Gormican 1988) For instance, the dimension "colour" of an object can be characterized by one of the features "red," "blue," "purple," and so on. After breaking down a visual object into these more basic elements, the question arises as to whether the above definitions of "dimensions" and "features" in vision constitute an appropriate basis for characterizing chunks in visual memory. In other words, one may ask (1) whether chunking in visual working indeed functions by binding features into integrated visual objects, and (2) analyse the rules and limitations of this binding.

With regard to the first question there is recent evidence from Luck and Vogel (1997, also referred to in sect. 3.1.1), suggesting that the capacity limit of visual STM indeed refers to feature bundles in the form of objects. In their experiments, subjects were required to retain simple geometrical visual objects made up of feature conjunctions such as of a certain colour, orientation, and length. The data showed that objects defined by conjunctions of two or more dimensions (e.g., a line of a certain orientation, colour, and length) can be retained as well as objects defined by only a single dimension (e.g., orientation only). For any of these combinations, the estimated memory capacity was about four objects.

The second question is strongly related to the problem of how many of these basic elements can be bound into a single chunk, and whether there exist limitations as to the possible combination of features.

In our own experiments, we attempted to replicate this surprising finding (Deubel et al., in preparation). We used the same stimuli as Luck and Vogel (1997), and identical experimental parameters such as presentation and retention times. Our experimental results (Fig. 1) clearly show that retention of objects defined by a conjunction of two colours leads to a strong drop in performance, as compared to the condition in which the objects consisted of one colour only. So, external objects with two colours seem to require two chunks for the internal coding. This finding is in obvious contrast to the result of Luck and Vogel (1997). The reason why we could not replicate their data is unclear to us, however, in an independent study, Wheeler and Treisman (submitted) recently reported a finding similar to ours.

These data are clear evidence that there exist prominent limitations to chunking in visual memory. As a possible, preliminary rule of thumb suggested by the result, one may assume that a visual chunk can consist of not more than one feature per dimension, that is, one colour, one shape primitive, and so on. A further, yet unresolved important issue in this context is the question whether there is also a limit in the number of possible dimensions that define a chunk. Luck and Vogel (1997) found no limit (i.e., no drop in memory performance) up to a conjunction of four different dimensions (colour, orientation, length, gap). However, it might be that a limitation larger than that can indeed be found.

The empirical task of the future will be to determine more precisely the limitations of chunking and how they relate to visual features and dimensions. Indeed, we think that the paradigm presented here offers a promising experimental approach to answer questions about the nature of chunks in vision. Measuring memory performance for a variety of stimuli and features could reveal the basic dimensions and features in vision in a very straightforward way: If adding the feature in question to the stimuli leaves the memory capacity (in terms of number of objects) unaffected, one may conclude that it is really a basic visual feature, forming an elementary part of a visual chunk.

Cowan defines a chunk as "a collection of concepts that have strong associations to one another and much weaker associations to other chunks currently in use." This definition does not impose any constraints on the nature and number of elements that can be bound into a chunk. Our experiment is a demonstration that such limitations exist, and that their analysis may lead to important insights into properties of visual memory.

The magical number 4 in vision

Brian J. Scholl[a] and Yaoda Xu[b]

[a]*Department of Psychology, Vision Sciences Laboratory, Harvard University, Cambridge, MA 02138;* [b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.*
scholl@wjh.harvard.edu **yaoda@psyche.mit.edu**
www.wjh.harvard.edu/~scholl

Abstract: Some of the evidence for a "magical number 4" has come from the study of visual cognition, and Cowan reinterprets such evidence in terms of a single general limit on memory and attention. We evaluate this evidence, including some studies not mentioned by Cowan, and argue that limitations in visual processing are distinct from those involved in other memory phenomena.

Cowan's discussion of the "magical number 4" synthesizes evidence from domains which are rarely discussed together. In particular, Cowan draws on work from the study of visual cognition – such as studies of subitizing (sect. 3.3.2) and multiple object tracking (sect. 3.3.3) and attempts to reinterpret such evidence in terms of a general memory limitation, which he suggests is a reflection of the underlying capacity of the "attentional focus" (a thesis which is discussed in Cowan 1995, but which he does not argue for in his target article). Here we note additional evidence for a limit of approximately 4 objects in certain types of visual processing, and discuss why these limits are probably distinct from those involved in other (e.g., verbal) tasks.

Additional evidence from visual cognition. Additional evidence for a "magical number 4" in visual processing comes from studies of infants, normal adults, and neuropsychological syndromes. Recent looking-time studies with infants have suggested that they are able to keep track of arrays of objects through additions and subtractions, but only if there are less than 4 objects in these arrays (e.g., Uller et al. 1999; Wynn 1992), and this evidence has been interpreted in terms of developing mechanisms of visual attention (e.g., Carey & Xu, in press; Scholl & Leslie 1999). In normal adults, there appears to be a limit of 4 on the number of objects which can receive prioritized processing due to attentional capture (Yantis & Johnson 1990), and the number of items which can be simultaneously examined in a visual search for a change (Rensink 2000).

Finally, it has been shown that bilateral lesions of the parietal lobes in Balint's syndrome can reduce visual processing capacity. Patients with Balint's syndrome have great deficits in perceiving complex visual scenes, although their ability to recognize individual objects is usually preserved (for a review, see Rafal 1997). Dehaene and Cohen (1994) studied visual enumeration in 5 Balint's patients and found that these patients could enumerate sets of 1, 2, and sometimes 3 items correctly, but not sets comprising more than 3 items. Reaction time slopes for these patients were flat for set sizes of 1 and 2 items, but increased sharply for set sizes of 3 or more items. Treisman and colleagues (Friedman-Hill et al. 1995; Robertson et al. 1997) reported another Balint's patient who could not correctly enumerate more than one or two objects even when he was aware that more were present. In rare and extreme cases, Balint's patients report seeing only one object when presented with multiple objects (e.g., Coslett & Saffran 1991).

Specific visual limits or general memory/attention limits? Cowan views such evidence as continuous with data concerning the number of chunks which can be simultaneously active in short term memory (STM). In contrast, we think there are good reasons to resist this reinterpretation, and to view the limits on visual processing as separate from those involving verbal and other non visual material. (In this respect we take a position similar to that of Miller 1956 who suspected that STM limits and subitizing limits were independent.) Given space restrictions, we will largely restrict our discussion of this issue to the evidence which Cowan does discuss in his target article: subitizing (wherein observers can determine the cardinality of sets with less than 5 items roughly in parallel and without errors) and multiple object tracking (MOT;

wherein observers can attentionally track up to 4–5 independently and unpredictably moving identical items in a field of identical distractors).

Cowan presents only a few arguments for interpreting these phenomena in terms continuous with general STM limits. For MOT he provides no arguments, simply stating that one could use a general STM-based theory to explain performance. (Such an explanation, it seems to us, could not easily account for the strong dependency of MOT performance on subtle visual details such as the type of accretion and deletion behind occluders; Scholl & Pylyshyn 1999). For subitizing, he notes the vision-based theory of Trick and Pylyshyn (1994a), and argues against it mainly by appeal to two phenomena. First, he suggests that the "pop-out" alluded to by Trick and Pylyshyn can also occur for larger numbers of items, for example "when all of the eggs [in a carton] pop out against the surrounding carton" (sect. 3.3.2). This, however, is clearly not the type of pop-out that Trick and Pylyshyn (and others who have investigated visual search) have in mind, since the eggs in this case do not pop out as individuals, but as a group. Second, Cowan suggests that focused central attention is more important to enumeration than is suggested by Trick and Pylyshyn's theory, since other researchers (Atkinson et a1. 1976; Simon & Vaishnavi 1996) studying the enumeration of dots in afterimages have claimed that observers cannot enumerate sets greater than 4 without eye movements. This claim is false, however, and the limits these investigators found were due to the confounding effects of crowding (He et al. 1997).

Beyond Cowan's arguments, we think there are several additional reasons to view these limits as distinct from those involved in verbal STM. First, viewing them as identical seems to necessitate a prediction that one should not be able to track 4 targets in the MOT task and simultaneously acquire and hold 4 verbally-presented items in STM. However, this is trivial to do, and such tasks seem not to interfere at all. (In an informal test, two observers tracked 4 in 8 items for 10 sec with an accuracy of 87.5% averaged over 10 trials. When they also had to remember 4 random digits presented auditorily as the targets were being specified, they tracked with an accuracy of 92.5%, and made no errors on the memory task.) Cowan notes in section 4.2 of the target article that such evidence against a single capacity limit could be explained away by appeal to attentional switching back and forth between the two tasks, but in this respect MOT is an ideal foil, since one can succeed in the task only by continuous tracking (Pylyshyn & Storm 1988). Second, an explanation based on a single general limitation of memory or attention predicts that these limits should stand or fall together in neuropsychological impairments, which they do not. For example, none of the Balint's patients mentioned above exhibited deficits in short-term memory span. There are patients who, after lesions in the left hemisphere language areas, exhibited reduced STM span despite normal speech production in some cases (e.g., Baddeley 1986; Shallice & Warrington 1970). However, none of these patients showed any signs of Balint's symptoms or deficits in visual processing. Moreover, although these patients showed very poor retention of auditorily presented digits, with a span in the region of two items, they usually showed better retention of visually presented digits, with a span in the region of 4 or 5. These double dissociations in lesion sites and patient performance argue strongly against the notion that a common capacity limitation underlies capacity limited performance in both verbal and visual tasks.

Visual objects vs. chunks in memory. The view that these limitations in visual processing are distinct from those involved in other memory phenomena is further strengthened by the fact that the "units" of processing in each case are quite different. The "chunks" of memory can be almost infinitely flexible in their composition, and are thus defined by Cowan and others simply in terms of association networks (see sect. 1.3). This flexibility is in marked contrast to the units of visual attention – visual objects – which appear to be characterized by highly constrained and inflexible rules (Scholl, in press). In MOT, for instance, observers can track 4 dots in a field of 8 dots, but completely fail when trying to track 4 line endpoints in a field of 4 lines (and thus, 8 endpoints). In general, very specific rules involving connectedness and part-structure seem to determine whether a feature cluster can be tracked in MOT (Scholl et al., in press). Similarly, in visual short-term memory studies using a change detection paradigm, color and orientation features are best remembered if they belong to the same part of an object and less well remembered if they belong to different parts of an object (Xu, submitted). All of these constraints are in marked contrast to the robustness and flexibility of potential STM chunks with verbal materials.

We think the considerations discussed here provide good reasons for thinking that the limits of approximately 4 involved in various types of visual processing are distinct from other similar STM limits. We remain agnostic on the question of why there should exist similar independent limits. It could be for the teleological and computational reasons discussed by Cowan (in sect, 4.1), or it could be – as George Miller (1956) suspected of the similarity of memory capacity and subitizing limitations – "nothing more than a coincidence."

How unitary is the capacity-limited attentional focus?

Torsten Schubert and Peter A. Frensch

Department of Psychology, Humboldt-University Berlin, D-10117 Berlin, Germany. **{torsten.schubert; peter.frensch}@psychologie.hu-berlin.de**

Abstract: Cowan assumes a unitary capacity-limited attentional focus. We argue that two main problems need to be solved before this assumption can complement theoretical knowledge about human cognition. First, it needs to be clarified what exactly the nature of the elements (chunks) within the attentional focus is. Second, an elaborated process model needs to be developed and testable assumptions about the proposed capacity limitation need to be formulated.

One of the main contributions of Cowan's important target article is the assumption of a unitary limitation of the attentional focus. Cowan's arguments in favor of this assumption should reinforce the current discussion about the nature of unitary (e.g., Baddeley 1986; Norman & Shallice 1986) or distributed attentional mechanisms (Allport 1987; Meyer & Kieras 1997; Neumann 1987). Although we agree that this assumption is intriguing, we are somewhat disappointed by its theoretical elaboration and by the absence of significant support advanced in its favor by Cowan.

As supporting evidence for his assumption, Cowan lists different studies that all yield performance restrictions of about 4 items in different experimental contexts and over a wide range of stimulus materials, for example, dots, digits, screen locations, auditory signals, and so forth. Because all these studies somehow yield the number "4," Cowan's main argument is that there should be a unitary mechanism underlying this limitation. However, for this argument to be convincing, it needs to be shown, first, that the items across the different reported experimental contexts are comparable entities. Second, one would need to describe an elaborated process model with a set of mechanisms formulated that allows an integrated understanding of the findings.

Unfortunately, the present work falls somewhat short on both of these points. For example, Cowan uses the concept of chunks in stressing the equality of items in different experimental contexts. However, he does not formulate a convincing operational definition of what exactly a chunk is. How do we know that 4 dots, screen locations, or digits correspond to 4 chunks as, for example, 4 words might? The main question is: How can we measure or define chunks independently of the experimental context we are dealing with in a concrete experimental situation? If there is no sufficiently constraining definition, what prevents us from arguing