Detailed sensory memory, sloppy working memory

Ilja G. Sligte*, Annelinde R. E. Vandenbroucke, H. Steven Scholte and Victor A. F. Lamme

Cognitive Neuroscience Group, Department of Psychology, University of Amsterdam, Amsterdam, Netherlands

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% of the iconic memory trials, on 71% of the fragile VSTM trials and merely on 53% 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.

Keywords: iconic memory, fragile VSTM, working memory, visual detail, resolution

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 analog 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 (Sperling, 1960; Averbach and Coriell, 1961), while visual working memory maintains a maximum of four objects for longer periods of time (Luck and Vogel, 1997; Vogel et al., 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 and Nobre, 2003; Landman et al., 2003; Lepsien et al., 2005; Lepsien and Nobre, 2007; Makovski and Jiang, 2007; Matsukura et al., 2007; Makovski et al., 2008; Sligte et al., 2008, 2009) compared to when the same cue is presented after the change (so-called *post-change cue*). Then, capacity is limited to four objects, which is the well-known limit of visual working memory

(Luck and 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 and Nobre, 2003; Lepsien et al., 2005), response biases (Griffin and Nobre, 2003), eye movements (Griffin and Nobre, 2003; Matsukura et al., 2007), or articulation (Makovski and Jiang, 2007; Makovski et al., 2008). The most surprising finding, however, is the fact that retro-cues boost performance even when they are presented 4 s after stimulus disappearance (Lepsien and 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 off-set of the memorized 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 off-set. When the retro-cue was presented 1,000 ms after off-set 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 objects 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 gage the capacity of three different forms of VSTM 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 and 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 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.

To probe the resolution or visual detail of VSTM representations, we adopted a method developed by Levin and colleagues (Beck and Levin, 2003; Mitroff et al., 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 (**Figure 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 distracters that were present in neither. In addition, cues were presented either (1) 10 ms after memory display off-set to measure

iconic memory (**Figure 1B**), (2) 1,000 ms after memory display off-set, but during the retention interval to measure fragile VSTM (**Figure 1C**), or (3) 1,000 ms after memory display off-set, but after the possible change had already occurred to measure visual working memory (**Figure 1D**).

MATERIALS AND METHODS
SUBJECTS
Twenty students (11 females) with normal or corrected-to-normal vision and no color 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 center of the screen. Phosphors returned to baseline activity 6.4 ms after their peak amplitude (see Sligte et al., 2008 for data). Stimuli were presented on screen with Presentation (NeuroBehavioral Systems, Inc.).

STIMULI
We selected 50 colored line drawn objects from a series of 260 objects created by Rossion and Pourtois (2004) that can by found on the web (titan.cog.brown.edu:8080/TarrLab/, courtesy of Michael J. Tarr). In addition, we created grayscale versions of these images with the use of Matlab (Mathworks, Inc.). All objects used can be found in **Figure S1** in Supplementary Material.

Subjects were shown memory and test displays containing eight (out of 50) randomly selected objects (about $1° \times 1°$ of visual angle) placed radially at $4°$ eccentricity around a red fixation dot ($0.1° \times 0.1°$ of visual angle; 13.52 cd/m²). All stimuli were presented on a pure white background (87.66 cd/m²). An example of a memory display is depicted in **Figure 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 center 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 eight objects that were either all colored or all in grayscale. 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; **Figure 1B**), 1,000 ms after off-set of the memory display, but before on-set of the test display (late retro-cue; **Figure 1C**), or 1,000 after off-set of the memory display, but 100 ms after on-set of the test display (post-change cue; **Figure 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 color 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 four simple items in post-cue conditions, about six in late retro-cue conditions, and about seven 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% (calculated with Cowan's K; see section 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) × change-present (2) × color/grayscale (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 6 min, the experiment was paused and subjects were required to take a few minutes rest. In total, the experiment lasted about 2 h. 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 (2001). The formula is K = (hit rate − chance + correct rejection − chance) × number of objects presented. This formula provides an estimate of the representational capacity and corrects for guessing trials. To calculate the amount 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 **Figure 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 grayscale (performance: 86.7%) and 6.3 (89.6%) colored objects in early retro-cue conditions, 4.3 (76.7%) grayscale and 4.8 (80.6%) colored objects in late retro-cue conditions, and 2.0 (62.4%) grayscale and 2.3 (64.3%) colored objects in post-change cue conditions (**Figure 2A**). Repeated measures ANOVAs revealed that people could retain slightly more items in memory when they were colored than when they were presented in grayscale ($F(1,19) = 33.51, p < 0.001, \eta^2 = 0.638$). Moreover, performance decreased significantly over conditions ($F(1,19) = 87.71, p < 0.001, \eta^2 = 0.907$). 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 (grayscale: $t(19) = 6.43, p < 0.001$; color: $t(19) = 7.38, p < 0.001$). In addition, memory capacity was higher when a late retro-cue compared to a post-change cue was shown (grayscale: $t(19) = 7.67, p < 0.001$; color: $t(19) = 8.12, p < 0.001$).

Change detection performance tended to improve over the course of the experiment (five bins of 10 trials per condition; $F(4,16) = 3.00, p = 0.050$), 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 grayscale conditions, but not for color conditions ($F(1,19) = 33.68, p < 0.001$). We did not observe significant differences in learning curves across VSTM stages.
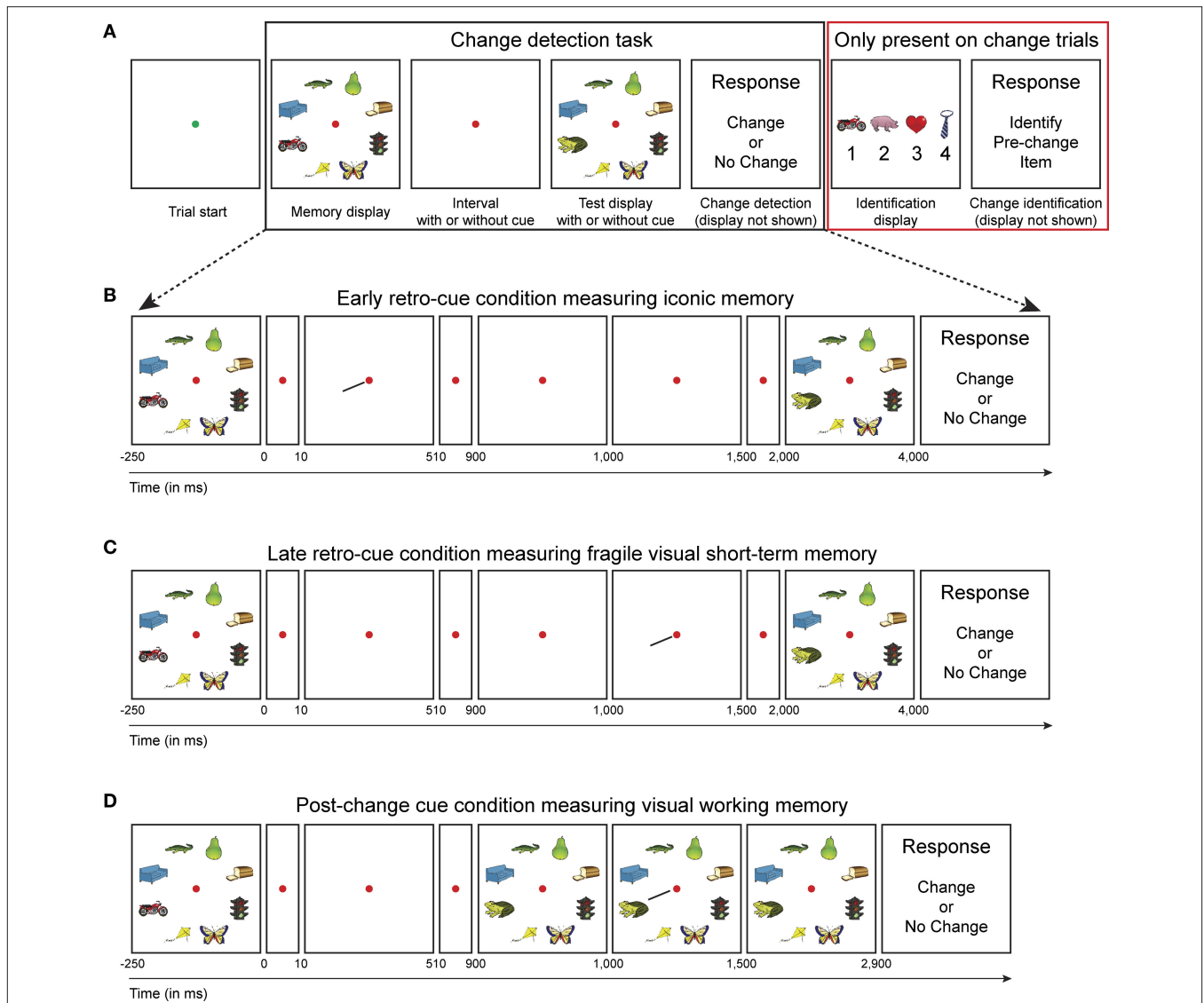
FIGURE 1 | Experimental design. (A) Subjects performed a change detection task to measure the capacity of short-term memory representations (black box). After each change trial, an identification display was presented that contained the item that was present in the memory display, but not anymore in 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-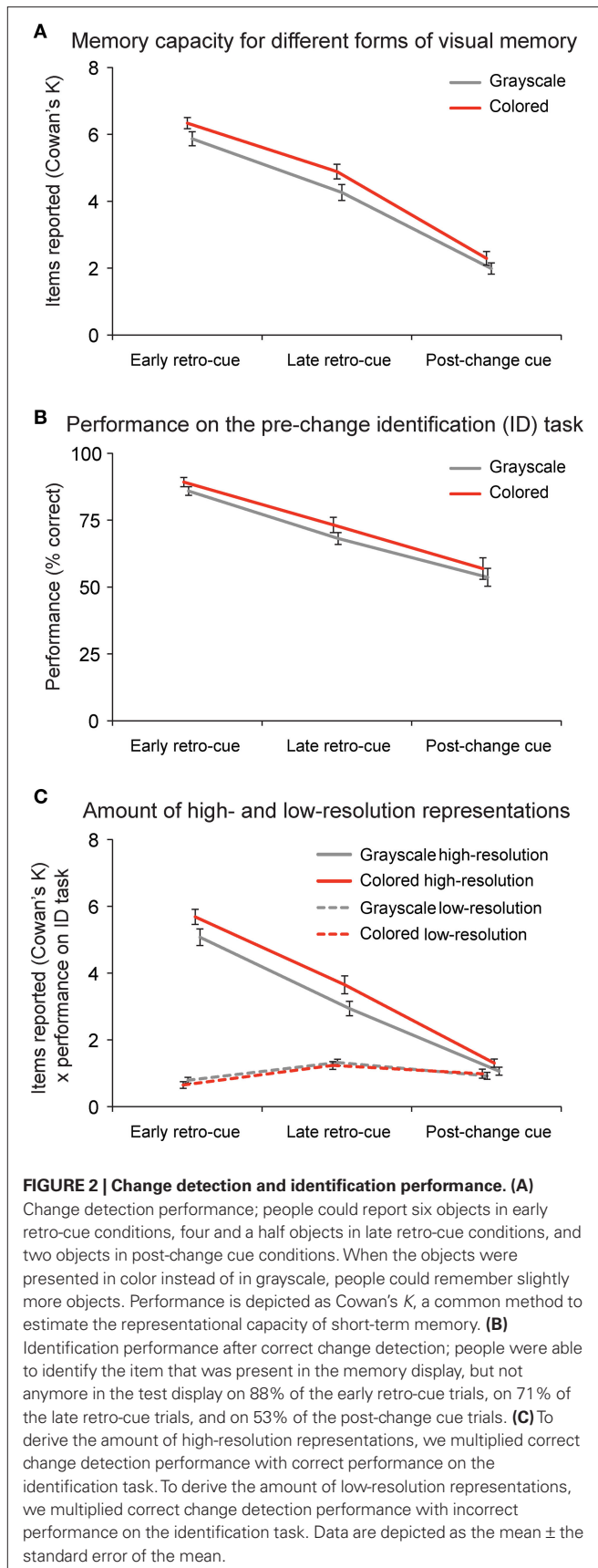resolution representations support change detection or pre-change identification only. (B) Early retro-cue condition; 10 ms after off-set of the memory display a spatial cue was presented that singled out the item that changed in 50% of the trials. Effectively, this condition measures iconic memory. (C) Late retro-cue condition; 1 s after off-set 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 on-set of the test display, a spatial cue was presented. This condition measures only visual working memory.

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.

AMOUNT OF HIGH-RESOLUTION REPRESENTATIONS

To derive the amount of high-resolution (or visually detailed) representations, we multiplied change detection performance (**Figure 2A;** expressed as Cowan's K) with correct performance on the subsequent change identification task (**Figure 2B**). On average,

subjects could report 5.1 (out of 5.8) detailed grayscale and 5.7 (out of 6.3) detailed colored representations in the early retro-cue condition, 2.9 (out of 4.3) detailed grayscale and 3.7 (out of 4.8) detailed colored representations in the late retro-cue condition, and only 1.1 (out of 2.0) detailed grayscale and 1.3 (out of 2.3) detailed colored representations in the post-change cue condition (**Figure 2C**). Repeated measures ANOVAs revealed that subjects could report more detailed representations when the objects were presented in color ($F(1,19) = 39.51, p < 0.001, \eta^2 = 0.675$), but we observed no benefit of color when people detected the change without being able to identify the pre-change item ($F(1,19) = 0.99, p = 0.33$) (see

FIGURE 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 color instead of in grayscale, 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 amount of high-resolution representations, we multiplied correct change detection performance with correct performance on the identification task. To derive the amount of low-resolution representations, we multiplied correct change detection performance with incorrect performance on the identification task. Data are depicted as the mean ± the standard error of the mean.
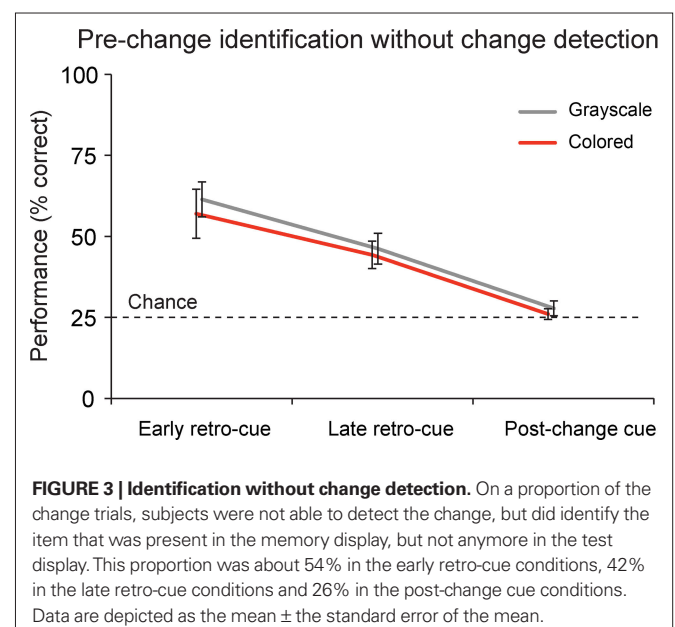
low-resolution representations in **Figure 2C**). In addition, memory capacity for detailed representations decreased over conditions ($F(1,19) = 118.66$, $p < 0.001$, $\eta^2 = 0.930$). Subsequent *post hoc* paired *t*-tests showed that the amount of detailed representations was highest when an early retro-cue was provided compared to a late retro-cue (color: $t(19) = 10.151$, $p < 0.001$; grayscale: $t(19) = 9.155$, $p < 0.001$), and the capacity for detailed representations was also higher when a late retro-cue compared to a post-change cue was shown (color: $t(19) = 7.800$, $p < 0.001$; grayscale: $t(19) = 7.622$, $p < 0.001$). 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.

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 **Figure 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

FIGURE 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.

could have chosen the right object by chance (one out of four). To be sure this is not the case, we first performed one sample *t*-tests against chance level of 25%.

When subjects did not detect the change, they still identified the pre-change item on 58.0% (grayscale) and 50.7% (color) of the early retro-cue trials, on 43.2% (grayscale) and 41.2% (color) of the late retro-cue trials, and on 27.9% (grayscale) and 26.2% (color) of the post-change cue trials. Statistically, performance exceeded chance levels in early retro-cue conditions for both grayscale ($t(19) = 6.727$, $p < 0.001$) and colored objects ($t(19) = 4.224$, $p < 0.001$). These same results apply to the late retro-cue conditions (grayscale: $t(19) = 4.498$, $p < 0.001$; colored: $t(19) = 4.594$, $p < 0.001$). However, performance in post-change conditions did not exceed chance levels for both grayscale ($t(19) = 1.256$, $p = 0.224$) and colored objects ($t(19) = 0.669$, p = 0.511).

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 1**, third column) with trials where people detected the change only (**Table 1**, fourth column) *or* identified the change only (**Table 1**, fifth column). This ratio was 4.07:1 (grayscale) and 6.02:1 (color) in the early retro-cue condition, 1.47:1 (grayscale) and 1.96:1 (color) in the late retro-cue condition, and 0.54:1 (grayscale) and 0.63:1 (color) 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 third to fifth column; **Table 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 color or in grayscale ($F(1,19) = 2.411$, $p = 0.137$). This might suggest that the absolute capacity of iconic memory, fragile VSTM and visual working memory is identical for color and grayscale stimuli, but that color adds to the resolution of the representation.

DISCUSSION

The standard model of 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 (six items), fragile VSTM (4.6 items), and visual working memory (2.2 items), and the capacity of all VSTM stages was higher for colored objects than for grayscale 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 and 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.

Table 1 | Performance on different VSTM conditions.

	No-change performance			Change performance			Detection and Identification			Detection only			Identification only		
	IC	fVSTM	VWM	IC	fVSTM	VWM	IC	fVSTM	VWM	IC	fVSTM	VWM	IC	fVSTM	VWM
Gray	84.6	79.7	88.0	88.8	73.6	36.9	76.5	50.6	19.2	12.3	23.0	17.7	6.5	11.4	17.6
	(1.6)	(1.9)	(1.7)	(1.9)	(2.5)	(2.7)	(2.5)	(2.8)	(1.6)	(1.4)	(1.4)	(2.1)	(1.0)	(1.3)	(1.6)
Color	86.9	82.2	88.8	92.3	78.9	40.0	82.5	58.0	21.6	9.8	20.9	18.4	3.9	8.7	15.7
	(1.3)	(2.2)	(2.5)	(1.1)	(2.0)	(3.1)	(2.1)	(3.0)	(1.6)	(1.5)	(2.3)	(2.5)	(0.6)	(0.9)	(1.4)

*Subjects performed a change detection task that measures iconic memory (**Figure 1B;** IC), fragile VSTM (**Figure 1C**; fVSTM), and visual working memory (**Figure 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 and 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 grayscale and color images apart.*

A key issue in "representation land" is whether features exist in a freely "floating" form or are bound into object representations (Treisman and Gelade, 1980; Treisman, 1996). The displays in **Figure 1A**, for example, consist of many features: there are the colors green, red, yellow, etc. Then there are forms that are primarily (i.e., in their low spatial frequency 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, colors, etc. Some objects fall into categories that may be detected in parallel, such as animal – non-animal (Thorpe et al., 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 et al., 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 1A**, change detection may have been possible according to the change in color (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 and 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 **Figure 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 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 et al., 2001). As a consequence, higher visual areas maintain less detailed representations (shown as blurred objects in **Figure 4**) than lower visual areas. In
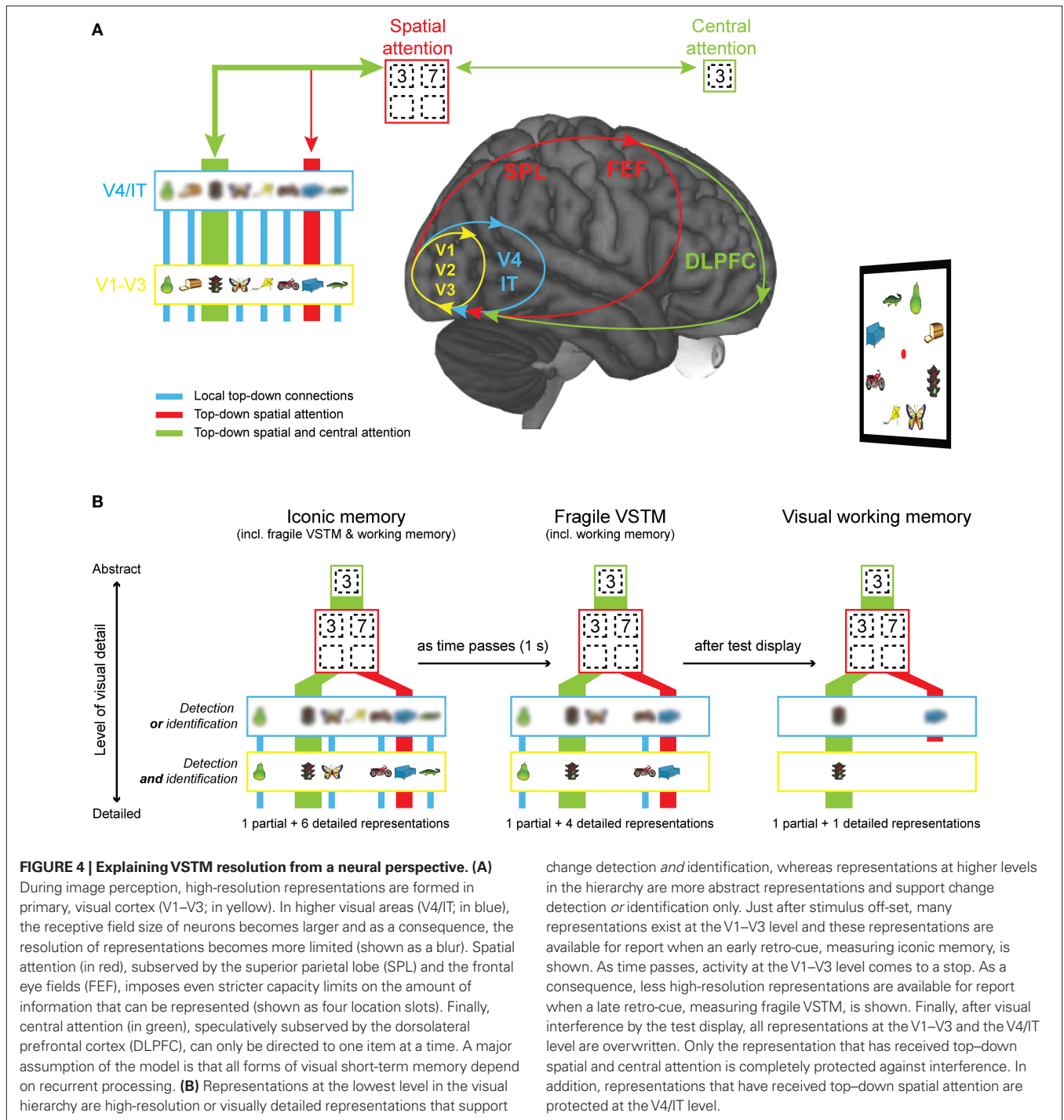
addition, higher visual areas have lower storage capacity as the (larger) receptive fields of neurons "see" multiple objects at the same time, while they can only represent one object at a time. Neurons thus have to "choose" which object to represent and this happens by means of biased competition that in turn causes a modulation in the firing pattern to represent one or the other object (Kastner and 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 **Figure 4A** in yellow), whereas low-resolution representations depend on visual areas higher up in hierarchy (V4/IT; see **Figure 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 (SPL) and prefrontal cortex (visual working memory; *in* red and green). The nodes in SPL and prefrontal cortex are special, as they control feedback signals related to spatial and central attention, respectively (Corbetta and Shulman, 2002; Vogel et al., 2005; Xu and Chun, 2006; Mcnab and Klingberg, 2008). 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 and Nobre, 2007; Matsukura et al., 2007; Makovski et al., 2008). 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 off-set (**Figure 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 amount 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–down attention from prefrontal cortex and the SPL 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 (**Figure 4B** left-most figure), and visual working memory is also measured in the fragile VSTM condition (**Figure 4B** middle figure).

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

FIGURE 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 amount of information that can be represented (shown as four 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 off-set, 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.

complexity (Alvarez and 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 amount of full representations that are maintained in short-term memory, but rather signals the amount of representations that are sufficiently detailed to detect the current change.

REFERENCES

Alvarez, G. A., and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111.

Averbach, E., and Coriell, A. S. (1961). Short-term memory in vision. *Bell Syst. Tech. J.* 40, 309–328.

Beck, M. R., and Levin, D. T. (2003). The role of representational volatility in recognizing pre- and postchange objects. *Percept. Psychophys.* 65, 458–468.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114; discussion 114–185.

Griffin, I. C., and Nobre, A. C. (2003). Orienting attention to locations in internal representations. *J. Cogn. Neurosci.* 15, 1176–1194.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files – object-specific integration of information. *Cogn. Psychol.* 24, 175–219.

Kastner, S., and Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39, 1263–1276.

Lamme., V. A. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18.

Landman, R., Spekreijse, H., and Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vis. Res.* 43, 149–164.

Lepsien, J., Griffin, I. C., Devlin, J. T., and Nobre, A. C. (2005). Directing spatial attention in mental representations: interactions between attentional orienting and working-memory load. *Neuroimage* 26, 733–743.

Lepsien, J., and Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cereb. Cortex* 17, 2072–2083.

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

Makovski, T., and Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychon. Bull. Rev.* 14, 1072–1078.

Makovski, T., Sussman, R., and Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 369–380.

Matsukura, M., Luck, S. J., and Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: protection or prioritization? *Percept. Psychophys.* 69, 1422–1434.

Mcnab, F., and Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.* 11, 103–107.

Mitroff, S. R., Simons, D. J., and Levin, D. T. (2004). Nothing compares 2 views: change blindness can occur despite preserved access to the changed information. *Percept. Psychophys.* 66, 1268–1281.

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236.

Sligte, I. G., Scholte, H. S., and Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS One* 3(2), e1699. doi: 10.1371/journal.pone.0001699.

Sligte, I. G., Scholte, H. S., and Lamme, V. A. (2009). V4 activity predicts the strength of visual short-term memory representations. *J. Neurosci.* 29, 7432–7438.

Smith, A. T., Singh, K. D., Williams, A. L., and Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* 11, 1182–1190.

Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.

Treisman, A. M. (1996). The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178.

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

Vogel, E. K., McCollough, A. W., and Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500–503.

Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 92–114.

Xu, Y. D., and Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95.

SUPPLEMENTARY MATERIAL

FIGURE S1 | We selected 50 line drawn objects from a series of 260 objects created by Rossion and Pourtois (2004). All objects used in our experiment are depicted in this image.