Table 1. (Wilding). *Time to initiate recall(s) for lists of 5 to 17 items in a visual digit span task*

List Length	5	6	7	8	9	10	11	12	13	14	15	16	17
MP	1.1	1.3	1.3	1.3	1.3	1.6	1.5	1.5	2.1	2.1	2.3	2.1	2.6
Rajan	—	—	0.2	0.3	0.2	0.3	0.7	0.7	1.4	4.2	3.0	8.4	11.2

The mean of the longest span achieved on each run was 14.96 (range 12–18, s.d. 1.4), less outstanding than Rajan, but about twice the norm. Table 1 gives the time to initiate recall at each list length (successful recalls only) for MP and corresponding figures for Rajan (estimated from Thompson et al., Fig. 2.1; absolute times are not comparable, due to the different methods of recording used). While Rajan certainly shows a steep rise in times for lists of more than 13 items, there is also some suggestion of an increase at 11 items; no data are given for lists of five or six items and presumably measurement of times by stopwatch was imprecise for short intervals. MP, however, shows a gradual increase, with possible stops at 10, 13, and 17 items. There is, however, no clear discontinuity, which would unambiguously indicate a basic span limitation. The other expert we tested, who overtly converted numbers into images, showed a sharp discontinuity at 9 items, followed by a continuous rise in times, like Thompson et al.'s (1993) subject GN, who also developed a mnemonic method, so the pattern shown by MP was not some artefact of the measurement method.

Inter-item latencies for successful recall, however, provided unambiguous evidence for chunking in MP's recall; however, the pattern obtained was only partly consistent with Cowan's argument. Figure 1 gives cumulative recall times for lists of 5 to 16 items. Once lists exceeded 7 items, a pause occurred after the fourth item, once lists exceeded 11 items, another pause occurred after the eighth item, and once lists exceeded 15 items, a third pause occurred after the twelfth item. MP apparently divided the lists initially into groups of four items (though lists of 14 items seem to be divided 4–4–3–3), but the final group could be as large as seven items before it was further subdivided. MP confirmed that, though he does not have a deliberately practised strategy for memorisation, he tried to find some mathematical pattern or relation, normally in each group of four.

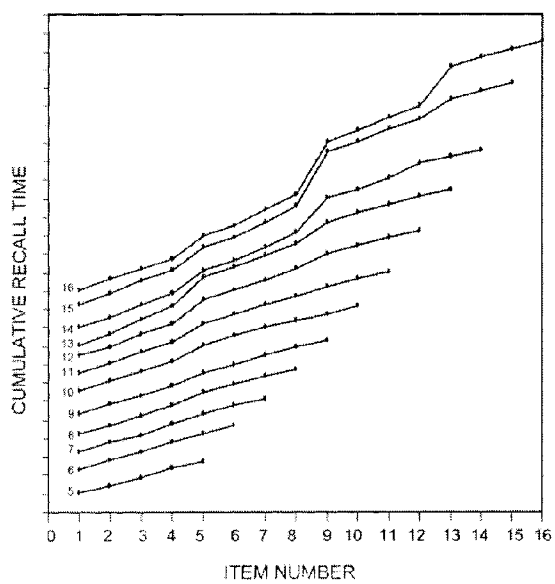

Figure 1 (Wilding). Mean cumulative recall times by MP for each item in lists of 5 to 16 items. Numbers on the left indicate list length for each function. As list length increases, each function has been displaced upward by 1 sec to avoid overlap.

While any conclusions clearly need data from other individuals before they can be firmly established, these results imply that: (1) Thompson et al.'s (1993) response lag measure is not decisive as an indication of basic span. Though no clear discontinuity in MP's data indicated a basic span limit, discontinuities in inter-item latencies suggested chunking was occurring. (2) items were in general grouped into fours, but such grouping was not mandatory and up to seven items were treated as a group in some circumstances. Such flexibility may explain the long running argument as to whether the limit is four or seven items and raises further questions about the mechanisms involved.

Attention is not unitary

Geoffrey F. Woodman, Edward K. Vogel, and Steven J. Luck

Department of Psychology, University of Iowa, Iowa City, IA 52242-1407.
{geoff-woodman; steven-luck}@uiowa.edu vogel@sdepl.ucsd.edu
www.psychology.uiowa.edu/Faculty/{woodman/woodman; luck/luck}.html

Abstract: A primary proposal of the Cowan target article is that capacity limits arise in working memory because only 4 chunks of information can be attended at one time. This implies a single, unitary attentional focus or resource; we instead propose that relatively independent attentional mechanisms operate within different cognitive subsystems depending on the demands of the current stimuli and tasks.

Cowan's model of working memory posits that the limited capacity of working memory is entirely caused by the limited capacity of attention. He further proposes that a unitary mechanism of attention operates across cognitive subsystems, with a single capacity that must be shared across subsystems. Here, we argue in favor of an alternative view in which the brain makes use of a collection of loosely interconnected attention mechanisms that operate in different cognitive subsystems and that reflect the properties of the individual subsystems. We will focus specifically on the operation of attention in three coarsely defined cognitive subsystems, namely perceptual analysis, working memory, and response selection.

We would expect that different mechanisms of attention would operate in visual perception and in visual working memory for the simple reason that these subsystems appear to operate at very different rates. For example, Potter (1976) showed that observers can identify pictures of complex real-world scenes at rates of up to 8 pictures per second, but much slower presentation rates were required for observers to store the scenes in working memory. More recent studies have used a combination of psychophysical and electrophysiological measures to demonstrate that attention shifts at different rates in visual perception and in visual working memory. Specifically, studies of the attentional blink phenomenon, which reflects the operation of attention in working memory, indicate that hundreds of milliseconds are required to shift working memory-level attention from one stimulus to another (Vogel et al. 1998). In contrast, studies of visual search, which emphasized the operation of attention in perception, have indicated that shifts of perceptual-level attention can occur every 50–100 msec (Woodman & Luck 1999). Horowitz and Wolfe (1998) and colleagues have also provided evidence that the operation of attention in perception is relatively independent of memory.

Given that attention operates at different speeds in perception and in working memory, it would be sensible if these mechanisms of attention operated asynchronously; otherwise, the fast perceptual system would be continually waiting for the slow working memory system to catch up. We have recently provided evidence for asynchronous operation of perception and working memory by showing that shifts of attention during visual search are not slowed when visual working memory is full (Woodman et al., in press). In this study, participants were required to maintain a set of object representations in visual working memory while performing a visual search task. We observed that subjects could perform a difficult visual search task just as efficiently when working memory was full as when working memory was empty, indicating that perceptual-level attention can be allocated and shifted very efficiently even when working memory-level attention is operating at maximal capacity.

It is also reasonable to suppose that visual perception and visual working memory might differ in their spatial properties as well as in their temporal properties. In particular, visual perception relies on topographically mapped representations, but there is no evidence that the representation of objects in working memory is topographic. Thus, perceptual-level attention might be expected to have various spatial properties that are absent from working memory-level attention. A recent study by Vogel (2000) supports this hypothesis. Specifically, Vogel found that working memory-level attention can be divided among multiple noncontiguous locations without being allocated to the regions between these locations, whereas perceptual-level visual attention must be focused on a contiguous region of the visual field. Thus, perceptual-level and working memory-level mechanisms of attention differ in both their spatial and temporal characteristics.

It also appears that the mechanisms of attention that operate during response selection and initiation are different from those that operate at earlier stages. The operation of attention during response-related processes has been studied by means of the psychological refractory period (PRP) paradigm. In this paradigm, two stimuli are presented to subjects in rapid succession and the subjects make separate speeded responses to each stimulus. Reaction time to the second stimulus is slowed when the delay between the two targets is short, and several studies indicate that this is due to capacity limitations at the stage of response selection (Pashler 1994). Pashler (1991) and Johnston et al. (1995) have provided evidence that shifts of visual-spatial attention do not operate at the same stage as the capacity limits that are observed in the PRP paradigm.

In conclusion, there are now many forms of evidence indicating that there are different mechanisms of attention that operate with different properties within different cognitive subsystems. Moreover, it is worthwhile to ask what is gained by proposing that the limited capacity of working memory arises from the limited capacity of attention. The real question is why there are limits at all, whether we call them limits of attention or limits of working memory-specific resources.

Author's Response

Metatheory of storage capacity limits

Nelson Cowan
Department of Psychology, University of Missouri, Columbia, MO 65211.
cowann@missouri.edu www.missouri.edu/~psycowan

Abstract: Commentators expressed a wide variety of views on whether there is a basic capacity limit of 3 to 5 chunks and, among those who believe in it, about why it occurs. In this response, I conclude that the capacity limit is real and that the concept is strengthened by additional evidence offered by a number of commentators. I consider various arguments why the limit occurs and try to organize these arguments into a conceptual framework or "metatheory" of storage capacity limits meant to be useful in future research to settle the issue. I suggest that principles of memory representation determine what parts of the representation will be most prominent but that limits of attention (or of a memory store that includes only items that have been most recently attended) determine the 3- to 5-chunk capacity limit.

R1. General reaction to commentaries

In *BBS*'s first round of refereeing of the target article, it seemed to several of the referees that the paper was not controversial enough to generate useful commentaries. I think that the present set of commentaries has proven to be very interesting, diverse, and thought-provoking after all.

By my count, about 15 of the 39 commentaries solidly accepted the hypothesis that some core working memory faculty is limited to about 4 separate chunks of information. Indeed, at least 8 commentaries (**Lane et al.; Nairne & Neath; Pothos & Juola; Rensink; Rypma & Gabrieli; Todt; Usher et al.; Wilding**) presented additional evidence for a 4-chunk limit in many circumstances. However, at least seven other commentaries seemed strongly opposed to that concept and the remaining 17 seemed less committed either way. Many of the commentators who more or less agreed with the 4-chunk hypothesis offered alternative theoretical explanations.

R1.1. Self-justification

I believe the target article has made three main contributions. First, principles were offered as guidelines for identifying experimental situations in which the number of separate chunks can be estimated. Second, a diverse field of evidence was shown to yield similar estimates of capacity in those identified situations; specifically, 3 to 5 chunks on the average, among normal adult humans. Third, various plausible theoretical explanations for the capacity limit were described.

The eligibility of evidence often depended upon the application of critical theoretical assumptions. For example, one theoretical assumption provides an answer to **Usher et al.**, who commented that "it is somehow paradoxical that recall of *unattended* material should provide a measure for the capacity of the focus of *attention*". It is not so paradoxical if one considers that, ordinarily, attention exerts effects broadly across the perceptual encoding, mnemonic storage, and retrieval processes so that any one particular effect of