Visual memory needs categories

Henrik Olsson* and Leo Poom

Department of Psychology, Uppsala University, Box 1225, SE-751 42 Uppsala, Sweden

Edited by Anne Treisman, Princeton University, Princeton, NJ, and approved April 29, 2005 (received for review January 31, 2005)

Capacity limitations in the way humans store and process information in working memory have been extensively studied, and several memory systems have been distinguished. In line with previous capacity estimates for verbal memory and memory for spatial information, recent studies suggest that it is possible to retain up to four objects in visual working memory. The objects used have typically been categorically different colors and shapes. Because knowledge about categories is stored in long-term memory, these estimations of working memory capacity have been contaminated by long-term memory support. We show that when using clearly distinguishable intracategorical items, visual working memory has a maximum capacity of only one object. Because attention is closely involved in the working memory process, our results add to other studies demonstrating capacity limitations of human attention such as inattentional blindness and change blindness.

capacity | vision | working memory

Demonstrations of limitations in the way humans process and store information have generated much controversy among researchers in the cognitive sciences. A major theme concerns the capacity of working memory. Separate verbal, spatial, and visual object working memory systems can be distinguished (1–3), but similar estimates of their capacities have been established. These estimates have sometimes been made in terms of "magical numbers," which have ranged from seven to about four words, numbers, or locations (4). In line with these results, recent studies suggest that it is possible to retain information of up to four objects in visual working memory (5–9).

Studies of visual working memory have routinely used a change detection paradigm (5–7, 10) where a number of distinct objects or colors are presented briefly in a sample array. After a short blank interval, a test array is presented that is identical to the sample array, except for one object that may have changed. The task is to indicate whether all objects in the sample and test arrays are identical or whether one of them has changed. The objects and colors typically used in visual working memory studies have been few, repeatedly presented, and without difficulty classified into distinct categories easily given verbal labels such as red, green, blue, square, disk, etc. Whether the categorical boundaries are rule-, prototype-, or exemplar-based, the categorization process needs support by long-term memory storage. Repeated presentations of a few items, initially not separated by category boundaries, may lead people to develop boundaries and consequently use long-term memory support. Also, because items separated by category boundaries are easily assigned verbal labels, verbal working memory may be activated, leading to overestimation of visual working memory capacity (11). Additional potential problems with the traditional change-detection paradigm are that relational coding influences the results (12) and that spatial memory may be used to assist performance (1). A task that is supposed to give a pure measure of visual working memory capacity of objects should not give the possibility of relationally coding several objects into chunks or be influenced by other memory systems such as spatial memory, verbal working memory, or long-term memory.

We investigated visual working memory capacity for items created with continuous feature dimensions making the items difficult to categorize. The influences of relational coding and spatial coding were reduced by using sample arrays consisting of one to four objects followed after 1,000 ms by a single centrally located test object that was a member of the sample array on 50% of the trials and not a member on the remaining 50% (Fig. 1). Although all objects within all arrays were highly discriminable, some arrays of objects were easily separated by category boundaries and others were difficult to categorize. The goal was to decide whether the test object was present in the sample array. Three different stimulus types were used to create the sample arrays and the test objects in three separate conditions that represented a progression from objects with a few discrete stimulus dimensions that are easily categorized (e.g., red square) to objects with continuous stimulus dimensions not easy to put in separate categories (e.g., ovals with varying aspect ratios and color mixtures crossing the natural boundaries for the perception of color) and with new values on the dimensions on every trial to prevent development of categories. The study presented here is similar to the work on discrimination in visual memory because it also emphasizes continuous dimensions (13, 14). The main difference is that the stimuli used in the discrimination studies are continuous low-level features such as spatial frequency and orientation, whereas the study presented here investigates object classes in a different way. Two object-discriminating features were used in our discrete color/shape condition (discrete colors and discrete shapes; Fig. 2*a*), continuous color/shape condition (color mixtures and oval shapes; Fig. 2*b*), and the continuous size-ratio/shape condition (small ovals inside larger ovals; Fig. 2*c*).

Methods

Eight undergraduate students and the authors participated in each condition in the main experiment. All participants received 50 trials in each condition where each condition was a combination of set size with one to four objects and with one of the three stimulus types. The authors and six other participants also completed three discrimination conditions with stimuli from the discrete color/shape, continuous color/shape, and continuous size-ratio/shape conditions. Eight other participants completed a categorization task that investigated the extent to which the participants saw the objects in each of the three conditions as belonging to the same category or to different categories.

In the discrete color/shape condition, discrete shapes and colors easily put in distinct categories were randomly selected from a set of five prespecified shapes (square, circle, bar, cross, and triangle) and eight colors (red, green, blue, yellow, black, white, magenta, and cyan). The combinations of shapes and colors were selected so that the same shape or the same color never appeared in the same trial. In the continuous color/shape condition, oval aspect ratios were selected from a Euclidian two-dimensional shape-space consisting of orthogonal length and width axes (i.e., aspect ratio space). For each trial, the objects were located at equal distances on a randomly located circle in this shape space. Similarly, the colors were selected from

---

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: CI, confidence interval.

*To whom correspondence should be addressed. E-mail: henrik.olsson@psyk.uu.se.

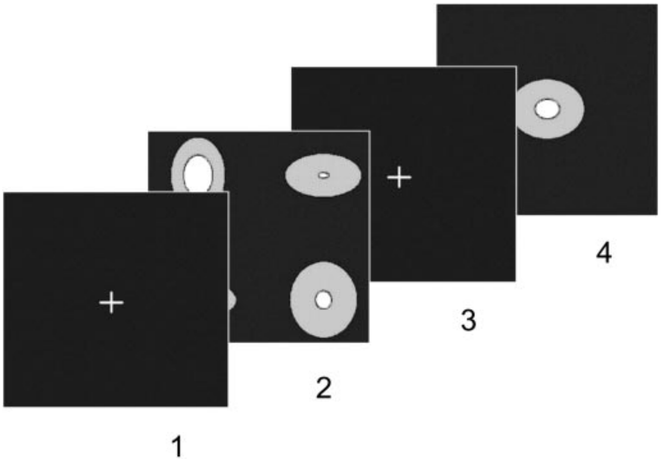

Fig. 1. The sequence of frames in the visual working memory task. For each trial, a fixation cross (frame 1) appeared until observers initiated the sample array (frame 2), which was presented for 500 ms. The fixation cross (frame 3) reappeared instantaneously after the sample array and was presented for 1,000 ms until one centrally presented test object was shown (frame 4). The test object was visible until observers indicated whether it was one of the objects in the sample array.

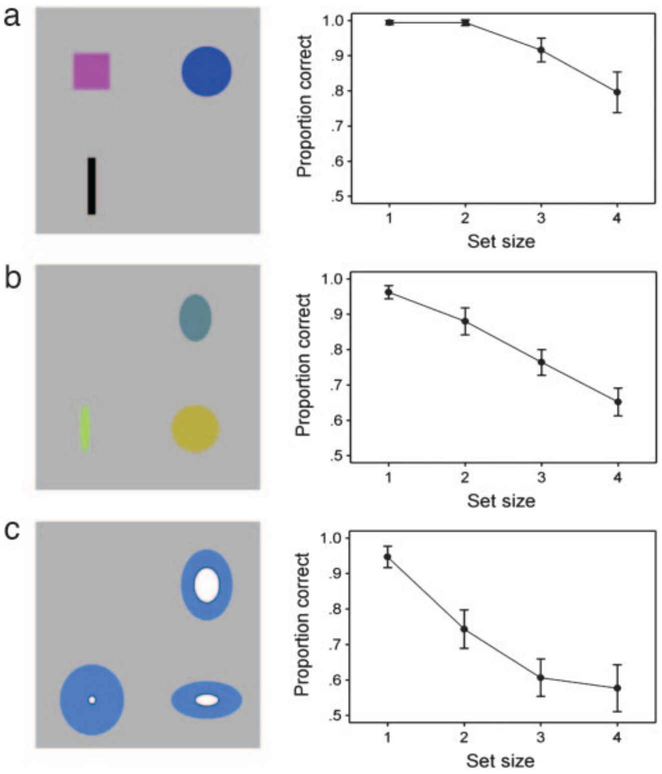

Fig. 2. Examples of stimulus arrays and performance for each condition. (*a*) Discrete color/shape condition. (*b*) Continuous color/shape condition. (*c*) Continuous size-ratio/shape condition. Error bars show 95% CI.

a two-dimensional color space with orthogonal red/green and blue/yellow color axis. The distance between objects in shape-space and color-space were chosen to create highly discriminable shapes and colors. In the continuous size-ratio/shape condition, the size ratios between inner and outer ovals and their aspects were used in conjunction as features specifying the shapes. Size-ratio/shape conjunction stimuli were created by a small oval embedded in a larger oval having the same aspect ratio specified as in the continuous color/shape condition. The size ratio between the small and large oval was between 0.1 and 0.9, and the main axis of the larger oval had a fixed length of about 3°.

All stimulus arrays were presented on a computer screen within a 13° × 13° area with gray background ($3\ \mathrm{cd/m^{-2}}$). This area was divided into four regions, each region occupied by at most one sample object. Each sample object was about 3° wide. When fewer than four objects were presented, their locations were randomized. One test object appeared in the center of the stimulus area 1,000 ms after the offset of the sample array (Fig. 1). The 1,000-ms gap between the sample array and the test object efficiently disrupt iconic memory (10). In 50% of the trials, the test object was identical to one of the objects in the previously presented sample array. The test object was always taken from the same category as the sample objects and appeared as distinct, compared with the objects in the sample array; the objects within the sample array were distinct to each other because their relative distances in the stimulus parameter spaces were the same. Examples of the resulting stimuli are shown in Fig. 2 *a–c*. After each test object, the participants were instructed to indicate whether the test object was one of the objects from the sample display. The sample arrays were visible for 500 ms in all conditions.

We used a formula improved from Pashler (15) by Cowan *et al.* (4, 16) to estimate the maximum number, k, of apprehended objects in visual arrays consisting of N objects. When the test object is a member of the sample array, the probability is k/N that it belongs to one of the apprehended objects and $(N - k)/N$ that it belongs to the not apprehended objects. If g is the guessing rate of responding "yes" (a member) in trials when the test object was a member of the sample array but not apprehended (accidental hits), then the total hit rate is $H = k/N + [(N - k)/N]g$. The guessing rate of responding "no" (not member) is then $1 - g$ when the test object was neither apprehended or a member of the sample array (accidental correct rejections), and the total correct rejection rate is $CR = k/N + [(N - k)/N](1 - g)$. By adding these equations to get the total proportion of correct responses and rearranging terms, we get the capacity estimate of $k = (H + CR - 1)N$. We calculated k separately for each set size. To get as close as possible to the maximum value of a participant's visual working memory capacity, we used the highest estimated capacity across set sizes.

To confirm that all objects were highly discriminable in all conditions, the authors and six other participants performed a same or different discrimination task. There were 30 pairs of objects created by the stimulus-generation procedure from each of the three conditions. Each pair of objects appeared side by side for 200 ms on the display area. The objects were identical in 50% of randomly assigned trials and dissimilar in the remaining trials. The mean (median) discrimination performance was 98.3% (98.3%) for the size-ratio/shape items, 98.8% (100%) for the continuous color/shape items, and 99.6% (100%) for the discrete items. The few errors probably arise because of slips of attention during the 200-ms presentation, and we therefore conclude that it is highly unlikely that low perceptual discrimination in the encoding will lead to underestimation of memory capacity.

To confirm that the objects used in the discrete color/shape condition were seen as belonging to different categories and that those in each of the other conditions were seen as belonging to the same category, eight participants rated to what extent they saw the objects used in each of the three conditions as belonging to the same category. The participants were told the following: "You are going to see four figures at a time on the computer screen, and your task is to rate to what extent you believe that the figures belong to the same category. You answer on a scale

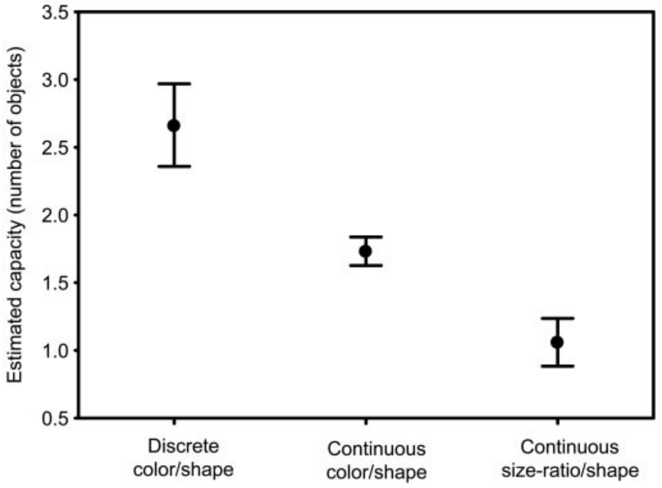

Fig. 3. The estimated maximum capacity (*k*) of visual working memory for each condition. Error bars show 95% CI.

from 1 to 9, where 1 indicates that the objects belong to different categories and 9 indicates that the objects belong to the same category. You answer with the number keys on the keyboard." On each trial, the participants saw four objects from each of the three conditions used in the main experiment. This procedure was repeated four times for each condition. The results showed that the participants were inclined to see the discrete color/shape objects as belonging to different categories and the objects in each of the other conditions as belonging to the same category. The differences between the mean rating of the discrete color/shape objects and the mean ratings of objects from each of the other conditions were large, with nonoverlapping confidence intervals (CIs). The mean values for the discrete color/shape objects, the continuous color/shape objects, and the continuous size-ratio/shape objects were 1.75 (95% CI, 0.88–2.62), 6.66 (95% CI, 5.38–7.93), and 7.28 (95% CI, 6.63–7.94), respectively.

Results

The proportion of correct answers over all conditions and set sizes in the main experiment was 0.84, and the corresponding mean estimated working memory capacity, *k*, was 1.46 objects for the authors. The corresponding values were 0.84 (95% CI, 0.80–0.88) and 1.49 (95% CI, 1.28–1.71) for the participants that were naïve to the purpose of the experiment. The inclusion of the means of the authors in the naïve participants' 95% CI indicated that there were little overall differences in performance between the authors and the naïve participants, and the results were therefore merged.

In the discrete color/shape condition, categorical colors were used, and the shapes were easily put into separate categories, and consequently, the performance in this condition was the highest for all set sizes (Fig. 2*a*) with an estimated memory capacity, *k*, slightly below three objects (Fig. 3). Studies using the change-detection paradigm with similar objects have generated capacity estimates of up to four objects (5, 6). Our smaller estimate may suggest that position coding or relational coding has been used in these previous studies to facilitate memory performance. That is, instead of judging whether all test items were identical to all sample items, the participants may simply judge the overall similarity between the two displays without comparing individual items. Here, only one centrally displayed test item was shown making relational coding a less efficient strategy. In the continuous color/shape condition, where the colors and the shapes are unlikely to reappear during an experimental session and the items were difficult to place in separate categories, the performance was significantly lower than in the discrete color/shape condition for set sizes above one object (Fig. 2*b*) and the estimated capacity was slightly below two objects (Fig. 3).

The continuous color/shape condition may have occasionally invited categorical separation, because sometimes the random assignment of colors made the shapes belong to separate natural color categories. It also is possible that shape and color are features that activate separate memory systems (17). We therefore examined working memory capacity for pure shape and size conjunctions. In the continuous size-ratio/shape condition, the stimuli were difficult to place in separate categories. Consequently, the performance for set sizes higher than one was lower than in the continuous color/shape condition (Fig. 2*c*), with an estimated memory capacity slightly above one object (Fig. 3).

In the continuous size-ratio/shape condition, no consistent differences were observed for *k* between set sizes one and two or between three and four. The value of *k* was somewhat higher for set sizes one and two than for set sizes three and four, indicating some interference in visual working memory. This difference (mean $\Delta k = 0.27$; 95% CI, −0.06–0.60), however, was not statistically reliable because the 95% CI included $\Delta k = 0$. With a set size of one object, the participants were 100% correct in discrete conditions (Fig. 2*a*) but declined to ≈95% correct when continuous stimulus dimensions were used to create the objects (Fig. 2 *b* and *c*). One interpretation of the drop in performance is that within category boundaries, people have only a coarse representation of the objects in visual working memory, although they are easily discriminated perceptually during very brief presentations. When a purely visual representation is no longer supported by the external object, it gradually degrades in quality over time and becomes more difficult to match to an identical test object when both objects belong to the same category.

Discussion

The results demonstrate that, for intracategorical items, visual working memory has an ability to store only one object. As suggested by the surprise the participants showed about the difficulty of the task, our introspective estimations of the capability of visual working memory for such items are typically overestimated. This overestimation may occur because our experience of using visual memory in natural settings outside the laboratory typically involves objects that belong to separate categories. Categorical structures kept in long-term memory may be required for the retention of up to four categorically distinct objects in visual working memory, but when the objects belong to the same category, only one object can be retained. Previous investigations may have arrived at higher estimates of visual working memory capacity, because the items used have been easily classified in separate categories. Also, because only a limited number of objects, repeatedly presented to the participants, have been used, it is possible that categorical structures have developed during the experiment.

Alvarez and Cavanagh (7) used objects that were hard to verbalize and most likely belonged to the same category (e.g., random polygons or Chinese characters) and yet observed memory capacity estimates of two to three objects, which is higher than our estimate of one object. The higher estimate may be caused by relational coding offered by their use of the change-detection method. It should be noted that the effects of relational coding on accuracy and capacity estimates are nontrivial. In experiment 1 in the work of Jiang *et al.* (12), accuracy increased significantly from 65% correct for a single probe to 74% correct when the traditional change detection paradigm was used. Furthermore, Vogel *et al.* (6) concluded from experiments using the traditional change detection paradigm that "the capacity of visual [working memory] appears to be rather consistent at about three to four items in naïve participants" (p. 109). We found an estimate that was significantly lower than three in our discrete color/shape condition that used similar stimuli to those used by Vogel *et al.* but a single probe that reduces the influence of relational coding. Finally, Alvarez and Cavanagh used repeated presentations of a few items, which could invite reliance on long-term memory and thereby inflate capacity estimates. Alvarez and Cavanagh also showed that visual working memory capacity varied across stimulus sets and that the capacity was inversely related to search rate, which was interpreted as a measure of visual information. In our view, no concluding empirical evidence exists that can dissociate between the categorization hypothesis and the visual information hypothesis. The two hypotheses may be intrinsically linked if categorization is seen as a recoding process that reduces the information load. As formulated by Miller (18), recoding means that the input given in a code with many chunks containing a few bits per chunk (such as the features of objects) is recoded into few chunks where each chunk contains a lot of information, as happens when classifying objects into categories.

It has been argued that nonvisual coding such as verbal labels may preserve binding so that the features are bound to the right object during retention (17). For example, when a red square and a green disk are presented, this also is what should be remembered, rather than a green square and a red disk. In line with this proposal, the close relation between labeling and classification suggests that category formation preserves binding. One problem with using objects that can be easily labeled is that verbal memory may become activated causing visual working memory capacity to be overestimated. A concurrent verbal load, such as holding two digits (5), often is used to prevent verbal memory interference in the main memory task. It cannot be ruled out, however, that the digits are stored in long-term memory or that the verbal load is insufficient (19). Morey and Cowan (20) showed that when memory load was two random digits, or familiar seven-digit telephone numbers, no interference was found with the retention of categorically colored squares. However, performance declined with a memory load of seven random digits, especially for trials when performance of recalling the digits was inaccurate. One possibility is that rehearsal mechanisms for maintaining numbers are recruited to maintain object attention (21).

Some results suggest that when the constituent features of an object are independent, such as color and shape, the memory capacity is higher than when the features belong to the same dimension (17). This hypothesis gains support from absolute judgment tasks, where people distinguish novel stimuli by assigning numbers to magnitudes of various characteristic of the stimuli. About 50 years ago, Miller (18) argued in his classical paper that in absolute judgment tasks, the capacity for distinguishing stimuli is restricted when the stimuli are separated in a one-dimensional feature space. The capacity increases, however, when the number of perceptually independent features of stimuli increases (18).

Care must be taken when counting the number of independent features of stimuli. For instance, whereas wavelength is a one-dimensional feature of light, its perceptual quality is described by a two-dimensional color space. It is unclear whether the sizes and shapes of the different features belonging to the same object are perceptually independent. It may be the case that each feature is stored as a separate object in a single memory system so that memory capacity decreases as the number of constituent features increases. Alternatively, if each feature is stored in a separate memory system, memory capacity should increase with the number of features. We cannot rule out the possibility that the participants coded single dimensions instead of complete objects. However, the generic principle of perceptual organization states that arrays of separate objects are unlikely by accident to form regular patterns in the image. When items such as those used in the size ratio/shape condition appear, where the smaller oval is perfectly centered on the larger one, the visual system has a tendency to merge the features into objects. On the other hand, in the continuous two-dimensional color/shape condition, we found larger capacity estimates than in the two-dimensional size ratio/shape condition. It is unclear whether color and shape are treated as independent features that are stored in separate memory systems or whether this enhancement of memory capacity is caused by the fact that the colors occasionally crossed the categorical boundaries. It may be that the capacity is not so much determined by the number of constituent features of objects (5), especially when they are perceptually independent, but rather by categorical separability.

It has been demonstrated that discriminations of low-level features such as spatial frequency, contrast, and orientation across time spans up to tens of seconds can be made almost as accurate as when the items are presented side by side. Even multiple discriminations of independent dimensions can be made without increase in thresholds. The discriminability decreases, however, when two discriminations are required on the same dimension (13). In other words, interference in multiple discriminations occurs within but not between dimensions, suggesting parallel memory processes in low-level visual memory (13, 14). These results further support Miller's claim that the capacity for distinguishing stimuli is restricted when the stimuli are separated in a one-dimensional feature space and increases when the number of perceptually independent features of stimuli increases. Although our experiments investigate capacity as number of objects rather than dimensions, the two approaches bear some similarities, at least to studies using sample arrays with multiple stimuli (22, 23). However, objects such as squares and crosses are of a different quality than dimensions such as contrast and orientation. Although memory discriminability for spatial frequency, contrast, and orientation is close to perceptual discriminability (13, 14), it is possible that perceptual discriminability of our objects was superior to discriminability in memory. Still, it is important that perceptual discriminability was close to 100% correct in all conditions, showing that the memory capacity limitations were not due to perceptual limitations in the encoding phase. When inspecting the stimulus arrays before the experiment, observers confirmed that the objects were considerably dissimilar, and they were very confident in being able to correctly store all four objects in the sample array in memory for later identification. Furthermore, an inherent property of categorical structures is that psychological similarities between objects within a category are likely to be higher than between objects taken from different categories.

Accurate visual working memory requires the preservation of attention over the retention time and fails when attention is directed to other objects (17, 21). This result suggests that inattentional blindness and change blindness (1), where humans occasionally fail to notice highly distinct events outside the focus of attention, may be related to the capacity limitations in visual working memory. Also, evidence from studies using serial presentations of letters suggests that the focus of attention can accommodate only a single item (24), although this accessibility expands to four items after practice (25). One possibility is that, to keep a purely visually represented object active in working memory requires that during retention, attention has not been directed at any other object. That is, when attention is directed to any new object belonging to the same category, the object representation in visual working memory is updated by an overwriting process (26, 27).

We thank Ben Newell and Peter Juslin for comments on previous drafts of this paper. This work was supported by grants from the Swedish Research Council and the Magnus Bergwall Foundation.

1. Simons, D. J. (1996) *Psychol. Sci.* **7,** 301–305.
2. Tresch, M. C., Sinnamon, H. M. & Seamon, J. G. (1993) *Neuropsychologia* **31,** 211–219.
3. Smith, E. E. & Jonides, J. (1997) *Cognit. Psychol.* **33,** 5–42.
4. Cowan, N. (2001) *Behav. Brain Sci.* **24,** 87–185.
5. Luck, S. J. & Vogel, E. K. (1997) *Nature* **390,** 279–281.
6. Vogel, E. K., Woodman, G. F. & Luck, S. J. (2001) *J. Exp. Psychol. Hum. Percept. Perform.* **27,** 92–114.
7. Alvarez, G. A. & Cavanagh, P. (2004) *Psychol. Sci.* **15,** 106–111.
8. Vogel, E. K. & Machizawa, M. G. (2004) *Nature* **428,** 749–751.
9. Todd, J. J. & Marois, R. (2004) *Nature* **428,** 751–754.
10. Phillips, W. A. (1974) *Percept. Psychophys.* **16,** 283–290.
11. Ceraso, J. (1985) in *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed. Bower, G. (Academic, New York), pp. 179–210.
12. Jiang, Y., Olson, I. R. & Chun, M. M. (2000) *J. Exp. Psychol. Learn. Mem. Cognit.* **26,** 683–702.
13. Magnussen, S. & Greenlee, M. W. (1997) *J. Exp. Psychol. Hum. Percept. Perform.* **23,** 1603–1616.
14. Magnussen, S. (2000) *Trends Neurosci.* **23,** 247–251.
15. Pashler, H. (1988) *Percept. Psychophys.* **44,** 369–378.
16. Cowan, N., Elliot, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A. & Conway, A. R. A., *Cognit. Psychol.*, in press.
17. Wheeler, M. E. & Treisman, A. M. (2002) *J. Exp. Psychol. Gen.* **131,** 48–64.
18. Miller, G. A. (1956) *Psychol. Rev.* **63,** 81–97.
19. Cowan, N. (1998) *Trends Cognit. Sci.* **2,** 77–78.
20. Morey, C. C. & Cowan, N. (2004) *Psychon. Bull. Rev.* **11,** 296–301.
21. Awh, E. & Jonides, J. (2001) *Trends Cognit. Sci.* **5,** 119–126.
22. Palmer, J. (1990) *J. Exp. Psychol. Hum. Percept. Perform.* **16,** 332–350.
23. Wright, M., Green, A. & Baker, S. (2000) *Visual Cognit.* **7,** 237–252.
24. McElree, B. (2001) *J. Exp. Psychol. Learn. Mem. Cognit.* **27,** 817–835.
25. Verhaeghen, P., Cerella, J. & Basak, C. (2004) *J. Exp. Psychol. Learn. Mem. Cognit.* **30,** 1322–1337.
26. Lakha, L. & Wright, M. J. (2004) *Vision Res.* **44,** 1707–1716.
27. Enns, J. T. & Di Lollo, V. (2000) *Trends Cognit. Sci.* **4,** 345–352.