Visual and auditory working memory capacity

Nelson Cowan

A considerable amount of cognitive, behavioral research has been conducted on working memory[1]. Definitions vary, but a theory-independent definition might state that working memory is the collection of mental processes that permit information to be held temporarily in an accessible state, in the service of some mental task. The nature of the task can vary widely and can include immediate recall, reading or listening comprehension, reasoning, or problem-solving. In listening comprehension, for example, it is often the case that the intended meaning of a word within a sentence is unclear until subsequent words in the sentence are presented. It is necessary to hold words in mind in some form until their meanings can be interpreted in light of the remainder of the sentence. In reasoning, assumptions and facts must be held in mind and considered together until conclusions can be deduced from them. It has been clear that the capacity of working memory is limited ever since George Miller[2] described various research studies suggesting that people can recall at most about seven independent, meaningful items or 'chunks' at a time.

Although limits to working memory are easily observed, it is much more difficult to determine what specific mental faculties underlie the observed limits. The observed limit depends upon details of the stimuli, suggesting that it is not a single, simple limit. For example, immediate memory for lists of words is better when the lists contain words that can be pronounced relatively quickly[3], and this 'word-length effect' occurs to some extent even when the short-word and long-word sets actually comprise the same words, but with instructions to pronounce them quickly versus slowly[4]. To account for stimulus-dependent working memory limits as well as age differences in working memory, various researchers have proposed working memory systems that include multiple components. Baddeley[1] proposed a system that includes a 'central executive' process that makes use of a passive, time-limited phonological store along with a 'covert rehearsal' process for verbal recall, and a passive, time-limited 'visuospatial store', possibly with another covert rehearsal process, for visual recall. In my own theoretical writing[5,6] I have proposed that working memory is composed of a capacity-limited focus of attention, along with a temporarily activated portion of the information in permanent memory, which extends beyond the focus of attention to include some automatically activated information (see Fig. 1). When researchers use the term 'working memory', some of them seem to be referring only to the focus of attention, whereas others seem to be referring to all of the temporarily activated information. It is also likely that certain inactive portions of memory can be stored in a way that allows them to be recalled (or reactivated) quickly[2,6,7]. For example, in a reasoning problem involving rainbow colors, encoding the seven colors of the rainbow as the name 'Roy G. Biv' and keeping that name in mind makes the color names easily accessible while using up perhaps only one to three items of working memory capacity.

There have been some attempts to go beyond the observed working memory limits to glean the limits of the underlying processing components. Broadbent[8] suggested, on the basis of past evidence, that the true capacity limit is about three items (presumably when the contributions of mnemonically useful processes such as rehearsal and long-term memorization have been eliminated). For example, this is about the number of items that can be recalled without error across many trials, and about the maximum number of items that can be grouped together into a single 'chunk', although the actual limit may be closer to four. Other researchers have proposed similar capacity limits of three or four items in the number of processing channels for visual search[9], the number of items that can be enumerated quickly, without a slow, serial counting process[10], and the number of moving visual items that can be tracked at the same time[10,11]. A similar limit of about four items has been found when subjects encounter a spatial array of printed characters[12] or a spatiotemporal array of spoken characters[13] and must report them all. It is not clear if all of these similar limits are related; if so, perhaps they reflect the capacity limit of the focus of attention[6].

Recently, Luck and Vogel[14] have contributed to this area in an important way. They first presented a spatial array of colored squares or rectangles on every trial. The second presentation was another array that could differ from the first array in the color of one item. Subjects were able to carry out the task well only if the first array, the one to be held in memory, contained four or fewer items. (The same pattern was obtained in an experiment in which a single item within the second array, the one that sometimes differed from the first array, was marked with a surrounding square in order to limit the decision to that one item.) These results extend the previously observed capacity limit to nonverbal visual stimuli, and to a situation in which there was only one decision to be made (in the case when there was a single probe item). A few other results warrant special mention and discussion.

First, the observed capacity of visual working memory was not reduced when subjects had to hold in mind two digits during a visual memory trial, to be recalled immediately afterwards[14]. One might expect a reduction of visual working memory if both verbal and visual representations were held in the same capacity-limited store. However, it is possible that the two verbal items could be held entirely in the form of a passive phonological store and rehearsal process[1] without taking up space in the capacity-limited store or focus of attention[5,6]. If the verbal memory load were increased further or accompanied by a rehearsal-blocking task[1], it might well be shown to reduce the observed capacity of visual working memory.

Second, the capacity limit was found to be the same (about four items) no matter whether the discrepancy between displays occurred in one feature, two features, or four features of each object. Thus, the capacity is apparently for integrated objects, not features per se. This at first may seem curious, in view of the fact that other research[15] suggests that links between features must be perceived one object at a time rather than for all objects in parallel. Possibly, subjects read the items into working memory one at a time but are still limited to about four such items on a trial. If this is true, despite the short durations of the arrays (≥100 ms), then items must be read out of a post-stimulus sensory memory[12] and, therefore, memory should be greatly curtailed by a pattern mask immediately following the first array in a trial.

For arrays larger than four it is not even absolutely clear whether subjects encode a specific four items from a visual array (either serially or in parallel) or do a partial encoding of all of the items (e.g. about half the features of each item in an eight-item array). In

N. Cowan is at the Department of Psychology, University of Missouri, Columbia, MO 65211, USA.

tel: +1 573 882 4232
fax: +1 573 882 7710
e-mail: psycowan@showme.missouri.edu

 1364-6613/98/$19.00 PII: S1364-6613(98)01144-9

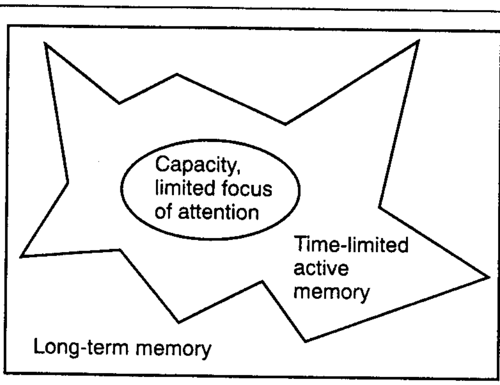

Fig. 1 An illustration of one simple theoretical model of working memory. (Modified from Refs 5,6.)

other words, the basis of the four-item limit is still unclear. Various research strategies could be of use here. For example, an item-by-item analysis could theoretically reveal that it is usually the four items closest to the fixation point that are encoded. If, instead, all items are partially encoded, then it should be possible to improve performance by changing more than one feature of the target object between presentations in the same trial, increasing the chances that at least one of the critical features had been encoded by the subject.

Working memory has also been a popular topic within recent neuroimaging studies. It is important to realize that there is still considerable behavioral work to be done before it will become clear what the behaviors are that might be explained through brain processes.

References

1 Baddeley, A. (1992) Working memory *Science* 255, 556–559

2 Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information *Psychol. Rev.* 63, 81–97

3 Baddeley, A.D., Thomson, N. and Buchanan, M. (1975) Word length and the structure of short-term memory *J. Verbal Learn. Verbal Behav.* 14, 575–589

4 Cowan, N. *et al.* (1997) There are two word length effects in verbal short-term memory: opposed effects of duration and complexity *Psychol. Sci.* 8, 290–295

5 Cowan, N. (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system *Psychol. Bull.* 104, 163–191

6 Cowan, N. (1995) *Attention and Memory: An Integrated Framework* (*Oxford Psychology Series No. 26*), Oxford University Press

7 Ericsson, K.A. and Kintsch, W. (1995) Long-term working memory *Psychol. Rev.* 102, 211–245

8 Broadbent, D.E. (1975) The magic number seven after fifteen years, in *Studies in Long-Term Memory* (Kennedy, A. and Wilkes, A., eds), pp. 3–18, John Wiley & Sons

9 Fisher, D.L. (1984) Central capacity limits in consistent mapping, visual search tasks: four channels or more? *Cognit. Psychol.* 16, 449–484

10 Trick, L.M. and Pylyshyn, Z.W. (1994) Why are small and large numbers enumerated differently?: a limited-capacity preattentive stage in vision *Psychol. Rev.* 101, 80–102

11 Yantis, S. (1992) Multi-element visual tracking: attention and perceptual organization *Cognit. Psychol.* 24, 295–340

12 Sperling, G. (1960) The information available in brief visual presentations *Psychol. Monogr.* 74, No. 498

13 Darwin, C.J., Turvey, M.T. and Crowder, R.G. (1972) An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage *Cognit. Psychol.* 3, 255–267

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

15 Treisman, A.M. and Gelade, G. (1980) A feature integration theory of attention *Cognit. Psychol.* 12, 97–136

Response from Luck and Vogel

Human abilities in cognitive tasks are clearly limited, and a central goal of cognitive science is to understand the processing restrictions that underlie limitations in the performance of these tasks. Perhaps the most thoroughly studied cognitive limitation is the highly restricted capacity of working memory. As discussed by Cowan in the preceding commentary, working memory can be defined as the temporary storage of information in an accessible state that permits the information to be used for ongoing mental operations.

Working memory is analogous to the internal memory registers of a computer's central processing unit (CPU). In standard computer architectures, the number of CPU registers is relatively small, just as the capacity of working memory is limited to a few items, and the registers are used to store the inputs and outputs of computational operations, just as working memory is used in the service of cognitive processes. In addition, many different types of data can be stored in the registers of a CPU, and this is paralleled by the ability of working memory to store information about sights, sounds, words and concepts. Moreover, just as the speed and flexibility of a computer is partially limited by the number of CPU registers, performance on cognitive tasks, such as reading and arithmetic, is correlated with an individual's working memory span[1]. Thus, it is clear that working memory plays a central role in cognitive processing and that limitations in working memory capacity are a significant source of limitations in the performance of cognitive tasks.

As discussed by Cowan, it appears that the capacity of working memory appears to be only about four items, although various 'mental tricks' can be used to retain larger sets of information[2]. Working memory capacity has been studied extensively in the context of verbal information[3], but has been addressed only rarely in the case of purely visual information. In addition, previous studies of visual working memory have primarily used letters and digits as stimuli[4], which is problematic because these stimuli could easily be encoded verbally as well as visually, and because they are relatively complex and highly overlearned visual forms. To overcome these limitations, we recently examined the capacity of visual working memory for simple features, such as color and orientation, as well as for combinations of these features[5]. This approach permits a more systematic exploration of factors such as stimulus complexity and similarity.

Our basic experimental paradigm is illustrated in Figure 1A. On each trial, subjects viewed two arrays of colored squares (called S1 and S2) and reported whether the two arrays were identical or differed in the color of one of the squares. A 900 ms gap separated S1 and S2, making it necessary for subjects to store S1 in memory so that it could be compared with S2. To assess the capacity of working memory, we varied the number of items in the stimulus arrays (the set size), assuming that performance would be nearly perfect when the displays were within the memory span and that performance would decline when the displays exceeded the memory span. Indeed, subjects were nearly perfect when the arrays contained one, two or three items, but accuracy began to decline as the set size increased to four or more items.

On the basis of an equation developed by Pashler[6], we estimated that subjects were able to hold an average of about four items in visual working memory at one time (because performance was not perfect with a set size of four items, it is likely that there was some variation from trial to trial in the number of items actually held in working memory). This estimate of working memory capacity is quite similar to the estimate derived for very different types of stimuli, as discussed in Cowan's commentary, and this is somewhat surprising because the information to be retained in this experiment was extremely simple.

To examine the role of stimulus complexity, we conducted an additional

S.J. Luck and E.K. Vogel are at the Department of Psychology, University of Iowa, 11 Seashore Hall E., Iowa City, IA 52242-1407, USA.

tel: 11 319 335 2422
fax: 11 319 335 0191
e-mail: steven-luck@uiowa.edu

 1364-6613/98/$19.00 PII: S1364-6613(98)01143-7

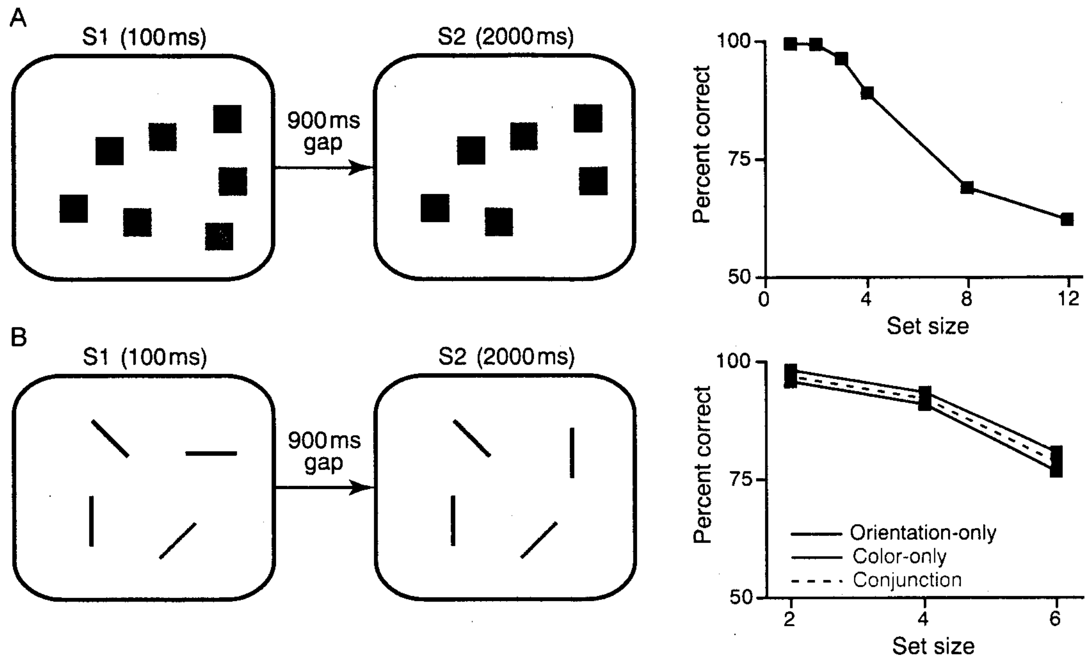

Fig. 1 Stimuli and results from a study of visual working memory[5]. Subjects are instructed to report whether the two stimulus arrays (S1 and S2) are identical or differ in terms of a single feature of one object. When color is the only feature **(A)**, accuracy declines systematically as the number of items in each array (the set size) is increased beyond three items. When each item has a color and an orientation **(B)**, subjects are just as accurate when they must remember both features of each object (the conjunction condition) as when they are instructed to remember a single feature (the orientation-only and color-only conditions).

experiment in which the stimuli were colored bars presented in one of four orientations, as illustrated in Figure 1B. Studies of visual perception with stimuli of this nature have shown that subjects frequently make errors when asked to combine features such as color and orientation[7,8], especially when the stimulus arrays contain a large number of closely-spaced objects[9,10]. Thus, to avoid perceptual errors, we used relatively small set sizes (two, four or six items) and spaced the objects widely. When subjects were instructed to look for changes in only one of the two features (the 'color-only' and 'orientation-only' conditions), memory capacity was again found to be four items. In the critical condition (the 'conjunction' condition), either the color *or* the orientation of an object could change, thus requiring subjects to retain in memory both features of each object in order to perform the task correctly. In other words, to perform this task perfectly when the array contained four items, eight features had to be held in working memory. Although this might have been expected to result in reduced accuracy, we found that subjects were just as accurate in the conjunction condition as in the color-only and orientation-only conditions, even though twice as many features had to be retained in working memory in the conjunction condition.

Moreover, in a follow-up experiment, we found that subjects could retain four features of each object just as easily as they could retain a single feature of each object, allowing 16 features to be retained – as long as they were distributed across four objects. Along with other results, these findings indicate that visual working memory does not store individual features, but instead stores integrated object representations. This is similar to the results of various studies of verbal working memory showing that complex 'chunks' can be stored as individual items[11].

As Cowan remarks, the finding of equivalent working memory performance for individual features and conjunctions of features seems curious, especially given the greater perceptual difficulty of identifying multi-feature objects. A likely explanation is that the binding together of features is the responsibility of the perceptual system, and that working memory simply stores the integrated object representations that are the result of perception. However, this does not fully explain why there is a severe limit on the number of objects that can be stored but no evident limitation in the number of features that compose each object. To explain this finding, we have borrowed from the literature on perceptual binding, in which several investigators have proposed that the separately coded features of an object are bound together by means of synchronous firing of the neurons that code the individual features[12,13].

As illustrated in Figure 2, a red vertical bar and a green horizontal bar can be represented by means of synchronization of the neurons that code red and vertical, with independent synchronization of the neurons that code green and horizontal. In models of this nature, synchronization plays a key role in binding features together, so that (in this example) the stimuli are not incorrectly coded as a red horizontal bar and a green vertical bar. Synchronization is also important for individual features, which are coded by the pattern of activity across a set of relatively broadly tuned neurons.

A synchronization-based model of working memory can explain both the existence of severe limits in the number of objects that can be stored and the lack of a limit on the number of features that can be stored. We hypothesize that the number of objects that can be maintained in working memory is limited by the possibility of spurious synchronizations between neurons that code different objects. If many objects are stored simultaneously, there will be a high probability that the neurons coding one object will fire at the same time as the neurons coding a different object, leading to interference between the representations and thus impaired performance. However, adding more features to each object will not lead to additional interference, because this will simply increase the number of neurons that will fire simultaneously as a group. On the other hand, this process of creating synchronization among multiple neurons might be quite difficult, which could make the encoding of objects into working memory very time-consuming. Indeed, as Cowan suggests, the insertion of a mask between S1 and

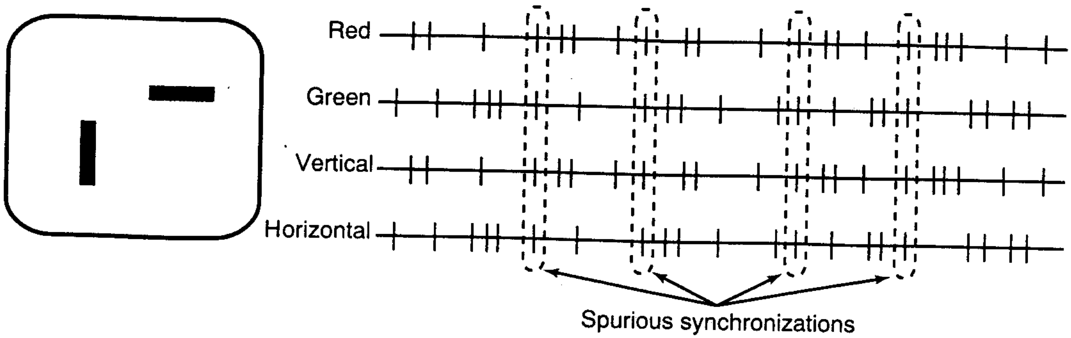

Fig. 2 Example of the use of synchronized neural firing to code combinations of features in working memory. The responses of four individual neurons are shown, separate neurons coding red, green, horizontal and vertical. Each vertical mark represents a neural response (i.e. an action potential), and the horizontal axis represents the passage of time. The red and vertical neurons are bound together by means of firing at the same time as each other, as are the green and horizontal neurons. Occasionally, both sets of neurons fire at the same time, and these spurious synchronizations will tend to degrade the representation. This will occur more frequently as more items are stored, and might underlie the capacity limits of visual working memory.

S2 in the paradigm shown in Figure 1 might severely disrupt the process of transferring perceptual information into working memory, especially for multi-feature objects. Some observations of this nature have already been reported[14].

The synchronization model also speaks to the last issue raised by Cowan, that of the coding of stimulus arrays that exceed the capacity of working memory. Specifically, if too many items are stored simultaneously, then interference between the items will lead to a degraded representation of each item. However, it is also possible that selective attention might be used to store only a subset of the items in working memory, allowing a high-fidelity representation of the selected items rather than a low-fidelity representation of the entire array. We believe that interactions of this nature between attention and working memory will become a significant area of research in the coming years.

References

1 Daneman, M. and Carpenter, P.A. (1980) Individual differences in working memory and reading *J. Verbal Learn. Verbal Behav.* 19, 450–466

2 Ross, J. and Lawrence, K.Q. (1968) Some observations on memory artifice *Psychonomic Science* 13, 107–108

3 Baddeley, A.D. (1986) *Working Memory*, Clarendon Press

4 Sperling, G. (1960) The information available in brief visual presentations *Psychological Monographs* No. 74

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

6 Pashler, H. (1988) Familiarity and visual change detection *Percept. Psychophys.* 44, 369–378

7 Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention *Cognit. Psychol.* 12, 97–136

8 Treisman, A. and Schmidt, H. (1982) Illusory conjunctions in the perception of objects *Cognit. Psychol.* 14, 107–141

9 Cohen, A. and Ivry, R. (1989) Illusory conjunctions inside and outside the focus of attention *J. Exp. Psychol. Hum. Percept. Perform.* 15, 650–663

10 Cohen, A. and Ivry, R.B. (1991) Density effects in conjunction search: evidence for a coarse location mechanism of feature integration *J. Exp. Psychol. Hum. Percept. Perform.* 17, 891–901

11 Simon, H.A. (1974) How big is a chunk? *Science* 183, 482–488

12 Singer, W. and Gray, C.M. (1995) Visual feature integration and the temporal correlation hypothesis *Annu. Rev. Neurosci.* 18, 555–586

13 von der Malsburg, C. (1996) The binding problem of neural networks, in *The Mind–Brain Continuum* (Llinás, R. and Churchland, P.S., eds), pp. 131–146, MIT Press

14 Enns, J.T. and Di Lollo, V. (1997) Object substitution: a new form of masking in unattended visual locations *Psychol. Sci.* 8, 135–139