

# Journal of Experimental Psychology: Learning, Memory, and Cognition

## **What Limits Working Memory Capacity? Evidence for Modality-Specific Sources to the Simultaneous Storage of Visual and Auditory Arrays**

Daryl Fougne and René Marois

Online First Publication, August 22, 2011. doi: 10.1037/a0024834

### CITATION

Fougne, D., & Marois, R. (2011, August 22). What Limits Working Memory Capacity? Evidence for Modality-Specific Sources to the Simultaneous Storage of Visual and Auditory Arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/a0024834

# What Limits Working Memory Capacity? Evidence for Modality-Specific Sources to the Simultaneous Storage of Visual and Auditory Arrays

Daryl Fougnie  
Harvard University

René Marois  
Vanderbilt University

There is considerable debate on whether working memory (WM) storage is mediated by distinct subsystems for auditory and visual stimuli (Baddeley, 1986) or whether it is constrained by a single, central capacity-limited system (Cowan, 2006). Recent studies have addressed this issue by measuring the dual-task cost during the concurrent storage of auditory and visual arrays (e.g., Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Fougnie & Marois, 2006; Saults & Cowan, 2007). However, studies have yielded widely different dual-task costs, which have been taken to support both modality-specific and central capacity-limit accounts of WM storage. Here, we demonstrate that the controversies regarding such costs mostly stem from how these costs are measured. Measures that compare combined dual-task capacity with the higher single-task capacity support a single, central WM store when there is a large disparity between the single-task capacities (Experiment 1) but not when the single-task capacities are well equated (Experiment 2). In contrast, measures of the dual-task cost that normalize for differences in single-task capacity reveal evidence for modality-specific stores, regardless of single-task performance. Moreover, these normalized measures indicate that dual-task cost is much smaller if the tasks do not involve maintaining bound feature representations in WM (Experiment 3). Taken together, these experiments not only resolve a discrepancy in the field and clarify how to assess the dual-task cost but also indicate that WM capacity can be constrained both by modality-specific and modality-independent sources of information processing.

*Keywords:* working memory, capacity limits, dual task

Working memory (WM) refers to the temporary storage and manipulation of information necessary for cognition. The ability to keep representations in an active and accessible state is critical for adaptive, intelligent behavior and is thought to underlie such diverse cognitive processes as language learning and problem solving (Baddeley, 1986). This ability, however, appears to be strikingly limited, as one can only store a small amount of information in WM at any one time (Cowan, 2001, 2006; Grimes, 1996; Irwin, 1992; Irwin & Andrews, 1996; Miller, 1956; Pashler, 1988; Rensink, 2000). By understanding the nature of this capacity limit, it is believed that researchers can learn how our WM system is organized (Baddeley, 1986; Baddeley & Hitch, 1974; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Cowan, 1988, 1995; Engle, Tuholski, Laughlin, & Conway, 1999; Fougnie & Marois, 2006; Just & Carpenter, 1992; Morey & Cowan, 2004, 2005). One critical, enduring question in regards to the capacity of

working memory, is whether it is mediated by subsystems specialized for maintaining specific types of representations (Baddeley & Logie, 1999) or whether it is limited by a single, domain-general store (Cowan, 1995).

According to the multicomponent model, WM contains several modality-specific subsystems capable of keeping task-relevant information in an active state (Baddeley, 1986; Baddeley & Logie, 1999; Baddeley & Hitch, 1974). Early instantiations of this model argued for phonological and visuospatial slave systems (e.g., Baddeley, 1986), but there is also strong support for distinct slave systems for spatial and nonspatial visual information (Logie, 1995; Logie & Marchetti, 1991). A main source of evidence for the model comes from dual-task (DT) paradigms that show *intermodal savings* or significantly reduced DT cost when two tasks draw on different modalities, compared with when the tasks originate from the same modality. In selective interference paradigms, participants have to perform a secondary task while concurrently maintaining information in WM. When the task-relevant information for the secondary task differs in modality from the contents of WM, DT cost is significantly less than if the two tasks tap the same modality (Baddeley & Hitch, 1974; Logie & Marchetti, 1991; Logie, Zucco, & Baddeley, 1990). Similarly, DT studies that require the concurrent maintenance of two stimulus sets in WM often reveal intermodal savings: When the stimulus sets differ in modality, little, if any, DT cost is typically observed (Cocchini et al., 2002; Fougnie & Marois, 2006; Luck & Vogel, 1997). In contrast, if two WM sets of the same modality have to be concurrently maintained, strong DT cost is observed (Fougnie & Marois, 2006).

---

Daryl Fougnie, Department of Psychology, Harvard University; René Marois, Department of Psychology, Vanderbilt Brain Institute, Vanderbilt University.

This work was supported by National Institute of Mental Health Grant MH70776 to René Marois and National Eye Institute Grant P30-EY008126 to the Vanderbilt Vision Research Center. We thank Doug Godwin for help with data collection.

Correspondence concerning this article should be addressed to Daryl Fougnie, Department of Psychology, Harvard University, 33 Kirkland Street, Harvard University, Cambridge, MA 02138. E-mail: darylougnie@gmail.com

In contrast to the multiple-component model, the embedded processes model suggests that a central, limited-capacity storage system underlies WM capacity (Cowan, 1988, 1995). According to this view, auditory and visual arrays should compete for limited WM storage capacity, and therefore, intermodal savings are not expected. Indeed, Saults and Cowan (2007) have suggested that evidence of intermodal savings in WM maintenance may be due to the contribution of modality-independent sensory memory to WM. Sensory memory, which is distinct from WM, refers to the temporary persistence of sensory information after a stimulus has ceased and is characterized as having an extremely large capacity but a duration that is too brief (visual sensory memory; 200–300 ms; auditory sensory memory; 1–2 s) to assist performance in typical WM tasks (Averbach & Coriell, 1961; Broadbent, 1958; Crowder, 1982; Crowder & Morton, 1969; Rostron, 1974; Sperling, 1960). However, Cowan (1988, 1995) has argued that in addition to this initial transient phase with unlimited capacity, sensory memory also has a capacity-limited phase lasting several seconds that would be able to aid performance in typical WM tasks.

Cowan's model predicts that intermodal savings in WM will only occur under conditions in which WM performance can be assisted by sensory memory. Consistent with this view, Saults and Cowan (2007) have recently argued that such savings disappear when sensory memory is disrupted with pattern masks. In their study, participants had to concurrently maintain auditory and visual WM loads. WM capacity was assessed by computing Cowan's  $K$  (an estimate of the number of object representations stored in WM; Pashler, 1988; Cowan, 2001; Cowan, Johnson, & Saults, 2005) from participant's accuracy at detecting whether a test probe was identical to or differed in one item from the WM sample (Luck & Vogel, 1997; Pashler, 1988). To test for intermodal savings, the combined auditory and visual WM capacity in the DT condition (3.49 items) was compared with the two single-task (ST) conditions—ST visual (3.62 items) and ST auditory (1.40 items).<sup>1</sup> Because the combined DT capacity was not different from the higher (vision) ST capacity, Saults and Cowan argued that audition and vision share a common WM limit. This reasoning follows from the theoretical assumption that WM is limited only by the maximum number of discrete object representations (approximately four) that can be maintained irrespective of modalities. According to this model, the higher ST capacity of the two tasks is the best approximation of a participant's capacity, whereas the lower ST capacity for the other task may stem from limitations during perception (Saults & Cowan, 2007). Thus, this maximum capacity analysis draws on the assumptions of theoretical models that propose that WM capacity is limited solely by the discrete representational quantity of items and that incorrect responses must reflect failures to store items (Cowan, 2001, 2006; Luck & Vogel, 1997).

However, discrete representation of stored information is not an assumption that is shared by all models of WM. For example, resource-based or signal-detection models of WM (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wilken & Ma, 2004) can account for limitations in the quality of WM representation by assuming that resources can be flexibly allocated to each object representation. Indeed, recent evidence suggests that errors in WM arise not just from failures to store items but also from limitations in the quality of stored representations (Bays, Catalao, & Husain,

2009; Fougnie, Asplund, & Marois, 2010; Jiang, Shim, & Makovski, 2008; Scott-Brown, Baker, & Orbach, 2000; Wilken & Ma, 2004). Limited-resource models can explain the lower performance in the ST auditory (1.40  $K$ ) condition relative to the ST visual (3.62  $K$ ) condition of Saults and Cowan's (2007) experiment without having to invoke perceptual limitations. For example, given that the  $K$  estimate for a given object is inversely correlated with its information load or complexity (Alvarez & Cavanagh, 2004), it is certainly possible that the low  $K$  estimate for the ST auditory condition reflects the greater amount of information stored per auditory item than per visual item, as fewer auditory items could be stored in a limited-capacity WM store. If so, then differences in ST capacities cannot be ignored when assessing DT capacity because both ST capacities will contribute to it. Consider a central WM store whose capacity is evenly divided across tasks; its DT capacity can be predicted by simply summing the halves of each of the ST capacities. Assuming the  $K$  per item of Saults and Cowan's auditory and visual tasks are, say, 0.5 and 1.0, respectively, then the total DT capacity of that central WM store would be 2.51  $K$  ( $0.7 [1.40/2]$  for the auditory WM task plus  $1.81 [3.62/2]$  for the visual WM task). In this particular case, Saults and Cowan's reported DT capacity of 3.49  $K$  would, in fact, suggest modality-specific contributions to WM capacity. Thus, when ST capacities are disparate, DT cost would be more appropriately estimated by comparing combined DT capacity with the summed halves of ST capacities ( $[\text{Task 1 ST capacity}/2] + [\text{Task 2 ST capacity}/2]$ ), a measure that is mathematically equivalent to the averaged ST capacity. Importantly, this equation is valid not only for resources-based models of WM but for all circumstances for which there is a disparity in measures of ST capacity, regardless of whether these disparities arise from differences in the information load of stimuli (Alvarez & Cavanagh, 2004), in the quality of the stored representations (Bays et al., 2009; Fougnie et al., 2010; Wilken & Ma, 2004; Zhang & Luck, 2008), in the rate of postretention errors (Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009), or in the number of discrete representations necessary to store each stimulus type (Luck, 2008).

The drawback of the average ST capacity method is that it works only if there is an equal division of resources across tasks. However, participants may fail to do so, even with their best intentions and explicit experimenter instructions. An alternate measure that is resilient with regard to differences in how participants allocate resources to tasks is to normalize the DT cost for each task to that task's ST capacity (Fougnie & Marois, 2006; see also Logie, Cocchini, Della Sala, & Baddeley, 2004; Salthouse, Fristoe, Lineweaver, & Coon, 1995; Salthouse, Rogan, & Prill, 1984, for analogous normalizing approaches applied to the cognitive aging and neuropsychology literature). This estimates the percentage change in capacity between ST and DT conditions for

<sup>1</sup> These are capacity estimates from Experiment 3 of Saults and Cowan (2007), averaged across the two visual WM loads. Similar results were found in other studies.

each task. The result can then be averaged across tasks to estimate an aggregate measure of DT cost,<sup>2</sup> termed  $\Delta K$ .

$$\Delta K = \left\{ \left( \frac{\text{Task 1}_{ST \text{ capacity}} - \text{Task 1}_{DT \text{ capacity}}}{\text{Task 1}_{ST \text{ capacity}}} + \frac{\text{Task 2}_{ST \text{ capacity}} - \text{Task 2}_{DT \text{ capacity}}}{\text{Task 2}_{ST \text{ capacity}}} \right) / 2 \right\} \times 100 \quad (1)$$

If both tasks tap into the same limited-capacity system,  $\Delta K$  would be 50%, or half of the combined ST capacity, *regardless of how capacity is allocated across tasks*. If both tasks tap into completely separate WM stores,  $\Delta K$  should instead be zero because capacity should not change. A downside to the  $\Delta K$  measure is that unless variance in performance estimates scale with performance, variance will have a greater (but unbiased) effect on  $\Delta K$  measures as performance nears the chance level (i.e., WM capacity is zero). This occurs because  $\Delta K$  variance is scaled by dividing by ST capacity (see Equation 1). Thus,  $\Delta K$  is not appropriate for tasks in which participants may have very low ST capacity. Here, we term the average ST capacity and  $\Delta K$  computations *normalized measures* because they normalize for differences in each task's ST capacity to estimate the expected DT capacity. Although both metrics have inherent limitations, when the normalized measures agree, they provide converging evidence for estimates of DT cost.

A critical difference between these normalized measures and the maximum capacity method used by Saults and Cowan (2007) concerns whether a disparity in performance between the two STs will affect the predicted DT cost. Fortunately, this is an empirical question; by manipulating the relative difference in capacity across auditory and visual WM tasks, we can determine the consistency of each measure. In the following experiments, we aimed to compare the two types of analytical methods using a DT WM paradigm similar to that of Saults and Cowan (2007), in which participants had to maintain just one set of auditory stimuli (ST auditory), just one set of visual stimuli (ST visual), or both sets (DT). To eliminate contributions from sensory memory, we presented pattern masks similar to those used by Saults and Cowan (2007) during the WM retention interval. In Experiments 1 and 2, we demonstrate that DT cost estimates from the maximum capacity method (as used by Saults & Cowan, 2007) are affected by the disparity in ST capacity, whereas cost estimates from the normalized measures are not. Such results help validate the theoretical assumptions of normalized measures and suggest that the maximum capacity method may, at least in some instances, overestimate DT cost.

In addition to comparing methods for measuring WM capacity, in the present study, we aimed at using these methods to assess whether DT performance is limited by a single, amodal WM store, as suggested by Cowan and colleagues (Saults & Cowan, 2007), or by two independent and modality-specific stores. The results of the first two experiments indicate that auditory and visual WM do not tap into a single limited-capacity process, nor do they indicate that they tap into entirely separate capacity-limited processes. Instead, the results reinforce our previous claim that WM maintenance is assisted by both central and modality-specific processes (Fougnie & Marois, 2006). Moreover, the results of Experiment 3 clarify one source of central capacity limitation, and that is the recruitment of common attentional processes to maintain bound featural

information in WM (Fougnie & Marois, 2009; Wheeler & Treisman, 2002).

## Experiment 1

Saults and Cowan (2007) had participants concurrently perform auditory and visual WM tasks and found that combined DT capacity was not greater than the higher ST capacity (visual task). Our goal in Experiment 1 was to replicate their findings, including the disparity in ST visual and auditory performance. We predicted that the maximum capacity method would show no intermodal savings. The key question is whether measures that normalize costs to each task's ST capacity would also fail to show intermodal savings. If so, this would provide converging evidence for Saults and Cowan's claim that WM capacity is limited by a single, amodal store. If, on the other hand, normalized measures reveal intermodal savings, this would suggest that the maximum capacity method overestimates DT cost when there is a disparity in ST performance.

In the visual WM task, participants had to remember the color and location of four squares briefly presented on a computer screen. With such a task, participants are generally able to store three to four items (Luck & Vogel, 1997; Todd & Marois, 2004; Vogel, Woodman, & Luck, 2001). For the auditory WM task, participants were presented with four digits, each spoken in a different voice (two distinct male voices, two distinct female voices) and asked to remember the numerical value of and speaker identity for each digit. In a pilot study with eight participants, we found that this task had a capacity of about two items, similar to the ST auditory WM task used by Saults and Cowan (2007).

## Method

**Participants.** Twelve young adults (4 male, 8 female) between the ages of 18 years and 25 years (mean age 20.4) participated for course credit or monetary reward.

**Stimuli.** Colors for the visual WM task were blue, green, red, or yellow, without replacement. The squares subtended 1.4°, and they were presented at horizontal and vertical axes positions that were 2.9° from fixation (Figure 1A). Visual pattern masks (1.4°) were formed by presenting a multicolored square (with four colored stripes, randomly assigned from the visual WM color set) at the four stimulus locations. Auditory WM stimuli were drawn from the digits 0–9, without replacement. Each digit was randomly assigned a distinct voice (from a set of two male and two female voices), with each audio file lasting 300 ms. Masks for the auditory WM task were formed by layering the 40 audio files. The mask sound was 300 ms in duration and was presented four times during the 1,200 ms mask interval.

**Procedure.** Prior to the main experiment, participants completed a practice session of 32 ST and DT trials that served to familiarize them with the tasks. In the main experiment, participants performed one ST auditory, one ST visual, and two DT blocks, with 40 trials in each block. Block order was counterbalanced across participants with the restriction that DT blocks were

<sup>2</sup> In Fougnie and Marois (2006), we did not average  $\Delta K$  across tasks. We chose to average across tasks here so that it would be a measure of the average capacity decrease between ST and DT conditions.

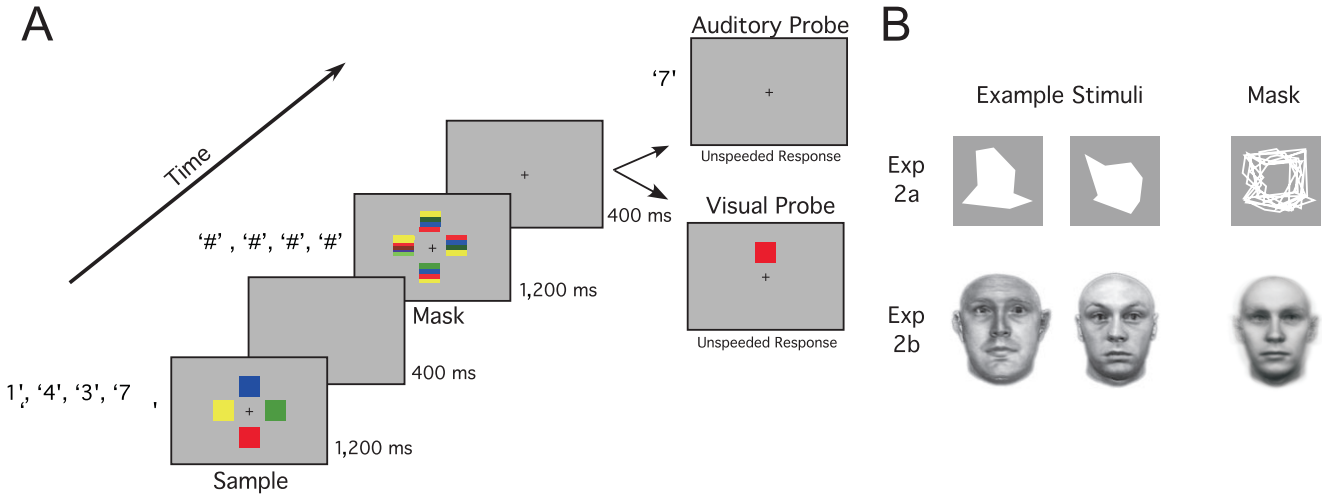


Figure 1. A: Trial timeline for Experiment 1; see text for details. B: Example stimuli and sensory masks used in Experiment (Exp) 2A and Experiment 2B. The # symbols represent the presentation of auditory masks.

performed consecutively. In ST trials, both the auditory and visual WM samples and masks were still presented to minimize perceptual differences between ST and DT conditions. However, the participants were instructed to encode only the task-relevant modality, and only that modality was probed. In DT trials, participants were tested on either the auditory or the visual WM sample. Because the tested modality was assigned on a per trial basis on DT trials, participants had no way of knowing which modality would be tested and, therefore, were required to maintain both samples until the probe appeared.

A trial began with the presentation of a fixation dot located in the center of a gray background presented on the computer screen. Participants were instructed to keep their gaze centered on the fixation for the duration of a trial. Presentation of the WM stimuli began 1,000 ms after fixation onset. Participants heard four digits over headphones spoken at a rate of 300 ms/item. Concurrent with digit presentation, four colored squares were presented, and they remained onscreen for 1,200 ms. A retention interval lasting 2,000 ms followed the WM sample presentation. Auditory and visual masks were presented 400 ms into the retention interval, with a 1,200 ms duration. Following an additional 400 ms of retention, a single-item change detection probe was presented. The visual WM probe was a colored square that was either in the same location as in the sample display or in one of the three other locations. The auditory WM probe (300 ms) was a digit from the sample spoken either in the same voice or in one of the three other voices. Participants indicated whether the probe item matched the sample (50% probability) by pressing one of two keys with their right hand. Accuracy was stressed, and participants were under no time pressure to respond. The visual

WM probe remained onscreen until a response was recorded. A 200 ms intertrial interval period separated trials.

Our experimental design was similar to that of Saults and Cowan's (2007), with the following modifications. In Saults and Cowan's study, the auditory stimuli were presented simultaneously from four different speakers. In the current study, we opted for sequential presentation of the auditory stimuli during visual WM array presentation. This should minimize perceptual confusability of the digits during encoding and preclude common spatial representation across the visual and auditory tasks from being a confounding source of DT interference while still allowing a concurrent presentation of auditory and visual information. Saults and Cowan's experiments manipulated the set size of the visual WM array (four or eight items). Because this set size manipulation did not impact their DT cost, the current study only used a set size of four. Also, the current study used a single-item probe in contrast to the whole probe display used by Saults and Cowan. The single-item probe minimized decision error and allowed us to probe participants on their storage of color–location pairings. On trials that required a different response, the response probe was an item from the sample presented in the wrong location or spoken in the wrong voice. Finally, in DT trials, we only probed participants on one of the two tasks to ensure that any DT cost was not due to additional retrieval cost (Cowan & Morey, 2007).

## Results

Task capacity ( $K$ ; Figure 2A, left) was calculated from accuracy data with Cowan's (2001) modification of Pashler's (1988) for-

Figure 2 (opposite). Left column of A–E: Working memory (WM) capacity for auditory WM (AWM, red) and visual WM (VWM, blue) tasks as a function of single-task or dual-task (DT) conditions for Experiments 1, 2A, 2B, 3A, and 3B, respectively). The combined (Comb; purple) DT capacity was calculated by summing each participant's auditory and visual DT WM capacity. Right column of A–E: Normalized DT costs ( $\Delta K$ ) for the AWM (red) task and the VWM (blue) task and the average of these two costs (purple). A normalized cost of 50% would represent no intermodal savings, whereas 0% indicates no interference across modalities. Note that the normalized cost is inappropriate for Experiment 3B because single-task AWM was near chance level. Error bars represent the standard error of the mean.

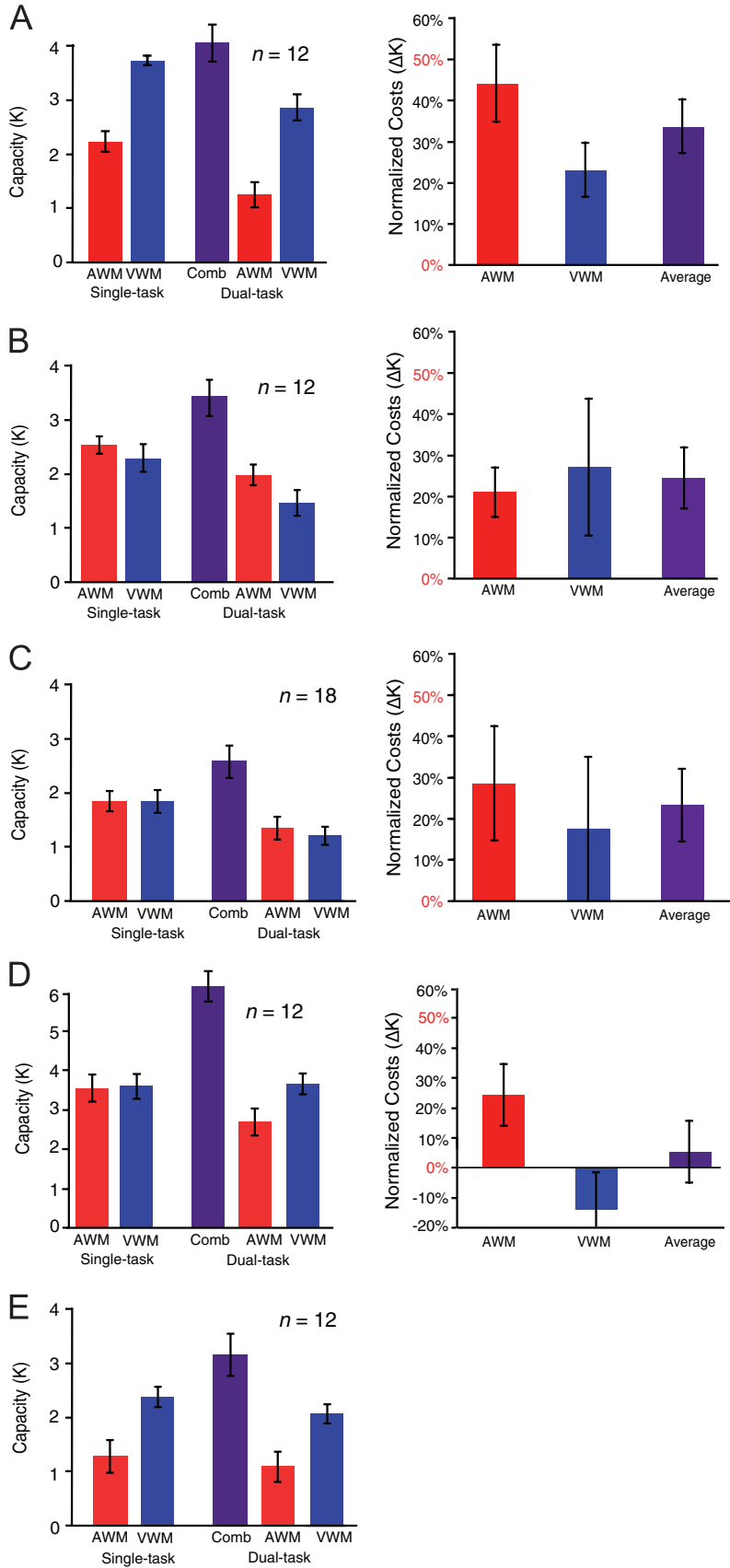


Figure 2 (opposite).

mula:  $K = (\text{hit rate} + \text{correct rejection rate} - 1) * \text{set size}$ . Only analyses on  $K$  are reported here to be consistent with Sauls and Cowan (2007), although analyses of variance (ANOVAs) were quantitatively the same whether  $K$  or change detection accuracy was the dependent measure because  $K$  is a linear transform of accuracy when there are equivalent numbers of same and different trials and set size is not varied (Morey & Cowan, 2005). Capacity data were entered in a within-subjects ANOVA with the factors of task modality (visual or auditory) and task condition (ST or DT). There was a main effect of modality,  $F(1, 11) = 55.37, p < .001$ , with visual WM having a higher capacity than auditory WM, and a main effect of task condition, with higher capacity in ST trials than in DT trials,  $F(1, 11) = 22.31, p < .001$ . There was no interaction between modality and task condition,  $F(1, 11) = 0.29, p = .6$ .

These results replicate two major aspects of Sauls and Cowan (2007). First, in their study, the capacity for visual WM was nearly twice that of auditory WM capacity. This was also the case in the present study, as a paired  $t$  test found that ST visual WM capacity (3.73) was significantly higher than ST auditory WM capacity (2.23),  $t(11) = 8.03, p < .001$ .<sup>3</sup> Second, the ANOVA results showed clear evidence of DT cost. We then used several methods to measure the amount of DT cost and determine whether it was indicative of a single, shared capacity across modalities.

**Maximum capacity method.** The maximum capacity method tests whether combined auditory and visual DT capacity is greater than the ST with the higher capacity. This was not the case. A paired  $t$  test found that combined DT capacity (4.11) was not greater than ST visual capacity (3.73),  $t(11) = 1, p = .33$ , thereby providing evidence against intermodal savings, in accordance with Sauls and Cowan's (2007) study.

**Normalized measures.** The average capacity comparison revealed that DT capacity (4.11) was greater than average ST capacity (2.98),  $t(11) = 3.11, p < .001$ , which is suggestive of significant intermodal savings. Similar results were found with the  $\Delta K$  computation, which measures the average percentage decrease in each task's DT capacity relative to its ST capacity (see Equation 1). With this measure, if two tasks tap into the same capacity-limited process, a  $\Delta K$  of 50% is expected (see Fougne & Marois, 2006, Experiment 2),<sup>4</sup> whereas if two tasks share no capacity, the  $\Delta K$  should be 0. We found a  $\Delta K$  of 34% (Figure 2A, right), which is significantly lower than 50%,  $t(11) = 2.37, p = .04$ , thus revealing significant intermodal savings. However,  $\Delta K$  was also significantly greater than zero ( $p < .001$ ), suggesting sizable competition for storage between the two WM arrays.

## Discussion

Significant DT cost was observed in Experiment 1 when participants had to concurrently perform an auditory and a visual WM task. The maximum capacity estimate (Sauls & Cowan, 2007) of DT cost showed no evidence of intermodal savings, consistent with the predictions of a single, shared capacity across modalities (Cowan, 2001, 2006). In contrast, quantification of DT cost with normalized metrics showed that the cost was less than was predicted by a single, shared capacity. We hypothesize that this discrepancy results from the maximum capacity method overestimating DT interference when there is a large disparity between the ST capacities of the two tasks. This hypothesis predicts that

making the ST capacities across tasks equal will affect the maximum capacity method but not the normalized metrics, such that all measures will now show evidence of intermodal savings.

To test this prediction, in Experiment 2, we paired the auditory WM task of Experiment 1 with visual WM tasks that had a ST capacity similar to the auditory WM task. Visual WM tasks that require storage of complex stimuli show reduced change detection performance and lower estimates of capacity (Alvarez & Cavanagh, 2004; Todd, Han, Harrison, & Marios, 2011). Although the explanation for these lower capacity estimates is a matter of debate (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Barton, Ester, & Awh, 2009; Eng, Chen, & Jiang, 2005; Jiang, Shim, & Makovski, 2008; Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010; Scolar, Vogel, & Awh, 2008), there is clear evidence that increasing the complexity of representations reduces visual WM accuracy and, therefore, measured  $K$  values. For example, Alvarez and Cavanagh (2004) found that WM capacity for colored squares was 4.4, whereas capacity for complex polygons was 2.0. In addition, there is evidence that WM capacity for faces is around two items (Curby & Gauthier, 2007; Eng et al., 2005; Todd et al., 2011). Hence, Experiments 2A and 2B paired auditory WM for digits with visual WM for polygons and faces, respectively, as the WM capacity for the latter two object categories appear to be similar to that of our auditory WM task (see Figure 2A).

## Experiment 2A

### Method

A separate set of 12 young adults (six male, six female) between the ages of 18 years and 23 years (mean age 20.3) participated for course credit or monetary reward. The color WM task was replaced with a visual WM task that required participants to memorize the shape of complex polygons. A set of 10 eight-sided polygons were randomly generated, such that their spatial extent did not exceed a  $1.6^\circ \times 1.6^\circ$  area. Polygons had a solid white color and were presented against a gray background. None of these polygons resembled any familiar shapes. Every trial, four random polygons were assigned to one of the four visual WM locations (see Experiment 1) without replacement. Participants were instructed to remember the pairing of shape and location. If the visual array was tested, the single-item probe was a polygon from the sample array presented at the correct location or one of the other possible locations (each outcome was equally likely). Participants made an unspeeded response to indicate whether the shape and location matched. Visual masks were constructed by layering the outline of the 10 polygons (Figure 1B) and presenting this stimulus at all four visual WM locations for the 1,200 ms mask duration. In all other respects, this experiment was the same as Experiment 1.

<sup>3</sup> All  $t$  tests were two tailed.

<sup>4</sup> By comparison, when Equation 1 was applied to Fougne and Marois' (2006) Experiment 2, which required storage of two visual WM tasks, a  $\Delta K$  value of 51% was observed, consistent with a complete division of resources between the two VWM tasks.

## Results and Discussion

Capacity data (Figure 2B, left) were entered into a within-subjects ANOVA with the factors of modality (visual or auditory) and task condition (ST or DT). The effect of modality was marginal,  $F(1, 11) = 3.31, p = .1$ , with a trend for higher capacity for the auditory stimuli than for the visual stimuli. This marginal effect appears to be influenced by the data in the DT condition because there is no difference in capacity between ST auditory (2.45) and ST visual (2.22) WM capacity (paired  $t$  test,  $p = .34$ ), suggesting that unlike Experiment 1, ST capacity is equated in the current study. The ANOVA also revealed a higher capacity for ST trials than DT trials,  $F(1, 11) = 18.76, p = .001$ . The interaction between modality and task condition was not significant,  $F(1, 11) = 0.55, p = .47$ .

In contrast to the results of Experiment 1, the maximum capacity method provided evidence for intermodal savings. Paired  $t$  tests indicate that combined DT capacity (3.33) was greater than both auditory ( $p = .005$ ) and visual ST capacity ( $p = .001$ ). Intermodal savings were also found when DT cost was quantified with normalized measures. Combined DT capacity (3.33) was greater than average ST capacity (2.35),  $t(11) = 4.11, p = .002$ . In addition,  $\Delta K$  (24%; Figure 2B, right) was significantly lower than 50%,  $t(11) = 3.16, p = .001$  (and significantly above zero)  $t(11) = 2.94, p = .013$ . Thus, as predicted, the maximum capacity method no longer shows evidence for a single, shared capacity across modalities when ST WM capacity was equated. Instead, it converges with the normalized measures in showing evidence of intermodal savings.

### Experiment 2B

Experiment 2A differed from Experiment 1 in two ways: The capacities for the two STs were matched, and the visual WM stimulus set involved complex polygons instead of colored squares. An additional experiment was conducted to show that the findings of Experiment 2A generalize to a different stimulus set. Experiment 2B has face stimuli for the WM task because this stimulus set has also been shown to have a capacity of around two items (Curby & Gauthier, 2007; Eng et al., 2005).

### Method

A separate set of 18 young adults (eight male, 10 female) between the ages of 18 years and 28 years (mean age 20.6) participated for course credit or monetary reward. Except for the stimuli, the design of Experiment 2B was identical to that of Experiment 2A. The visual WM stimuli were 10 male faces obtained from the Max-Planck face database (Troje & Bühlhoff, 1996). These images were presented in grayscale, against a white background and subtended  $1.6^\circ \times 3.2^\circ$  degrees of visual angle. Four random faces were selected per trial, without replacement, to occupy one of the four stimulus locations. A mask was formed by averaging the luminance values of each pixel across the 10 face stimuli (Figure 1B). This mask was presented at each of the four stimulus positions for the duration of the 1,200 ms mask interval.

## Results and Discussion

Capacity data (Figure 2C, left) were entered in a within-subjects ANOVA with the factors of modality (visual or auditory) and task

condition (ST or DT). There was no main effect of modality,  $F(1, 17) = 0.23, p = .64$ : Both ST auditory and visual WM capacity were 1.8 items. There was a main effect of task condition revealing that ST trials had a higher capacity than did DT trials,  $F(1, 17) = 21.07, p < .001$ . The interaction between modality and ST or DT condition was not significant,  $F(1, 17) = 0.16, p = .69$ .

The maximum capacity method provided evidence for intermodal savings: Paired  $t$  tests indicate that combined DT capacity (2.51) was greater than both ST auditory and visual WM capacity ( $ps < .01$ ). Similarly, the average capacity comparison revealed that DT capacity was also greater than the average ST capacity (2.33),  $t(11) = 4.1, p = .004$ . When DT cost was quantified with the  $\Delta K$  method, there was also evidence for intermodal savings: The  $\Delta K$  (24%; Figure 2C, right) was significantly lower than 50%,  $t(17) = 2.69, p = .01$ . These results replicate the findings of Experiment 2A with a distinct stimulus set and strongly suggest that the differences in results between Experiment 1 and Experiment 2 are due to the change in ratio of ST capacities rather than stimulus-specific factors.

Together with Experiments 1 and 2A, the results of this experiment point to two principal conclusions. First, normalized metrics are a more reliable measure of DT cost than is the maximum capacity method because they are not affected by unequal ST capacities. Differences in ST capacity can bias DT metrics that assume equivalency of storage units across tasks, and such differences may account for Saults and Cowan's (2007) finding of no intermodal savings. The second conclusion is that intermodal savings can be reliably detected, regardless of which analysis metric is used, a finding that suggests that auditory and visual WM tasks draw on at least partially dissociable storage systems.

Although the DT cost reported in the present study is significantly less than that predicted by a single, shared capacity limit across modalities, the cost is still larger than the DT cost reported in similar studies (e.g., Cocchini et al., 2002; Fougne & Marois, 2006; Morey & Cowan, 2004). What could account for the relatively large DT cost in the present study? In Experiment 3, we investigated the possibility that such cost could originate from the requirements for the maintenance of bound representations of task-relevant features into integrated objects. In Experiments 1 and 2, participants were required to remember color–location and digit–voice pairings. Similarly, in Saults and Cowan's (2007) experiments, colors and digits could appear more than once in a WM sample, and therefore, participants would be encouraged to maintain color–location and digit–voice bindings in order to identify all potential changes (Vogel et al., 2001). Given that attention has been considered a central, capacity-limited process (Cowan, 1995) and that there is evidence that it is involved in maintaining feature bindings in WM (Brown & Brockmole, 2010; Fougne & Marois, 2009), it is possible that at least some of the DT cost in the present study resulted from limited central resources required for the maintenance of bound representations in WM (Depoorter & Vandierendonck, 2009; Oberauer & Lange, 2009; Wheeler & Treisman, 2002). That is to say, we hypothesized that the concurrent storage of modality-independent stimuli in largely independent WM systems may take place with little or no interference as long as the tasks do not engage additional common capacity-limited processes (Dutta, Schweickert, Choi, & Proctor, 1995; Navon, 1984; Navon & Miller, 1987).



To test this possibility, we measured in Experiment 3 the amount of DT cost when the WM tasks did not involve feature binding of stored representations. This experiment combined a visual WM task for the identity of colored squares presented sequentially at fixation with an auditory WM task for the identity of auditory tones presented sequentially over headphones. The WM stimuli differed from each other only in the task-relevant feature, thereby eliminating any potential benefit for storing bound representations. We predicted that there would be significantly less DT cost in this experiment than in Experiment 1.

### Experiment 3A

#### Method

A separate set of 12 young adults (four male, eight female) between the ages of 19 years and 24 years (mean age 20.8) participated for course credit or monetary reward in Experiment 3A.

A pilot study on eight participants showed that ST capacity for the auditory and visual WM tasks was greater than four but was no higher than six. Therefore, the set size for each task was increased from four in Experiment 1 to six in Experiment 3A. The visual WM stimuli were six colored squares ( $1.4^\circ$ ) presented sequentially at fixation for 300 ms/item. Sample display colors were randomly selected from blue, orange, purple, brown, dark green, black, white, yellow, light blue, pink, light green, or red without replacement. To prevent sensory memory from presentation of the sample items to carry over during working memory retention and affect WM performance, we followed the sample colors at fixation with a multicolored pattern mask resembling that of Experiment 1, except that it contained 12 distinctly colored stripes. As in Experiment 1, the mask was presented 400 ms into the retention interval, with a 1,200 ms duration. Participants indicated whether the single-item visual WM probe was the same color as one of the sample items (50% likelihood). The auditory WM task consisted of a series of six tones (300 ms) presented sequentially over headphones. The tones were selected, without replacement, from a set of 12 possible tones with frequencies varying from 220–1,100 Hz, in 80 Hz steps. A tone mask was formed by layering all tone stimuli and was presented for the entire 1,200 ms mask interval. Participants indicated whether a single-item probe tone was the same frequency as one of the sample items (50% likelihood). Note that the auditory and visual samples were presented concurrently, such that a colored square was on screen for the duration of a single tone stimulus. In other respects, this study was the same as Experiment 1.

#### Results and Discussion

Capacity data (Figure 2D, left) were entered in a within-subjects ANOVA with the factors of modality (visual or auditory) and task condition (ST or DT). There was a marginal effect of modality,  $F(1, 11) = 3.84, p = .07$ . However, this effect appears to be influenced by performance in DT conditions. Indeed, there is no difference between ST auditory (3.53) and ST visual (3.60) WM capacity (paired  $t$  test,  $p = .89$ ). Therefore, the auditory and visual WM tasks were matched for capacity. Most important, and unlike Experiments 1 and 2, there was no longer evidence of DT cost because the ANOVA revealed that ST trials did not have a higher

capacity than did DT trials,  $F(1, 11) = 2.6, p = .13$ . There was a significant interaction between modality and ST or DT condition,  $F(1, 11) = 4.8, p = .05$ , driven by the fact that although auditory and visual ST performance was equivalent, there was a difference in auditory (2.4) and visual (3.8) DT performance.

The maximum capacity method provided evidence for intermodal savings: Paired  $t$  tests indicate that combined DT capacity (6.18) was greater than both auditory and visual ST capacity ( $ps < .005$ ). Correspondingly, combined DT capacity was also larger than the average ST capacity ( $p < .001$ ). Intermodal savings were also found when DT cost was quantified with the  $\Delta K$  measure: The  $\Delta K$  (5%; Figure 2D, right) was lower than 50%,  $t(11) = 5.05, p < .001$ . Indeed, because  $\Delta K$  was not significantly above zero,  $t(11) = 0.49, p = .64$ , and the ANOVA results found no difference between ST capacity and DT capacity, there is no evidence of DT cost in the present study. However, this latter conclusion is based on a null finding that trended in the direction of showing DT cost, and it is therefore possible that they could have reached significance with a larger sample size. Nevertheless, we can affirm that there is less DT cost in Experiment 3 than in Experiment 1: A between-subjects ANOVA with the factors of experiment (Experiment 1 vs. Experiment 3A), task condition (ST or DT), and modality (visual or auditory) found that there was a greater effect of task condition in Experiment 1 than in Experiment 3A,  $F(1, 11) = 6.36, p = .02$ . Also, an independent samples  $t$  test on  $\Delta K$  values across studies found that  $\Delta K$  was lower in the current study,  $t(22) = 2.63, p = .01$ . Because the main distinction between these two experiments was the requirement to form and maintain integrated representations, this result suggests that interference between auditory and visual arrays in Experiment 1 may have been largely due to the costs associated with the maintenance of feature bindings (Depoorter & Vandierendonck, 2009; Fougny & Marois, 2009; Oberauer & Lange, 2009; Wheeler & Treisman, 2002). Thus, caution is necessary in the use of a DT paradigm to assess interference between auditory and visual WM loads: Cross-modal storage costs may be overestimated if there is processing overlap between the tasks.

Unlike in Experiment 1, in which all the color stimuli were presented simultaneously at different locations, Experiment 3A presented the color stimuli sequentially at fixation. Sequential presentation was adopted here to prevent participants from implicitly binding the color–location pairings. However, it is possible that the sequential presentations of colored stimuli, which coincided with the sequential presentations of the auditory tones, allowed participants to encode and store multimodal objects.<sup>5</sup> Moreover, this procedural modification also complicates the interpretation of the cause(s) of the differences in the results between Experiment 1 and Experiment 3A. To provide further evidence that it was the elimination of the need to store bindings—rather than other methodological differences—that was the source of the reduced DT cost between these two experiments, we conducted an additional study on six volunteers that paired a visual working memory task for six colors presented simultaneously along an imaginary circle around fixation, with an auditory working memory task for the identity of six sequentially spoken digits (total visual and auditory stimulus presentation duration was the same at

<sup>5</sup> We thank Nelson Cowan for this suggestion.

1,800 ms). Critically, the task did not require participants to explicitly remember feature bindings because there were no overt experimenter instructions to do so and the probe color was presented at fixation. Average DT capacity (4.44,  $SE = 0.30$ ) was equivalent to the average ST capacity (4.62,  $SE = 0.46$ ),  $t(5) = 0.5$ ,  $p = .7$ , and DT cost is significantly less than in Experiment 1,  $t(16) = 2.4$ ,  $p = .03$ , providing strong evidence that it was the removal of feature bindings that lowered the DT cost in Experiment 3A.

### Experiment 3B

In Experiment 3A, we found negligible DT cost. A potential concern with that experiment is that participants may have relied on subvocal rehearsal to minimize competition in working memory. For example, if participants were subvocally rehearsing the color labels of the visual array and not storing that information in visual WM, then this could explain such low DT cost. This potential confound could also provide an explanation for the difference in DT cost between Experiment 1 and Experiment 3A if the requirement to bind features to spatial positions in Experiment 1 made this subvocal strategy less effective. To address this issue, we conducted a control experiment in which participants were required to perform an articulatory suppression task for the entire trial duration. Because this additional task requirement made performance in the task more difficult (particularly for the auditory WM task), the sample array set size was reduced from six to five items. In all other respects, this study was the same as Experiment 3A.

### Method

A separate set of 12 young adults (three male, nine female) between the ages of 19 years and 22 years (mean age 19.3) participated for course credit or monetary reward in Experiment 3B. For the articulatory suppression task, participants were required to repeat the word “the” at a 3 Hz rate, starting 1,000 ms prior to the sample arrays and ending after a response was collected. Participants’ verbalizations were monitored remotely by an experimenter to confirm that they performed the articulatory suppression task. This articulatory suppression task is often used to minimize verbal encoding and rehearsal of stimuli (Allen, Hitch, & Baddeley, 2009; Logie, Brockmole, & Jaswal, 2011).

### Results and Discussion

Capacity data (see Figure 2E) were entered into a within-subjects ANOVA with the factors of modality (visual or auditory) and task condition (ST or DT). Unlike in Experiment 3A, there was a main effect of modality,  $F(1, 11) = 12.83$ ,  $p = .004$ , with better performance for visual WM than for auditory WM (this was also true when comparing just the ST conditions,  $p < .05$ ). We attribute this difference to the use of the articulatory suppression task, which likely interfered with the auditory WM task. This reasoning can also explain the lack of interaction between modality and task condition,  $F(1, 11) = 0.19$ ,  $p = .67$ , in this experiment. As in Experiment 3A, however, there was no difference between ST and DT conditions,  $F(1, 11) = 1.81$ ,  $p = .20$ .

Aside from the large cost to auditory WM caused by the articulatory suppression task, the results appear remarkably similar to those of Experiment 3A. To directly compare the two studies, we entered the capacity data from both studies into a between-subjects ANOVA with factors of articulation (absent vs. present), modality (visual or auditory), and task condition (ST or DT). There was a main effect of modality,  $F(1, 11) = 17.05$ ,  $p < .001$ , with participants having lower capacity for the auditory WM task than for visual WM task. Interestingly, the main effect of task condition was significant,  $F(1, 11) = 4.34$ ,  $p = .05$ , suggesting that the lack of significant DT cost when Experiments 3A and 3B were analyzed separately was due to low sample sizes. There was also a main effect of articulation,  $F(1, 11) = 11.13$ ,  $p = .003$ , with worse performance during articulation. More important, there was no interaction between task condition and articulation,  $F(1, 11) = 0.09$ ,  $p = .76$ , indicating that the amount of DT cost did not differ across the two studies.

To test whether the current study still showed significantly less DT cost than in Experiment 1, even after minimizing the potential for subvocal rehearsal, the task capacities for Experiment 1 and Experiment 3B were entered in a between-subjects ANOVA. Critically, there was an interaction between study condition and task condition (ST versus DT),  $F(1, 22) = 6.25$ ,  $p = .02$ , revealing that even under conditions of articulatory suppression, there is still less DT cost when participants are not required to integrate multiple features into a single object.

The capacity results of Experiment 3B suggest that articulatory suppression cannot explain the negligible DT cost in Experiment 3A, compared with Experiment 1. Also, because an equivalent amount of DT cost was observed regardless of whether the experimental conditions deterred subvocal rehearsal, we can conclude that even if participants were engaging in this strategy, it was not allowing them to minimize their load in DT conditions.

Consistent with the capacity results, the maximum capacity analysis revealed greater DT capacity than the higher ST capacity ( $p = .004$ ), therefore hinting at intermodal savings. Unfortunately, the  $\Delta K$  measure could not be applied to the present study because the auditory WM performance was too low: When performance nears chance level, variance in task performance has a much greater effect on the measure. For the several participants who had ST auditory  $K$  values near zero, differences in measured DT capacity would consequently translate into very large changes in capacity. Indeed, the 95% confidence intervals for auditory  $\Delta K$  values ranged from a 155% drop in capacity to a 111% increase in capacity. These values are well outside the range of theoretically meaningful values.

### General Discussion

What is the nature of our limited storage capacity in WM? The current experiments explore whether these limits arise from a single, central capacity (Cowan, 1995, 2001, 2006; Saults & Cowan, 2007) or whether there are distinct subsystems for auditory and visual information (Baddeley, 1986; Baddeley & Hitch, 1974; Baddeley & Logie, 1999). This question was addressed with a DT approach in which participants were required to concurrently store auditory and visual arrays in WM. Past studies have yielded widely different estimates of DT cost, with some suggesting strong divergence in limitations across modalities (Cocchini et al., 2002;

Scarborough, 1972) and others instead suggesting convergence of memory systems (Saults & Cowan, 2007). The present study hints at one potential explanation for these discrepancies in findings, and that is the way such costs are measured.

Decisions on how to measure WM costs reflect the theoretical assumptions of the structure of WM limits. Drawing on an object-based account of WM limitations, a recent article (Saults & Cowan, 2007) concluded that there was a single limit in the *number* of object representations that can be stored across modalities. This conclusion was drawn from the finding that combined DT capacity was no greater than the higher (visual) of the two ST capacities because that capacity sets the upper limit on the number of items that can be stored in WM if that store is used by both modalities. Saults and Cowan (2007) argued that performance limitations for the lower task capacity (the auditory task) likely originated during perception and, therefore, was not germane to WM capacity limitations. Thus, for this maximum capacity measure of DT capacity, the only aspect of interest is the maximum number of representations that can be maintained, and visual and auditory items are assumed to be of equal weight regardless of ST performance.

Very different conclusions about WM capacity are drawn, however, when assuming a resource-based account of WM. Such an account suggests that stimuli compete for limited resources and that the task performance per unit of this commodity will depend on performance in both ST conditions. Taking Saults and Cowan's (2007) study as an example, it is possible that the lower performance for auditory WM occurred because each auditory WM item loaded more on WM's resources than each visual item did (e.g., one auditory item's resource load may have been equivalent to two visual items' load). Were this to be the case, then one cannot estimate the DT capacity of a single central WM store by ignoring the load imposed by one of the two tasks. The merit of taking into account both ST capacities is not unique to resource-based accounts of WM, however. Differences between ST capacities are also important in estimating the predicted DT capacity of object-based models if those differences arise during retention or postretention stages of memory (Awh et al., 2007; Barton et al., 2009; Luck, 2008).

Accordingly, the measures of DT capacity that we propose treat disparities in performance between the individual tasks as relevant to the final DT estimates. First, we compared combined DT capacity to the average ST capacity. If differences in ST performance occur due to differences in the amount of a resource or process expended per unit of  $K$  performance, then evidence for at least partially distinct stores would be provided by higher combined DT capacity than the average ST capacity. This is because each task's ST capacity indicates the number of items that can be stored with full WM capacity but, under the assumption of a common single store, only 50% of that capacity will be devoted to each task in DT condition, and therefore, DT capacity will not be higher than the sum of half of each ST's capacity. A down side of this measure, however, is that it assumes that participants will allocate resources to both tasks evenly under DT conditions. Because of this limitation, we also measured the percentage decrease in DT capacity relative to each task's ST capacity (i.e.,  $\Delta K$ , Fougne & Marois, 2006). This measure is more resilient to differences in capacity allocation across tasks because it normalizes changes in capacity in DT conditions relative to each task's ST

capacity, but it is inappropriate for experiments in which performance is low or near chance level. Except for Experiment 3B, in which auditory WM was too low for the  $\Delta K$  measure, both normalized measures converged on the same result.

To compare the validity of Saults and Cowan's (2007) maximum capacity measure of WM with our normalized metrics (average capacity and  $\Delta K$  methods), we manipulated the disparity between auditory and visual ST capacity between Experiment 1 and Experiment 2. The maximum capacity method pointed to a single, shared capacity for auditory and visual arrays in Experiment 1, but to at least partially dissociable capacities in Experiment 2, when ST capacities were equated across tasks. In contrast, the normalized measures found evidence for partially dissociable capacities regardless of the capacity disparity in ST conditions. Therefore, we conclude that the normalized measures are a more appropriate estimate of DT capacity than the maximum capacity metric. It is important, however, to emphasize that our conclusion that WM capacity is at least partly determined by modality-specific stores is based not just on the two normalized metrics but also on the maximum capacity metric, for all three measures reveal modality savings in Experiment 2, the experiment that had equivalent ST capacities.

We have recently argued that WM capacity is set both by task/modality-specific and central, amodal processing limitations (Fougne & Marois, 2006). The principal finding of that study was that DT cost was significantly smaller than that predicted by a single, shared capacity limit across modalities. Consistent with this conclusion, here we show clear evidence for modality-specific contributions to WM capacity.

The modality-specific source of WM capacity may be self-sustaining neural activity (Funahashi & Inoue, 2000; Hebb, 1949) in brain regions specialized for the maintenance of a specific type of information (e.g., auditory or visual; Gruber & von Cramon, 2001, 2003; Kirschen, Chen, Schraedley-Desmond, & Desmond, 2005; Rämä & Courtney, 2005; Romanski & Goldman-Rakic, 2002; Schumacher et al., 1996; Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006). Such modality-specific contributions to WM may even reside in the sensory cortical regions that process visual or auditory perceptual inputs (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009). Moreover, the negligible interference observed between the auditory and visuospatial WM displays in Experiments 3A or 3B, when there were no demands for stored bound representations, suggests that such representations may be maintained largely independently of each other and without the need for a central, active rehearsal mechanism (Washburn & Astur, 1998). This possibility is consistent with neurocomputational models that suggest that representations can be sustained by recurrent excitation in a neural network in the absence of top-down signals (Hopfield, 1982; Amit, Brunel, & Tsodyks, 1994).

In addition to modality-specific WM processes, our study also provides evidence for amodal sources of WM capacity because we observed significant DT cost in Experiments 1 and 2. What is the source of this DT cost? Experiment 3 suggests that one source of interference occurs when two WM tasks overlap in other capacity-limited processes, such as in the requirement for binding feature representations in WM. Significantly more DT cost was observed when the WM tasks required participants to encode integrated objects (Experiments 1 and 2) than when individual features were

tested (Experiment 3). These results are consistent with the finding that maintenance of integrated representations in WM requires constant attention (Brown & Brockmole, 2010; Fougne & Marois, 2009; but see Allen, Baddeley, & Hitch, 2006; Johnson, Hollingworth, & Luck, 2008), which may be of central origin. The present results also serve a cautionary note about how critical it is for DT paradigms to eliminate sources of interference ancillary to the process of interest (Cowan & Morey, 2007; Navon, 1984).

The notion that costs between auditory and visual arrays may occur due to overlap in nonmnemonic processes is consistent with the multicomponent model, which suggests that a central executive may be involved in coordinating two tasks (Baddeley, 2000; Baddeley & Logie, 1999; Cocchini et al., 2002) but that it is not directly involved in storing information in WM (Duff & Logie, 2001). To be sure, the current study cannot rule out the contribution of an amodal WM system capable of storing both auditory and visual information, as evidenced by the small DT cost in Experiment 3. However, our findings clearly suggest that an amodal WM store cannot be the sole, or even primary, factor in limiting capacity on auditory and visual WM tasks, as advocated by the embedded process model (Cowan, 1995, 2001, 2006; Sauls & Cowan, 2007). That being said, our findings do agree with the embedded process model in one respect—the involvement of attention. But, whereas the embedded process model proposes that central storage in WM is set entirely by an amodal attentional capacity, we suggest that attention's capacity-limiting contribution to cross-modal WM may largely consist in the maintenance of integrated object representations.

## References

- Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*, 298–313. doi:10.1037/0096-3445.135.2.298
- Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2009). Cross-modal binding and working memory. *Visual Cognition*, *17*, 83–102. doi:10.1080/13506280802281386
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111. doi:10.1111/j.0963-7214.2004.01502006.x
- Amit, D. J., Brunel, N., & Tsodyks, M. V. (1994). Correlations of cortical Hebbian reverberations: Theory versus experiment. *Journal of Neuroscience*, *14*, 6435–6445.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*, *40*, 309–328.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628. doi:10.1111/j.1467-9280.2007.01949.x
- Baddeley, A. D. (1986). *Working memory*. New York, NY: Oxford University Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. doi:10.1016/S1364-6613(00)01538-2
- Baddeley, A. D., & Hitch, D. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Baddeley, A. D., & Logie, R. (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). New York NY: Cambridge University Press.
- Barton, B., Ester, E. F., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1359–1367. doi:10.1037/a0015792
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 1–11.
- Bays, P. M., & Husain, M. (2008, August). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854
- Broadbent, D. E. (1958). *Perception and communication*. London, England: Pergamon Press. doi:10.1037/10037-000
- Brown, L. A., & Brockmole, J. R. (2010). The role of attention in binding visual features in working memory: Evidence from cognitive aging. *Quarterly Journal of Experimental Psychology*, *63*, 2067–2079.
- Cocchini, G., Logie, R. H., Della Sala, S. D., 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. doi:10.3758/BF03194326
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system, *104*, 163–191.
- Cowan, N. (1995). *Attention and memory*. New York, NY: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114. doi:10.1017/S0140525X01003922
- Cowan, N. (2006). *Working memory capacity*. New York, NY: Psychology Press.
- Cowan, N., Johnson, T. D., & Sauls, J. S. (2005). Capacity limits in list item recognition: Evidence from proactive interference. *Memory*, *13*, 293–299. doi:10.1080/09658210344000206
- Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science*, *18*, 686–688. doi:10.1111/j.1467-9280.2007.01960.x
- Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica*, *50*, 291–323. doi:10.1016/0001-6918(82)90044-0
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*, 365–373. doi:10.3758/BF03210660
- Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*, 620–628. doi:10.3758/BF03196811
- Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *Quarterly Journal of Experimental Psychology*, *62*, 531–549. doi:10.1080/17470210801995002
- Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology*, *54*, 31–48.
- Dutta, A., Schweickert, R., Choi, S., & Proctor, R. W. (1995). Cross-task cross talk in memory and perception. *Acta Psychologica*, *90*, 49–62. doi:10.1016/0001-6918(95)00021-L
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127–1133. doi:10.3758/BF03206454
- 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. doi:10.1037/0096-3445.128.3.309
- Fougne, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*, 1–11. doi:10.1167/10.12.27
- Fougne, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, *17*, 526–534. doi:10.1111/j.1467-9280.2006.01739.x
- Fougne, D., & Marois, R. (2009). Attentive tracking disrupts feature

- binding in visual working memory. *Visual Cognition*, *17*, 48–66. doi:10.1080/13506280802281337
- Funahashi, S., & Inoue, M. (2000). Neuronal interactions related to working memory processes in the primate prefrontal cortex revealed by cross-correlation analysis. *Cerebral Cortex*, *10*, 535–551. doi:10.1093/cercor/10.6.535
- Grimes, J. (1996). On the failure to detect changes across scenes. In K. Atkins (Ed.), *Vancouver studies in cognitive science: Vol. 5. Perception* (pp. 89–110). New York, NY: Oxford University Press.
- Gruber, O., & von Cramon, D. Y. (2001). Domain-specific distribution of working memory processes along human prefrontal and parietal cortices: A functional magnetic resonance imaging study. *Neuroscience Letters*, *297*, 29–32. doi:10.1016/S0304-3940(00)01665-7
- Gruber, O., & von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited. Evidence from 3-t fmri studies using classical domain-specific interference tasks. *NeuroImage*, *19*, 797–809. doi:10.1016/S1053-8119(03)00089-2
- Harrison, S. A., & Tong, F. (2009, April). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635. doi:10.1038/nature07832
- Hebb, D. O. (1949). *Organization of behavior: A neuropsychological theory*. New York, NY: Wiley.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2558. doi:10.1073/pnas.79.8.2554
- Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 307–317. doi:10.1037/0278-7393.18.2.307
- Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.
- Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Perception & Psychophysics*, *70*, 1581–1591. doi:10.3758/PP.70.8.1581
- Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 41–55. doi:10.1037/0096-1523.34.1.41
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi:10.1037/0033-295X.99.1.122
- Kirschen, M. P., Chen, S. H., Schraedley-Desmond, P., & Desmond, J. E. (2005). Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: An fMRI study. *NeuroImage*, *24*, 462–472. doi:10.1016/j.neuroimage.2004.08.036
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ: Erlbaum.
- Logie, R. H., Brockmole, J., & Jaswal, S. (2011). Feature binding in visual working memory is unaffected by task-irrelevant changes of location, shape and color. *Memory & Cognition*, *39*, 24–36. doi:10.3758/s13421-010-0001-z
- Logie, R. H., Cocchini, G., Della Sala, S., & Baddeley, A. D. (2004). Is there a specific executive capacity for dual task co-ordination? Evidence from Alzheimer's disease. *Neuropsychology*, *18*, 504–513. doi:10.1037/0894-4105.18.3.504
- Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? In R. H. Logie & M. Denis (Eds.), *Mental images in human cognition* (pp. 105–115). New York, NY: North-Holland. doi:10.1016/S0166-4115(08)60507-5
- Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*, 55–74. doi:10.1016/0001-6918(90)90066-O
- Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 43–86). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780195305487.003.0003
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846
- Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, *22*, 496–512. doi:10.1162/jocn.2009.21214
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. doi:10.1037/h0043158
- 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. doi:10.3758/BF03196573
- Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 703–713. doi:10.1037/0278-7393.31.4.703
- Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, *91*, 216–234. doi:10.1037/0033-295X.91.2.216
- Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 435–448. doi:10.1037/0096-1523.13.3.435
- Oberauer, K., & Lange, E. B. (2009). Activation and binding in verbal working memory: A dual-process model for the recognition of non-words. *Cognitive Psychology*, *58*, 102–136. doi:10.1016/j.cogpsych.2008.05.003
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378. doi:10.3758/BF03210419
- Rämä, P., & Courtney, S. M. (2005). Functional topography of working memory for face or voice identity. *NeuroImage*, *24*, 224–234. doi:10.1016/j.neuroimage.2004.08.024
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42. doi:10.1080/135062800394667
- Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, *5*, 15–16. doi:10.1038/nn781
- Rostron, A. B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica*, *38*, 471–482. doi:10.1016/0001-6918(74)90007-9
- Salthouse, T. A., Fristoe, N. M., Lineweaver, T. T., & Coon, V. E. (1995). Aging of attention: Does the ability to divide decline? *Memory & Cognition*, *23*, 59–71. doi:10.3758/BF03210557
- Salthouse, T. A., Rogan, J. D., & Prill, K. A. (1984). Division of attention: Age differences on a visually presented memory task. *Memory & Cognition*, *12*, 613–620. doi:10.3758/BF03213350
- Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, *136*, 663–684. doi:10.1037/0096-3445.136.4.663
- Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology*, *3*, 408–429.
- Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppel, R. A. (1996). PET evidence for an amodal verbal working memory system. *NeuroImage*, *3*, 79–88. doi:10.1006/nimg.1996.0009
- Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, *15*, 215–222. doi:10.3758/PBR.15.1.215
- Scott-Brown, K. C., Baker, M. R., & Orbach, H. S. (2000). Comparison blindness. *Visual Cognition*, *7*, 253–267. doi:10.1080/135062800394793

- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*, 207–214. doi:10.1111/j.1467-9280.2009.02276.x
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1–29.
- Todd, J. J., Han, S. W., Harrison, S., & Marois, R. (2011). The neural correlates of visual working memory encoding: A time-resolved fMRI study. *Neuropsychologia, 49*, 1527–1536.
- Todd, J. J., & Marois, R. (2004, April). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751–754. doi:10.1038/nature02466
- Troje, N. F., & Bulthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771. doi:10.1016/0042-6989(95)00230-8
- Vogel, E. K., & Machizawa, M. G. (2004, April). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748–751. doi:10.1038/nature02447
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92–114. doi:10.1037/0096-1523.27.1.92
- Washburn, D. A., & Astur, R. S. (1998). Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad? *Memory & Cognition, 26*, 277–286. doi:10.3758/BF03201139
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*, 48–64. doi:10.1037/0096-3445.131.1.48
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120–1135. doi:10.1167/4.12.11
- Xu, Y., & Chun, M. M. (2006, March). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*, 91–95. doi:10.1038/nature04262
- Zhang, W., & Luck, S. J. (2008, May). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233–235. doi:10.1038/nature06860

Received August 6, 2010

Revision received May 17, 2011

Accepted May 29, 2011 ■