Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 4, 703–713

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.4.703

When Do Visual and Verbal Memories Conflict? The Importance of Working-Memory Load and Retrieval

Candice C. Morey and Nelson Cowan
University of Missouri—Columbia

Examinations of interference between verbal and visual materials in working memory have produced mixed results. If there is a central form of storage (e.g., the focus of attention; N. Cowan, 2001), then cross-domain interference should be obtained. The authors examined this question with a visual-array comparison task (S. J. Luck & E. K. Vogel, 1997) combined with various verbal memory load conditions. Interference between tasks occurred if there was explicit retrieval of the verbal load during maintenance of a visual array. The effect was localized in the maintenance period of the visual task and was not the result of articulation per se. Interference also occurred when especially large silent verbal and visual loads were held concurrently. These results suggest central storage along with code-specific passive storage.

Keywords: working memory, attention, dual task, cross-domain, storage

A simple question that has yielded a complex answer is whether memories in different domains can be actively represented in working memory at the same time. Baddeley and Hitch (1974) and Baddeley (1986) reported only mild interference between various types of cognitive tasks and a verbal memory load. The working-memory theory of Baddeley (1986; see also Baddeley & Logie, 1999) includes separate, passively held storage faculties for verbal and visuospatial forms of information, but some conflict is said to be possible if both forms of information are demanding enough to require the involvement of central executive processes at the same time for rehearsal and/or processing. Given that this central executive involvement is not always necessary, it is perhaps not surprising that examinations of the extent of conflict between two tasks in different domains have yielded mixed results (e.g., Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Duncan, Martens, & Ward, 1997; Jolicoeur, 1999; Luck & Vogel, 1997; Morey & Cowan, 2004; Sanders & Schroots, 1969; Stevanovski & Jolicoeur, 2003).

Another theoretical approach to working memory seems to imply somewhat different reasons why visual and verbal tasks would or would not conflict. Whereas the model of Baddeley (1986) includes central executive mechanism thought to manipulate information that is held completely within passive types of storage, Cowan (1995, 2001) suggested that some information can be held also in the focus of attention in addition to passively held forms of storage. This information in the focus of attention was said to be subject to a capacity limit, as opposed to the temporal limits and interference factors that are prominent for the passively held stores. A recent amendment of the working-memory model of Baddeley (1986) includes an episodic buffer (Baddeley, 2000), which could similarly be limited in capacity (Baddeley, 2001). With this type of approach, as well, it is still an open question as to when visual and verbal maintenance will conflict.

We investigated the question of amodal storage in working memory with a dual task involving retention of visual arrays and verbal sequences concurrently, a type of procedure that has been used in at least two other recent studies (Cocchini et al., 2002; Morey & Cowan, 2004). Whereas Cocchini et al. (2002) found only mild interference between modalities, Morey and Cowan (2004) found more severe interference. We believe that these seemingly divergent findings can be reconciled with an empirical study that focuses on one difference between their procedures: That the regimen of rehearsal was left up to the participant in the study of Cocchini et al., whereas Morey and Cowan required that the verbal task be rehearsed aloud. The hypothesis is that silent maintenance can be accomplished with little continual demand on attention, whereas overt rehearsal requires that each stimulus be retrieved into the focus of attention.

To observe conflict between visual and verbal working-memory tasks, we find it necessary to present task conditions that make it impossible to rely exclusively on the passively held stores, or on rapidly formed long-term memory representations, to hold the information. If participants were able to hold all of the information presented in at least one of the modalities in a dedicated, passive store, then there would be no competition between modalities for the general capacity-limited storage faculty. A key question for our approach is how one can ensure that capacity-limited storage must be used for both tasks.

For visual, nonverbal stimuli, use of attention may be ensured relatively easily. A task as simple as comparing two sequentially presented arrays of colored squares, which are identical or differ in the color of only one square, can be accomplished reliably only if there are four or fewer squares in the array (Luck & Vogel, 1997). This severe limit is not what one would expect if participants could

---

This research was supported by National Institutes of Health Grant R01 HD-21338, awarded to Nelson Cowan. Some of this research was performed in partial fulfillment of the requirements of the degree of Master of Arts by Candice C. Morey. We thank J. Scott Saults for making digit stimuli and Mike Carr for helping collect data.

Correspondence concerning this article should be addressed to Candice C. Morey or Nelson Cowan, Department of Psychological Sciences, University of Missouri—Columbia, 210 McAlester Hall, Columbia, MO 65211. E-mail: ccmorey@mizzou.edu or cowann@missouri.edu

draw freely from a richer source of memory at the time of recall, such as a passively held visual-spatial representation of the first array. The observed capacity limit resembles what is found in many other situations (see Cowan, 2001) in which retention must take place without the likelihood of systematic rehearsal or grouping, such as recall from verbal lists with rehearsal blocked, which tends to be limited to three to five items, and recall from a category in long-term memory, which tends to occur in bursts of three to five items at a time.

In the visual-array comparison task of Luck and Vogel (1997), given the recall limit, it appears that a capacity-limited store must be used instead of participants relying solely on a passively held visual-spatial store. That may be the case because the second array overwrites the visual-spatial representation of the first before a comparison can be made. Cowan (2001) reported that a simple formula for capacity, taking into account guessing rates, typically provides a stable capacity estimate of about 3.5 items for all set sizes above 4. Reinforcing the notion that it is a central type of capacity that is needed for two-array comparisons, Stevanovski and Jolicoeur (2003) found that a task as simple as tone identification imposed during the interarray period seriously interfered with array comparisons. Two recent articles pinpointed posterior brain areas involved in the capacity-limited memory used in this sort of task (Todd & Marois, 2004; Vogel & Machizawa, 2004).

Inasmuch as the visual working-memory task was to be combined with a verbal working-memory task (e.g., Cocchini et al., 2002; Morey & Cowan, 2004), a critical question is whether the verbal task could be constructed in such a way that it could not be carried out primarily by using a passive store, such as the *phonological loop*, in which an automatic rehearsal process is said to refresh a passively held sequence of phonemes with very little commitment of central executive resources (Baddeley, 1986). We hypothesized that this factor might explain differences in the outcomes of the recent studies by Cocchini et al. and Morey and Cowan (2004). In both of these studies, in some experimental conditions, participants were to hear a digit sequence, examine a spatial pattern, recall the spatial pattern, and then recall the digit sequence. Given that no stimuli intervened between the spatial stimulus and response, effects of the digit sequence could be attributed to the memory load imposed by the verbal stimuli during the execution of the spatial working-memory task. The effect of this memory load appeared much milder in the case of Cocchini et al., who ultimately argued that the mild effects were incompatible with the notion of a central storage faculty. In the appropriate condition within Cocchini et al.'s Experiment 1, visual arrays were remembered with 90% accuracy in a single task versus 84% with a digit-span load (a 6% effect). The condition of Morey and Cowan's experiment that was psychometrically most comparable (with the most similar level of single-task performance) was six-item visual arrays, which were remembered with 91% accuracy in a single task versus 78% with a seven-digit load (a 13% effect). Because two similarly motivated studies reached such different conclusions, we designed the present experiments to explore possible reasons for the discrepancy in effect magnitudes.

Cocchini et al.'s (2002) conclusion that storage in working memory occurs only in dedicated within-domain buffers rests on the inability to find a discrepancy such as the one observed by Morey and Cowan (2004), but several procedural differences between these studies may account for the differences in results. First, whereas Morey and Cowan separated their data into trials in which the digit sequence was correctly versus incorrectly recalled, Cocchini et al. (2002) collapsed across this variable. Morey and Cowan found more severe forgetting of the visual arrays on trials in which the digit sequence was recalled incorrectly. This might be attributed to the greater difficulty of some random sequences of digits than of others. It also is noteworthy that the participants of Morey and Cowan correctly recalled the digit sequences only 45% of the time but much more often in Cocchini et al., who chose digit-list lengths corresponding to each participant's ability (e.g., in the relevant condition of their Experiment 1, .81). The greater difficulty of the verbal task of Morey and Cowan, who used list lengths of two and of seven random digits for all participants, might account for much of the discrepancy in results.[1]

Nevertheless, given the speculative nature of comparisons between different procedures, we conducted follow-up studies to examine the role of another potentially important factor: digit-list articulation. Whereas the participants of Cocchini et al. (2002) were to rehearse the digit load silently (if at all), Morey and Cowan (2004) required that the digit load be recited aloud during the retention interval between the visual array to be remembered and a comparison array. Given an assumption that explicit recall requires that information be retrieved from passive storage back into the focus of attention (Cowan, 1995) or requires the involvement of central executive processes (Baddeley, 1986), articulating the digit loads should increase the possibility of taxing faculties that are needed across domains and modalities. This hypothesis was tested in Experiment 1. Then, in Experiment 2, we aimed to show that some central storage, such as the focus of attention (see Cowan, 1995, 2001) or the episodic buffer (Baddeley, 2001), is necessary to account fully for observed cross-modal interference.

The suggested importance of overt recitation of the memory load also is consistent with the results of a study by Barrouillet, Bernardin, and Camos (2004), who found that simply reading numerals aloud interfered with short-term memory (STM) for consonants, to an extent that depended on the rate of reading numerals. They proposed a "time-based resource sharing model" in which both storage and processing depend on attention. Presumably, according to their approach, it was the need to retrieve numerals to read them aloud that interfered with the storage of consonants.

In addition to our hope of reconciling the findings of Cocchini et al. (2002) with those of Morey and Cowan (2004), we designed Experiment 1 with another important question in mind: Does cross-domain interference occur during encoding or maintenance of stimuli in working memory? We considered the possibility that rehearsing aloud might compete with encoding of the first visual array rather than its maintenance. To examine that possibility, we included conditions in which the recital of the digit load was to

---

[1] If we take the percentages of correct visual-array performance in the data of Morey and Cowan for six-item visual arrays when digit recall was correct (84% array performance) versus when it was incorrect (73% array performance) and assume that these same percentages apply to a situation in which the digit recall is correct 81% of the time (the proportion found by Cocchini et al., 2002), it leads to an estimate of a 9% digit-load effect on visual-array performance (i.e., $.91 - [.84 \times .81 + .73 \times .19] = .09$), not much higher than the 6% effect that Cocchini et al. obtained.

begin before versus after the presentation of the first visual array (i.e., early vs. late recital), in either case continuing until the response to the second visual array. In the early recital condition, active maintenance of the list presumably had to share any central resources needed for either the encoding or the maintenance of that visual array. In contrast, in the late recital condition, given that digit-list articulation was delayed until after the offset of the first visual array, active maintenance of the list presumably had to share central resources needed only for the maintenance of the visual array and not for its encoding. If the latter condition produces at least as much interference, then it will support the conclusion that central resources are needed for maintenance, per se.

Theoretically, this interpretation of the data might not be valid if encoding continued after the offset of the first array, in the form of a memory-consolidation process. However, recital in the late condition of our study never began sooner than 750 ms after the onset of the array. It has been shown that the consolidation process, measured by varying the time of presentation of a mask to interrupt consolidation, does not continue beyond 200–300 ms after the onset of the array (Vogel, Woodman, & Luck, in press; for convergent evidence, see Eriksen & Eriksen, 1971; Shaffer & Shiffrin, 1972).[2]

Interference during encoding has a powerful effect on long-term recall accuracy (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996), but presumably, encoding for short-term recall is simpler (Healy & McNamara, 1996) and may not be as vulnerable to interference from a dual task. Moreover, the question of the locus of cross-domain interference in working memory is, at least to our knowledge, untested. In previous studies (including Cocchini et al., 2002; Luck & Vogel, 1997; Morey & Cowan, 2004), the digit list was always administered before the first visual array, in effect dividing attention at the time of encoding and continuing through maintenance of the first array, making any judgment about the locus of interference in working memory speculative. Localizing interference is a critical test for models of working memory that posit central storage (models such as Baddeley, 2001; Cowan, 1995): If reduced resources during encoding alone could account for cross-domain dual-task deficits, then the supposition of shared storage, as opposed to a shared controlling mechanism such as the central executive, would be unsupported.

Experiment 1

Method

Participants

The participants were 26 students from the University of Missouri—Columbia introductory psychology subject pool (12 women) between the ages 18 and 24, participating to fulfill course requirements. One participant was excluded because of insufficient English language comprehension in understanding the instructions, and 1 participant failed to complete the experiment. Two other participants were unable to recall any digit lists correctly and were omitted from the analyses, leaving a final *N* of 22 (9 women and 13 men).

Apparatus and Stimuli

The experiment was executed with E-Prime (Schneider, Eschman, & Zuccolotto, 2002) software. Stimuli were displayed on 17-in. color monitors. The visual arrays included four, six, or eight squares (0.65° × 0.65°) arranged randomly within a 7.3° × 9.8° viewing area on a neutral gray background, each square of a color randomly selected from one of the following seven easily discriminable colors: red, blue, violet, green, yellow, black, or white. The items in the arrays were separated by at least 2° of visual angle, measured from the upper left corner of the squares. Participants were situated approximately 50 cm from the monitors. Array presentations were separated by a 2,000-ms interstimulus interval (ISI).

Phonological stimuli were lists of seven random digits recorded in a male voice, selected without replacement from the numbers 1–9. Digits were presented through headphones at a rate of 1 digit per second and fell between 74 and 80 dB as measured by a sound level meter and earphone coupler. On some trials, the presentation of a tone signaled that the participant should begin repeating the verbal load aloud.

Procedure

Participants compared two sequentially presented visual arrays, making judgments of *same* or *different* for each. In some trials, participants also maintained a seven-digit verbal load, which they reported at the end of the trial by typing it. We manipulated the onset of articulation of the verbal load. In trials requiring articulation, participants were instructed to begin repeating the digits upon hearing a tone and to continue repeating the digits until the probe array appeared. An experimenter monitored each session to ensure that this occurred. Occasionally, the experimenter needed to remind a participant to speak aloud when the tone sounded. This occurred no more than once during a session. Participants were told that they could stop articulating when the probe array appeared, although no effort was made to enforce stopping. Participants performed 30 trials each of the following five trial types.

Verbal task control. There was no visual task in this condition. Participants heard a list of seven digits and were immediately prompted to recall the list by typing it on a keyboard. In all conditions, including a verbal digit task, participants were allowed to change their answers by using the backspace key until they designated the list complete with a final keypress. The responses appeared on the screen as they were typed.

Visual-array comparison no-load control. After the appearance of a 1,000-ms fixation cross in the center of the screen, participants saw a sample array of colored squares for 500 ms. After a 2,000-ms ISI, a test array appeared with one square encircled. This test array remained on the screen until a response was entered. Participants were instructed to judge whether the encircled square was the same color in the second (*test*) array as it was in the first (*sample*) array by typing the *S* key for same or the *D* key for different. Participants were aware that the unindicated squares always remained the same color in the test array as they had been in the sample array. Parameters of the visual-array task were the same in all conditions that included it.

Silent memory load. Each trial began with the presentation of a list of seven randomly selected digits. The digits sounded at a rate of one per second. After the offset of the seventh digit, a fixation cross appeared in the center of the screen to remind the participant where to look. This fixation cross remained on the screen for 1,000 ms and was followed by 1,000 ms of a blank gray screen. The sample array then appeared for 500 ms. After a 2,000-ms ISI, the test array appeared and remained on the screen until a response was registered. After responding to the test array, participants

---

[2] In another recent study of cross-domain interference in a visual-array comparison task, Stevanovski and Jolicoeur (2003) found that the reaction time to a task interpolated between two arrays to be compared (a tone-identification task) decreased with the temporal position of the interpolated task in the interarray period only until it was 400 ms after the onset of the first array, after which it stabilized. This again suggests that encoding and consolidation of the sample arrays occurred prior to 400 ms after the onset of the array.

were prompted to type the seven digits they had heard at the beginning of the trial.

Aloud early. The aloud early trials followed a similar course of events as the silent memory load trials. After presentation of the digits, a fixation cross appeared and simultaneously a tone sounded, both lasting 1,000 ms. Recitation of the seven-digit list was to begin at the tone. After a 1,000-ms blank, gray screen, the sample array appeared for 500 ms. After a 2,000-ms ISI, the test array appeared; participants responded to the test array and were subsequently prompted to type the digit list.

Aloud late. The aloud late condition included all of the components of the aloud early condition, except that the tone that signals digit-list recitation occurred after the offset of the sample array. Participants heard a list of seven digits, followed by a 1,000-ms fixation cross, 1,000-ms blank screen, and the 500-ms sample array. Two hundred and fifty milliseconds after the offset of the sample array the tone sounded, signaling participants to begin articulating the digit list. One thousand seven hundred and fifty milliseconds after the tone onset, the test array appeared (making the interarray interval 2,000 ms as in the other conditions). Once participants responded to the probe array, they were prompted to type the digit list.

Results

The visual task accuracies for all conditions in the experiment are shown in Table 1. A two-way analysis of variance (ANOVA) with the visual-array size (four, six, or eight squares) and articulation condition (no-load control, silent, aloud early, or aloud late) as within-participant factors revealed significant main effects of visual-array size, $F(2, 42) = 50.37$, $MSE = 0.01$, $p < .01$, and onset of articulation, $F(3, 63) = 24.15$, $MSE = 0.02$, $p < .01$. The interaction between these variables was nonsignificant, $F(6, 126) = 0.39$, $MSE = 0.01$, $p > .05$. Post hoc Neuman–Keuls pairwise comparisons showed that visual-array comparison accuracy was significantly lower in the aloud early and aloud late groups than in all other conditions ($p \leq .02$ in all such comparisons); furthermore, accuracy was significantly lower in the aloud late group than in the aloud early group. No differences were observed between the silent memory load and the no-load control conditions.

Table 1
Experiment 1: Mean Visual Task Accuracy by Verbal Task Condition, Verbal Task Accuracy, and Visual-Array Size

Verbal task condition	Visual-array size 4	6	8
No verbal task	.95 (.02)	.87 (.02)	.80 (.03)
Silent load			
Digits correct	.96 (.02)	.87 (.05)	.67 (.06)
Digits incorrect	.90 (.03)	.83 (.06)	.72 (.04)
All data	.93 (.02)	.85 (.03)	.75 (.03)
Aloud early			
Digits correct	.94 (.03)	.78 (.04)	.71 (.05)
Digits incorrect	.80 (.04)	.77 (.05)	.64 (.06)
All data	.85 (.03)	.75 (.03)	.65 (.03)
Aloud late			
Digits correct	.89 (.03)	.72 (.04)	.64 (.06)
Digits incorrect	.68 (.05)	.70 (.05)	.59 (.04)
All data	.79 (.02)	.69 (.03)	.62 (.03)

Note. Means for correct and incorrect digit rows include only 18 participants with both correct and incorrect verbal trials in each visual-array size. The rows marked "no verbal task" and "all data" include all 22 valid participants. Standard errors of the mean are in parentheses.

A one-way ANOVA of verbal task accuracy, defined as entering each digit in correct order, revealed a significant main effect of articulation condition, $F(3, 63) = 8.41$, $MSE = 0.01$, $p < .01$. Verbal task accuracy was highest in the verbal control condition, in which there was no visual-array task ($M = 0.53$, $SEM = 0.05$). According to post hoc Neuman–Keuls pairwise comparisons, performance was significantly higher in this condition than in the other conditions ($p < .01$ in all cases), indicating that performing the visual-array task interfered with digit-list recall. Performance in the other conditions was not statistically different (silent: $M = 0.42$, $SEM = 0.05$; aloud early: $M = 0.44$, $SEM = 0.05$; aloud late: $M = 0.40$, $SEM = 0.04$).

As Morey and Cowan (2004) noted, when the digit stimuli were recalled incorrectly, accuracy on the visual-array task was lower than when the digits were not recalled correctly. Results of a comparable analysis for the present data are shown in Figure 1, collapsed across set size. (The no-load control condition, which was not included in the analysis, is also shown in Figure 1 for the sake of comparison.) For the 18 participants with correct and incorrect verbal responses on trials at each array size, a three-way ANOVA of response to the visual task, including accuracy on the verbal task, articulation condition, and visual-array size as factors, revealed significant main effects for each factor: verbal task accuracy, $F(1, 17) = 8.29$, $MSE = 0.04$, $p < .02$; articulation condition, $F(2, 34) = 17.21$, $MSE = 0.02$, $p < .01$; visual-array size, $F(2, 34) = 21.78$, $MSE = 0.05$, $p < .01$. None of the interactions reached significance: Verbal Task Accuracy × Recital Condition, $F(2, 34) = 1.38$, $MSE = 0.03$, $p > .05$; Verbal Task Accuracy × Visual-Array Size, $F(2, 34) = 3.16$, $MSE = 0.03$, $p = .06$; Recital Condition × Visual-Array Size, $F(4, 68) = 0.48$, $MSE = 0.03$, $p > .05$; Verbal Task Accuracy × Recital Condition × Visual-Array Size, $F(4, 68) = 0.60$, $MSE = 0.04$, $p > .05$.

To quantify the sharing of a central resource between the two tasks, the effects of silent and articulated memory loads can be examined in terms of visual item capacity estimates under each memory load condition. A discussion on generating capacity estimates for simultaneously presented visual stimuli can be found in Cowan (2001) and Morey and Cowan (2004). These capacity estimates refer to the number of items separately held in a capacity-limited portion of working memory at the same time (e.g., in the focus of attention or a capacity-limited episodic buffer) and are calculated in a manner that takes guessing into account. The formula can be expressed as $k = N \times [p(\mathrm{h}) - p(\mathrm{f})]$, where k is the number of objects held; N is the set size; $p(\mathrm{h})$ is the probability of a hit, that is, correctly responding that there was a change in the array; and $p(\mathrm{f})$ is the probability of a false alarm, that is, incorrectly responding that there was a change in the array. Capacity estimates provide more specific information than proportion correct statistics: They allow direct comparison across the set size of the number of items recalled.

Table 2 shows the mean capacity estimates for each condition. An ANOVA of individual's visual capacity estimates with visual-array size and recital condition as factors revealed a significant effect of recital condition on the capacity estimate, $F(3, 54) = 22.88$, $MSE = 2.24$, $p < .01$. The effect of visual-array size approached significance, $F(2, 36) = 3.11$, $MSE = 2.00$, $p = .06$, and it appears to be due to differences between capacity estimates

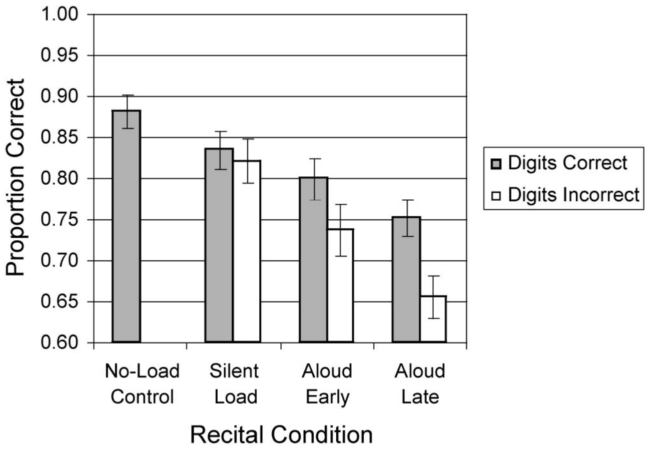

Figure 1. Experiment 1: Visual task accuracy by articulation condition and verbal task accuracy for only the 18 participants who had trials with both correct (shaded bars) and incorrect (white bars) recall of digit lists in all conditions. Error bars represent standard errors of the mean.

at Array Size 4 and capacity estimates at Array Sizes 6 and 8 (presumably because capacity estimates at Array Size 4 could not exceed 4). The interaction was nonsignificant, $F(6, 108) = 1.88$, $MSE = 1.98$, $p > .05$. Post hoc Neuman–Keuls pairwise comparisons show that recital conditions were all significantly different from each other (all ps $< .02$) Because we needed to compare these data with visual capacity estimates generated for trials in which participants correctly recalled the digit list, the analysis was restricted to the 19 participants who correctly recalled some digit lists in each condition.

An ANOVA of visual capacity estimates was carried out for trials in which the digits were correctly recalled and therefore presumably imposed a known memory load, with recital condition and visual-array size as factors. As seen in Table 2, this analysis revealed a significant main effect of recital condition, $F(3, 54) = 6.04$, $MSE = 3.95$, $p < .01$. The effect of visual-array size and the interaction were nonsignificant: visual-array size, $F(2, 36) = 0.69$, $MSE = 3.50$, $p > .05$; Recital Condition × Visual-Array Size, $F(6, 108) = 1.59$, $MSE = 3.87$, $p > .05$. Post hoc Neuman–Keuls analyses indicated that visual capacity was significantly lower in the late articulation condition than in the silent or no-load conditions and estimates in the aloud early condition were significantly lower than in the no-load condition (ps $< .05$).

Using the capacity measure in the case of lists correctly recalled, we can quantify the difference between verbal conditions by items recalled. Table 2 shows that, compared with silent maintenance of the digits, late recitation (during the visual maintenance period) cost 0.54 items from four-item arrays, 1.9 items from six-item arrays, and 0.52 items from eight-item arrays. Considering that the estimated visual capacity is 3–4 items, as is typical (Cowan, 2001; Luck & Vogel, 1997), these losses are substantial.

An apparent exception to the claim that only articulated, or retrieved, memory loads cause interference is the case in which a large load is held silently by the participant while performing an array discrimination including eight squares, the largest array size we examined. A dependent samples t test comparing visual capacity estimates at Array Size 8 in the no-load and silent load conditions for trials in which the digits were correctly reported revealed a significant difference, $t(18) = 2.15$, $p < .05$. As Table 2 shows, the difference was 1.67 visual items. If, as we suspect, attention must be shared between verbal lists and visual arrays to supplement any nonattentional forms of storage, then it would be expected that interference is most severe when the total demand on memory is largest. Silent retention of a seven-digit list may require only a small amount of attention, given the use of passive storage, rehearsal, or phonological buffers (e.g., Baddeley & Logie, 1999), but even that small amount of attention may make a difference for recall of arrays with as many as eight objects.

Capacity estimates are least informative for four-item arrays because of ceiling effects in some participants. Overall, Table 2 suggests a pattern in which maintaining items aloud drastically interferes with memory for six-item arrays (costing about 1.9 items), whereas maintaining items either silently or aloud drastically interferes with eight-item arrays (costing 1.7 items for silent recall and 2.2 items for late recitation).

Discussion

In sum, the addition of articulation of the digit lists to the verbal and visual memory task demands produced a dramatic difference in the results. We posit that this difference is because of forced retrieval of the digit lists. When allowed to maintain the digits silently, participants may be able to switch between the verbal and visual tasks by holding the digits in phonological storage, or perhaps by recruiting long-term memory to store all or some of the digits (Ericsson & Kintsch, 1995). However, when participants are forced to articulate the verbal load, attention presumably is called away from the visual task and engaged with recalling the list. This withdrawal of attention from the visual memoranda is the apparent cause of a loss of between 0.5 and 2.0 items, depending on the array size. Presumably, this depends on use of a central attentional resource for the maintenance of visual items as well as for the recitation of spoken items.

In contrast to the striking effect of overt retrieval of the digit load during the interarray interval, there was no observable effect

Table 2
Experiment 1: Mean Visual Capacity Estimates by Verbal Task Condition, Verbal Task Accuracy, and Visual-Array Size

Verbal task condition	Visual-array size 4	6	8
No verbal task	3.54 (0.19)	4.61 (0.21)	5.14 (0.53)
Silent load			
Digits correct	3.67 (0.16)	4.58 (0.44)	3.47 (0.64)
All trials	3.33 (0.14)	4.23 (0.38)	3.45 (0.43)
Aloud early			
Digits correct	3.51 (0.21)	3.22 (0.48)	3.55 (0.71)
All trials	2.86 (0.21)	3.03 (0.37)	2.86 (0.47)
Aloud late			
Digits correct	3.13 (0.25)	2.68 (0.43)	2.95 (0.64)
All trials	2.23 (0.17)	2.27 (0.35)	2.19 (0.43)

Note. Participants missing data in some cells were excluded. Included $N = 19$. Correct digit recall occurred on 47%, 50%, and 46% of the trials in the silent load, aloud early, and aloud late conditions, respectively. Standard errors of the mean are in parentheses.

of overt retrieval of the digit load during the encoding of the first array. Visual-array performance was superior in the aloud early condition (which involved articulation of the load during encoding of the first array) compared with the aloud late condition (which did not involve articulation during encoding). If the locus of cross-domain interference in working memory were during encoding, then visual capacity should have been lowest in the aloud early condition in which attention was thought to be shared during both encoding and maintenance of the sample visual array.

According to evidence that led to the maintenance rehearsal theory of Naveh-Benjamin and Jonides (1984), the processes involved in the first iteration of the digit-list rehearsal demand a great deal of attention. Successive iterations require progressively less attention as repetition of the list becomes automatic. Accordingly, more interference during maintenance of the sample visual array was to be expected in the late articulation condition, as the most attention-demanding aspect of verbal load retrieval certainly occurred during maintenance in this condition. The finding of the lowest visual-array memory in that condition supports this theoretical analysis.

Experiment 2

Although articulation of a memory load during maintenance of a visual array was shown in Experiment 1 to be important, there are two possible explanations for why this effect occurred. As we suggested above, it may have to do with forced retrieval of the digits. Alternatively, though, it may be that overt pronunciation itself interferes with maintenance of the visual array. This could occur if verbal means are used to encode the visual array, as is often the case for pictorial stimuli (e.g., Hitch, Halliday, Dodd, & Littler, 1989). We do not believe that to be the case for the visual arrays, given that Morey and Cowan (2004) found no effect of the overt recitation of the participant's own telephone number during the array. However, we have yet to examine the effect of unrelated articulation in the presence of a verbal memory load. Perhaps it is the combination of these tasks that requires attention and results in decreased accuracy on a concurrent visual-array comparison task. Toward this end, we conducted a second experiment in which, in two blocks of trials, participants performed the visual-array comparison task while maintaining lists of two, four, or six digits. In one of these blocks, additionally, articulatory suppression (Baddeley, 1986) was added to the procedure. This design enables us to compare visual-array comparison accuracy under varying levels of verbal load and under a silent load to which articulatory suppression was added.

Our concern about the demand of carrying out three tasks concurrently (visual-array memory, digit-list memory, and articulatory suppression), and about the potent effects of articulatory suppression on serial recall (Baddeley, 1986) even for spoken lists (Cowan, Cartwright, Winterowd, & Sherk, 1987), led us to use slightly less demanding stimuli in this experiment. We used lists of two, four, and six digits instead of seven digits; a 900-ms period between visual arrays instead of a 2,000-ms period; and a simple one-repeated-word articulation requirement. Thus, the two experiments are not procedurally comparable. However, in this experiment the proportion of six-digit lists with articulatory suppression correctly recalled was just slightly lower than the Experiment 1 proportion of seven-digit lists recited aloud and then correctly recalled, making these manipulations psychometrically more comparable than if we had adhered to the Experiment 1 parameter values. Moreover, articulatory suppression always began before, rather than after, the presentation of the first visual array (providing a counterpart to the aloud early, but not the aloud late, condition of Experiment 1), because we were concerned that there might be attention requirements of beginning the articulatory process and wished to exclude that factor as a source of interference with visual memory maintenance.

This study also provides an additional test of the notion that it is central storage rather than central executive processes defined in terms of task scheduling (see Baddeley, 1986) that are taxed by the requirement of simultaneous visual-spatial and auditory-phonological memory tasks. The concurrent performance of three unrelated tasks (articulation of the word *the*, visual-array comparisons, and maintenance of random digit lists) should pose a task-scheduling challenge. If we observe cross-domain interference only because demanding concurrent cross-modal tasks require central executive-type coordination, then we should observe interference between visual and verbal memoranda in this study, even though retrieval of the digit lists during array comparison is not forced. However, if interference is caused by retrieval of verbal items into central storage as proposed above, unrelated articulation and silently held (and thus, presumably unretrieved) memory loads should have little if any effect on visual-array comparison accuracy.

Method

Participants

Sixteen students (14 women, 18–23 years old) from the introductory psychology subject pool at the University of Missouri—Columbia with normal or corrected-to-normal vision and normal hearing participated in the study to fulfill course requirements.

Apparatus and Stimuli

The apparatus for Experiment 2 was the same as that used in Experiment 1, with the addition of a Psychology Software Tools serial response box and a microphone. The visual stimuli were generated with the same parameters as those used in Experiment 1. The verbal stimuli were the same as those used in Experiment 1 except that list lengths of two, four, and six digits were used in this study (because it was not known which length would provide the most appropriate level of recall when combined with articulatory suppression).

Procedure

Participants compared two visual arrays and determined whether they were identical or different while maintaining a verbal load of two, four, or six digits to be recalled. In addition to these tasks, in one of two trial blocks, participants were asked to repeat the word "the" at a rate of about two repetitions per second.

Each trial began with the presentation of digits at a rate of one per second. After the offset of the final digit of the list, a fixation screen appeared with a "+" in the center as a fixation cross. In the articulatory suppression block, participants were to begin saying "the" when the "+" appeared; speaking triggered a voice key, which allowed the trial to continue. In the no-suppression block, the trial simply continued after the "+" remained onscreen for 1,000 ms. After activation of the voice key or offset of the fixation screen, the sample array appeared, remaining onscreen for 500 ms. After a 900-ms ISI, the test array appeared, prompting participants to make a response. In the articulatory suppression blocks, articulation was to cease when this second array appeared. Once a response to the test array was recorded, participants were prompted to recall the digit list.

To standardize the effects of articulatory suppression, the trial would restart if speech was not detected within 1,000 ms of the appearance of the fixation cross. An experimenter listened outside of the private testing booth to ensure that participants continued articulating until a response to the test array was registered. Trials that were restarted were generally because of insufficient loudness of articulation rather than because participants neglected to follow the articulation instructions, and restarts were infrequent after the practice session. None of the participants needed to be chastised for failing to articulate altogether, although some participants occasionally were reminded to continue to articulate throughout the presentation.

Two blocks of 90 experimental trials each were administered to each participant, one of which included the articulatory suppression procedure and one of which omitted it. These blocks were counterbalanced across participants to control for order effects. Each block began with three practice trials, including one practice trial at each digit-list length.

Results

Table 3 presents the mean proportions correct for every condition in the Experiment 1. In Experiment 1, large differences were observed between visual capacity in articulated versus silent verbal load conditions. In the present experiment, in contrast, similar effects were not observed for trials with versus without articulatory suppression added to a silent load. Confirming this description of the results, a three-way ANOVA including visual-array size, digit-list length, and suppression condition as factors revealed only a significant effect of visual-array size, $F(2, 30) = 78.78$, $MSE = 0.01$, $p < .01$, and a significant interaction between visual-array size and digit-list length, $F(4, 60) = 2.71$, $MSE = 0.01$, $p < .05$. All other factors and interactions were nonsignificant: digit-list length, $F(2, 30) = 3.19$, $MSE = 0.01$, $p = .06$; suppression condition, $F(1, 15) = 1.90$, $MSE = 0.03$, $p > .05$; Digit-List Length × Suppression Condition, $F(2, 30) = 1.34$, $MSE = 0.01$, $p > .05$; Suppression Condition × Visual-Array Size, $F(2, 30) = 0.21$, $MSE = 0.01$, $p > .05$; Visual-Array Size × Digit-List Length × Suppression Condition, $F(4, 60) = 1.10$, $MSE = 0.01$, $p > .05$. As shown in Table 3, visual-array performance was less affected by array size in the presence of two-digit lists than in the presence of four- or six-digit lists. It is of most importance to note that neither the effect of articulatory suppression nor any other main effects or interactions reached significance, indicating that an unrelated articulatory suppression task combined with a verbal memory load does not negatively affect visual-array task performance.

Table 3
Experiment 2: Visual Task Accuracy as a Function of Verbal Task Condition and Visual-Array Size

	Visual-array size		
Verbal task condition	4	6	8
2-digit load, silent	.96 (.01)	.88 (.03)	.81 (.03)
2-digit load, with suppression	.96 (.03)	.83 (.03)	.77 (.04)
4-digit load, silent	.98 (.01)	.94 (.02)	.80 (.03)
4-digit load, with suppression	.91 (.04)	.88 (.03)	.76 (.05)
6-digit load, silent	.96 (.02)	.83 (.04)	.76 (.03)
6-digit load, with suppression	.91 (.03)	.88 (.03)	.76 (.03)

Note. $N = 16$. All trials are included, regardless of digit recall accuracy. Standard errors of the mean are in parentheses.

To assess the risk of Type II error in accepting the null effect of articulatory suppression on visual-array comparisons, we performed a power analysis on visual-array task data collapsed across array set size, using only trials with six-digit memory loads. To detect a .09 effect on the proportion of correct recall of visual arrays (which, in Experiment 1, is the magnitude of the difference between the aloud early and silent load conditions), we found that the power was .95. The power was .80 to detect a difference of .07. Thus, a moderate-sized effect of suppression in the presence of a memory load would have been detected. Across all list lengths, the power to detect an effect of suppression was even greater (a power of .98 to detect a .09 effect; a power of .80 to detect a .06 effect). Thus, if there is any actual effect of suppression on visual-array performance, it is probably a 6% effect or less.

An analysis of verbal task accuracy was carried out as a three-way ANOVA, including suppression condition, visual-array size, and digit-list length as factors. It revealed significant effects of suppression condition, $F(1, 15) = 47.61$, $MSE = 0.04$, $p < .01$, and digit-list length, $F(2, 30) = 87.47$, $MSE = 0.07$, $p < .01$, but not visual-array size, $F(2, 30) = 0.93$, $MSE = 0.01$, $p > .05$. Verbal accuracy was significantly higher in the no-suppression condition than in the suppression condition. The interaction between suppression condition and digit-list length was significant, $F(2, 30) = 17.21$, $MSE = 0.02$, $p < .01$ (no suppression: two-digit $M = 0.98$, $SEM = 0.01$; four-digit $M = 0.95$, $SEM = 0.02$; six-digit $M = 0.64$, $SEM = 0.05$; suppression: two-digit $M = 0.95$, $SEM = 0.02$; four-digit $M = 0.80$, $SEM = 0.04$; six-digit $M = 0.36$, $SEM = 0.06$). Thus, it is not surprising that the recall of long lists was impaired by articulatory suppression. All other interactions were nonsignificant: Suppression Condition × Visual-Array Size, $F(2, 30) = 0.46$, $MSE = 0.01$, $p > .05$; Visual-Array Size × Digit-List Length, $F(4, 60) = 0.11$, $MSE = 0.01$, $p > .05$; Suppression Condition × Visual-Array Size × Digit-List Length, $F(4, 60) = 0.66$, $MSE = 0.01$, $p > .05$.

Finally, we asked whether the effect of suppression might be different for trials in which the digits were correctly versus incorrectly recalled. These analyses were carried out only among participants who had data at all visual-array sizes with correct (or with incorrect) recall of six-digit lists. However, as shown in Figure 2, there was no effect of suppression either for trials with correct list recall, $t(12) = 0.41$, $p > .05$, or with incorrect list recall, $t(13) = -0.62$, $p < .05$. It seems clear that articulatory suppression has no effect on visual-array task performance.

Discussion

In sum, the effect of unrelated articulatory suppression on visual-array comparison task performance is different from that of articulating a verbal load. Articulating a to-be-remembered verbal load (in Experiment 1) caused an obvious decline in visual-task performance, especially when verbal memoranda were incorrectly recalled. In contrast, when visual-array comparisons were carried out with a silent, two-, four-, or six-digit load plus a separate articulation task (in Experiment 2), visual performance was not statistically different from a control condition in which the load

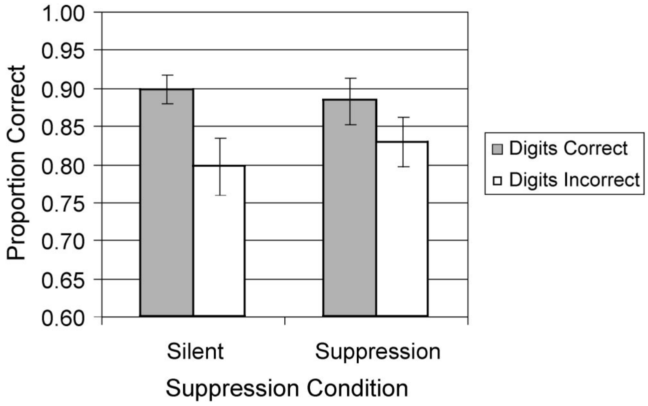

Figure 2. Experiment 2: Visual-array task accuracy for trials with correctly and incorrectly reported six-digit lists, with and without an unrelated articulatory suppression task for participants with data at each visual-array size and suppression condition. For digits correct, $N = 13$; for digits incorrect, $N = 14$. Error bars represent standard errors of the mean.

was retained silently. On the basis of this evidence, it does not appear to be merely any articulation, but specifically articulation of a memory load, and therefore retrieval of the load into the focus of attention that interferes with visual-array memory maintenance during a delay.

General Discussion

We have observed results suggesting that the disruption of attention by a verbal memory load during the period of maintenance of a visual array in working memory interferes with this visual working memory. There are three unique empirical contributions of this study. The first is the finding in Experiment 1 that reciting the spoken memory load aloud has a much larger effect on visual-array comparisons than maintaining it silently, even though both types of memory maintenance are equally effective for maintenance of the spoken digit list. The second is the finding in Experiment 1 that recitation beginning before the presentation of the first visual array is not more damaging to visual-array comparisons (and, indeed, is slightly less damaging) than recitation beginning 750 ms or more after the first visual array. Compared with a silently maintained verbal memory load, articulation of the verbal memory load beginning prior to the sample visual array resulted in a 9% decrease in visual-array comparison accuracy, whereas articulating the verbal memory load beginning after the offset of the sample array resulted in a 14% decrease. Finally, there is the finding in Experiment 2 that silent maintenance plus a separate, simple articulatory suppression task is no more damaging to visual-array comparisons than silent maintenance alone.

Taken together, these results suggest that the effect of reciting a memory load aloud occurs because central resources are taken away from the maintenance of the visual array in working memory. To the best of our knowledge, we have found that no previous published study shows this. It is suggested by Stevanovski and Jolicoeur (2003) but the role of verbal processing in their interpolated task, tone identification, was not investigated. In other previous studies (including Cocchini et al., 2002; Luck & Vogel, 1997; Morey & Cowan, 2004) verbal memory loads were imposed during both the encoding and the maintenance of the visual stimuli, making it unclear whether the interference occurred during maintenance or encoding. Moreover, there is little evidence in previous studies as to whether it matters that the memory load be recited aloud (though see Baddeley, 1986, for prior suggestions that it does).

These results clarify the processes taking place in an experiment reported by Morey and Cowan (2004). That study showed that visual-array comparisons were disrupted by overt recitation of a seven-digit memory load, but not by recitation of the participant's own seven-digit telephone number. The latter condition ruled-out speech, per se, as the source of the interference, confirming that the visual-array comparison task does not depend on the ability to encode items verbally in the array (e.g., by color name). However, it left open several theoretical possibilities that can be eliminated on the basis of the present evidence.

One possibility is that attention might be used solely for maintenance of both visual-spatial and acoustic-phonological information. Although this is likely to be a factor, the effect of a silent phonological memory load (Experiment 1) was small, and the effect of a memory load recited aloud was much larger. Yet, silent maintenance was just as effective as overt recitation of the digits. Therefore, pure maintenance activities in the two modalities do not seem to share much in the way of resources. The larger part of the resource sharing seems to be between visual-spatial maintenance, on one hand, and verbal retrieval, on the other hand.

Another possibility is that there could be sharing of resources at the time when visual information is encoded into memory. If that were the case, however, then we should have found that aloud recitation of a memory load beginning just before the first visual array and continuing throughout the maintenance period would have a larger effect on visual comparisons than aloud recitation of the memory load beginning only 750 ms after the onset of the array. Instead, effect was larger in the latter, aloud late condition in our Experiment 1. It is thus visual maintenance, not encoding, that appears to be vulnerable to interference.

Also, it is theoretically possible that there are at least two routes to visual maintenance that can be compromised: an attentional route that is compromised by a verbal memory load and a phonological route that is compromised by speaking aloud. If that were the case, however, then maintaining a load while carrying out articulatory suppression (affecting both routes) should be more effective than maintaining a load silently (affecting only one route). However, Experiment 2 showed that they are equally ineffective. Instead, it must be the process of retrieving digits for overt recitation that is important in interfering with visual memory maintenance.

These experiments, taken together, thus suggest that one needs some resource that is not domain- or modality-specific to maintain items in working memory.

This common resource may be the focus of attention (Cowan, 1995, 2001) or perhaps the episodic buffer of Baddeley (2000, 2001). The results certainly do not rule out the involvement of passive, code-specific storage devices such as phonological and visual-spatial storage (e.g., Baddeley & Logie, 1999), but they do suggest that these cannot serve as the only means of memory storage.

Different investigators have developed different theoretical views regarding what goes on in the focus of attention, some of which are not entirely compatible with the present view. Garavan (1998) suggested that the focus of attention holds only one item at a time. From that viewpoint it would seem implausible for a resource shared across modalities to hold information from a visual array. If would be possible, however, if the cross-modal resource were something other than the focus of attention. This could be the case if there were a capacity-limited fringe surrounding the focus of attention (Oberauer, 2002) or if an episodic buffer (Baddeley, 2000, 2001) held the information and proved to be a mechanism other than the focus of attention. We would caution, though, that the studies of Garavan (1998) and Oberauer (2002) were not designed to test the upper limits of a focus of attention, but they simply distinguished between one item on which the most recent action was to be carried out and other items without the same privileged status.

Another view compatible with the present findings is that the focus of attention encompasses a single item at a time for list stimuli, with other items recalled from outside of the focus of attention, but that the focus of attention encompasses several items at a time for spatial arrays (McElree & Dosher, 2001). This view could explain why the resource conflict between visual-array item maintenance and digit-list maintenance is no larger than an item or two. Making this sort of view more compatible with the overall theoretical notions of Cowan (2001), Verhaeghen, Cerella, and Basak (2004) extended the work of McElree and Dosher (2001) to find that, with several hours of practice, the focus of attention in list recall appeared to expand from one to four items, a limit similar to what is found for arrays.

Kane et al. (2004) agreed with the present approach in that they have used latent variable analyses of various working-memory tasks to demonstrate that there is a resource, presumably attention related, that is general across modalities. They differ in assuming that the general resource is the control of attention and is often most necessary not to hold items, per se, but to maintain a task goal and relevant stimuli in the presence of conflicting response tendencies and interfering stimuli. Cowan (in press) argued that the focus of attention serves both of these purposes, zooming out to encompass the maximal number of items in an array or zooming in, to some extent compromising its storage capacity, to deal with difficulties such as difficult goal maintenance and interference. In our study, the focus of attention could zoom out to encompass up to about four items in a visual array, except that difficulties in the maintenance and retrieval of digits would detract from that storage to some extent.

According to conventional views of what takes place during the silent maintenance of a verbal list in memory, it is puzzling that there should be an effect of overt articulation of the list. Silent maintenance has been assumed to be a process in which the words are articulated silently in a repeating loop (e.g., Baddeley, 1986). If that were the case, it should involve the same processes as reciting the list aloud, except for the resulting acoustic and motor-system feedback, which would not be predicted to have an effect on visual memory. Consequently, the effect of overt articulation suggests that, in the silent load condition, participants may not have maintained the digits solely through a rehearsal loop. There are several alternative possibilities. One possibility is the contribution from echoic (auditory sensory) memory, which is automatically held for some seconds even without active articulation of the stimuli (for a review, see Cowan, 1995). It is also possible that portions of the list are rapidly memorized, so that retrieval of the verbal information after the response to the visual probe requires retrieval from the same memory mechanism that operates in the long term (i.e., the long-term working memory mechanism described by Ericsson & Kintsch, 1995; see also Cowan, 1995) and the use of lexical knowledge to reconstitute or “redintegrate” fragmentary phonological representations into known words (Hulme et al., 1997; Schweickert, 1993).

Yet another possibility is that covert rehearsal does not involve the regular, unrelenting repetition that was required for overt recitation, but instead can be carried out in a more varied, intermittent, and flexible fashion. That some sort of covert phonological rehearsal and some sort of central resource were used for verbal memory maintenance is suggested by the finding that, in Experiment 1, digit loads were recalled best in the absence of the visual-array task; and by the finding that, in Experiment 2, six-digit lists were recalled 64% of the time with no articulatory suppression but only 36% of the time in the presence of concurrent suppression.

In Experiment 1, the requirement to articulate the verbal load forced immediate retrieval of the load from whatever mechanism stored it, apparently compelling active rather than passive maintenance. It is unlikely that rehearsal uniformly requires attention. In fact, the research of Naveh-Benjamin and Jonides (1984) indicates that verbal maintenance rehearsal includes two stages. The first stage is an attention-demanding retrieval of the verbal material and initiation of rehearsal. In the second stage, rehearsal becomes automatic with repetition. For example, in overt rehearsal of a word pair, the response to a concurrent, secondary probe stimulus that coincided with the first, fourth, or tenth repetitions of the word pair was 507 ms, 495 ms, and 473 ms, respectively. There were other converging types of evidence in support of their hypothesis of the increasing automaticity of rehearsal (e.g., the diminishing value of repetitions after the first few for long-term recognition of the word pair; the increasing difficulty of interrupting rehearsal). Applied to our studies, this research suggests that the reason that visual-array performance was interrupted more heavily in the condition in which overt recitation had to begin during maintenance of the sample visual array, rather than during its encoding, was that this caused the most attention-demanding portion of the rehearsal process to occur during maintenance of the visual array.

The data of Experiment 1 also suggest that when participants are overloaded with stimuli (eight colored squares and seven digits), there is interference even when the digit load is to be held silently. It may well be that under such circumstances, participants use additional attention-demanding central executive processes to try to recode the pattern of colored squares (e.g., to notice patterns) and that this detracts from retention of the digit lists. The application of attention-demanding executive processes to digit recall could take several forms, perhaps including a mixture of memorization and the initial programming of phonological rehearsal processes.

The reason why effects of a concurrent verbal load recited aloud are obtained with lower visual-array sizes than are effects of a verbal load silently held may be that participants have more flexibility when attention is applied in the case of a silent verbal load. This is certainly consistent with the finding of Barrouillet et al. (2004), mentioned above, that the rate of reading off digits from the screen was inversely related to working-memory performance. Reading off digits from working memory may have a similar effect in our study. In both cases, when retrieval must occur on a demanding, experimenter-determined schedule, it draws resources away from maintenance of items in visual working memory. In at least one way, however, our result goes further. In the case of Barrouillet et al., the possibility of interference between similar, verbal stimuli (consonants to be recalled and numbers to be read off) cannot be ruled out. In contrast, we observed interference between much less similar stimuli (spoken digit lists and visual arrays). Showing that the key process is retrieval, in Experiment 2 we found that simple articulation unrelated to the retrieval process did not interfere with visual-array memory.

Other researchers' findings also are consistent with the notion that retrieving one stimulus from memory can disrupt the maintenance of another. Recent work by Stevanovski and Jolicoeur (2003) shows that performing a tone discrimination task during the ISI of a visual comparison task similar to the one used here results in much lower accuracy than performing the array comparison task alone. The tone task itself is not very difficult, but it does require retrieval of a remembered tone for comparison.

The work of Craik and others (Craik et al., 1996; Naveh-Benjamin, Craik, Guez, & Dori, 1998) might appear to contradict our assertion that attention is necessary during maintenance rather than during encoding of memoranda. Craik and colleagues' work shows that divided attention during encoding is more harmful than divided attention at retrieval for long-term paired associate learning. However we think that Craik and colleagues' findings and ours are not fundamentally in disagreement. In their studies, they manipulated attention during encoding and/or retrieval and occupied attention during an interpolated retention interval by using an attention-demanding rehearsal-prevention task. In contrast, we manipulated attention specifically during the retention interval so our procedures are not closely comparable. It is true that in Experiment 1, we found that imposing a memory load during encoding and retention of a visual array was no more damaging than a load imposed first during retention. However, this could mean that encoding processes for STM might be less attention demanding than encoding for long-term retention.

We consider our results as unfavorable to the notion of independent modality-specific attentional capacities (e.g., Duncan et al., 1997) and unfavorable to the notion that there is no use of central attentional capacity for storage purposes (e.g., Cocchini et al., 2002). Instead, it favors the alternative notion of a central attentional capacity (e.g., Jolicoeur, 1999). This central capacity presumably is one means of storing information about the visual arrays, as well as being involved in the maintenance and overt retrieval of the verbal lists. It would supplement other means of storage, such as the visual-spatial and phonological buffers of Baddeley and Logie (1999) or activated features from long-term memory (Cowan, 1995).

It seems relevant that studies of visual retention have suggested that visual-spatial retention is closely linked to visual processing, using central executive resources. For example, Miyake, Friedman, Rettinger, Shah, and Hegarty (2001) found a close relation between visual working-memory tasks that required processing and those that required only retention. This finding was in contrast to the typical finding with verbal stimuli, which is that tasks requiring only storage versus those requiring both storage and processing together are dissociable, with only the storage-and-processing tasks being strongly related to fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999). To explain this sort of difference between modalities, Kane et al. (2004) suggested the possibility that "without the well-practiced strategies people have developed for maintaining linguistic information in verbal tasks, participants in spatial STM tasks must rely solely on attentional/executive processes to maintain target information" (p. 212). Further work would be needed to determine whether the requirement for attention applies to spatial or nonspatial aspects of visual memory, which seem separable (Baddeley & Logie, 1999; Klauer & Zengmei, 2004).

To summarize, these data are consistent with the notion that some shared attentional resource is necessary during maintenance of to-be-remembered visual stimuli. Some resource must be shared because verbal maintenance and recital both affected visual maintenance. It is, nevertheless, unnecessary to posit that all storage is shared. Because the costs of maintaining distracting stimuli do not constitute a perfect trade-off, it is likely that some within-domain storage media, such as the storage buffers of Baddeley and Logie (1999), Cowan's (1995) activated features from long-term memory, and/or sensory memories, are able to contribute. More research is necessary to further specify the relationship between attention and passive storage in working memory.

References

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. (2001). Comment on Cowan: The magic number and the episodic buffer. *Behavioral and Brain Sciences, 24,* 117–118.

Baddeley, A., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in 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.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83–100.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30,* 1086–1095.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163–183.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford, England: Oxford University Press.

Cowan, N. (2001). The magical number four in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114.

Cowan, N. (in press). Working-memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (Eds.), *The 4th Tsukuba International Conference on Memory: Human learning and memory: Advances in theory and application* (pp. 155–175). Mahwah, NJ: Erlbaum.

Cowan, N., Cartwright, C., Winterowd, C., & Sherk, M. (1987). An adult model of preschool children's speech memory. *Memory & Cognition, 15,* 511–517.

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.

Duncan, J., Martens, S., & Ward, R. (1997, June 19). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808–810.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Eriksen, C. W., & Eriksen, B. A. (1971). Visual perceptual processing rates and backward and forward masking. *Journal of Experimental Psychology, 89,* 306–313.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26,* 263–276.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual Review of Psychology, 47,* 143–172.

Hitch, G. J., Halliday, M. S., Dodd, A., & Littler, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology, 7,* 347–362.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1217–1232.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review, 6,* 87–92.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Klauer, K. C., & Zengmei, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133,* 355–381.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

McElree, B., & Dosher, B. A. (2001). The focus of attention across space and across time. *Behavioral and Brain Sciences, 24,* 129–130.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories conflict: Evidence of cross-domain interference in working memory. *Psychonomic Bulletin & Review, 11,* 296–301.

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Dori, H. (1998). Effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1091–1104.

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 369–385.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421.

Sanders, A. F., & Schroots, J. J. F. (1969). Cognitive categories and memory span: III. Effects of similarity on recall. *Quarterly Journal of Experimental Psychology, 21,* 21–28.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 168–175.

Shaffer, W. O., & Shiffrin, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology, 92,* 292–296.

Stevanovski, B., & Jolicoeur, P. (2003, November). Attentional limitations in visual short-term memory. Poster presented at the annual meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.

Todd, J. J., & Marois, R. (2004, April 15). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428,* 751–754.

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working-memory workout: How to expand the focus of serial attention from one to four items, in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1322–1337.

Vogel, E. K., & Machizawa, M. G. (2004, April 15). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428,* 749–751.

Vogel, E. K., Woodman, G., & Luck, S. J. (in press). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance.*

Received June 16, 2004
Revision received December 15, 2004
Accepted January 3, 2005 ■