A new definition of visual short-term memory
Sligte, I.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
3. Detailed sensory memory, sloppy working memory

Visual short-term memory (VSTM) enables us to actively maintain information in mind for a brief period of time after stimulus disappearance. According to recent studies, VSTM consists of three stages - iconic memory, fragile VSTM, and visual working memory - with increasingly stricter capacity limits and progressively longer lifetimes. Still, the resolution (or amount of visual detail) of each VSTM stage has remained unexplored and we test this in the present study. We presented people with a change detection task that measures the capacity of all three forms of VSTM, and we added an identification display after each change trial that required people to identify the “pre-change” object. Accurate change detection plus pre-change identification requires subjects to have a high-resolution representation of the “pre-change” object, whereas change detection or identification only can be based on the hunch that something has changed, without exactly knowing what was presented before. We observed that people maintained 6.1 objects in iconic memory, 4.6 objects in fragile VSTM and 2.1 objects in visual working memory. Moreover, when people detected the change, they could also identify the pre-change object on 88 percent of the iconic memory trials, on 71 percent of the fragile VSTM trials and merely on 53 percent of the visual working memory trials. This suggests that people maintain many high-resolution representations in iconic memory and fragile VSTM, but only one high-resolution object representation in visual working memory.

Introduction

Look around you and consider the richness of the visual world revealing itself anew with each eye movement you make. Then close your eyes for a brief period of time and try to bring back an internal image of what you have just seen. You will probably realize that you can remember little of what you have just seen, with the exception of a few visual “hotspots” or objects that seem to last in your mind’s eye. This distinction between the richness of your immediate perception and the impoverished image you keep in memory finds its analogue in different forms of visual short-term memory (VSTM); for a fraction of a second after image disappearance, iconic memory maintains a high-capacity representation of the outside world (Averbach & Coriell, 1961; Sperling, 1960), while visual working memory maintains a maximum of four objects for longer periods of time (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001).

Recent studies have suggested another form of VSTM that operates in between iconic memory and visual working memory. In the design of these studies, a partial-report cue is presented during the delay of a change detection task and the cue retrospectively singles out the item to change before the potential change occurs (so-called retro-cue). To be effective, a retro-cue requires people to search their memory for the identity of the object that was presented at the signaled location before. Using this procedure, several studies have shown that retro-cues dramatically boost change detection performance (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Lepsien, Griffin, Devlin, & Nobre, 2005; Lepsien & Nobre, 2007; Makovski & Jiang, 2007; Makovski, Sussman, & Jiang, 2008; Matsukura, Luck, & Vecera, 2007; Sligte, Scholte, & Lamme, 2008, 2009) compared to when the same cue is presented after the change (so-called post-change cues). Then, capacity is limited to 4 objects, which is the well-known limit of visual working memory (Luck & Vogel, 1997; Vogel, et al., 2001). Moreover, increases in change detection performance caused by a retro-cue are not due to grouping processes (Sligte, et al., 2008), speed-accuracy trade-offs (Griffin & Nobre, 2003; Lepsien, et al., 2005), response biases (Griffin & Nobre, 2003), eye movements (Griffin & Nobre, 2003; Matsukura, et al., 2007), or articulation (Makovski & Jiang, 2007; Makovski, et al., 2008). The most surprising finding, however, is the fact that retro-cues boost performance even when they are presented four seconds after stimulus disappearance (Lepsien & Nobre, 2007; Sligte, et al., 2008, 2009), which is far beyond the lifetime of iconic memory.

In a previous study (Sligte, et al., 2008), we systematically evaluated whether this late boost in retro-cue performance taps into the same form of sensory memory as early retro-cues do. We found that when a retro-cue was shown 10 ms after offset of the memorized
Detailed sensory memory, sloppy working memory

display, people could report 30 items (out of 32 items shown) when the memorized display contained high-contrast stimuli, but only 20 (out of 32 items shown) when the display contained isoluminant stimuli. In addition, when light flashes were presented before this early retro-cue, the difference in performance between high-contrast and isoluminant stimuli disappeared. This suggests that retinal afterimages are partially responsible for the increased retro-cue performance just after stimulus offset. When the retro-cue was presented 1,000 ms after offset of the memorized display, we observed that people could report a maximum of 15 items (out of 32 items shown). Importantly, we found no differences in late retro-cue performance between high-contrast and isoluminant stimuli. Moreover, light flashes before the late retro-cue did not influence performance, whereas the presence of new and irrelevant object before the cue greatly reduced performance. Finally, from other work of our lab it was evident that late retro-cues tap into a memory store wherein features are bound to form coherent objects (Landman, et al., 2003). These combined results indicate that late retro-cues tap into a high-capacity form of VSTM that is different from the classic notion of iconic memory. Altogether, it seems that a cued change detection task with early retro-cues, late retro-cues and post-change cues is a robust way to gauge the capacity of three different forms of visual short-term memory in a single experiment, using the same stimuli and cues for each of the three memory types. We will refer to these three forms of memory as iconic memory, fragile VSTM and visual working memory in the rest of this paper.

While VSTM thus seems to consist of three stages with large differences in capacity, it is unclear how detailed objects are represented in each form of VSTM. According to the current consensus, sensory memory is a raw snapshot of the features in a visual scene and these floating features are not bound together to form coherent objects. It is only in visual working memory where (a limited set of) integrated object representations are retained (Luck & Vogel, 1997; Vogel, et al., 2001). Based on these ideas, one would expect that iconic memory and fragile VSTM contain many low-resolution object representations, while working memory contains a limited set of high-resolution object representations. To make clear what we mean by saying high- or low-resolution representations, please take a look at Figure 3.1A. In the trial shown, the motorcycle changes into a frog, so there is a clear color change. People could decide to press change, because they have noticed this color change, but this does not necessarily mean that they maintained the object “motorcycle” in short-term memory. In that sense, low-resolution representations are just as useful as high-resolution representations in supporting change detection performance and measuring change detection performance alone does not reveal the resolution of object representations.
Figure 3.1 Experimental design A Subjects performed a change detection task to measure the capacity of short-term memory representations (black box). After each change trial, an identification display was presented that contained the item that was present in the memory display, but not anymore in the match display (so-called pre-change item) in addition to three distracter items that were present in neither memory nor test display (red box). We assume that high-resolution representations support both change detection and identification, whereas low-resolution representations support change detection or pre-change identification only. B Early retro-cue condition; 10 ms after off-set of the memory display a spatial cue was presented that singled out the item that changed in 50% of the trials. Effectively, this condition measures iconic memory. C Late retro-cue condition; 1s after offset of the memory display, but before the on-set of the test display, a spatial cue was presented. Effectively, this condition measures fragile VSTM. D Post-change cue condition; 100 ms after onset of the test display, a spatial cue was presented. This condition measures only visual working memory.
To probe the resolution or visual detail of VSTM representations, we adopted a method developed by Levin and colleagues (Beck & Levin, 2003; Mitroff, Simons, & Levin, 2004). In their approach, an identification line-up is shown after each change detection trial that asks people to identify the pre-change object, the post-change object and/or one of the non-changing objects among one or more distracter objects that were presented in neither display. It was observed that post-change object identification was relatively good, but pre-change object identification was far worse. These results thus seem to suggest that standard change detection tasks measure a mix of high- and low-resolution representations.

In the present study, we aimed to measure both the capacity and the resolution of iconic memory, fragile VSTM and visual working memory. In the general set-up of our task (Fig. 3.1A), a memory display containing multiple objects was shown, followed by a retention interval, after which a test display was shown and subjects had to indicate on each trial whether a particular (cued) object changed between memory and test display. To probe the resolution of VSTM representations, we introduced an identification display after each change trial. This identification display contained four objects; one object that was present in the memory display, but not in the test display (so-called pre-change item) and three distractors that were present in neither. In addition, cues were presented either 1) 10 ms after memory display off-set to measure iconic memory (Fig. 3.1B), 2) 1,000 ms after memory display off-set, but during the retention interval to measure fragile VSTM (Fig. 3.1C), or 3) 1,000 ms after memory display off-set, but after the possible change had already occurred to measure visual working memory (Fig. 3.1D).

Methods

Subjects
20 students (11 females) with normal or corrected-to-normal vision and no colour deficiencies participated in this study. Subjects were rewarded with course credits for their participation. All subjects gave their written informed consent to participate in the experiment, which was approved by the local ethics committee of the department of Psychology of the University of Amsterdam.

Equipment
The experiment was done on a 19 inch LG CRT-display (type FB915BP) at a refresh rate of 100 Hz. We measured phosphor persistence of the display using a photo-cell placed at the centre of the screen. Phosphors returned to baseline activity 6.4 ms after their peak
A new definition of visual short-term memory

amplitude (see Sligte, et al., 2008 for data). Stimuli were presented on screen with Presentation (NeuroBehavioral Systems, Inc.).

Stimuli
We selected 50 coloured line drawn objects from a series of 260 objects created by Rossion and Pourtois (Rossion & Pourtois, 2004) that can by found on the web (titan.cog.brown.edu:8080/TarrLab/, courtesy of Michael J. Tarr). In addition, we created greyscale versions of these images with the use of Matlab (Mathworks, Inc.). All objects used can be found in Supplementary File 1.

Subjects were shown memory and test displays containing eight (out of 50) randomly selected objects (about 1° by 1° of visual angle) placed radially at four degrees eccentricity around a red fixation dot (0.1° by 0.1° of visual angle; 13.52 cd/m$^2$). All stimuli were presented on a pure white background (87.66 cd/m$^2$). An example of a memory display is depicted in Figure 3.1A.

After each change trial, an identification display containing four objects was shown. Objects in this display were placed horizontally at -1.5°, -0.5°, 0.5° and 1.5° of visual angle with respect to the centre of the screen. One object in this display had been presented in the memory display, but not in the test display (so-called pre-change item). The other three objects were neither shown in the memory nor the test display and were randomly chosen from all objects that were not used in the trial (N = 41).

Task
On each trial, the red fixation dot in the middle of the screen turned green for 1,000 ms to indicate the start of the trial. Thereafter, we showed a 250-ms memory display containing 8 objects that were either all coloured or all in greyscale. Subjects were instructed to remember as many objects of this memory display as possible. On each trial, one object was cued to indicate which item was the one to report. After a retention interval in which no stimulation was provided, a test display was shown and subjects were asked to indicate by button press whether the cued item was the same (50 % of the trials) or a different (50 % of the trials) object than was shown at the same location in the memory display. Test displays were present for 2,000 ms or until the subject made a response.

Spatial cues were introduced at different latencies during the trial; either 10 ms after off-set of the memory display (early retro-cue; Fig. 3.1B), 1,000 ms after off-set of the memory display, but before on-set of the test display (late retro-cue; Fig. 3.1C), or 1,000 after off-set of the memory display, but 100 ms after on-set of the test display (post-change
Detailed sensory memory, sloppy working memory

cue; Fig. 3.1D). The interval between memory and test display was 2,000 ms for the early and late retro-cue conditions, and 900 ms for the post-change cue conditions. In effect, late retro-cues and post-change cues were provided at the same latency after memory display off-set ruling out differences in capacity due to a different time interval in which subjects had to remember all objects before knowing which object was relevant for detecting a change. All conditions were presented randomly intermixed and subjects received auditory feedback on whether they had responded correctly or not.

After each change trial, irrespective of whether the subject detected the change or not, an identification display was shown. This identification display contained the pre-change object, i.e. the object that was in the memory display but changed to another object in the test display, and three distracter objects that were in neither displays. We chose to present identification display only on change trials, because subjects know during the test display which item is relevant for detecting the change. If this single (non-changed) item is then repeated during the identification display, the task will be trivially easy. The identification display was shown until the subject made a response. Again, subjects received auditory feedback about the correctness of their response.

Procedure

First, we tested subjects on visual acuity and colour blindness. Thereafter, they were trained for a maximum of three blocks of 60 trials on a basic version of the task containing simple oriented rectangles instead of line drawings. We did this on the one hand for participants to learn the task and on the other hand to have an objective criterion for when participants had learned the task. In previous experiments (Sligte, et al., 2008), we consistently found that subjects could remember about 4 simple items in post-cue conditions, about 6 in late retro-cue conditions, and about 7 to in early retro-cue conditions. On average, subjects would then maintain about 5.6 objects in memory over conditions corresponding to a performance level of 85 percent (calculated with Cowan’s K; see Data Analysis for details). After subjects had reached this performance level, they were trained for one block (60 trials) on the actual experiment containing line drawings and the identification display.

Subjects performed 50 trials in each condition, cue-timing (3) x change-present (2) x colour/grey-scale (2), resulting in a total of 600 trials. Subjects were asked to keep fixating the dot in the middle of the screen, at least until the (potential) identification display appeared. Every six minutes, the experiment was paused and subjects were required to take a few minutes rest. In total, the experiment lasted about two hours. At the end of the
experiment, subjects received course credits for their participation and they were debriefed about the goal of the experiment.

Data analysis
We computed memory capacity using a formula developed by Cowan (Cowan, 2001). The formula is \(K = (\text{hit rate} - \text{chance} + \text{correct rejection} - \text{chance}) \times \text{number of objects presented}\). This formula provides an estimate of the representational capacity and corrects for guessing trials. To calculate the number of high-resolution object representations, we multiplied Cowan's K with performance on the identification task. We do this on the assumption that high-resolution representations support both change detection and subsequent identification of the pre-change item. Object representations are lower in resolution, however, when they support only change detection or identification (but not both). All statistical analyses were performed with repeated measures ANOVAs. In some cases, we tested specific differences with paired t-tests.

Results
In the present study, we aimed to assess the capacity and resolution of iconic memory, fragile VSTM, and visual working memory. Capacity was estimated by a linear transformation of change detection performance into the capacity estimate \(K\) (Cowan, 2001). Resolution (or amount of visual detail) was estimated by presenting an identification display after each change trial that required people to identify the item that was present in the initial memory array, but not anymore in the test array (see Fig. 3.1A). We presume that representations supporting change detection and subsequent identification of the pre-change item are more detailed or higher-resolution representations than representations that support change detection or identification only. In the following section, we will first present capacity estimates for each form of VSTM. Then, we will present differences in representational resolution between VSTM stages.

VSTM capacity
Subjects could report on average 5.8 greyscale (performance: 86.7%) and 6.3 (89.6%) coloured objects in early retro-cue conditions, 4.3 (76.7%) greyscale and 4.8 (80.6%) coloured objects in late retro-cue conditions, and 2.0 (62.4%) greyscale and 2.3 (64.3%) coloured objects in post-change cue conditions (Fig. 3.2A). Repeated measures ANOVAs revealed that people could retain slightly more items in memory when they were coloured than when they were presented in greyscale \(F(1,19) = 33.51, p < .001, \eta^2 = .638\). Moreover,
Detailed sensory memory, sloppy working memory

performance decreased significantly over conditions ($F(1,19) = 87.71, p < .001, \eta^2 = .907$). As revealed by subsequent post-hoc paired t-tests, memory capacity was highest when an early retro-cue was provided compared to a late retro-cue (greyscale: $t(19) = 6.43, p < .001$; colour: $t(19) = 7.38, p < .001$). In addition, memory capacity was higher when a late retro-cue compared to a post-change cue was shown (greyscale: $t(19) = 7.67, p < .001$; colour: $t(19) = 8.12, p < .001$).

Change detection performance tended to improve over the course of the experiment (5 bins of 10 trials per condition; $F(4,16) = 3.00, p = .050$), but the improvement was not large (bin 1; 0.9% below mean performance over bins; bin 5; 1.1% above mean performance over bins). Surprisingly, performance only got better for greyscale conditions, but not for colour conditions ($F(1,19) = 33.68, p < .001$). We did not observe significant differences in learning curves across VSTM stages.

In sum, there seem to be large differences in capacity between iconic memory and fragile VSTM, and between fragile VSTM and visual working memory. In addition, it seems that the availability of an extra feature (color) boosts change detection performance, and thus capacity, of all forms of VSTM.

Number of high-resolution representations

To derive the number of high-resolution (or visually detailed) representations, we multiplied change detection performance (Fig. 3.2A; expressed as Cowan’s K) with correct performance on the subsequent change identification task (Fig. 3.2B). On average, subjects could report 5.1 (out of 5.8) detailed greyscale and 5.7 (out of 6.3) detailed coloured representations in the early retro-cue condition, 2.9 (out of 4.3) detailed greyscale and 3.7 (out of 4.8) detailed coloured representations in the late retro-cue condition, and only 1.1 (out of 2.0) detailed greyscale and 1.3 (out of 2.3) detailed coloured representations in the post-change cue condition (Fig. 3.2C). Repeated measures ANOVAs revealed that subjects could report more detailed representations when the objects were presented in colour ($F(1,19) = 39.51, p < .001, \eta^2 = .675$), but we observed no benefit of colour when people detected the change without being able to identify the pre-change item ($F(1,19) = .99, p = .33$) (see low-resolution representations in Fig. 3.2C). In addition, memory capacity for detailed representations decreased over conditions ($F(1,19) = 118.66, p < .001, \eta^2 = .930$). Subsequent post-hoc paired t-tests showed that the number of detailed representations was highest when an early retro-cue was provided compared to a late retro-cue (colour: $t(19) = 10.151, p < .001$; greyscale: $t(19) = 9.155, p < .001$), and the capacity for detailed representations was also higher when a late retro-cue compared to a post-change cue was
A new definition of visual short-term memory

shown (colour: $t(19) = 7.800, p < .001$; greyscale: $t(19) = 7.622, p < .001$). We did not observe significant learning effects over the course of the experiment.

To summarize, these results suggest that people initially build up many detailed object representations in iconic memory. When no new stimulation is provided, people tend to forget some of these representations over time. Yet, the major factor for diminished performance is the fact that new stimulation, such as the test display, overwrites all but one detailed representation.

Figure 3.2 Change detection and identification performance

A Change detection performance; people could report six objects in early retro-cue conditions, four and a half objects in late retro-cue conditions, and two objects in post-change cue conditions. When the objects were presented in colour instead of in greyscale, people could remember slightly more objects. Performance is depicted as Cowan’s K, a common method to estimate the representational capacity of short-term memory. B Identification performance after correct change detection; people were able to identify the item that was present in the memory display, but not anymore in the test display on 88% of the early retro-cue trials, on 71% of the late retro-cue trials and on 53% of the post-change cue trials. C To derive the number of high-resolution representations, we multiplied correct change detection performance with correct performance on the identification task. To derive the number of low-resolution representations,
**Availability of low-resolution information in VSTM**

At some trials, subjects were able to detect a change without being able to identify the pre-change item. At other trials, subjects did not detect the change, but did successfully identify the correct pre-change item on the subsequent identification task (see **Fig. 3.3**). We propose that both trials signal the availability of information in VSTM, but the information does not have the same representational quality as information that supports change detection and change identification. Correct identification without change detection might occur because of the certainty of the subject’s response; if a subject is not certain whether a change occurred, he/she might press no-change during change detection. If subsequently the identification display is shown, the change is confirmed and subjects might then rely on the low-resolution representation that was at first not strong enough for them to select the change response. Nevertheless, we have to be cautious to express identification without change detection in terms of the number of objects remembered as subjects could have chosen the right object by chance (1 out of 4). To be sure this is not the case, we first performed one sample t-tests against chance level of 25%.

![Figure 3.3 Identification without change detection](image)

**Figure 3.3 Identification without change detection**

On a proportion of the change trials, subjects were not able to detect the change, but did identify the item that was present in the memory display, but not anymore in the test display. This proportion was about 54% in the early retro-cue conditions, 42% in the late retro-cue conditions and 26% in the post-change cue conditions. Data are depicted as the mean +/- the standard error of the mean.

When subjects did not detect the change, they still identified the pre-change item on 58.0% (greyscale) and 50.7% (colour) of the early retro-cue trials, on 43.2% (greyscale) and 41.2% (colour) of the late retro-cue trials, and on 27.9% (greyscale) and 26.2% (colour) of the post-change cue trials. Statistically, performance exceeded chance levels in early retro-cue conditions for both greyscale ($t(19) = 6.727$, $p < .001$) and coloured objects ($t(19) = 4.224$, $p < .001$). These same results apply to the late retro-cue conditions (greyscale: $t(19) = 4.498$, $p < .001$; coloured: $t(19) = 4.594$, $p < .001$). However, performance in post-change conditions did not exceed chance levels for both greyscale ($t(19) = 1.256$, $p = .224$) and coloured objects ($t(19) = .669$, $p = .511$).
A new definition of visual short-term memory

To further explore the ratio of high-resolution versus low-resolution representations between VSTM stages, we compared trials where people detected and identified the change (Table 3.1, 3rd column) with trials where people detected the change only (Table 3.1, 4th column) or identified the change only (Table 3.1, 5th column). This ratio was 4.07:1 (greyscale) and 6.02:1 (colour) in the early retro-cue condition, 1.47:1 (greyscale) and 1.96:1 (colour) in the late retro-cue condition, and .54:1 (greyscale) and .63:1 (colour) in the post-change cue condition. This suggests that almost all representations in iconic memory are high-resolution representations, that fragile VSTM contains slightly more high-resolution than low-resolution representations, yet visual working memory consists mostly of low-resolution representations.

A final interesting observation is that on the majority (96%) of early retro-cue change trials, people were able to detect something of a change (combining 3rd – 5th column; Table 3.1), somewhat less so on late retro-cue change trials (86%), and even less on post-change cue trials (55%) and this combined change performance is indifferent for whether stimuli were presented in colour or in greyscale ($F(1,19) = 2.411, p = .137$). This might suggest that the absolute capacity of iconic memory, fragile VSTM and visual working memory is identical for colour and greyscale stimuli, but that colour adds to the resolution of the representation.

### Table 3.1 Performance on different VSTM conditions

Subjects performed a change detection task that measures iconic memory (Fig. 3.1B; IC), fragile VSTM (Fig. 3.1C; fVSTM), and visual working memory (Fig. 3.1D; VWM) in a single experiment. In addition, after each change trial an identification display was shown that required subjects to identify the pre-change item. Here, we present the proportion correct on no-change trials (No-change performance) and the proportion correct on change trials (Change performance). In addition, we present the proportion of change trials where subjects correctly detected the change and identified the pre-change item (Detection & Identification), correctly detected the change without correct identification (Detection only), and correctly identified the change without change detection (Identification only). Reported values are averages (+ SEM) for greyscale and color images apart.

<table>
<thead>
<tr>
<th></th>
<th>No-change performance</th>
<th>Change performance</th>
<th>Detection &amp; Identification</th>
<th>Detection only</th>
<th>Identification only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC</td>
<td>fVSTM</td>
<td>VWM</td>
<td>IC</td>
<td>fVSTM</td>
</tr>
<tr>
<td>Grey</td>
<td>84.6</td>
<td>(1.6)</td>
<td>79.7</td>
<td>(1.9)</td>
<td>88.0</td>
</tr>
<tr>
<td>Color</td>
<td>86.9</td>
<td>(1.3)</td>
<td>82.2</td>
<td>(2.2)</td>
<td>88.8</td>
</tr>
</tbody>
</table>

To further explore the ratio of high-resolution versus low-resolution representations between VSTM stages, we compared trials where people detected and identified the change (Table 3.1, 3rd column) with trials where people detected the change only (Table 3.1, 4th column) or identified the change only (Table 3.1, 5th column). This ratio was 4.07:1 (greyscale) and 6.02:1 (colour) in the early retro-cue condition, 1.47:1 (greyscale) and 1.96:1 (colour) in the late retro-cue condition, and .54:1 (greyscale) and .63:1 (colour) in the post-change cue condition. This suggests that almost all representations in iconic memory are high-resolution representations, that fragile VSTM contains slightly more high-resolution than low-resolution representations, yet visual working memory consists mostly of low-resolution representations.

A final interesting observation is that on the majority (96%) of early retro-cue change trials, people were able to detect something of a change (combining 3rd – 5th column; Table 3.1), somewhat less so on late retro-cue change trials (86%), and even less on post-change cue trials (55%) and this combined change performance is indifferent for whether stimuli were presented in colour or in greyscale ($F(1,19) = 2.411, p = .137$). This might suggest that the absolute capacity of iconic memory, fragile VSTM and visual working memory is identical for colour and greyscale stimuli, but that colour adds to the resolution of the representation.
Discussion

The standard model of visual short-term memory (VSTM) distinguishes between iconic memory, a brief and high-capacity store, and visual working memory, a sustained store with limited capacity. Recently, we found evidence for an intermediate store in between iconic memory and working memory, both in terms of capacity and in terms of lifetime (Sligte, et al., 2008). Based on the fragile nature of this intermediate store, we have termed it fragile VSTM. While it is evident that there are large capacity differences between all VSTM stores, it remains unclear how detailed representations are stored in each form of VSTM.

In the present paper, we measured capacity and visual detail (or resolution) of iconic memory, fragile VSTM and visual working memory. There were large capacity differences between iconic memory (6 items), fragile VSTM (4.6 items), and visual working memory (2.2 items), and the capacity of all VSTM stages was higher for coloured objects than for greyscale objects. While the observed capacity estimates seem to be relatively low compared to previous studies, we used complex stimuli in the present study that usually yield lower capacity estimates than simple objects (Alvarez & Cavanagh, 2004; Sligte, et al., 2008).

In addition, we found that the majority of iconic memory representations were visually detailed or high-resolution representations (i.e. supporting change detection and pre-change identification). Also, fragile VSTM representations were mostly high-resolution representations. However, visual working memory seemed to contain only one high-resolution object representation in addition to one low-resolution representation. Thus, representations are numerous and rich in detail before visual interference (constituting sensory memory), but after visual interference capacity and resolution of VSTM is limited (constituting visual working memory).

What is the exact nature of a ‘high-resolution’ representation?

We used operational definitions of high- and low-resolution representations: a representation that supports both change detection and identification is ‘visually detailed’ or high-resolution, when it supports change detection or identification only, we consider it to be ‘abstract’ or low-resolution. One might wonder, however, what this means in terms of the nature of that representation.

A key issue in ‘representation land’ is whether features exist in a freely ‘floating’ form or are bound into object representations (Treisman, 1996; Treisman & Gelade, 1980). The displays in Figure 3.1A, for example, consist of many features: there are the colours green, red, yellow, etc. Then there are forms that are primarily (i.e. in their low spatial frequency
A new definition of visual short-term memory

content) vertical, horizontal or diagonal. As far as details of the objects go (i.e. the high spatial frequency domain) there are even more orientations, colours, etc. Some objects fall into categories that may be detected in parallel, such as animal - non-animal (Thorpe, Fize, & Marlot, 1996). In an unbound representation, all these features would be freely floating, meaning that they would not be bound to any specific location in the visual field, nor would they be explicitly linked to each other. In other words, it would not be known whether the butterfly is yellow or green, the pear is standing or lying, or whether the crocodile is somewhere up or below the fixation spot. It is the prerogative of higher level object representations to have the features ‘green’, ‘horizontal’, and ‘animal’ bound into a single ‘object file’ to represent the small crocodile at 11:00 ‘o clock (Kahneman, Treisman, & Gibbs, 1992).

What degree of such binding would be necessary to support both change detection and identification? That is difficult to determine in this study, as the objects that changed, as well as the objects that were used for identification were randomly selected from the set of objects available. They may have differed in any feature dimension, sometimes with large differences in one feature but not in another. In the example of Figure 3.1A, change detection may have been possible according to the change in colour (going from the red motorcycle to the green frog), but identification would not, as all four objects of the identification array are red. But in other cases, the reverse might have been the case, or other features may have played a role. The experiment is not explicit about which features play a role.

Whether high-resolution representations indeed have a higher degree of feature binding than low-resolution representations remains an open question, but we believe they do. As change detection and identification more often than not will depend on different feature dimensions, a higher degree of feature binding would be necessary to perform correctly on change detection and change identification. Moreover, previous research has shown that elementary feature binding is present in iconic memory (Landman, et al., 2003). Still, for a more definitive answer to this question, it might be sensible to combine the present experimental design with a conjunction change detection task (Luck & Vogel, 1997) that is able to assess the degree of feature binding.

Explaining VSTM resolution from a neural perspective

In this section, we present a neural model that might explain how low-resolution representations differ from high-resolution representations (see Fig. 3.4A). The basic idea of the model is that high-resolution representations are formed in primary and secondary visual cortex (V1-V3) during image perception. In lower levels in the visual hierarchy, the receptive
Detailed sensory memory, sloppy working memory

Field size is relatively small (V1 0.5°; V2 1.5°; V3 2.8°; at 4° eccentricity) compared to higher visual areas (V4 4.3°; at 4° eccentricity) (Smith, Singh, Williams, & Greenlee, 2001). As a consequence, higher visual areas maintain less detailed representations (shown as blurred objects in Fig. 3.4) than lower visual areas. In addition, higher visual areas have lower storage capacity as the (larger) receptive fields of neurons “see” multiple objects at the same time, while they can only represent one object at a time. Neurons thus have to “choose” which object to represent and this happens by means of biased competition that in turn causes a modulation in the firing pattern to represent one or the other object (Kastner & Ungerleider, 2001). Thus, when going up in the visual system both capacity and resolution become more limited.

On the basis of these neural attributes, we suggest that high-resolution representations depend on activity in visual areas low in hierarchy (V1-V3; see Fig. 3.4A in yellow), whereas low-resolution representations depend on visual areas higher up in hierarchy (V4/IT; see Fig. 3.4A in blue). In addition, we assume that VSTM maintenance is accomplished by reverberating activity within and between brain regions, or so-called recurrent processing (RP; Lamme, 2003). We propose that the major difference between VSTM stages is whether RP is confined to V1-V3 (iconic memory; in yellow), spreads to include V4 (fragile VSTM; in blue), or even includes key nodes in superior parietal lobe and prefrontal cortex (visual working memory; in red and green). The nodes in superior parietal lobe and prefrontal cortex are special, as they control feedback signals related to spatial and central attention, respectively (Corbetta & Shulman, 2002; Mcnab & Klingberg, 2008; Vogel, McCollough, & Machizawa, 2005; Xu & Chun, 2006). As a consequence of these feedback signals, activity in posterior parts of the brain is boosted and this protects representations against interference by new visual stimulation, such as the test display (Lepsien & Nobre, 2007; Makovski, et al., 2008; Matsukura, et al., 2007). We suggest that when a representation receives top-down spatial attention only, top-down amplification is less strong than when the representation receives both top-down spatial and central attention.

Our model predicts that just after stimulus offset (Fig. 3.4B left-most figure), many representations exist at a low level in the visual hierarchy and these representations support change detection and identification of the pre-change item (iconic memory). As time passes, progressively less items are represented at the lowest level in the visual hierarchy and thus the number of detailed representations supporting change detection and identification will diminish (fragile VSTM). Finally, when new visual stimulation (such as the test display) is shown all representations that have not received top-down amplification are overwritten (visual working memory). The model assumes that representations that have received top-
A new definition of visual short-term memory

down attention from prefrontal and superior parietal lobe are sufficiently protected against visual interference to survive at the V1-V3 level, but representations that have received feedback from the SPL alone are not protected at the V1-V3 level, but do persist at the V4/IT level. In this model, fragile VSTM and visual working memory are also measured in the iconic memory condition (Fig. 3.4B left-most figure), and visual working memory is also measured in the fragile VSTM condition (Fig. 3.4B middle figure).

**Figure 3.4 Explaining VSTM resolution from a neural perspective**

A During image perception, high-resolution representations are formed in primary, visual cortex (V1-V3; in yellow). In higher visual areas (V4/IT; in blue), the receptive field size of neurons becomes larger and as a consequence, the resolution of representations becomes more limited (shown as a blur). Spatial attention (in red), subserved by the superior parietal lobe (SPL) and the frontal eye fields (FEF), imposes even stricter capacity limits on the number of information that can be represented (shown as 4 location slots).
Finally, central attention (in green), speculatively subserved by the dorsolateral prefrontal cortex (DLPFC), can only be directed to one item at a time. A major assumption of the model is that all forms of visual short-term memory depend on recurrent processing. B Representations at the lowest level in the visual hierarchy are high-resolution or visually detailed representations that support change detection and identification, whereas representations at higher levels in the hierarchy are more abstract representations and support change detection or identification only. Just after stimulus offset, many representations exist at the V1-V3 level and these representations are available for report when an early retro-cue, measuring iconic memory, is shown. As time passes, activity at the V1-V3 level comes to a stop. As a consequence, less high-resolution representations are available for report when a late retro-cue, measuring fragile VSTM, is shown. Finally, after visual interference by the test display, all representations at the V1-V3 and the V4/IT level are overwritten. Only the representation that has received top-down spatial and central attention is completely protected against interference. In addition, representations that have received top-down spatial attention are protected at the V4/IT level.

Limitations of the standard change detection task
The change detection paradigm is a currently often-used method for measuring the capacity of visual working memory. With the use of this task, many authors have shown that people can retain a maximum of four items in visual working memory, although memory capacity tends to decrease with increases in stimulus complexity (Alvarez & Cavanagh, 2004; Sligte, et al., 2008). The current study suggests that we have to be cautious to express performance on a change detection task in terms of short-term memory representations, as only in half of the working memory trials, people were able to detect a change and identify the item that was presented before it changed into another item. This implies that change detection performance cannot be equated to the number of full representations that are maintained in short-term memory, but rather signals the number of representations that are sufficiently detailed to detect the current change.