A new definition of visual short-term memory

Sligte, I.G.

Publication date
2011

Citation for published version (APA):
Sligte, I. G. (2011). *A new definition of visual short-term memory*. [Thesis, fully internal, Universiteit van Amsterdam].

5. V4 activity predicts the strength of visual short-term memory representations

Recent studies have shown the existence of a form of visual memory that lies intermediate of iconic memory and visual short-term memory (VSTM), both in terms of capacity (up to 15 items) and in terms of the duration of the memory trace (up to four seconds). As new visual objects readily overwrite this intermediate visual store, we believe that it reflects a weak form of VSTM with high capacity that exists alongside a strong, but capacity-limited form of VSTM. In the present study, we isolated brain activity related to weak and strong VSTM representations with the use of functional MRI. We found that activity in V4 predicted the strength of VSTM representations; activity was low when there was no VSTM, medium when there was a weak VSTM representation regardless whether this weak representation was available for report or not, and high when there was a strong VSTM representation. Altogether, this study suggests that the high capacity, yet weak VSTM store is represented in visual parts of the brain. Allegedly, only some of these VSTM traces are amplified by parietal and frontal regions and as a consequence reside in traditional or strong VSTM. The additional weak VSTM representations remain available for conscious access and report when attention is redirected to them, yet are overwritten as soon as new visual stimuli hit the eyes.

Introduction

Traditional models of visual memory distinguish between iconic memory and visual short-term memory (VSTM). Iconic memory is usually considered to be a high-capacity, but rapidly decaying form of visual memory, more or less like a brief (and degraded) internal snapshot of what is just seen (Sperling, 1960). VSTM, on the other hand, contains a maximum of four integrated visual objects (Luck & Vogel, 1997) and only these object representations can be maintained for longer durations. Nevertheless, recent change detection tasks using spatial cues at different moments during the trial (see **Fig. 5.1A**) show that this two-stage model of visual memory is too simplistic.

Generally, in these cued change detection tasks a memory display containing multiple objects is shown, followed by a blank retention interval, after which a test display is shown. When a spatial cue is directed to the location of the change at test (*post-cue* trial) subjects perform poorly, reflecting the limited capacity of VSTM (Matsukura et al., 2007; Makovski et al., 2008; Sligte et al., 2008). However, when cues are presented during the blank interval (*retro-cue* trial), people can report many more objects, even up to 15 items (Griffin and Nobre, 2003; Landman et al., 2003; Sligte et al. 2008).

This high capacity store seems to exist for at least four seconds after stimulus disappearance (Lepsien et al. 2005; Sligte et al. 2008), which is far longer than the duration of iconic memory. Based on these findings, we suggest that people represent many items in a weak form of VSTM and that only a few of these items receive enough attention to be stored in traditional or strong VSTM. Weak VSTM representations can be accessed when people direct attention to them during retention (as in retro-cue trials), but they are erased by subsequent stimulation, such as the test display, when they are not attended (as in post-cue trials). In this latter case, only representations in strong VSTM remain available for report.

In the present study, we aimed to unravel the neural substrate of weak VSTM and how it differs from strong VSTM. With the use of a cued change detection task, we were able to determine on each trial whether the cued item was in weak VSTM, in strong VSTM or not represented in VSTM at all. We used an array of eight objects, as we needed an array that overflows the capacity of strong VSTM (i.e. is larger than four); else all items could be in strong VSTM. We then related the representational status of the cued item to BOLD activity at its retinotopic location. This effectively restricted our analysis to visual areas V1, V2, V3, and V4 in which we could delineate eight separate retinotopic locations corresponding to the eight objects of the memory array (see **Fig. 5.2**). To anticipate, we found that V4 activity was low when the cued item was not represented, medium when it was in weak VSTM and high

when it was in strong VSTM. These results suggest that V4 activity predicts the representational status of a VSTM representation.
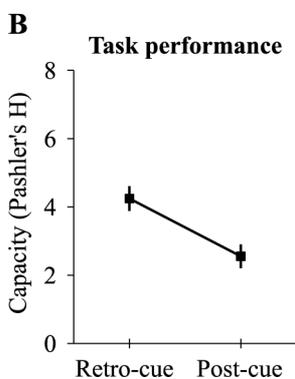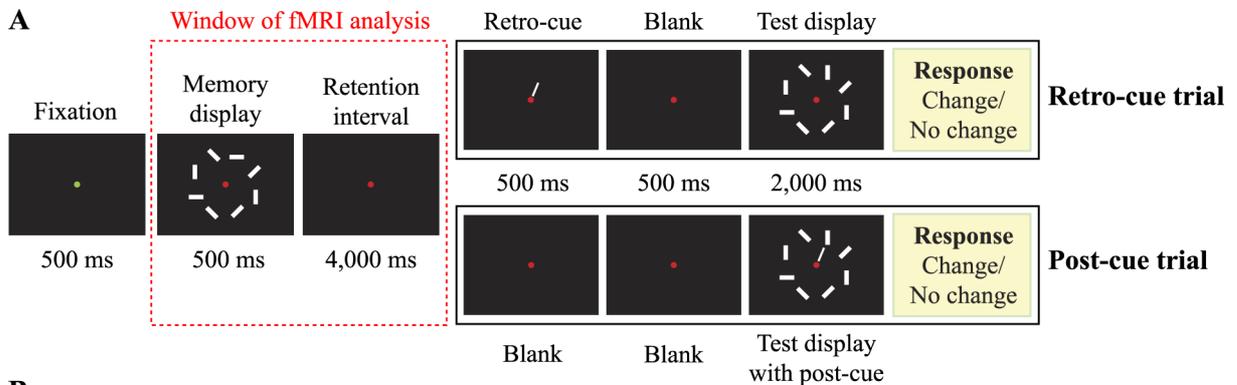

Figure 5.1 *Trial design and task performance **A** Observers performed a VSTM task while cues were presented before the test display appeared (retro-cue trial) or during the test display (post-cue trial). Performance on post-cue trials reflects the capacity of strong VSTM, and the increase in performance on retro-cue trials reflects the capacity of weak VSTM. BOLD-fMRI was analyzed up to the delivery of the earliest cue (indicated as the red dotted rectangle). Up to that time point, trials are identical, but differences in subsequent cue delivery and the response of the subject reveals the perceptual status of the cued item during retention. **B** Task performance on both kinds of trials is depicted as Pashler's H (Pashler, 1988). Plotted data are the mean SE performance across subjects.*

Methods

Subjects

20 right-handed young adults (14 females) with normal or corrected-to-normal vision participated for financial compensation. The local ethics committee had approved the experiment.

Stimuli

Memory and test displays were made up of eight white (46.4 cd/m^2) oriented rectangles on a black (0.2 cd/m^2) background. Individual rectangles (1.56° by 0.39°) were presented at an eccentricity of 5° of visual angle and could either be horizontal, vertical, 45° to the vertical or 135° to the vertical. Each orientation was present at least once in the display and at most three times (to prevent chunking). Throughout the entire experiment, a red fixation dot (1.56°

by 0.39°; 4.78 cd/m^2) was present in the centre of the screen, and it only turned green (21.95 cd/m^2) for 500 ms to indicate the start of a new trial. Spatial cues consisted of 3-pixel thick white lines (46.4 cd/m^2; 2.5° by 0.05°) which were at one end close (0.5°) to the fixation point and at the other end close (2°) to the center of a single rectangle.

Task design main experiment

The trial design of the main experiment is shown in Fig. 5.1. At the start of each trial, the fixation dot in the middle of the screen turned green for 500 ms. Then a 500-ms memory display appeared. After off-set of the memory display, a blank retention interval of 4,000 ms was shown. Up to this time point in the trial, all stimulation is the same between conditions, which is critical for our fMRI analysis. In the retro-cue condition, we presented a spatial cue for 500 ms that indicated retrospectively which item was the one to potentially change (so-called *retro-cue*) followed by a 500-ms blank display to allow the retro-cue to take effect. In the post-cue condition, we presented a 1,000-ms blank instead of a 500-ms cue and a 500-ms blank display. Finally, the test display was shown for 2,000 ms (or until response) in which the cued item had rotated by 90° compared to the memory display in 50 percent of the trials. In the post-cue condition, a spatial cue was shown for 500 ms on top of the test display. During the test display, subjects were required to indicate by button press whether the cued item had changed or not, and it was stressed not to press any button when uncertain to prevent guessing to confound activity on correct and incorrect trials, thereby occluding effects of interest. After offset of the test display, a blank display was shown for 1,000-3,000 ms in steps of 1,000 ms to vary the delay between trials (not shown in **Fig. 5.1**). One difference between retro-cue and post-cue trials is that the effective retention interval up to the presentation of the cue is longer on post-cue trials (5s) than on retro-cue trials (4s). It is not likely that this explains the difference in performance in the post-cue condition compared to the retro-cue condition. In previous work (Landman et al., 2003; Sligte et al., 2008), we used an effective retention interval that was identical for both conditions and still differences between retro- and post-cue conditions were large. Moreover, we found that post-cue performance was stable over various lengths of the retention interval, whereas retro-cue performance decreased with increases of the retention interval.

Besides these retro-cue and post-cue trials, we also presented trials in which only a spatial cue was shown in the absence of other stimulation. The timing details of these trials are identical to those of the retro-cue trials. We presented these trials to estimate when the *retro-cue* start to affect the BOLD response (see **Fig. 5.3** for data).

Task design cortical mapping experiment

We performed a separate cortical mapping experiment in which we rotated rectangles at all eight possible memory locations of the main experiment (see **Fig. 5.2E** for an example). Two rectangles were rotated at the same time at locations that were maximally separated; rectangles at locations 1 and 5, locations 2 and 6, locations 3 and 7 and locations 4 and 8. Rotation speed was 90° per second. We rotated rectangles for 16 seconds followed by 8 seconds of no stimulation. In effect, each location was rotated four times for 16 seconds. During the entire mapping experiment, subjects performed a task in which they had to detect small changes in the orientation of a cross within the fixation dot, which requires accurate fixation and attention. Once every 2 seconds, this fixation cross rotated and subjects detected on average 96% (SD = 3%) of these changes.

In a separate session prior to the experiment, we presented each of the subjects with stimuli to map the polar angle and eccentricity of their visual BOLD responses to determine the locations of areas V1, V2, V3, and V4 (for details see Scholte, Witteveen, Spekreijse, & Lamme, 2006).

Procedure

Each fMRI run consisted of 16 *retro-cue* trials (change/no change x 8 locations), 16 *post-cue* trials (change/no change x 8 locations), and 8 cue trials (8 locations). The trial order was completely randomized within runs and each subject completed four runs (in addition, one run was performed outside of the scanner to practice the experiment). Each run started and ended with a 16-second lasting baseline in which nothing was shown.

At the end of these four runs, we presented the cortical mapping experiment. In a separate session prior to the experiment, we mapped the polar angle and eccentricity of visual areas.

Data acquisition and stimulus presentation

MRI data were acquired from a Philips 3T scanner (Eindhoven, The Netherlands) and analyzed with BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). For each subject, we acquired an anatomical high-resolution image with conventional parameters (T1 turbo field echo, 182 coronal slices, FA 8, TE 4.6 ms, TR 9.6 s, slice thickness 1.2 mm, FOV 256*256 mm, inplane voxel resolution 1x1 mm). For polar and eccentricity mapping, BOLD-MRI was recorded using these settings: GE-EPI, transversal slice orientation, TR 2,000 ms, TE 28 ms, FOV 200 mm, matrix size of 112x112, slice-thickness 2.5, slice gap 0.3, 24 slices and a sense factor of 2.5. For our main experiment as well as for the location mapping

A new definition of visual short-term memory

BOLD-MRI was recorded using these settings: fourteen 2.5-mm-thick (1.56 X 1.56 in-plane, 0.25 mm skip) coronal slices were taken that covered the back of the head. T2*-weighted parameters: echo time 28 ms; flip angle 90; FOV 200 mm; 128 X 128 matrix; repetition time 1,000 ms. These settings provided a detailed scope on visual areas of the brain, but we were not able to do whole-brain analyses. The start of a run was triggered by scanner pulses and trials were presented with Presentation (Neurobehavioral Systems, Inc.). Stimuli were front-projected from a LCD projector on to a screen at the feet of the supine subject. The subject in the MR scanner viewed the screen through a mirror just above the eyes. Total scanning time was approximately 80 min, and we acquired eccentricity and polar angle mappings and an additional high-resolution scan in a separate scan session. We immobilized the subject's head using foam pads to reduce motion artifacts and used earplugs to moderate scanner noise.

Eye tracker measurements

We recorded the eye movements (Resonance Technology Inc. / Arrington Research Inc, Scottsdale, USA) of eight of the subjects during the main experiment and analyzed the number of saccades and the number of eye blinks each of the subjects made. Eye movements were recorded at a speed of 60 Hz by digitizing video images. Data were analyzed by smoothing the recorded data (detrend data set & smooth adjacent time points) and classifying visual activity as blinking (i.e. loss of a visible pupil, eye-movement faster than 30° per second) or making a saccade to one of the locations where stimuli were presented (i.e. no loss of a visible pupil, an eye-movement faster than 30° per second & a shift in location of more than 4° of visual angle).

Data analysis behavior

For behavioral data, the number of objects encoded was estimated using Pashler's H formula (Pashler, 1988). In simplified form, capacity = (hit rate – false alarm rate) * set size / (1 – false alarm rate). This formula does not take trials where subjects guessed into account. However, we strongly encouraged subjects not to press when uncertain, making it unnecessary to control for guessing trials.

Data analysis cortical mapping

Image analysis was performed with BrainVoyager QX. Data preprocessing included image realignment, three-dimensional motion correction, linear detrending, correction for slice scan acquisition order, spatial smoothing with a 4-mm Gaussian kernel (full width at half-

maximum), and temporal smoothing with a high-pass filter of 0.01 Hz and a low-pass filter of 2.8s. One might question the use of a low-pass temporal filter, but it is not unusual to use a low-pass filter in blocked designs such as our mapping experiment (8 locations; stimulus on for 16s; off for 8s). High frequencies only cause noise on top of the slow frequencies of a block design. We used a filter of 2.8s based on early fMRI literature (Friston et al., 1995).
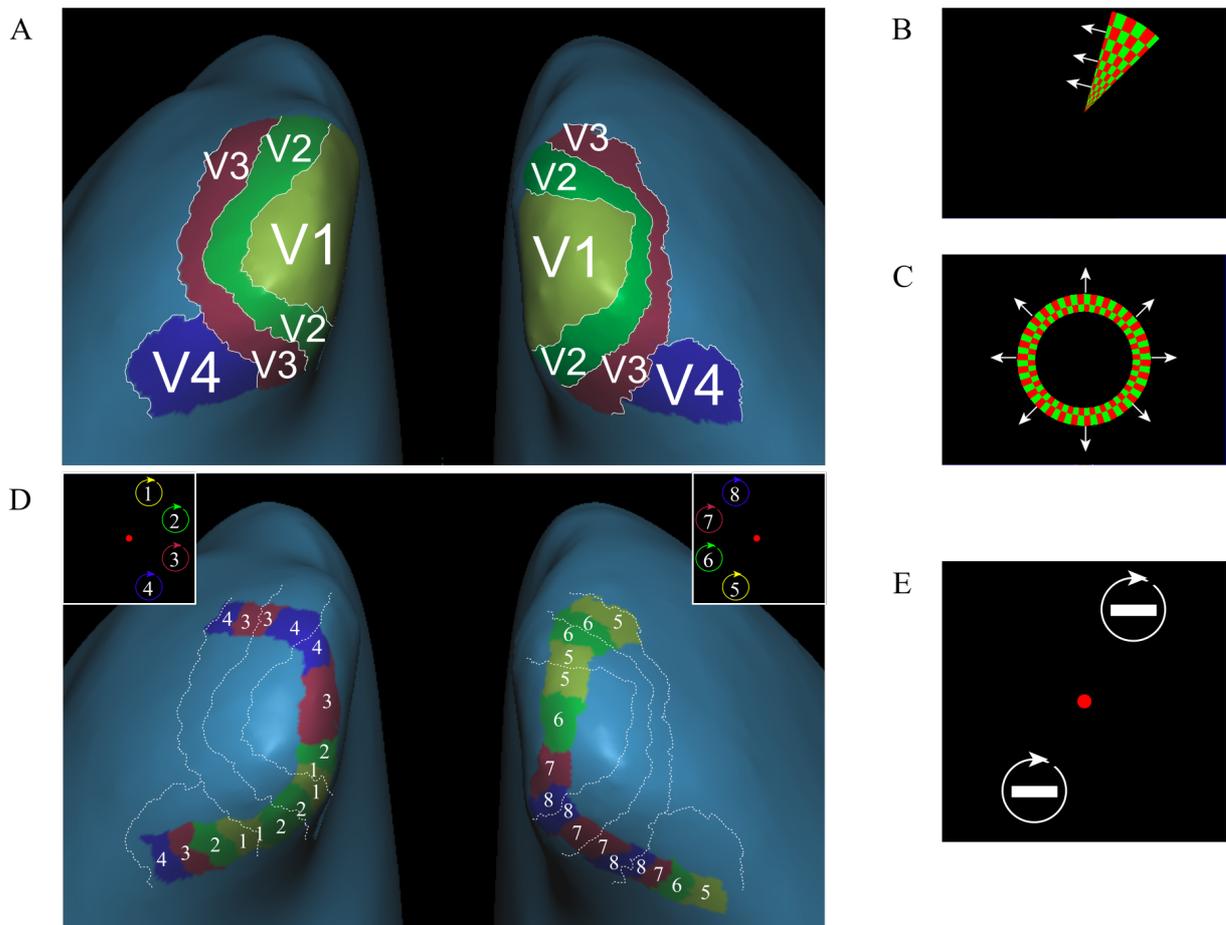

Figure 5.2 *Delineating regions of interest. First, we mapped V1–V4 (**A**) by showing a rotating wedge (**B**) to delineate the boundaries between areas and an expanding and contracting annulus (**C**) to delineate eccentricity boundaries of specific areas. Next, we outlined eight retinotopic locations in V1–V4 (**D**) by showing rotating white rectangles at eight different radial positions, at an eccentricity of 5° of visual angle (**E**).*

After we had delineated V1 to V4 using polar angle and eccentricity mapping (see **Fig. 5.2**; for details see Scholte et al., 2006), we created statistical parametric maps of BOLD activation for the rectangle mapping experiment by using a multiple regressors analysis, with regressors for each trial type convolved with a canonical hemodynamic function. We

contrasted each individual location with all other locations, and projected False Discovery Rate-corrected contrasts onto the inflated surface model of the brain. By overlaying the visual areas (found with eccentricity and polar mapping), we were able to distinguish 4 separate spatial locations in V1, V2, V3, and V4 in both hemispheres.

Data analysis BOLD main experiment

Image preprocessing was performed with BrainVoyager QX using the same parameters as specified in the previous section, only we did not apply a low-pass filter. After preprocessing, we exported the time series for all 32 regions of interest (4 areas x 8 locations) to Matlab (Mathworks, Inc.). In Matlab, we first removed the mean from each timeseries and we performed a z-transformation on the entire data set. This is a standard procedure that is implicitly used in some fMRI analysis software such as FSL (Smith et al., 2004) to avoid having predictors for separate runs. In this experiment, we used this procedure explicitly to avoid having predictors for separate runs and for separate functional regions of interest within the same brain area. Next, we performed a full deconvolution analysis based on predictors for cued location (n = 8) x *retro-cue* or *post-cue* (n = 2) x correct or incorrect response (n = 2). We also added 8 predictors for trials in which only a cue was shown and predictors for miss trials (no response). This adds up to a minimum of 40 predictors and a maximum of 56 predictors per subject (the number of miss trials did vary substantially between subjects). For eight subjects, we also recorded eye movements. We found no differences in the number of saccades and blinks between conditions. In the time window in which the memory array was present, eye blinks occurred on 2.13% (SD 2.39) of the trials, and saccades on 1.28% (SD 1.55) of the trials. Trials that contained saccades or eye blinks were collapsed into one additional noise predictor.

The deconvolution analysis provided us with deconvolved timeseries from three seconds before trial start until 20 seconds after trial start. Each deconvolved timeseries was corrected for baseline differences (baseline correction from -3s to -1s). Next, we computed the mean activity at the cued location (collapsed over cued locations) for retro-cue trials with an incorrect response (on average 16.6 trials per subject, SD 5.6), for retro-cue trials with a correct response (43.1 trials per subject, SD 6.3), for post-cue trials with an incorrect response (17.5 trials per subject, SD 4.7), for post-cue trials with a correct response (36.0 trials per subject, SD 6.8) and for cue trials (32 trials per subject). There were very few miss trials (retro-cue: 4.4 trials per subject, SD 3.6; post-cue: 10.6 trials per subject, SD 5.9). This makes statistical testing for miss trials hard and unreliable, especially in the retro-cue condition in which some subjects did not have any miss trial at all.

Results

Behavioral performance

We found that people could report 4.24 items (calculated from the percentage of hits and false alarms using Pashler's H formula, see Methods) when an attention-directing cue was presented during the retention interval (**Fig. 5.1**; *retro-cue* trial) and 2.55 items when the cue was presented after on-set of the test display (**Fig. 5.1**; *post-cue* trial). This difference in performance was highly significant ($t(19) = 5.17$, $p < .001$) and it illustrates the capacity differences between weak and strong VSTM. Note that capacity differences are much larger when using shorter retention interval (Sligte et al., 2008), but given the low temporal resolution of the BOLD response we need to have a long delay between memory and test display to isolate memory-related activity.

Restricting the window of fMRI analysis

For two reasons, our analyses need to be restricted to epochs that are not yet influenced by any attention-directing cue. First, before the cue is presented a few items will be represented in strong VSTM, additional items will be represented in weak VSTM, while some items will not be represented at all. The delivery of a retro-cue, however, may transfer weak VSTM representations to strong VSTM and this could change memory-related activity. Second, in retro-cue trials the cue is delivered four seconds after off-set of the memory display and no other stimulation is presented at the same time. In post-cue trials, the cue is presented five seconds after off-set of the memory display together with the test display. Differences in activity on retro-cue trials compared to post-cue trials may therefore be related to differences in physical stimulation, yet not before the earliest cue is delivered.

To reveal when the cue started to affect the BOLD signal, we sometimes delivered trials in which only an attention-directing cue was presented. The activity on these trials is shown in Figure 5.3. Note that at all locations, the BOLD signal in these cue-only trials immediately decreases after trial start. In the present study, cue-only trials are presented on average on 20 percent of the trials randomly intermixed with the memory trials. Therefore, it is likely that subjects expect to see a memory display on each trial, but once every five trials on average this expectation is violated. Other studies have also reported a decrease in BOLD signal when the expectancy of a subject is violated (Davidson et al., 2004; Munneke, Heslenfeld, & Theeuwes, 2008). The BOLD response evoked by the cue is riding on top of this general decrease in activity. See Figure 5.3, blue lines, for the cue specific effect. Paired t-tests between the activity at the cued location and the mean activity at all non-cued

locations revealed that the cue took effect 6 seconds after trial start in V2 (t(19) = 2.84, p = .011), V3 (t(19) = 2.23, p = .038), and V4 (t(19) = 2.63, p = .017) and 7 seconds after trial start in V1 (t(19) = 6.13, p < .001). In the main experiment, we will therefore restrict our analysis to the first five seconds after trial start.
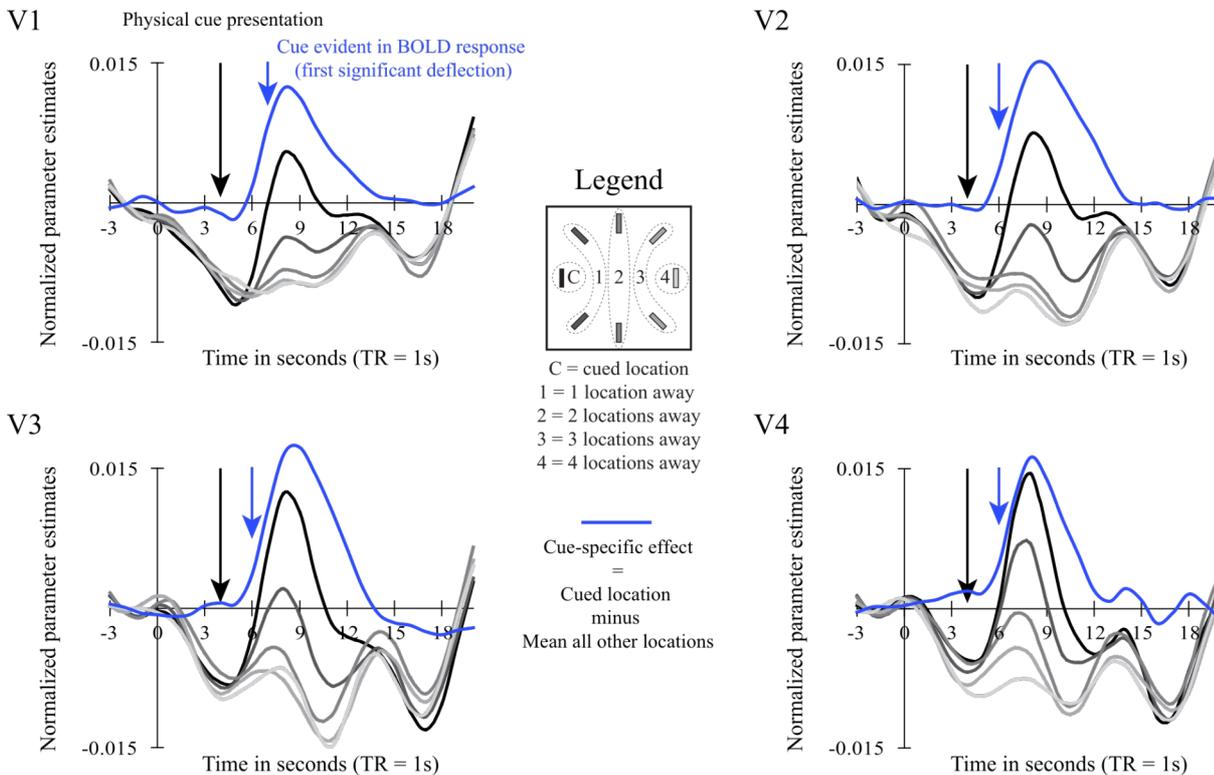

Figure 5.3 *Effect of attention-directing cue. In 20% of the trials, we delivered a cue without presenting the memory and the test array. This was done to measure when the cue took effect during the VSTM task. The cue was presented 4 s after trial start (black arrow) and started to be evident in the BOLD response 6 s after trial start (blue arrow). The delivery of a cue had an effect that was specific to the cued location (and the immediately adjacent ones) yet was riding on top of a general decrease of the BOLD response at all locations, which is probably attributable to anticipation of the subject. The cue-specific effect (BOLD response at cued location - BOLD response at other locations) is shown in blue. All data were computed with a full deconvolution analysis and baseline corrected from 3 s until 1 s before trial start. Depicted data are the mean activity across subjects (n = 20).*

VSTM effects in visual cortex

Based on the behavioral response of the subject, we can determine whether the cued location was in strong VSTM, in weak VSTM or not in VSTM during retention (so before the cue was delivered) in the following way. An **incorrect** response on a *retro-cue* trial suggests that the cued item was neither in weak nor in strong VSTM: if it would have been in weak

VSTM, the cue would have transferred it to strong VSTM, thereby making the representation available for report. An **incorrect** response on a *post-cue* trial suggests that the cued item was not in strong VSTM. It could have been in weak VSTM, however, considering that the capacity of weak VSTM is larger than that of strong VSTM. Still, new stimulation overwrites weak VSTM, so it will not show up in behavior when the cue is delivered after the change has already occurred. A **correct** response on a *retro-cue* trial suggests that the cued item was in weak VSTM (and upon cueing is transferred to strong VSTM). We are ignorant about whether it was already in strong VSTM before the cue: this could have been the case (simply by chance) or not. A **correct** response on a *post-cue* trial suggests that the cued location was represented in strong VSTM (otherwise change blindness would have occurred). With these aspects of our analysis in place we can now turn to the results.

Repeated measures ANOVAs were used to test whether there were differences in neural activity between conditions (activity at the cued location was used for testing). Note that the maximum BOLD response between trial start and five seconds after trial start was used for testing (after that, the cue takes effect; see previous section). We observed no differential activity in V1, V2 and V3 (**Fig. 5.4A;** small panels on the right). In V4 (**Fig. 5.4A**; large panel on the left), however, we did observe significant differences (F (3,17) = 5.44, p = .008). Paired t-tests revealed that activity was lowest on retro-cue trials with an incorrect response (compared to other conditions, lowest t-value: t(19) = 2.31, p = .032) and highest on post-cue trials with a correct response (compared to other conditions, lowest t-value: t(19) = 2.15, p = .045). Both retro-cue trials with a correct response and post-cue trials with an incorrect response evoked similar activity (t(19) = .32, p = .75) that was intermediate between activity on retro-cue trials with an incorrect response (lowest t-value: t(19) = 2.31, p = .032) and activity on post-cue trials with a correct response (lowest t-value: t(19) = 2.15, p = .045). From this, we can conclude that activity in V4 is highest when an item is in strong VSTM, lower when it is in weak VSTM and lower still when no memory trace is present whatsoever. It seems that for weak VSTM representations this pattern of activity is not related to whether behavioral responses are correct or incorrect, so that these differences cannot be attributed to fluctuations in arousal or other factors influencing correct or incorrect behavior. One caveat, however, is that in some conditions the status of the stimulus representation in question is uncertain (weak or strong VSTM in correct retro-cue trials; weak or no VSTM in incorrect post-cue trials). Graded effects observed in this case could be carried by one of the two underlying states. It may therefore not be correct to assume that activations increase gradually over no-weak-strong VSTM.

A   V4 activity predicts strong, weak or no representation in VSTM (cued location shown)
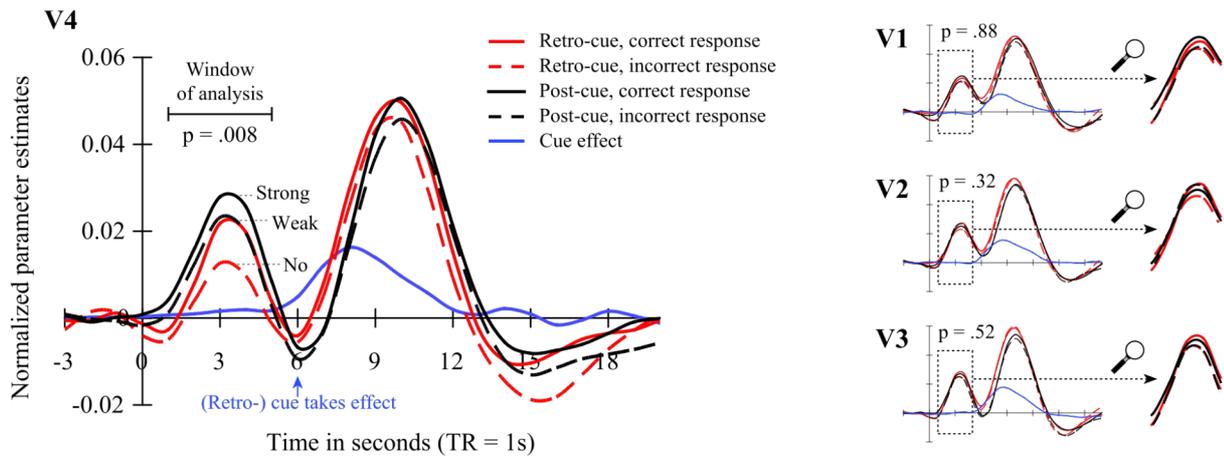
B   V4 effects related to no VSTM or strong VSTM are confined to the cued retinotopic location
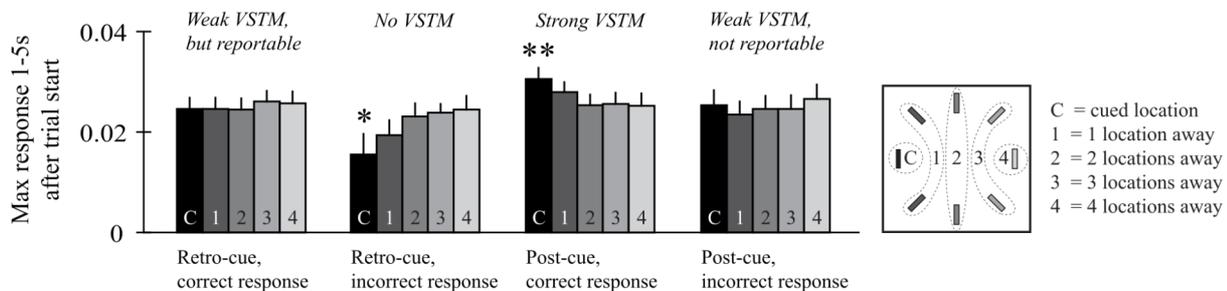

Figure 5.4 VSTM effects in visual cortex. **A**, Activity in V1–V3 did not vary over conditions (small panels on the right; area of interest is enlarged). V4 activity, however, did vary with the representational status in VSTM (large panel on the left). When there was no VSTM representation at the cued location (retro-cue trial with an incorrect response), activity was low. When there was a weak VSTM representation at the cued location, activity was intermediate regardless of whether this weak VSTM representation was reportable (retro-cue trial with a correct response) or not reportable (post-cue trial with an incorrect response). Finally, when there was a strong VSTM representation at the cued location (post-cue trial with a correct response), activity was high. Depicted data are the mean activity across subjects (n = 20). Depicted p values were computed using repeated-measures ANOVAs. For illustrative purposes, we also plotted the effect of a cue presented in isolation (subtraction between cued retinotopic location and all noncued locations). **B**, We observed that the absence of VSTM (retro-cue trial with an incorrect response) was characterized by low activity at the cued location, whereas activity at the noncued locations was higher. Similarly, the presence of strong VSTM (post-cue trial with a correct response) was characterized by high activity at the cued location, whereas the activity at all noncued locations was lower. This suggests that the absence of VSTM and the presence of strong VSTM are confined to the cued retinotopic location and that weak VSTM representations exist at all other (noncued) locations. Depicted data are the maximum activity between

*1 s after trial until 5 s after trial start (after that, the cue takes effect) averaged across subjects (n=20). *p=0.05, **p=0.01; p values were computed using repeated-measures ANOVAs testing for linear effects across cued locations.*

In addition to the previous analyses, we compared activity at the cued location with activity at all non-cued locations (**Fig. 5.4B**). We found that activity was similar at all locations on retro-cue trials with a correct response and on post-cue trials with an incorrect response. On retro-cue trials with an incorrect response, however, we observed that activity was low at the cued location, but was increasingly higher when the retinotopic distance to the cued item increased ($F(1,19) = 5.85$, $p = .026$). Similarly, we found that on post-cue trials with a correct response activity was high at the cued location, but was increasingly lower when the retinotopic distance to the cued item increased ($F(1,19) = 9.69$, $p = .006$). Note that these effects were recorded *prior* to any cue, so that they reflect the status of the representation during and shortly after disappearance of the memory display. Therefore, the activity of V4 neurons representing each object individually determines whether that object is in weak or strong VSTM, or has no VSTM representation at all; it is *not* a matter of general V4 activity, for instance fluctuating due to lapses of arousal.

Note that the differences between conditions are maximal before the cue is delivered (**Fig. 5.4A**), but tend to decrease after on-set of the second display. There is one reason why we did not analyze the reduced difference between conditions after the cue took effect; the physical stimulation of the two trial types (retro-cue/post-cue) starts to differ four seconds after trial start. In retro-cue trials, a cue is shown 4 seconds after trial start and the match display appears 5 seconds after trial start. In post-cue trials, on the other hand, the cue is shown 5 seconds after trial start together with the on-set of the match display. As differences in BOLD signal are so early, they are likely differences in the encoding phase of VSTM.

Discussion

Recent studies have suggested that people build up many representations in visual short-term memory (VSTM), but that only a sub-set of these VSTM representations is stored in a strong way that is protected against visual interference (Griffin and Nobre, 2003; Landman et al., 2003; Sligte et al., 2008). As long as new visual stimulation (for instance the test display) has not erased unprotected VSTM representations, weak VSTM representations remain available for report when people direct attention to them. In the present study, we aimed to reveal how weak and strong VSTM representations are expressed in visual cortex. We found that weak VSTM depended on sufficient activity in visual area V4, but not on activity in visual

areas V1, V2 and V3. When an item was represented in strong VSTM, we observed a spatially specific boost in V4 activity. Based on other studies, we can ascribe this boost in activity to the involvement of selective attention (Beck, Rees, Frith, & Lavie, 2001; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002).

The role of V4 in weak VSTM

The best starting point to explain the observed pattern of activity is provided by a recent functional MRI adaptation study (Konen & Kastner, 2008). In that study, it was found that many visual brain areas along the ventral and dorsal stream show object-specific adaptation. Only V1, V2 and V3 did not show object-specific adaptation, whereas V4 did. This suggests that the activity we observe here is related to the build-up of persistent object representations and not to something like persistent feature-specific information.

V4 seems at a low enough level in the visual hierarchy to allow for the (retinotopic) representation of the relatively large number of objects that is found in weak VSTM. Higher areas, such as the posterior parietal cortex (Todd & Marois, 2004; Vogel & Machizawa, 2004; Vogel, McCollough, & Machizawa, 2005) and (maybe) the frontal eye fields seem to impose the capacity limits that are found in strong VSTM. As V4 is strongly connected with these higher level areas involved with spatial attention and the direction of eye movements (Ungerleider, Galkin, Desimone, & Gattass, 2008), it seems like the ideal location to store object specific information that is not currently attended, yet that can be used whenever attention needs to be redirected.

However, we cannot be sure that V4 activity is related to the maintenance of weak VSTM. We found that V4 activity greatly differed between conditions just after trial start, but tended to decrease afterwards. As short-term memory processes can be divided into encoding, maintenance, and retrieval, it is likely that these early differences in V4 activity reflect the encoding phase of VSTM. It is conceivable that V4 activity reflects the amount of attention at a specific retinotopic location and that this determines whether the item is represented in strong VSTM (full attention), weak VSTM (some attention) or not represented in VSTM (no attention). Alternatively, it can be that strong VSTM representations are only formed when attention is at the right location and that the formation of weak VSTM representations is not dependent on attention at all. The difference between weak VSTM and no VSTM is then determined by other factors, such as sudden lapses in neural activity.

When V4 activity is not (or to a little extent) related to maintenance of weak VSTM, these representations have to be maintained in other areas of the brain and likely candidates are category-specific areas higher up in visual hierarchy. These category-specific visual

areas have large receptive field, but they are still able to code spatial location in a more distributed-code fashion. In a recent study (Lepsien & Nobre, 2007), it was indeed found that activity in the PPA (place stimuli) and the FFA (face stimuli) was modulated by the presentation of a retro-cue and this modulation could reflect a transition from weak VSTM to strong VSTM and back again.

A form of iconic memory or VSTM?

Traditional models of immediate memory distinguish between sensory memory (very high capacity; duration <500 ms) and short-term memory (very limited capacity; long duration). Here, we have shown that people maintain more representations than can fit into VSTM for more than four seconds, and that retrieval of these additional representations depends on the amount of activity in visual area V4. This seems to place these additional representations intermediate of sensory memory and standard notions of VSTM. Nonetheless, one could still question whether we are not measuring some kind of long-lasting iconic memory rather than a form of VSTM with high capacity.

Weak VSTM (as we have defined it) bears many resemblances to iconic memory. For instance, the experimental design for measuring weak VSTM is strikingly similar to iconic memory designs. In both, subjects are shown images containing multiple items and subjects have to remember as much items as possible up to the presentation of the cue. After the presentation of the retro-cue/partial-report cue memory load is reduced to one/few items. Differences between the experimental set-ups are that subjects are shown additional test images in this new task and that response options are limited. This difference (stimuli shown once vs. twice; free recall vs. two alternative forced-choice) could explain why a cued VSTM task could capture iconic memory effects for longer periods of time.

Also the characteristics of iconic memory are similar to the characteristics of weak VSTM: first, the capacity of iconic memory scales with the number of items to remember (Sperling, 1960), and we observed similar effects in retro-cue conditions with cues delivered 1,000 ms after image disappearance (Sligte et al., 2008). Second, iconic memory is very volatile and is easily overwritten by subsequent visual stimulation. Strong VSTM representations, on the other hand, are resistant to the presentation of new and/or distracting information. We found that weak VSTM is also easily overwritten when new stimulation is presented just before arrival of the retro-cue (Sligte et al., 2008) and in that sense it resembles iconic memory. Third, iconic memory decays rapidly over a period of about 500 ms when using traditional paradigms. Weak VSTM shows a similar decay curve (Sligte et al., 2008), only decay rate is slower.

On the other hand, there are also important dissimilarities. Weak VSTM seems to be a 'higher level' representation than traditional iconic memory, which is often conceived of as a raw 'snapshot' of the features of a visual scene. In weak VSTM, perceptual properties such as feature binding and figure-ground organization are encoded (Landman et al. 2003, 2004). Moreover, even when using the same change detection paradigm as used here, differences can be found between weak VSTM and something that is more like traditional iconic memory: traditional iconic memory traces are wiped out by any new stimulus, even a homogenous luminance screen (Sligte et al., 2008), whereas weak VSTM is not erased by homogenous or even textured screens, only by a new scene that contains object information (Landman et al., 2004; Sligte et al., 2008).

In sum, we are neutral to grouping weak VSTM into the domain of either iconic memory or VSTM, and the denominator used here (weak VSTM) is only intended to indicate its intermediate stage.

A fleeting form of consciousness?

The idea of additional object representations outside of VSTM finds its parallel in several theories of consciousness. Block (2005, 2007), for instance, argues for a distinction between phenomenal awareness (or "what we see") and access awareness (or "what we keep in mind") based on the enormous amount of information we experience seeing and the little information we can remember from one moment to the next (for congruent ideas, see Crick and Koch, 1990; Lamme, 2003, 2006). The transition between phenomenal and access awareness, and thus the availability for conscious report, is caused by the involvement of selective attention. From this theory, it follows that as long as representations exist in phenomenal awareness, they can be made available for conscious report by directing attention to them.

A similar, but neurally inspired theory distinguishes between unconscious, preconscious and conscious forms of visual processing (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). Unconscious processing is characterized by the absence of recurrent processing (RP), and visual awareness indeed never arises in the absence of RP (Hupé et al., 1998; Lamme, Zipser, & Spekreijse, 1998, 2002; Lamme & Roelfsema, 2000; Pascual-Leone & Walsh, 2001; Super, Spekreijse, & Lamme, 2001; Jolij & Lamme, 2005; Fahrenfort, Scholte, & Lamme, 2007). In some cases, a stimulus will evoke enough activation for conscious access, but top-down attention is simply not available at the moment or occupied elsewhere [as in attentional blink paradigms (Raymond, Shapiro, & Arnell, 1992) or the current experiment]. As a consequence, the stimulus will evoke RP confined to visual

and temporal areas. A long as this local RP is present, the stimulus representation will still be available for conscious access once top-down attention is re-deployed. When that happens, activity in visual and temporal areas is boosted and RP includes frontoparietal regions occupied with top-down or selective attention. These theories seem to qualify weak VSTM as what remains of phenomenal awareness (or preconscious processing) after the stimulus has been removed, and strong VSTM as what remains of access awareness.