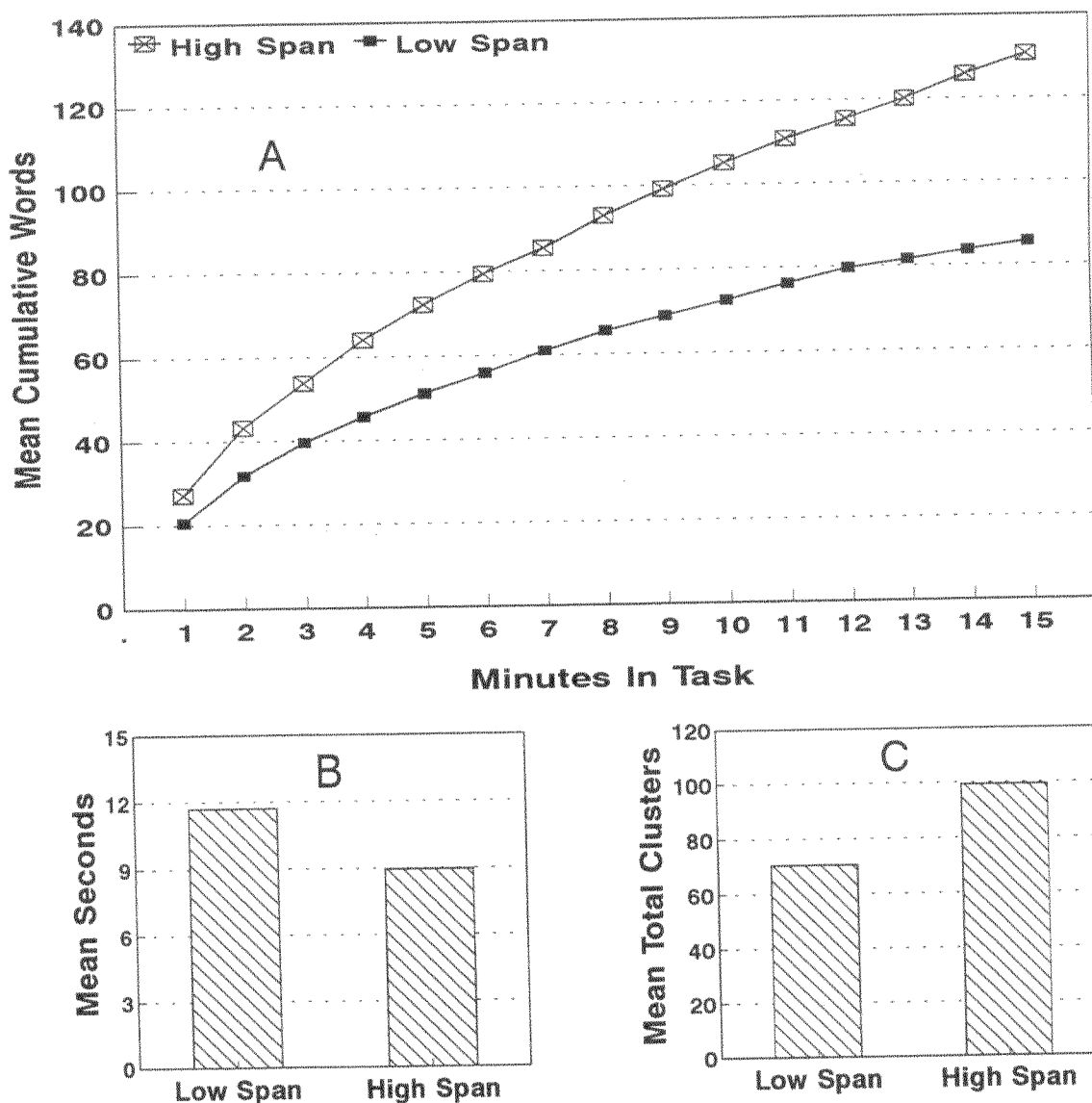
Figure 1. A: Mean number of cumulative words retrieved. B: Mean between-cluster retrieval times. C: Mean number of clusters retrieved.

Results

Cluster-Boundary Judging

We first examined the consistency of the semantic-based judgments of cluster boundaries among our three judges. The interrater reliability analysis produced a coefficient of .56 as an index of the level of cluster-boundary correspondence between the three judges. We then compared the semantic-based boundaries shared by all three judges with the algorithm-based cluster boundaries. A Pearson product-moment correlation showed a significant relationship between the two cluster boundary judging methods, $r(690) = .61$, $p < .0001$. We also looked at the consistency of this relationship across each of the 10 min. We found that this relationship was maintained for all but the 1st min of retrieval.

Digit Tracking

We examined digit-tracking performance for the low- and high-span groups in the load conditions. For the participants in the load condition, we compared the percentage of correctly detected digits in the 2-min practice session without the concurrent category generation with the percentage of correctly detected digits in the 10 min of digit tracking with the concurrent generation. A 2 (span) × 2 (type of task) mixed-design ANOVA, with span as a between-subjects variable and type of task as a within-subjects variable, showed that low- and high-span participants did not differ in the mean percentage of digits that were correctly detected, $F(1, 32) = .54$, $p < 1$ (mean of 80% vs. 83% correctly detected). However, a main effect of type of task showed that significantly more digits were detected when not con-

currently generating animal names, $F(1, 32) = 71.33$, $p <$ .0001 (mean of 95% vs. 68% correctly detected). There was no Span × Type of Task interaction, indicating that the two span groups showed comparable performance on the digit-tracking task, with and without concurrent generation of animal names. Finally, a Pearson product-moment correlation showed that there was no relationship between the percentage of correctly detected critical digits and the number of animal names retrieved, $r(34) = .18$, $p < 1$. This last finding indicated that there was no attentional trade-off between the two tasks.[1]

Retrieval of Names

Span × Load interaction. Table 2 contains the F values and mean square errors (*MSEs*) for the 2 (span) × 2 (load) ANOVA, with span and load as between-subjects variables. Figure 2A shows the Span × Load interaction that occurred for the number of animal names retrieved, $F(1, 64) = 5.55$, $p < .05$. The concurrent digit-tracking task only reduced retrieval performance for the high-span participants. This Span × Load interaction was borne out by a post hoc comparison of simple main effects. High-span participants retrieved significantly fewer animal names under load, $F(1, 32) = 13.60$, $p < .001$ (mean of 87 vs. 121 names, load and no load, respectively). In contrast, the low-span participants showed no significant effect of load on the number of animal names retrieved, $F(1, 32) < 1$ (mean of 61 vs. 67 names, load and no load, respectively). In addition to the effect of load on the number of names retrieved, post hoc comparisons of simple main effects showed an effect of load on the remaining measures for the high-span participants. Figure 2B shows that high-span participants were slower to access clusters when concurrently tracking digits, $F(1, 32) = 10.27$, $p < .01$ (mean of 7.97 s vs. 6.43 s between clusters) than they were when not tracking digits. Slower access times led to the retrieval of fewer clusters, $F(1, 32) = 6.70$, $p < .01$ (mean of 72 vs. 88 clusters), as shown in Figure 2C. Finally, high-span participants were faster to access names within a cluster, $F(1, 32) = 13.97$, $p < .001$ (mean of .85 s vs. 1.09 s between words in a cluster) when concurrently tracking digits. In contrast, the low-span participants showed no effect of load on any of the measures.

First-minute analyses. The Span × Load interaction for the number of animal names retrieved was evident even in the 1st min of retrieval, $F(1, 64) = 5.19$, $p < .05$ (mean of 24 vs. 18 names, high-span no load and load; mean of 17 vs. 16 names, low-span no load and load). The low-span participants showed no effect of load in the 1st min, whereas the high-span participants showed an effect of load in the 1st min.[2] A Span × Load interaction also occurred for cluster size. The high-span participants retrieved significantly smaller sized clusters when concurrently tracking digits, $F(1, 64) = 3.86$, $p < .05$ (mean of 1.19 vs. 1.39 names, load and no load, respectively). In contrast, the low-span participants showed no effect of load on cluster size (mean of 1.17 vs. 1.20 names, load and no load, respectively).

Main effects of span. The span effects in the present experiment mimicked the span effects in the first experiment. High-span participants retrieved significantly more animal names, $F(1, 64) = 45.72$, $p < .0001$ (mean of 104 vs. 64 names, high and low span, respectively) and significantly more clusters, $F(1, 64) = 42.48$, $p < .0001$ (mean of 80 vs. 53 clusters, high and low span, respectively) and were faster to access the clusters, $F(1, 64) = 38.16$, $p < .0001$ (mean of 7.2 s vs. 10.92 s between clusters, high and low span, respectively). High-span participants also showed larger sized clusters, $F(1, 64) = 5.64$, $p < .05$ (mean of 1.29 vs. 1.18 names, high and low span, respectively) and slower within-cluster retrieval times, $F(1, 64) = 23.62$, $p < .0001$ (mean of .97 s vs. .76 s between words in a cluster, high and low span, respectively).

Main effects of load. There was a main effect of load for all of the fluency measures. Participants retrieved significantly more animal names, $F(1, 64) = 12.04$, $p < .001$ (mean of 94 vs. 74 names) and significantly more clusters, $F(1, 64) = 6.75$, $p < .05$ (mean of 72 vs. 61 clusters) and were faster to access the clusters, $F(1, 64) = 8.16$, $p < .01$ (mean of 8.2 s vs. 9.92 s between clusters) when not concurrently tracking digits. Participants also showed larger sized clusters, $F(1, 64) = 6.14$, $p < .05$ (mean of 1.29 vs. 1.18 names) and slower within-cluster retrieval times, $F(1, 64) = 15.63$, $p < .001$ (mean of .95 s vs. .78 s between words in a cluster) when not concurrently tracking digits. These effects primarily reflected the performance of the high-span participants.

Repetitions

Repetitions amounted to less than 2% of the data. Generating animal names while tracking digits resulted in twice as many repetitions as generating animal names when not tracking digits (mean of 4.09 vs. 2.09 repetitions). Overall, more repetitions were made by the low-span than the high-span participants (mean of 3.88 vs. 2.29 repetitions). However, both span groups showed an increase in the number of repetitions as a result of concurrently tracking digits (mean of 2.35 vs. 5.29 repetitions, low span; mean of 1.6 vs. 2.94 repetitions, high span). As with the first experiment, there were too few repetitions to submit to a statistical analysis.

Discussion

The question addressed in the second experiment was how an attention-demanding concurrent task would affect the retrieval of high- and low-span participants. Only the high-span participants showed a reduction in the number of animal names retrieved due to the concurrent digit-tracking task. In fact, post hoc comparisons showed that the concurrent task reduced performance on all five measures for the

---

[1] We would like to thank Dan Fisk for suggesting this analysis.
[2] We would like to thank Morris Moscovitch for suggesting this analysis of load effects over the 1st min.

Table 2
Experiment 2: Analysis of Variance With Span and Load as Between-Subjects Variables and Cell Means of Measures

| | | | | | Span (N = 17) | | | |
| | Source | | | | Low | | High | |
Measures	Span (A) (df = 1)	Load (B) (df = 1)	A × B (df = 1)	MSE (df = 64)	No load	Load	No load	Load
No. of names retrieved	45.72****	12.04***	5.55*	594.91	67	61	121	87
No. of clusters retrieved	42.48****	6.75*	2.19	285.41	56	51	88	72
Time between clusters (s)	38.16****	8.16**	.09	6.18	9.97	11.87	6.43	7.97
Size of cluster	5.64*	6.14*	3.86*	.03	1.20	1.17	1.39	1.19
Time within a cluster (s)	23.62****	15.63***	3.29	.03	.81	.72	1.09	.85

*p < .05.   **p < .01.   ***p < .001.   ****p < .0001.

high-span participants. In contrast, the low-span participants showed no effect of the concurrent load on any of the measures.

One possible explanation for the lack of load effects for the low-span participants may be that the load effects were washed out across the 10 min. Both span groups may have begun the task actively searching for cues and names, but retrieval for the low-span participants may have become substantially slowed after the 1st min as they began to exhaust their knowledge base. If this were true, we should have found an equivalent load effect for the low- and high-span participants in the 1st min of the task. However, as in the first experiment, span differences in the number of names retrieved was evident in the 1st min of the task. More importantly, the concurrent task reduced the number of names retrieved in the 1st min for the high, but not the low-span participants. This finding suggested that the low-span participants used the same retrieval process throughout the entire 10 min, whether they retrieved names with or without a concurrent load. Further, this retrieval process was most likely automatic in nature. In contrast, the effect of load in the 1st min for the high-span participants suggested that retrieving animal names was not automatic for them.

Although a concurrent load did not affect the number of names retrieved for the low-span participants, the category-generation task was not completely automatic for them. Both span groups showed a comparable reduction in accuracy on the digit-tracking task, suggesting that some component of category generation was attention demanding for both the low- and high-span participants. Because of our instructions to avoid repetitions, participants were required to monitor for repetitions. Both span groups made few repetitions, but a greater number of repetitions were made by the low-span participants. Also, both span groups showed an increase in repetitions under load, which suggests that both span groups used their working memory capacity for monitoring. The more probable explanation for the lack of load effects on retrieval for the low-span participants is that they did not have sufficient working memory capacity for both monitoring for repetitions and performing the other controlled processes required to perform this task.

The negative impact of load on retrieval for the high-span participants suggested that they used their working memory capacity for retrieving animal names. In contrast, the lack of load effects for the low-span participants suggested that they did not use their working memory capacity for generating animal names. The most likely reason why the low-span participants did not use their working memory capacity for generating names was because monitoring may have been so demanding for the low-span participants that little capacity was left over to generate animal names. In contrast, the high-span participants had sufficient working memory capacity to both monitor and generate animal names.

If the search for animal names involved sampling with replacement, then there may have been a greater probability that individuals resample previously retrieved animal names with each retrieval attempt (Rundus, 1973). The probability of resampling may have been greater for the low-span participants, as reflected in their greater number of repetitions. Consequently, the greater resampling may have somehow interfered with generating cues to access new names. On the basis of the number of repetitions made when not concurrently tracking digits, it appears that both low- and high-span participants were relatively successful at identifying and restraining the vocalization of repetitions. Therefore, the longer between-cluster retrieval times for the low-span participants may have reflected greater resampling and the resultant vocal restraint of repetitions.

If we think of resampling as perseveration of a response, then we can take two different perspectives on resampling. The first perspective, disengaging attention, views resampling as the inability to disengage attention from previously retrieved responses. The second perspective, suppression, views resampling as the inability to suppress a previously retrieved response. These two views of resampling really say opposite things about the role of attention in perseveration. The first view, disengaging attention, implies that an individual is unable to direct their attention away from the previously retrieved response. The second view, suppression, implies that an individual is unable to direct their attention to the previously retrieved response, in order to suppress it. The two views make different predictions about resampling for the low- and high-span participants.
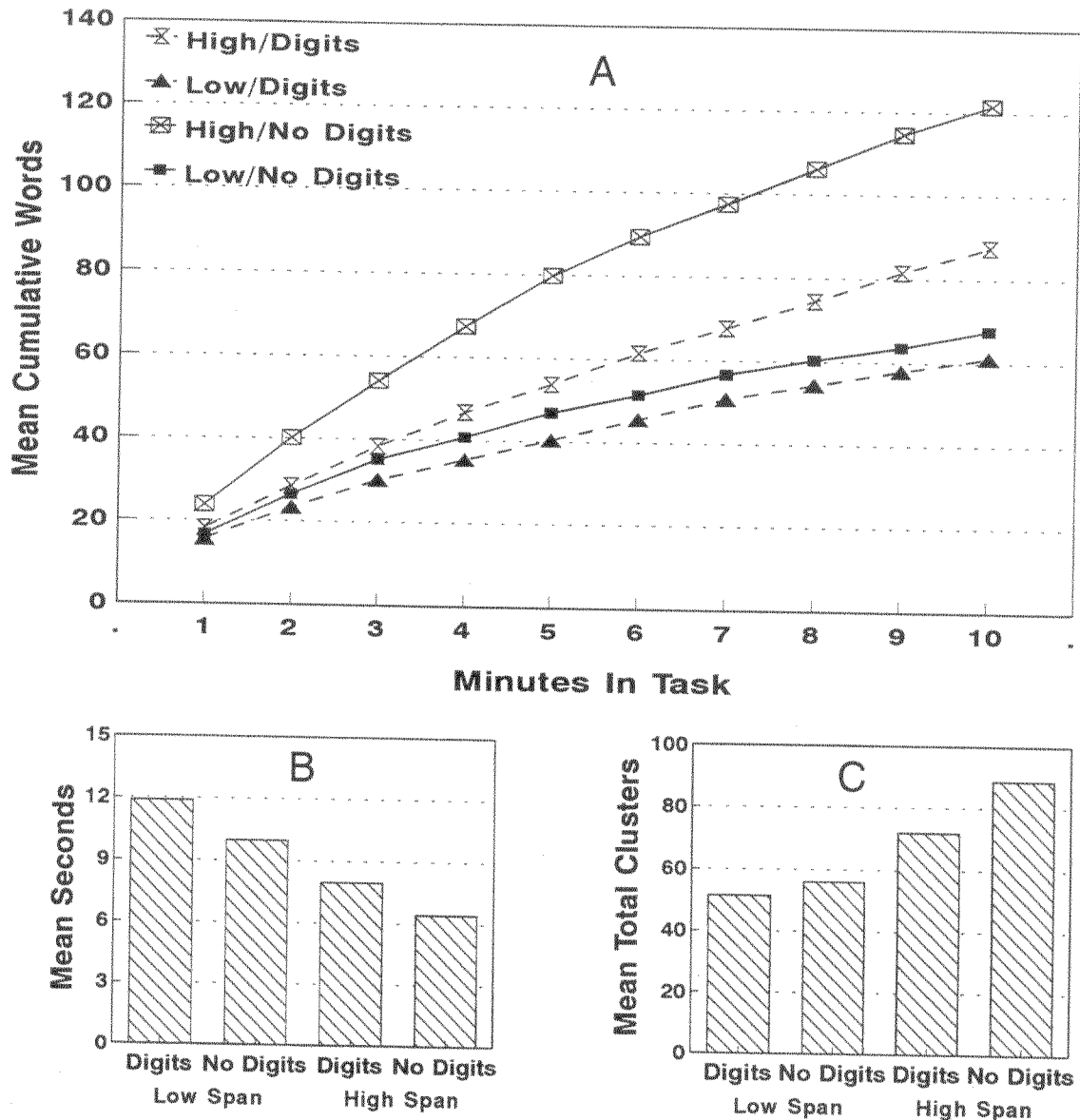
Figure 2. A: Mean number of cumulative words retrieved. B: Mean between-cluster retrieval times. C: Mean number of clusters retrieved.

The disengaging attention view predicts that if low-span participants experience greater resampling because they are unable to disengage attention from a previously retrieved response, then dividing their attention between generation and a concurrent task results in a lower probability of resampling as their attention is directed away from the previously retrieved response. Consequently, this view predicts that an increase in the number of new names retrieved by the low-span participants is found when their attention is divided during generation. However, our findings show that the low-span participants did not exhibit an increase in the number of new names retrieved when concurrently tracking digits. In fact, they did not show any effect of the concurrent task on any of the fluency measures but did show an increase in the number of repetitions made under load. In

contrast to the low-span participants, the high-span partici-pants were expected to be less likely to perseverate on previously retrieved responses. Therefore, the high-span participants were expected to show little change in resam-pling when under load. In fact, the high-span participants showed an effect of load on all of the fluency measures, as well as an increase in repetitions made when concurrently tracking digits. Therefore, the results do not support the disengaging attention view of resampling.

The suppression view predicts that if low span partici-pants experience greater resampling because they are unable to direct their attention to previously retrieved clusters in order to suppress them, then dividing their attention be-tween generation and a concurrent attentional task has little effect on resampling. In contrast, this view predicts that the

high-span participants experience a reduction in the ability to suppress previously retrieved responses when their attention is divided between generation and the concurrent task. The result is an increase in the probability that previously retrieved animal names are resampled by the high-span participants. In fact, as mentioned previously, only the high-span participants showed an increase in resampling under load. Therefore, the results support the suppression view of resampling.

The findings show that the low-span participants did not exhibit an increase in the number of new names retrieved when concurrently tracking digits. In fact, they did not show any effect of the concurrent task on any of the fluency measures but showed an increase in the number of repetitions made under load. In contrast, the high-span participants showed an effect of load on all of the fluency measures, as well as an increase in repetitions made when concurrently tracking digits. These results support our suppression perspective of resampling. Unless previously retrieved responses are suppressed, they continue to be resampled, thus blocking the generation of cues to access new animal names.

We use the term *block* in the same sense as Anderson and Bjork (1994) did when discussing occlusion mechanisms in their chapter on retrieval inhibition. In remaining above some conscious threshold, a highly activated response can act on a competing response in such a way as to prevent its retrieval, without dampening the activation level of the competing response. The highly activated response in this case was the previously retrieved response.

To review, the results of the second experiment indicate that low-span participants did not have sufficient working memory capacity to allocate to all of the components of the generation task, unlike the high-span participants. Two findings, in particular, suggested this interpretation. First, there were span differences in the effect of load on the number of names retrieved. Second, there was an increase in the number of repetitions made by both span groups under load. These findings together suggest that suppressing old names to generate new names, and monitoring for repetitions, are both attention-demanding components of retrieval. Because of our instructions to avoid repetitions, the low-span participants used their working memory capacity for detecting and restraining the vocalization of repetitions. As a result, the low-span participants most likely experienced greater resampling because of insufficient working memory capacity to suppress previously retrieved responses, in addition to monitoring. Therefore, it may be that the low-span participants retrieved animal names relatively automatically because they were unable to suppress previously retrieved responses in order to generate cues to access new names. In contrast, the high-span participants had sufficient working memory capacity to both monitor for repetitions and generate cues to access new animal names. Consequently, it appears that generating cues to access new names, in our task, was dependent on suppressing previously retrieved responses. We explore this further in the next experiment.

Experiment 3

The question addressed in the third experiment was how a memory preload would affect retrieval. We had all of the participants memorize and recall a 12-word list prior to generating animal names. The list contained either animal names (related) or building-part names (unrelated). If repeated resampling of previously retrieved animal names somehow blocks access to new animal names, then memorizing and recalling a list of animal names reduces the number of new animal names retrieved. In contrast, memorizing a list of building part names has no effect on retrieving new animal names because the two categories are unrelated.

Method

Participants

Seventy-six undergraduate students from the University of South Carolina participated in exchange for course credit in their psychology classes. Participants were assigned to low- and high-span groups based on their operation-span score. There were 19 participants in each of the two low-span groups, 18 participants in the high-span group that memorized animal names, and 20 participants in the high-span group that memorized building-part names. The mean operation-span score was 8 for the low-span and 24 for the high-span groups.

Procedure

The duration of category generation in Experiment 3 was 10 min. All of the participants generated animal names after first memorizing and recalling a list of 12 words. Half of the participants in each span group generated animal names after memorizing and recalling a list of 12 animal names. The other half of the participants in each span group generated animal names after memorizing and recalling a list of 12 building-part names. Both lists are contained in the Appendix.

The names for both lists were based on the Battig and Montague (1969) category norms and were presented on a computer using a program that was written with Micro Experimental Laboratory software (Schneider, 1988). The lists of words were randomly ordered for each participant, and presentation time for each word was controlled by the participant. The time to memorize a list was recorded for each participant. The participants in the two high-span groups, and the low-span group that memorized animals, took an average of 6 min to memorize a list. The participants in the low-span group that memorized building-part names took an average of 7 min to memorize the list. All participants were required to recall the 12 words perfectly three times in a row before generating animal names. After successfully recalling the list of 12 words, participants generated animal names for 10 min and were told not to say any of the words from the memorized list during the retrieval period. However, they were allowed to use the words on the list if they were part of a compound name. For example, if the word *lion* was on the list, then they were allowed to say the animal name *mountain lion* during retrieval.

Results

The list factor in these analyses refers to the type of list that was memorized. The list of animal names was consid-

ered the related condition, whereas the list of building-part names was considered the unrelated condition.

Retrieval of Names

Table 3 contains the F values and mean square errors ($MSEs$) for the 2 (span) $\times$ 2 (list) ANOVA, with span and list as between-subjects variables. Neither the ANOVA, nor the post hoc comparisons of simple main effects, showed an effect of relatedness for any of the four groups on any of the measures. The effects of span in the present experiment mimicked the span effects that were found in the previous experiments. Figure 3A shows that significantly more animal names were retrieved by the high, than the low-span participants, $F(1, 72) = 22.17, p < .0001$ (mean of 94 vs. 65 names).

Experiment 2 and 3 comparisons. Included in Figure 3A are the high- and low-span no-load conditions from the second experiment to serve as control groups for the four conditions in the third experiment. The six conditions in Figure 3A were subjected to an ANOVA, and post hoc comparisons of simple main effects were conducted. The comparisons showed that the mean number of animal names retrieved by the two low-span conditions in the third experiment did not differ from each other (mean of 63 vs. 66 animal names, related and unrelated list conditions) or from the control (mean of 67 animal names). The two high-span conditions in the third experiment did not differ from each other (mean of 89 vs. 99 animal names, related and unrelated list conditions), but both high-span groups retrieved significantly fewer animal names when compared to their control (mean of 121 animal names). For the related list condition compared to the control, $F(1, 52) = 9.83, p < .01$ (mean of 89 vs. 121 animal names); for the unrelated list condition compared to the control, $F(1, 52) = 5.0, p < .05$ (mean of 99 vs. 121 animal names).

Main effects of span. Figure 3B shows that the high-span participants were faster to retrieve clusters, $F(1, 72) = 13.24, p < .001$ (mean of 7.98 s vs. 10.69 s between clusters, high and low span, respectively), which led to their retrieval of more clusters, $F(1, 72) = 12.21, p < .001$ (mean of 73 vs. 58 clusters, high and low span, respectively), as

seen in Figure 3C. Also, cluster sizes were larger, $F(1, 72) = 12.29, p < .001$ (mean of 1.27 vs. 1.11 names, high and low span, respectively), whereas within-cluster retrieval times were slower for the high-span participants, $F(1, 72) = 8.08, p < .01$ (mean of .88 s vs. .77 s between words in a cluster, high and low span, respectively).

Repetitions

Repetitions amounted to less than 3% of the data. The low-span participants made more repetitions compared to the high-span participants. However, the number of repetitions made by both span groups did not appear to be a function of whether participants memorized building-part or animal names (mean of 1.84 vs. 1.76 repetitions for the low spans, and a mean of 1.60 vs. .94 repetitions for the high spans). As in the previous two experiments, there were too few repetitions to submit to statistical analysis. Span differences in the number of repetitions made are examined in the fourth experiment.

Intrusions

We examined each participant's output for intrusions from the list of words that was memorized. Intrusions amounted to less than 1% of the data. Low- and high-span participants who memorized animal names did not differ in the total number of intrusions made (total of 10 intrusions each). Also, both span groups did not differ in the total number of intrusions made when comparing the first 5 min of the task to the second 5 min of the task (total of 8 vs. 2 intrusions, first 5 min and second 5 min, respectively). However, there were too few intrusions to submit to a statistical analysis.

Discussion

The question addressed in the third experiment was how would a memory preload affect retrieval. We argued that if repeated resampling of previously retrieved responses

Table 3

Experiment 3: Analysis of Variance With Span and List as Between-Subjects Variables and Cell Means of Measures

					Span			
	Source				Low ($N = 19$)		High	
Measures	Span (A) ($df = 1$)	List (B) ($df = 1$)	A $\times$ B ($df = 1$)	MSE ($df = 72$)	Related	Unrelated	Related ($N = 18$)	Unrelated ($N = 20$)
No. of names retrieved	22.17****	1.14	.22	757.13	63	66	89	99
No. of clusters retrieved	12.21***	.07	.02	372.33	58	58	72	74
Time between clusters (s)	13.24***	.05	.30	10.57	10.98	10.41	7.86	8.10
Size of cluster	12.29***	3.66	.29	.04	1.08	1.14	1.21	1.32
Time within a cluster (s)	8.08**	.03	.47	.02	.78	.76	.86	.90

** $p < .01$. *** $p < .001$. **** $p < .0001$.

Figure 3 — Graph A (Mean Cumulative Words vs Minutes In Task):

Legend:
- High/Related
- High/Unrelated
- High/Exp 2 No Digits
- Low/Related
- Low/Unrelated
- Low/Exp 2 No Digits

Y-axis: Mean Cumulative Words (0–140)
X-axis: Minutes In Task (1–10)

Graph B: Y-axis: Mean Seconds (0–15); categories: Related, Unrelated (Low Span), Related, Unrelated (High Span)

Graph C: Y-axis: Mean Total Clusters (0–100); categories: Related, Unrelated (Low Span), Related, Unrelated (High Span)
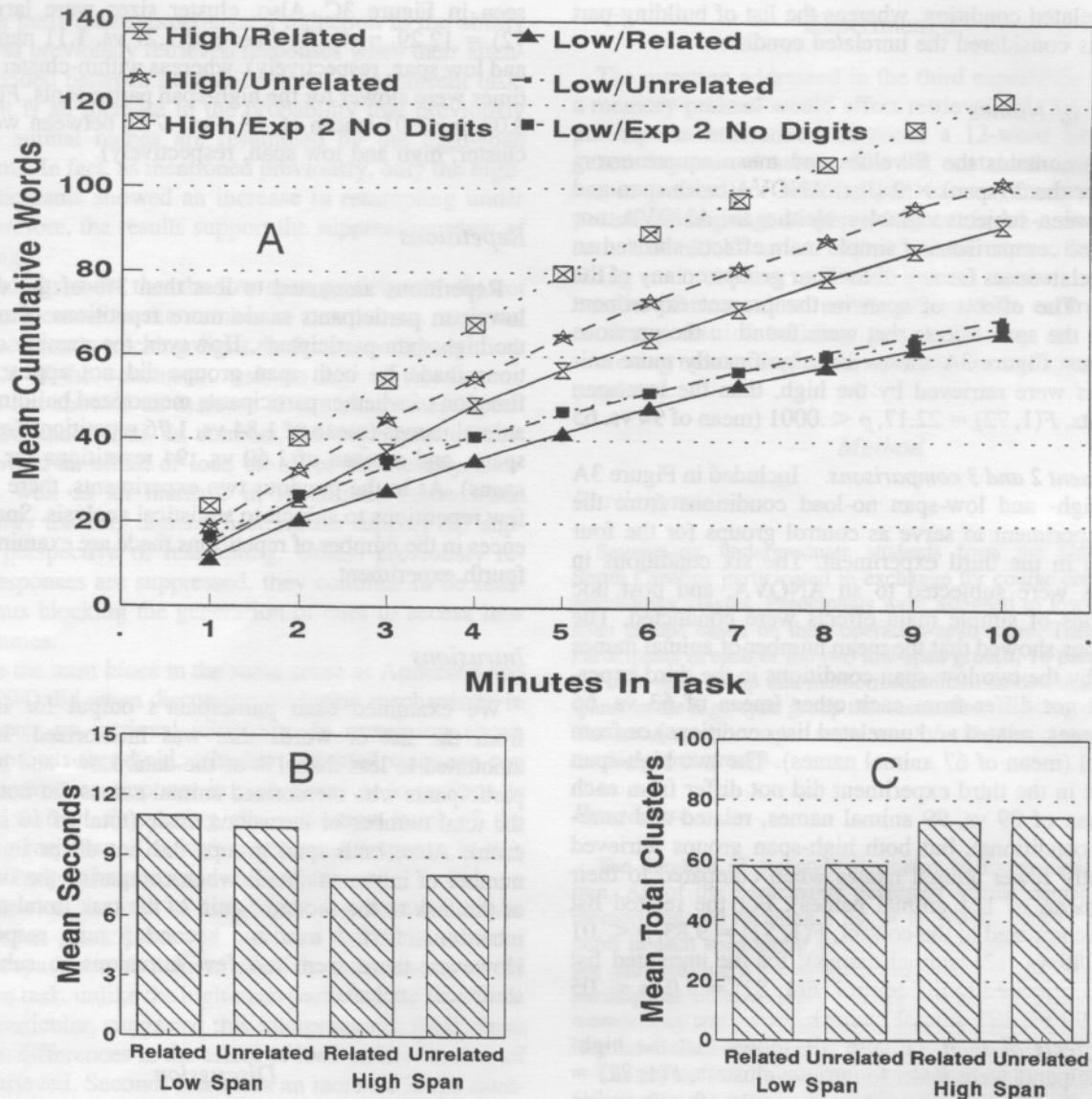
Figure 3. A: Mean number of cumulative words retrieved. Exp = Experiment. B: Mean between-cluster retrieval times. C: Mean number of clusters retrieved.

somehow blocks access to new names, then memorizing and recalling a list of animal names should reduce the number of animal names generated. In contrast, memorizing a list of building-part names should have no effect on generating animal names because the two categories are unrelated. This argument was based on the assumption that any retrieval block that occurred due to resampling was semantic based.

In the third experiment, we found no effect of list type on any of the measures, for any of the four groups. As in the previous two experiments, high-span participants retrieved more animal names and clusters, and were faster to access the clusters, when compared to the low-span participants on the same measures. However, unlike the low-span participants, the high-span participants showed a reduction in the number of animal names retrieved when compared to their control group. We found that memorizing both animal and

building-part names reduced the number of animal names retrieved by the high-span participants. In contrast, neither type of list had an effect on the number of animal names retrieved by the low-span participants.

The finding that memorizing either animal or building-part names reduced generation for high-span participants suggests that the retrieval block was not semantic based. It may be that highly activated responses were produced from memorizing and retrieving the names on either list, irrespective of semantic relatedness. The combination of memorization and the repeated retrieval of the list names prior to category generation produced highly activated responses that competed for retrieval and somehow blocked access to new animal names for the high-span participants (see Roediger & Neely, 1982, for a discussion of retrieval blocks). Once again, the lack of an effect of load on the number of

names retrieved by the low-span participants suggested that retrieving names was relatively automatic for them.

Although the findings of the first three experiments suggested that low-span participants experienced more resampling of previously retrieved responses, these findings are inconclusive without sufficient numbers of repetitions to analyze. We argued previously that the instructions given in the first three experiments prevented participants from making repetitions. Therefore, Experiment 4 was conducted in order to address span differences in resampling, when participants are not discouraged from making repetitions.

Experiment 4

The question addressed in the fourth experiment was whether low- rather than high-span participants are more likely to resample previously retrieved responses. In the fourth experiment, we compared the likelihood of low- and high-span participants making a repetition when concurrently generating animal names and tracking digits. We included a secondary task in order to increase the number of repetitions. We altered the instructions from the previous three experiments such that participants were encouraged to say any previously retrieved animal names that "popped" into their head. However, participants were still encouraged to keep retrieving new animal names throughout the 10 min.

Method

Participants

Thirty-four undergraduate students from the University of South Carolina participated in exchange for course credit in their psychology classes. Participants were assigned to low- and high-span groups based on their operation-span score. There were 17 participants in each group. The mean operation-span score was 7 for the low-span and 24 for the high-span groups.

Procedure

In Experiment 4, participants generated animal names for 10 min while concurrently tracking digits on a computer screen. The digit-tracking task and procedure was the same as that used in the second experiment. Participants were instructed to keep retrieving new animal names throughout the 10 min. However, unlike the first three experiments, participants were instructed to say any previously retrieved animal names that popped into their head. They were told that they should say the intrusion in order to clear their mind of it, which would then enable them to continue on with retrieving new animal names.

Results

The dependent measure was the ratio of repetitions made to unique names. This ratio gave us the likelihood of making a repetition. The number of unique names was the difference between the total number of names retrieved and the number of repetitions. The low-span participants retrieved a mean total of 100 animal names with a mean total of 45 repetitions, resulting in a mean total of 55 unique names. The high-span participants retrieved a mean total of 110 animal names with a mean total of 28 repetitions, resulting in a mean total of 82 unique names. We make note of the fact that the comparable mean totals for unique names in the second experiment, when concurrently tracking digits, was 56 for the low-span participants compared to 55 in the present study and 84 for the high-span participants compared to 82 in the present study. We submit that the similarities of these two sets of numbers is not just an attractive coincidence.

The likelihood of resampling a previously retrieved response was .95 for the low-span and .35 for the high-span participants. An ANOVA with span as a between-subjects variable showed a main effect of span for the likelihood of resampling a previously retrieved name, $F(1, 32) = 5.04$, $p < .05$, $MSE = .6109$. Low-span participants were more likely to resample previously retrieved names. More important, for the low-span participants, the likelihood of retrieving a previously retrieved name was about the same as retrieving a new name.

Discussion

The present study was conducted in order to determine whether low-span participants are more likely to resample previously retrieved responses. The mean number of repetitions made in the present experiment was substantially higher than the mean number of repetitions made in the previous three experiments. The only methodological difference between the two load conditions in the second experiment and the present experiment was the type of instructions given. Therefore, we attribute the differences in the number of repetitions made in the two experiments to the change in instructions.

Previous instructions to avoid repetitions meant that individuals had to restrain the vocalization of previously retrieved responses. In the present experiment, the low-span participants showed a significantly larger ratio of repetitions to unique names when compared to the same ratio for the high-span participants. In other words, low- rather than high-span participants were more likely to resample previously retrieved animal names. We further discuss the implications of these findings in the General Discussion.

General Discussion

The primary question asked in this paper was whether working memory capacity was important to retrieval. The results of the first experiment showed us that a relationship does indeed exist between a measure of working memory capacity and a measure of strategic retrieval. The results of the remaining three experiments reinforced this conclusion. Experiment 2 showed that a concurrent attentional load reduced the number of animal names retrieved by the high-span participants only. The low-span participants showed no effect of a concurrent load on category generation. Experiment 3 showed that a memory preload reduced the

number of names retrieved by the high-span participants, irrespective of semantic relatedness of the preload to the retrieval task. In contrast, the low-span participants showed no effect of the memory preload. Experiment 4 showed that with appropriate instructions and under load, low-span participants were far more likely to resample previously retrieved animal names than were high-span participants.

One possible and uninteresting explanation of our results is that low working memory individuals simply have a smaller knowledge base than do high working memory individuals. However, three findings suggest that span differences in retrieval were not due to knowledge-base differences. First, analysis of the initial minute of Experiment 2 showed that although load hurt all of the fluency measures for the high-span participants, it had no impact on the same measures for the low-span participants. Second, the fourth experiment showed that the low-span participants exhibited about the same likelihood of making a repetition as retrieving a new name. Third, if span differences in retrieval reflect knowledge-base differences, then richer semantic connections should have led to faster within-cluster retrieval times for the high-span participants. In fact, the low-span participants showed faster within-cluster retrieval times in the second and third experiments. There were no span differences in the within-cluster retrieval times in the first experiment. Greater resampling is the more logical explanation for the longer between-cluster retrieval times for the low-span participants. Span differences in resampling are further discussed in the next section.

Working Memory Capacity and Retrieval

The findings across all four experiments suggest that constraints on working memory capacity or controlled attention, caused by either individual differences in capacity or by the addition of a memory load, had a profound effect on retrieval. The reduction in digit-tracking accuracy due to concurrent generation, in the second experiment, indicated that both span groups most likely used working memory capacity for monitoring. However, high-span participants showed an effect of load on all of the fluency measures, whereas the low-span participants showed no effect of load on any of the measures. In fact, the span difference in the effect of load was evident even in the first minute. These findings suggested that only the high-span participants had sufficient working memory capacity to both monitor for repetitions and generate new names.

For the low-span participants, the lack of load effects on the number of names retrieved indicated that retrieving new animal names was relatively automatic. We do not believe that the lack of load effects reflect a floor effect. The extra 5 min in the first experiment allowed those low-span participants to retrieve an average of 86 animal names over 15 min. This number was 18 more names than the average of 68 animal names retrieved by the low-span participants in the second and third experiment over 10 min. In addition, the fourth experiment showed us that when they were not encouraged to avoid repetitions, the low-span participants

were almost as likely to retrieve a previously retrieved name as a new name. This suggests that in the first three experiments low-span participants were using the resources available to them to monitor their output in order to prevent repetition and error. Because their resources were directed to monitoring, they were unable to suppress previously retrieved responses or generate cues to access new names. As a result, the low-span participants may have experienced greater resampling of previously retrieved responses.

There were three findings that suggested span differences in resampling of previously retrieved animal names. The first finding was the span differences in between-cluster retrieval times. These times were consistently shorter for the high-span participants resulting in more clusters and animal names retrieved by them. In contrast, the low-span participants most likely experienced greater resampling of previously retrieved names, which resulted in longer between-cluster retrieval times and fewer animal names retrieved. The second finding was that the between-cluster retrieval times increased for the high-span participants when under load. Further, there was a trend toward more repetitions under load for the high-span participants. These findings suggested that the high spans may have experienced greater resampling because the concurrent task used some of the capacity that would have been used for suppressing previously retrieved responses in order to generate cues to access new names. The third finding was the span differences in the likelihood of making a repetition in the fourth experiment when concurrently tracking digits. Low-span participants were far more likely to resample previously retrieved animal names, when compared to the resampling of high-span participants.

A Component Model of Retrieval

The findings across all four experiments suggested that there were four retrieval components involved in our generation task: (a) activation automatically spreading from the cue, (b) self-monitoring of output to prevent repetition and error, (c) suppression of previously retrieved responses, and (d) generation of cues to access new names. In our model, the first component requires little in the way of attention, whereas the remaining three components are attention demanding. We argue that automatic spreading activation was important in retrieval for both span groups as was self-monitoring when instructions discouraged repetitions. However, only the high-span participants had sufficient working memory capacity to both suppress competing responses and generate cues to access new names, in addition to self-monitoring to prevent errors. In contrast, the low-span participants may have accessed names relatively automatically because they did not have sufficient working memory capacity to allocate to all three of the retrieval components that required controlled attention. Therefore, retrieval for the high-span participants, when not under load, involved an active search for animal names. In contrast, retrieval for the low-span participants was most likely associative and passive in nature.

Although speculative in nature, the findings across all four experiments fit well with a recent model of retrieval based on frontal lobe functioning. Morris Moscovitch (1992, 1994) presented a model of controlled and automatic retrieval that involved central systems and modules. His central system is the controlled processing component of the model and corresponds to Baddeley's (1986) central executive. This system is responsible for mediating strategic retrieval processes, including the monitoring of output. The automatic processing component of his model consists of the various cortical input modules (e.g., auditory and visual), in addition to a medial temporal lobe–hippocampal module (MTL–H module) that mediates both encoding and retrieval. Retrieval that is mediated by the frontal lobes is more strategic in nature, most often involves a search, and is susceptible to concurrent task interference. In contrast, retrieval mediated by the MTL–H module is cue dependent and episodic in nature and is not subject to task interference. More recently, Moscovitch has further suggested that the lateral temporal lobe (LTL) mediates retrieval from semantic memory (personal communication, November 1995).

On the basis of assumptions of Moscovitch's (1995) model, our findings suggested that retrieval may have been mediated by different components in the model for high- and low-span participants. For the high-span participants, generation may have been mediated by the frontal lobes because retrieval of new animal names was affected by concurrent task interference. In contrast, category generation for the low-span participants may have been mediated by the LTL module because retrieval of new animal names was not affected by concurrent task interference. In other words, retrieval was strategic for the high-span participants and relatively automatic for the low-span participants, except for monitoring. Reduced accuracy on the digit-tracking task, in the second experiment, showed us that monitoring was attention demanding for both span groups.

Strategic retrieval in Moscovitch's model includes monitoring and searching for cues to retrieve a memory trace. The decrease in the number of clusters and names retrieved by the high-span participants, when under load, suggested that they used strategic retrieval. In contrast, for the low-span participants, each retrieved animal name may have become a cue to access other animal names so that retrieval became more associative in nature for them. Also, the lack of load effects for the low-span participants in the first minute of the second experiment suggested that they retrieved animal names relatively automatically from the very beginning of the task.

Because the high-span participants had sufficient working memory capacity to suppress previously retrieved responses, they were able to generate cues to access new names. As a result, a greater number of clusters and animal names were retrieved by the high- than by the low-span participants. In contrast, our instructions required that the low-span participants use the working memory capacity they had available for monitoring instead of generating cues. Consequently, in the present set of experiments, the successful performance of the high-span participants in the no-load conditions resulted from their ability to use con-

trolled attention to suppress previously retrieved animal names and to strategically generate their own cues for further recall.

These studies, along with those of Baddeley et al. (1984) and Craik et al. (1996) suggest that not all retrieval may be thought of as equivalent in the demand for attention. Both Baddeley et al. and Craik et al. used briefly presented events such as a single trial of free recall or arbitrary paired associates. It may be that the level of learning involved in procedures such as theirs do not lead to the kind of controlled processes we have argued for here. It is, however, an all too common occurrence that we meet someone on the street or in a store and know that we know them. The act of desperately trying to recall the name or even what context of our life they are associated with is most certainly a controlled, effortful, and strategic search that involves generating cues for possible contexts, monitoring and rejecting various names and contexts, and suppressing, if possible, those names that keep coming erroneously to consciousness.

We argue that increases in working memory capacity resulting from frontal lobe development over childhood leads to increased ability to do controlled strategic search. If there are corresponding declines in working memory capacity with aging (Salthouse, 1991), then there are corresponding declines in the ability to do controlled strategic search. Retrieval that is mediated by the hippocampal system presumably does not show such developmental changes.

We make one other speculative inference about the relationship between our work and work on the frontal lobes. Verbal fluency has been long used as a diagnostic for frontal lobe damage. However, Reitan and Wolfson (1994) have pointed out the inconsistency and low validity of so-called frontal tasks. One extension of the arguments we have presented here is that there are large premorbid differences in performance on the verbal fluency task. Further, the effect of frontal damage may roughly correspond to the effects of a concurrent load. In that case, a low working memory individual would not show large pre- and postmorbid differences on verbal fluency, and possibly, other frontal tasks as well. In other words, this task would only be a good diagnostic of frontal damage for individuals with sufficient working memory capacity to do the task in a controlled, strategic manner prior to brain damage. These ideas are, as stated, speculative.

References

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265–325). San Diego, CA: Academic Press.

Baddeley, A. (1986). *Working memory.* London: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower, (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113,* 518–540.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition, 7*, 212–230.

Baldwin, J. M. (1894). *Mental development in the child and the race.* New York: Macmillan.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Pt. 2).

Benton, A. L. (1968). Differential effects of frontal lobe disease. *Neuropsychologia, 6*, 53–60.

Benton, S. L., Kraft, R. G., Glover, J. A., & Plake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology, 76*, 820–834.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*, 229–246.

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., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561–583.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1–18.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84*, 253–262.

Engle, R. W., & Oransky, N. (in press). The evolution from short-term to working memory: Multi-store to dynamic models of temporary storage. In R. J. Sternberg (Ed.), *The concept of cognition.* Boston: Cambridge University Press.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition, 53*, 129–153.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 225–240.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115–125.

Kiewra, K. A., & Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemporary Educational Psychology, 13*, 33–44.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*, 389–433.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2*, 129–160.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56–98.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 313–334). New York: McGraw-Hill.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4*, 257–267.

Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology, 8*, 524–534.

Moscovitch, M. (1995). Models of consciousness and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1341–1356). Cambridge, MA: MIT Press.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwarts, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.

Ormrod, J. E., & Cochran, K. F. (1988). Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Reading Research and Instruction, 28*, 33–43.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141–183). New York: Wiley.

Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Reitan, R. M., & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4*, 161–198.

Roediger, H. L., & Neely, J. H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology, 36*, 213–242.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12*, 43–50.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Erlbaum.

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 507–516.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM-PC compatibles. *Behavior Research Methods, Instrumentation, and Computers, 20*, 206–217.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7*, 1–24.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*, 138–146.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

West, R. L. (1996). An application of prefrontal cortex theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomics Bulletin & Review, 1,* 89–106.

Appendix

Experiment 3: Lists That Were Memorized Prior to Category Generation

Animals	Building parts
bee	attic
bird	ceiling
cat	closet
cow	door
dog	floor
elephant	hall
fish	roof
fly	room
horse	stair
lion	steps
snake	wall
turtle	window